1. Introduction

All we need is love, but all we get is work

1. Introduction

1.1 General Remarks

What follows is no longer relevant as far as lecture courses are concerned.
From here to the end of chapter 4 the content is more or less identical to the same chapters in the Hyperscript "Electronic Materials".

Some Important Links

- For a detailed *table of contents* use the link
- The *organization, timetable, etc.* for the running term can be found in the link.
- If you like to read *prefaces*, just click.
- For book recommendations: Consult the *list of books*

What is Special About this Course

- The lecture course "Electronic Materials" has a somewhat *special status* for two reasons:
  - **1.** It is far to short to really cover the topic appropriately, but yet, *it overlaps with other courses*. The reason for this is the mix of students who are required to take this course (see below).
  - **2.** It *had* a special format for the exercise part [1)].
  - Unfortunately, in the fall term of **2004**, this exercise format had to be abandoned for various reasons in favor of the more classical format.

Relation to Other Courses

- This graduate course "**Electronic Materials**" (in German: Werkstoffe der Elektrotechnik und Sensorik I) is a *required* course for

Study Course
1. *All* Materials Science Diploma students
2. *All* Master of Mat. Science and Engineering students.
3. *All* Electrical Engineering Diploma students.
4. *All* "Wirtschafts-Ingenieur ET&IT" Diploma students.

- Exactly what "required" means depends on your study course - look up your "Prüfungsordnung". Essentially the following rules obtain:
  - The first three study courses must pass the written examination, the last one must obtain the "Schein" for the exercise class
  - Even if you are *not* required to obtain the exercise "Schein" or the 1.5 ECTS "Credit Points", it is highly recommended to participate in the exercise class since it is a preparation for the examination!
- It interacts with several other courses in the materials science and electrical engineering curriculum. There is considerable overlap with the following courses
- **Silicon Technology I + II** (In German: Halbleitertechnologie I + II)
  - This course is required for Matwiss students in the Diploma track and electrical engineers specializing in solid state electronics.

  - It contains everything taught in the **Si**-technology section of "Electronic Materials". However, since the bulk of the electrical engineers will not be exposed to **Si**-technology anywhere else, "Electronic Materials" will cover the subject briefly. For all others, this part can be seen as an introduction to "Silicon Technology I + II"
- **Solid State Physics for Engineers II**
  - This course is required for Matwiss students in the Diploma and Master track and electrical engineers specializing in solid state electronics.
  - Dielectrics and magnetic materials will be covered in depth and from a more theoretical background. Again, the relevant chapters in "Electronic Materials" may be seen as introduction by those students enrolling in "Solid State II"; for the others it is an essential component of electrical engineering.
- The course has a very special relation to "**Introduction to Materials Science I + II**", which is a required course for all engineering (undergraduate) students.
  - "Electronic Materials" can be seen as part III of this series, because it covers the two major subjects left open in "Introduction to Materials Science I + II": *dielectrics* and *magnetic materials*. Moreover, the **Si**-technology part begins where the semiconductor part of "Introduction to Materials Science I + II" ends.
  - However, "Electronic Materials" is fully self-contained and can be taken by itself, provided the basic requirements are met.
  - For details of the contents of "Introduction to "Materials Science I + II" refer to the Hyperscripts (in German)
    MaWi I
    MaWi II
- **Sensors I** (In German: "Werkstoffe der Elektrotechnik und Sensorik II")
  - Required for all Materials Science students in the diploma track.
    (Used to be required for all electrical engineers).
  - Continues "Electronic Materials" with emphasize on sensor applications and ionic materials, but is self-contained and can be taken by itself.
  - "Electronic Materials" will include a brief chapter concerning ionic materials for those who do not take "Sensors I"
- **Semiconductors**
  - This course overlaps a little bit with "Electronic Materials", but essentially continues where Electronic Materials ends for Semiconductors.

Background Knowledge

- **Mathematics**
  - The course does not employ a lot of math. You should be familiar, however, with complex numbers, Fourier transforms and differential equations.
- **General Physics and Chemistry**
  - A general undergraduate level of basic physics should be sufficient. You should be comfortable with units and conversion between units.
- **General Materials Science**
  - You must know basic crystallography, quantum theory and thermodynamics.

1.1.2 How to use the Hyperscript

You'll figure it out. Otherwise use this link.

1.1.3 Electronic Materials and Products

- So what are "Electronic Materials"? Ask Google and you get an answer!

- Progress in Electrical Engineering was always dependent on progress in materials. For quite some time, electrical engineering meant electro*mechanical* engineering, and electrical products were made from "trivial" materials, as seen from a modern point of view. What was needed were cables, insulators, ferromagnetic sheet metal for transformers and generators, and a lot of metal for the general mechanics. A few applications centered around some mysterious materials - out of that grew *electronics* and electronic materials. But even then there were key materials:
    - *Cu wires* of all kinds. Not so trivial - how do you make a insulated but still flexible wire?
    - Insulating materials - plastics didn't quite exist yet. ***Mica*** was one of the key materials - there were mines for it!
    - *Graphite* and ***tungsten*** were important, whenever things got hot, like the filament in the light bulb or in a **vacuum tube**.
    - The "tube of **Braun**" - the "*Braunsche Röhre*" as it was known in Europe - the first **cathode ray tube** (***CRT)*** in other words - needed complicated glass work and some *ZnS* as electroluminescent material
    - Strange compounds like "*phosphor bronze*" were developed for contacts.
    - And **Selenium** (**Se**) was important for rectifiers, although nobody quite understood how it worked.
- The essential break through in the thirties was the **vacuum tube**; with it came electronics: Rectifiers, amplifiers, radio, black-and white **TV**, colour **TV**. It's not that long ago, but obviously long enough for some not to remember!
- The next break-through was called **transistor**; it happened in **1947**. **Integrated circuits** followed around **1970**, and since then we witness exponential growth with growth rates in the complexity of electronics (at constant prices) of up to **40%** a year!
    - A good (german) book covering this development in some detail is Hans **Queissers** "*Kristallne Krisen*".

1.2.2 Electronic Materials and Electronic Products

Electronic Products

- *Electronic Materials* are what you find inside the *components* making up *electronic products*. They consist of some stuff that you cannot easily exchange with something else - not even in principle - without losing the function.
    - What you *can* change easily for example, is the material for the box, the housing. Use *Al* instead of plastic or vice versa for your video recorder - it would still work, needing at most some minor adjustments.
    - You also may change (in principle) the metal for real wires. Using *Au*, *Ag*, or *Al* instead of - let's say - *Cu*, makes little difference for the function.
    - But exchange *any* material in a "chip" (i.e. in an **integrated circuit**) with something else (even allowing for minor adjustments) - and that definitely will be the end of your product.
- Let's look at some typical products or product groups that contain electronic materials:
    - *Electronics* in general (Computer, **TV**, Radio, Radar, Microwave, ...).
    - **Flat panel displays** (***FPD***).
    - **Micromechanics** and **Microsystems** (***MEMS***).
    - **Solar cells**.
    - **Lasers** (in particular semiconductor Lasers).
    - **Batteries**, **Accumulators**; energy storage systems in general.
    - **Sensors**, in particular solid state sensors, that convert whatever they sense directly into a current or a voltage.
    - **Fuel Cells**.
    - **Magnetic Memories**.

Looking at Components

- Consider, e.g., a **laptop** or **notebook** in more detail. If you take it apart, you find the "high tech" stuff:
  - Any number of *chips*, i.e. integrated circuits.
  - Some **quartz oscillators**.
  - A **hard disc**, i.e. a magnetic memory for the bulk memory.
  - A reading head for the hard disc that uses the "**giant magnetoresistance effect**"
  - A ***CD ROM***, i.e. an optical memory and a semiconductor ***Laser***
  - A **flat-panel display** (**FPD**) using "liqiud crystals", whichis pretty big as a single component, but cannot be subdivided in smaller pieces.
- But there is also "low tech" - or so it seems:
  - **Capacitors** and **inductors**.
  - *Switches*, connectors, the keyboard as a unit.
  - *Insulation*.
  - Mechanical stuff like the disk drive, but also the *housing.*
- Some components betray their key material in their name ("*quartz*" oscillator) or by common knowledge (who, besides some so-called intellectuals, does not know that the word "chip" is almost a synonym for Silicon?), but for most components we have to look deeper - we must open them up (which will almost always destroy them). What do we find?

Electronic Materials

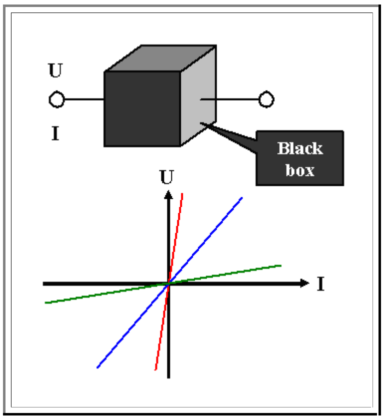
- Lets open up a chip. We find
  - **Packaging material** - either some *polymer* blend or *ceramics*.
  - A "chip" mostly consisting of **Si**, but interlaced in an intricate pattern with other materials like **P**, **B**, **As**, **SiO_2**, **Si_3N_4**, **$MoSi_2$**, **W**, **TiN**, **Al**, **Cu**....
  - A **lead frame** - the little pins sticking out of the package - made of some *metal alloys*.
  - Tiny **wires** connecting the leads to the chip or some pretty sophisticated stuff doing this job.
- Now open up the **FPD**. You will find many materials, the most suspicious beyond what we already found in chips are:
  - **Liquid crystals**, some strange liquid stuff.
  - **Amorphous Si**.
  - **Indium tin oxide *ITO***, a transparent electrode.
  - *Plastic foils* acting as *polarizers*.
  - A plastic (or *glass*) front and end plate.
- Now lets look at the **Laser** coming with the **CD** drive :
  - You find a complex mix of **GaAs**, **GaAlAs**, some other elements, as well as wires and packaging materials.
  - And all of this is quite different from what you find in the **Si** chips!
  - Soon you would find **GaN** in your Laser diode - and the capacity of your **CD** memory will quadruple!
- We could continue this, *but by now you got the idea*:

Progress in Electronic and Communication Technology is driven by **Progress in Material Science** **(and almost nothing else)**

2.1 Ohm's Law and Theory of Charge Transport

Note: **Chapter 2.1** here is a repetition of chapter 2.1 from "Einführung in die Materialwissenschaft II". Since it contains the scientific essentials about Ohm's law it is included in this Hyperscript

2.1.1. Ohms Law and Materials Properties



In this subchapter we will give an outline of how to progress from the simple version of **Ohms "Law"**, which is a kind of "electrical" definition for a black box, to a formulation of the same law from a *materials point of view* employing (almost) first principles.

- In other words: The *electrical engineering* point of view is: If a "black box" ***exhibits*** a linear relation between the (dc) current ***I*** flowing through it and the voltage ***U*** applied to it, it is an **ohmic resistor**.
- That is illustrated in the picture: As long as the voltage-current characteristic you measure between two terminals of the black box is linear, the black box is called an (ohmic) resistor).
- Neither the slope of the ***I-U***-characteristics matters, nor the material content of the box.

The*Materials Science* point of view is quite different. Taken to the extreme, it is:

- Tell me what kind of material is in the black box, and I tell you:

  1. If it really is an *ohmic* resistor, ***i.e.*** if the current relates *linearly* to the voltage for reasonable voltages and *both* polarities.
  2. What its (specific) resistance will be, including its temperature dependence.
  3. And everything else of interest.

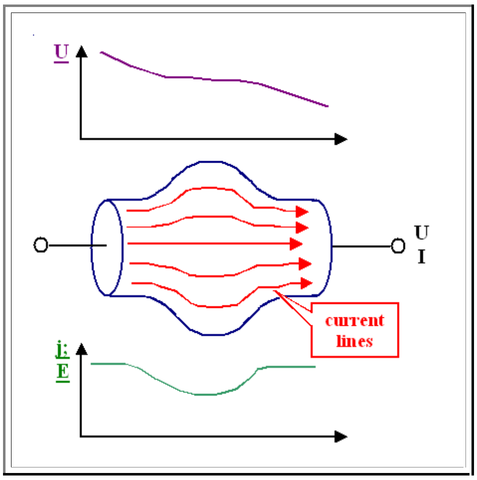
In what follows we will see, what we have to do for this approach. We will proceed in **3** steps

- In the first two steps, contained in this sub-chapter we simply reformulate Ohms law in physical quantities that are related to material properties. In other words, we look at the properties of the moving charges that produce an electrical current. But we only *define* the necessary quantities; we do not calculate their numerical values.
- In the third step - which is the content of many chapters - we will find ways to actually *calculate* the important quantities, in particular for semiconductors. As it turns out, this is not just difficult with classical physics, but simply impossible. We will need a good dose of quantum mechanics and statistical thermodynamics to get results.

1. Step: Move to *specific* quantities

First we switch from current ***I*** and voltage ***U*** to the **current density** $\underline{j}$ and the **field strength** $\underline{E}$, which are not only independent of the (uninteresting) size and shape of the body, but, since they are *vectors*, carry much more information about the system.

- This is easily seen in the schematic drawing below.



- *Current density* $\underline{j}$ and *field strength* $\underline{E}$ may depend on the coordinates, because ***U*** and ***I*** depend on the coordinates, e.g. in the way schematically shown in the picture to the left. However, for a homogeneous material with constant cross section, we may write

$$j = \frac{I}{F}$$

- with ***F*** **=** cross sectional area. The direction of the *vector* $\underline{j}$ would be parallel to the normal vector $\underline{f}$ of the reference area considered: it also may differ locally. So in full splendor we must write

$$\underline{j}(x,y,z) = \frac{I(x,y,z)}{F}\cdot\underline{f}$$

- The "global" field strength is

$$E = \frac{U}{l}$$

  - With l = length of the body. If we want the *local* field strength $\underline{E}(x,y,z)$ as a vector, we have, in principle, to solve the Poisson equation

$$\nabla\cdot\underline{E}(x,y,z) = \frac{\rho(x,y,z)}{\epsilon\epsilon_0}$$

  - With $\rho(x,y,z)$ = ***charge density***. For a homogeneous materials with constant cross section, however, $\underline{E}$ is parallel to $\underline{f}$ and constant everywhere, again which is clear without calculation.
- So. to make things easy, for a homogenous material of length l with constant cross-sectional area F, the field strength $\underline{E}$ and the current density $\underline{j}$ do not depend on position - they have the same numerical value everywhere.
  - For this case we can now write down Ohms law with the new quantities and obtain

$$j\cdot F = I = \frac{1}{R}\cdot U = \frac{1}{R}\cdot E\cdot l$$

$$\underline{j} = \frac{l}{F\cdot R}\cdot\underline{E}$$

- The fraction $l/F\cdot R$ *obviously* (think about it!) has the *same numerical value* for *any* homogeneous cube (or homogeneous whatever) of a given material; it is, of course, the **specific conductivity** σ

$$\sigma = \frac{1}{\rho} = \frac{l}{F\cdot R}$$

  - and ρ is the **specific resistivity**. In words: A **1 cm^3** cube of homogeneous material having the specific resistivity ρ has the resistance $R = (\rho\cdot l)/F$
  - Of course, we will never mix up the *specific resistivity* ρ with the *charge density* ρ or *general densities* ρ, because we know from the context what is meant!
  - The *specific resistivity* obtained in this way is necessarily identical to what you would define as specific resistivity by looking at some rectangular body with cross-sectional area F and length l.
  - The *specific conductivity* has the dimension $[\sigma] = \Omega^{-1}\mathrm{cm}^{-1}$, the dimension of the *specific resistivity* is $[\rho] = \Omega\mathrm{cm}$. The latter is more prominent and you should at least have a feeling for representative numbers by remembering

$$\begin{aligned}\rho\,(\text{metal}) &\approx 2\ \mu\Omega\text{cm}\\ \rho\,(\text{semicoductor}) &\approx 1\ \Omega\text{cm}\\ \rho\,(\text{insulator}) &\approx 1\ \text{G}\Omega\text{cm}\end{aligned}$$

- Restricting ourselves to isotropic and homogenoeus materials, restricts σ and ρ to being *scalars* with the *same numerical value* everywhere, and Ohms law now can be formulated for any material with weird shapes and being quite inhomogeneous; we "simply" have

$$\underline{j} = \sigma \cdot \underline{E}$$

- Ohms law in this *vector form* is now valid at *any point* of a body, since we do not have to make assumptions about the shape of the body.
  - Take an arbitrarily shaped body with current flowing through it, cut out a little cube (with your "mathematical" knife) at the coordinates **(*x,y,z*)** without changing the flow of current, and you must find that the local current density and the local field strength obey the equation given above *locally*.

$$\underline{j}(x,y,z) = \sigma \cdot \underline{E}(x,y,z)$$

  - Of course, obtaining the external current ***I*** flowing for the external voltage ***U*** now needs summing up the contributions of all the little cubes, i.e. integration over the whole volume, which may not be an easy thing to do.
- Still, we have now a much more powerful version of Ohms law! But we should now harbor a certain suspicion:
  - There is no good reason why $\underline{j}$ must always be *parallel* to $\underline{E}$. This means that for the most general case σ is not a *scalar* quantity, but a *tensor*; $\sigma = \sigma_{ij}$.
    (There is no good way to write tensors in html; we use the $_{ij}$ index to indicate tensor properties.
  - Ohms law then writes

$$\begin{aligned}j_x &= \sigma_{xx}\cdot E_x + \sigma_{xy}\cdot E_y + \sigma_{xz}\cdot E_z\\ j_y &= \sigma_{yx}\cdot E_x + \sigma_{yy}\cdot E_y + \sigma_{yz}\cdot E_z\\ j_z &= \sigma_{zx}\cdot E_x + \sigma_{zy}\cdot E_y + \sigma_{zz}\cdot E_z\end{aligned}$$

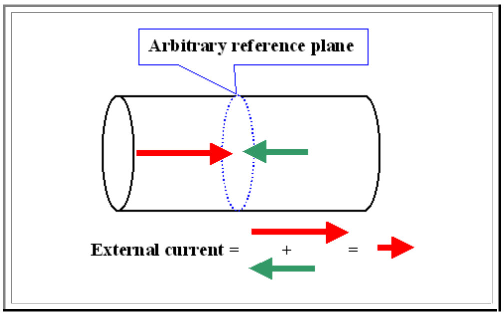
- For anisotropic inhomogeneous materials you have to take the tensor, and its components will all depend on the coordinates - that is the most general version of Ohms law.
  - Note that this is *not* so general as to be meaningless: We still have the basic property of Ohms law: The local current density is directly proportional to the local field strength (and not, for example, to $\exp-[\text{const.}\cdot\underline{E}]$).
  - We have a new thing. however: The current density vector $\underline{j}$ points no longer in the direction of the electrical field ***E***. In other words: The vector response of an *anisotropic material* to some disturbance or "driving force" still produces a vector but with a direction and amplitude that is determined by a tensor that describes the material properties. While this used to be a somewhat exotic material behavior for practitioners or engineers in the past, it is quickly becoming mainstream now., So you might as well acquaint yourself with tensor stuff right now. This link gives a first overview.
- Our goal now is to find a relation that allows to calculate σ_{ij} for a given material (or material composite); i.e. we are looking for
  - $\sigma_{ij} = \sigma_{ij}$**(material, temperature, pressure, defects...)**

2. Step: Describe σ_{ij} in Terms of the Carrier Properties

- Electrical current needs *mobile* charged "things" or *carriers* that are *mobile*. Note that we do not automatically assume that the charged "things" are *always* electrons. *Anything* charged and mobile will do.
- What we want to do now is to express σ_{ij} in terms of the *properties* of the carriers present in the material under investigation.
  - To do this, we will express an electrical current as a "mechanical" stream or current of (charged) particles, and compare the result we get with Ohms law.
- First, lets define an electrical current in a wire in terms of the carriers flowing through that wire. There are *three* crucial points to consider

1. The external electrical current as measured in an Ampèremeter is the result of the *net* current flow through any cross section of an (uniform) wire.

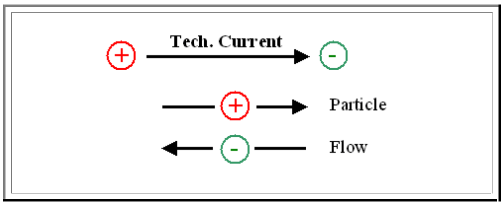
- In other words, the measured current is proportional to the *difference* of the number of carriers of the same charge sign moving from the *left to right* through a given cross sectional area *minus* the number of carriers moving from the *right to the left*.
- In short: the *net* current is the difference of two *partial* currents flowing in opposite directions:



- Do not take this point as something simple! We will encounter cases where we have to sum up **8** partial currents to arrive at the externally flowing current, so keep this in mind!

2. In summing up the individual current contributions, *make sure the signs are correct*. The rule is simple:

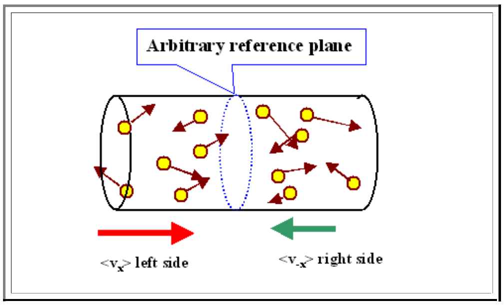
- The *electrical* current is (for historical reasons) defined as flowing from **+** to **–**. For a *particle* current this means:



- In words: A technical current ***I*** flowing from **+** to **–** may be obtained by *negatively* charged carriers flowing in the *opposite* direction (from **–** to **+**), by *positively* charged carriers flowing in the *same* direction, or from both kinds of carriers flowing at the same time in the proper directions.
- The particle currents of *differently* charged particles then must be *added*! Conversely, if negatively charged carriers flow in the same directions as positively charged carriers, the value of the partial current flowing in the "wrong" direction must be subtracted to obtain the external current.

3. The flow of particles through a reference surface as symbolized by one of arrows above, say the arrow in the **+*x*** -direction, must be seen as an *average* over the ***x*** -component of the velocity of the individual particles in the wire.

- Instead of *one* arrow, we must consider as many arrows as there are particles and take their *average*. A more detailed picture of a wire at *a given instant* thus looks like this



- An instant later it looks entirely different *in detail*, but exactly the same *on average*!
- If we want to obtain the net flow of *particles* through the wire (which is obviously proportional to the net *current* flow), we could take the average of the velocity components $\langle \mathbf{v}_{+x} \rangle$ pointing in the **+*x*** direction (to the right) on the left hand side, and subtract from this the average $\langle \mathbf{v}_{-x} \rangle$ of the velocity components pointing in the **–*x*** direction (to the left) on the right hand side.
- We call this *difference* in velocities the **drift velocity** $\underline{\mathbf{v}}_D$ of the *ensemble* of carriers.
- If there is no driving force, e.g. an electrical field, the velocity vectors are randomly distributed and $\langle \mathbf{v}_{+x} \rangle = \langle \mathbf{v}_{-x} \rangle$; the drift velocity and thus net current is zero as it should be.

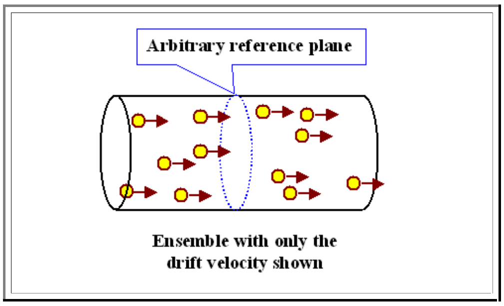
Average properties of ensembles can be a bit tricky. Lets look at some properties by considering the analogy to a localized **swarm of summer flies** "circling" around like crazy, so that the ensemble looks like a small cloud of smoke. This link provides for a more detailed treatment about averaging vectors.

- First we notice that while the *individual* fly moves around quite fast, its *vector* velocity $\underline{\mathbf{v}}_i$ averaged over time ***t***, $\langle \underline{\mathbf{v}}_i \rangle_t$, must be zero as long as the swarm as an ensemble doesn't move.

  - In other words, the flies, *on average*, move just as often to the left as to the right, etc. The net current produced by *all* flies at any given instance or by *one* individual fly after sufficient time is obviously zero for *any* reference surface.

- In real life, however, the fly swarm "cloud" often moves *slowly* around - it has a finite *drift velocity* which must be just the difference between the average movement in drift direction minus the average movement in the opposite direction.
  - The *drift velocity* thus can be identified as the proper average that gives the net current through a reference plane perpendicular to the direction of the drift velocity.
  - This drift velocity is usually much smaller than the average magnitude of the velocity $\langle \mathbf{v} \rangle$ of the individual flies. Its value is the difference of two large numbers - the average velocity of the *individual* flies in the drift direction minus the average velocity of the *individual* flies in the direction opposite to the drift direction.

- Since we are only interested in the drift velocity of the ensemble of flies (or in our case, carriers) we may now simplify our picture as follows:



- We now equate the *current density* with the *particle flux density* by the basic law of current flow:
  - Current density $\underline{j}$ = Number N of particles carrying the charge q flowing through the cross sectional area F (with the normal vector $\underline{f}$ and $|\underline{f}| = 1$) during the time interval t, or

$$\underline{j} = \frac{q \cdot N}{F \cdot t} \cdot \underline{f}$$

  - In scalar notation, because the direction of the current flow is clear, we have

$$j = \frac{q \cdot N}{F \cdot t}$$

- The problem with this formula is N, the *number* of carriers flowing through the cross section F every second.
  - N is not a basic property of the material; we certainly would much prefer the carrier *density* $n = N/V$ of carriers. The problem now is that we have to chose the volume $V = F \cdot l$ in such a way that it contains just the right number N of carriers.
  - Since the cross section F is given, this means that we have to pick the length l in such a way, that all carriers contained in that length of material will have moved across the internal interface after **1** second.
  - This is easy! The trick is to give l just that particular length that allows *every* carrier in the defined portion of the wire to reach the reference plane, i.e.

$$l = v_D \cdot t$$

  - This makes sure that *all* carriers contained in this length, will have reached F after the time t has passed, and thus *all* carriers contained in the volume $V = F \cdot v_D \cdot t$ will contribute to the current density. We can now write the current equation as follows:

$$j = \frac{q \cdot N}{F \cdot t} = \frac{q \cdot n \cdot V}{F \cdot t} = \frac{q \cdot n \cdot F \cdot l}{F \cdot t} = \frac{q \cdot n \cdot F \cdot v_D \cdot t}{F \cdot t}$$

- This was shown in excessive detail because now we have the *fundamental law of electrical conductivity* (in obvious vector form)

$$\underline{j} = q \cdot n \cdot \underline{v}_D$$

- This is a very general equation relating a *particle current* (density) via its *drift velocity* to an *electrical current* (density) via the charge q carried by the particles.
  - Note that it does not matter at all, *why* an ensemble of charged particles moves on average. You do not need an electrical field as driving force anymore. If a concentration gradient induces a particle flow via diffusion, you have an electrical current too, if the particles are charged.
  - Note also that electrical current flow *without* an electrical field as primary driving force as outlined above is *not* some odd special case, but at the root of most electronic devices that are more sophisticated than a simple resistor.
  - Of course, if you have different particles, with different density drift velocity and charge, you simply sum up the individual contributions as pointed out above.
- All we have to do now is to compare our equation from above to Ohms law:

$$\underline{j} = q \cdot n \cdot \underline{v}_D := \sigma \cdot \underline{E}$$

  - We then obtain

$$\sigma = \frac{q \cdot n \cdot v_D}{E} := \text{constant}$$

- If Ohms law holds, σ must be a constant, and this implies *by necessity*

$$\frac{v_D}{E} = \text{constant}$$

  - And this is a simple, but far reaching equation saying something about the driving force of electrical currents (= electrical field strength E) and the drift velocity of the particles in the material.
  - What this means is that *if* **v_D/E = const.** holds for *any* (reasonable) field E, the material will show ohmic behavior. *We have a first condition for ohmic behavior expressed in terms of material properties*.
  - If, however, **v_D/E** is constant (in time) for a *given* field, but with a value that depends on **E**, we have **$\sigma = \sigma(E)$**; the behavior will *not be ohmic*!
- The requirement **v_D/E = const.** for *any* electrical field thus requires a drift velocity in field direction for the particle, which is directly proportional to E. This leads to a simple conclusion:
  - This is actually a rather strange result! A charged particle in an electrical field experiences a constant force, and Newtons first law tells us that this will induce a constant accelerations, i.e. its velocity should increase all the time! Its velocity therefore would grow to infinity - if there wouldn't be some kind of friction.
  - We thus conclude that there *must* exist some mechanism that acts like a frictional force on all accelerated particles, and that this frictional force in the case of ohmic behavior must be in a form where *the average drift velocity obtained is proportional to the driving force*.
- Since **v_D/E = constant** must obtain for all (ohmic) materials under investigation, we may give it a name:

$$\frac{v_D}{E} = \mu = \textbf{Mobility} = \textbf{Material constant}$$

  - The **mobility** μ of the carriers has the unit
    $[\mu] = (m/s)/(V/m) = m^2/V \cdot s$.
  - The ***mobility*** μ (Deutsch: **Beweglichkeit**) then is a *material constant*; it is determined by the "friction", i.e. the processes that determine the average velocity for carriers in different materials subjected to the same force $q \cdot \underline{E}$.
  - **Friction**, as we (should) know, is a rather unspecified term, but always describing energy transfer from some moving body to the environment.

  - Thinking ahead a little bit, we might realize that μ is a basic material constant *even in the absence of electrical fields*. Since it is tied to the "friction" a moving carrier experiences in its environment - the material under consideration - it simply expresses how fast carriers give up surplus energy to the lattice; and it must not matter how they got the surplus energy. It is therefore no suprise if μ pops up in all kinds of relations, e.g. in the famous Einstein - Smoluchowski equation linking *diffusion coefficients* and *mobility* of particles.

- We now can write down the *most general form of Ohms law* applying to all materials meeting the two requirements: n **= const.** and μ **= const.** everywhere. It is expressed completely in particle (= material) properties.

$$\sigma = q \cdot n \cdot \mu$$

  - The task is now to calculate n and μ from first priciples, i.e. from only knowing what atoms we are dealing with in what kind of structure (e.g. crystal + crystal defects)
  - This is a rather formidable task since σ varies over a extremely wide range, cf. a short table with some relevant numbers.

- In order to get acquainted with the new entity "mobility", we do a little exercise:

Exercise 2.1-1

Derive and dicuss numbers for μ

- Since we like to give σ as a positive number, we always take only the magnitude of the charge q carried by a particle.
  - However, if we keep the sign, e.g. write $\sigma = -e \cdot n \cdot \mu_e$ for electrons carrying the charge $q = -e$; e = elementary charge, we now have an indication if the particle current and the electrical current have the *same* direction ($\sigma > 0$) or opposite directions $\sigma < 0$) as in the case of electrons.
  - But it is entirely a matter of taste if you like to ***schlepp*** along the signs all the time, or if you like to fill 'em in at the end.

- Everything more detailed then this is no longer universal but specific for certain materials. The remaining task is to calculate n and μ for given materials (or groups of materials).
  - This is not too difficult for simple materials like metals, where we know that there is one (or a few) free electrons per atom in the sample - so we know n to a sufficient approximation. Only μ needs to be determined.
  - This is fairly easily done with classical physics; the results, however, are flawed beyond repair: They just do not match the observations and the unavoidable conclusion is that classical physics must not be applied when looking at the behavior of electrons in simple metal crystals or in any other structure - we will show this in the immediately following subchapter **2.1.3**.

- We obviously need to resort to **quantum theory** and solve the **Schrödinger equation** for the problem.
  - This, surprisingly, is also fairly easy in a simple approximation. The math is not too complicated; the really difficult part is to figure out what the (mathematical) solutions actually *mean*. This will occupy us for quite some time.

Questionaire

Multiple Choice Questions to 2.1.1

2.1.2 Ohms Law and Classical Physics

- In this subchapter we will look at the *classical* treatment of the movement of electrons inside a material in an electrical field.
  - In the preceding subchapter we obtained the most basic formulation of *Ohms law*, linking the specific conductivity to two fundamental material parameters:

$$\boxed{\sigma = q \cdot n \cdot \mu}$$

- For a homogeneous and isotropic material (***e.g.*** polycrystalline metals or single crystal of cubic semiconductors), the *concentration* of carriers n and their *mobility* μ have the same value everywhere in the material, and the specific conductivity σ is a *scalar*.
  - This is boring, however. So let's look at useful complications:
- In general terms, we may have more than one kind of carrier (this is the common situation in semiconductors) and n and μ could be functions of the temperature T, the *local* field strength E_{loc} resulting from an applied *external* voltage, the detailed structure of the material (e.g. the defects in the lattice), and so on.
  - We will see that these complications are the essence of advanced electronic materials (especially semiconductors), but in order to make life easy we first will restrict ourselves to the special class of **ohmic materials**.
  - We have seen before that this requires n and μ to be independent of the local field strength. However, we still may have a temperature dependence of σ; even commercial ohmic resistors, after all, do show a more or less pronounced temperature dependence - their resistance increases roughly linearly with T.
- In short, we are treating *metals*, characterized by a *constant* density of *one* kind of carriers (= electrons) in the order of **1 ...3** electrons per atom in the metal.

Basic Equations and the Nature of the "Frictional Force"

- We consider the electrons in the metal to be "free", i.e. they can move freely in any direction - the atoms of the lattice thus *by definition* do not impede their movement
  - The (local) electrical field $\underline{E}$ then exerts a force $\underline{F} = - e \cdot \underline{E}_{loc}$ on any given electron and thus accelerates the electrons in the field direction (more precisely, opposite to the field direction because the field vector points from **+** to **–** whereas the electron moves from **–** to **+**).
  - In the fly swarm analogy, the electrical field would correspond to a steady airflow - some wind - that moves the swarm about with constant drift velocity.
- Now, if a single electron with the (constant) mass **m** and momentum $\underline{p}$ is subjected to a force $\underline{F}$, the equation of motion from basic mechanics is

$$\boxed{\underline{F} = \frac{d\underline{p}}{dt} = \frac{m \cdot d\underline{v}}{dt}}$$

  - Note that $\underline{p}$ does not have to be zero when the field is switched on.
- If this would be all, the velocity of a given electron would acquire an ever increasing component in field direction and eventually approach infinity. This is obviously not possible, so we have to bring in a mechanism that destroys an unlimited increase in $\underline{v}$.
  - In classical mechanics this is done by introducing a **frictional force** $\underline{F}_{fr}$ that is proportional to the velocity.

$$\boxed{\underline{F}_{fr} = - k_{fr} \cdot \underline{v}}$$

  - with k_{fr} being some friction constant. But this, while mathematically sufficient, is *devoid of any physical meaning* with regard to the moving electrons.
  - There is no "*friction*" on an atomic scale! Think about it! Where should a friction force come from? An electron feels *only* forces from two kinds of fields - electromagnetic and gravitational (neglecting strange stuff from particle physics). So we have to look for another approach.
- What friction does to big classical bodies is to ***dissipate*** ordered kinetic energy of the moving body to the environment. Any ordered movement gets slowed down to zero (surplus) speed, and the environment gets somewhat hotter instead, i.e. *unordered* movement has increased.

  - This is called **energy dissipation**, and that is what we need: Mechanisms that take kinetic energy away from an electron and "give" it to the crystal at large. The science behind that is called *(Statistical) Thermodynamics* - we have encounterd it before.

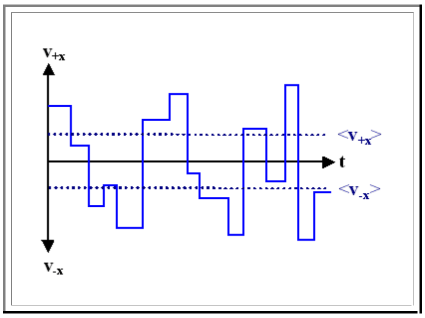
- The best way to think about this, is to assume that the electron, flying along with increasing velocity, will *hit something else* along its way every now and then; it has a *collision* with something else, it will be **scattered** at something else.
  - This collision or scattering event will change its *momentum*, i.e. the magnitude and the direction of $\underline{v}$, and thus also its kinetic energy E_{kin}, which is always given by

$$\boxed{E_{kin} = \frac{m \cdot \underline{v}^2}{2} = \frac{\underline{p} \cdot v}{2}}$$

- In other words, we consider collisions with something else, i.e. other particles (including "pseudo" particles), where the total energy and momentum of all the particles is preserved, but the individual particle looses its "memory" with respect to its velocity before the collision, and starts with a new momentum after every collision.
- What are the "partners" for collisions of an electron, or put in standard language, what are the **scattering mechanisms**? There are several possibilities:
  - Other ***electrons***. While this happens, it is not the important process in most cases. It also does not decrease the energy contained in the electron movement - the losses of some electron are the gains of others.
  - ***Defects***, e.g. foreign atoms, point defects or dislocations. This is a more important scattering mechanism and moreover a mechanism where the electron can transfer its surplus energy (obtained through acceleration in the electrical field) to the lattice, which means that the material heats up
  - ***Phonons***, i.e. *"quantized" lattice vibrations* traveling through the crystal. *This is the most important scattering mechanism*.
- Now that is a bit strange. While we (hopefully) have no problem imagining a crystal lattice with all atoms vibrating merrily, there is no immediate reason to consider these vibrations as being *localized* (whatever this means) and *particle-like*.
  - You are right - but nevertheless: The lattice vibrations indeed are best described by a bunch of particle-like **phonons** careening through the crystal.
  - This follows from a quantum mechanical treatment of lattice vibrations. Then it can be shown that these vibrations, which contain the thermal energy of the crystal, are quantized and show typical properties of (quantum) particles: They have a *momentum*, and an *energy* given by $h\nu$ (h = Plancks constant, ν = frequency of the vibration).
- *Phonons* are a first example of "pseudo" particles; but there is no more "pseudo" to phonons than there is to photons.
  - We will not go into more details here. All we need to know is that a hot crystal has more phonons *and* more energetic phonons than a cold crystal, and treating the interaction of an electron with the lattice vibration as a collision with a phonon gives not only *correct* results, it is the *only* way to get results at all.
- At this point comes a crucial insight: It would be far from the truth to assume that only ***accelerated*** electrons scatter; scattering happens all the time to all the electrons moving randomly about because they all have some thermal energy. Generally, scattering is the mechanism to achieve thermal equilibrium and equidistribution of the energy of the crystal.
  - If electrons are accelerated in an electrical field and thus gain energy in excess of thermal equilibrium, scattering is the way to transfer this surplus energy to the lattice which then will heat up. If the crystal is heated up from the outside, scattering is the mechanism to turn heat energy contained in lattice vibrations to kinetic energy of the electrons.
  - Again: Even without an electrical field, scattering is the mechanism to transfer thermal energy from the lattice to the electrons (and back). Generally, scattering is the mechanism to achieve *thermal equilibrium* and equidistribution of the energy of the crystal.
  - Our free electrons in metals behave very much like a gas in a closed container. They ***careen*** around with some average velocity that depends on the energy contained in the **electron gas**, which is - in classical terms- a direct function of the temperature.

Averaging over Random Scattering Events

- Lets look at some figures illustrating the scattering processes.



  - Shown here is the *magnitude* of the velocity $\underline{v}_{\pm x}$ of an electron in **+*x*** and **–*x*** direction *without* an external field. The electron moves with constant velocity until it is scattered, then it continues with some new velocity.
  - The scattering processes, though unpredictable at single events, must lead to the averages of the velocity, which is characteristic for the material and its conditions.
  - The plural in "average*s*" is intentional: there *are* different averages of the velocity
  - Whereas $\langle \underline{v} \rangle = 0$, $\langle v \rangle$ has a finite value; this is also true for $\langle \underline{v}_x \rangle = - \langle \underline{v}_{-x} \rangle$. Consult the "fly swarm modul" if you are unsure about this.

- From classical thermodynamics we know that the (classical) electron gas in thermal equilibrium with the environment contains the energy $E_{kin} = (1/2)kT$ per particle and degree of freedom, with ***k*** = Boltzmanns constant and ***T*** = absolute temperature. The three degrees of freedom are the velocities in ***x***-, ***y***- and ***z***-direction, so we must have

$$E_{kin,x} = \tfrac{1}{2} \cdot m \cdot \langle v_x \rangle^2 = \tfrac{1}{2} \cdot kT$$

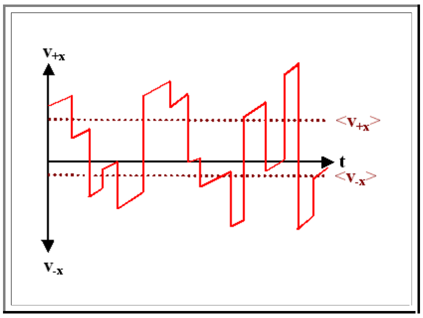
$$\langle v_x \rangle = \left(\frac{kT}{m} \right)^{1/2}$$

- For the other directions we have exactly the same relations, of course. For the total energy we obtain

$$E_{kin} = \frac{m \cdot \langle v_x^2 \rangle}{2} + \frac{m \cdot \langle v_y^2 \rangle}{2} + \frac{m \cdot \langle v_z^2 \rangle}{2} = \frac{m \cdot \langle v^2 \rangle}{2} = \frac{m \cdot (v_0)^2}{2} = \frac{3kT}{2}$$

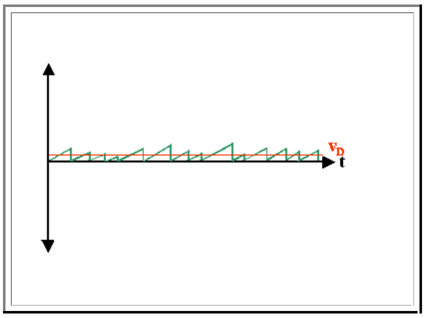
- with $v_0 = \langle v \rangle$. v_0 is thus the **average velocity of a carrier** careening around in a crystal.
- At this point you should stop a moment and think about just how fast those electrons will be careening around at room temperature (**300K**) without plugging numbers in the equation. Got a feeling for it? Probably not. So look at the exercise question (and the solution) further down!.
- Now you should stop another moment and become very aware of the fact that this equation is from purely *classical* physics. It is absolutely true for *classical* particles - which electrons are actually not. Electrons obey the Pauli principle, i.e. they behave about as non-classical behavior as it is possible. This should make you feel a bit uncomfortable. Maybe the equation from above is not correct for electrons then? Indeed - it isn't. Why, we will see later; also how we can "repair" the situation!

- Now lets *turn on an electrical field*. It will accelerate the electrons *between* the collisions. Their velocity in field direction then increases linearly from whatever value it had right after a collision to some larger value right before the next collision.
  - In our diagram from above this looks like this:



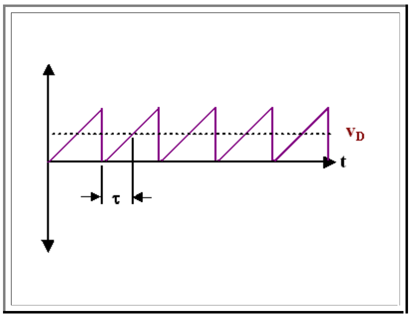
  - Here we have an electrical field that accelerates electrons in in ***x***-direction (and "brakes" in **–*x*** direction). Between collisions, the electron gains velocity in **+*x***-direction at a constant rate (= identical slope).
  - The average velocity in **+*x*** directions, $\langle v_{+x} \rangle$, is now larger than in **–*x*** direction, $\langle v_{-x} \rangle$.
  - However, beware of the pitfalls of schematic drawings: For real electrons the difference is very small as we shall see shortly; the slope in the drawing is very exaggerated.

- The drift velocity is contained in the difference $\langle \mathbf{v}_{+x}\rangle - \langle \mathbf{v}_{-x}\rangle$; it is completely described by the velocity gain between collisions. For obtaining a value, we may neglect the instantaneous velocity right after a scattering event because they average to zero anyway and just plot the *velocity gain* in a simplified picture; always starting from zero after a collision.



- The picture now looks quite simple; but remember that it contains some not so simple averaging.
- At this point it is time to define a very meaningful *new* average quantity:
- The *mean time between collisions*, or more conventional, the mean time τ for reaching the drift velocity **v** in the simplified diagram. We also call τ the **mean scattering time** or just *scattering time* for short.

This is most easily illustrated by simplifying the scattering diagram once more: We simply use just *one* time - the average - for the time that elapses between scattering events and obtain:



- This is the *standard diagram* illustrating the scattering of electrons in a crystal usually found in text books; the definition of the scattering time τ is included
- It is highly idealized, if not to say just wrong if you compare it to the correct picture above. Of course, the average velocity of both pictures will give the same value, but that's like saying that the average speed $\mathbf{v_a}$ of all real cars driving around in a city is the same as the average speed of ideal model cars all going at $\mathbf{v_a}$ all the time.
- Note that τ is only *half* of the average time between collisions.

So, while this diagram is not wrong, it is a highly abstract rendering of the underlying processes obtained after several averaging procedures. From this diagram only, no conclusion whatsoever can be drawn as to the average velocities of the electrons without the electrical field!

New Material Parameters and Classical Conductivity

With the scattering concept, we now have two new (closely related) material parameters:

- The **mean (scattering) time** τ between two collisions as defined before, and a directly related quantity:
- The **mean free path *l*** between collisions; i.e. the distance travelled by an electron (on average) before it collides with something else and changes its momentum. We have

$$\boxed{l = 2\tau \cdot (v_0 + v_D)}$$

- Note that $\mathbf{v_0}$ enters the defining equation for ***l***, and that we have to take twice the scattering time τ because it only refers to half the time between collisions!

After we have come to this point, we now can go on: Using τ as a new parameter, we can rewrite Newtons equation from above:

$$\boxed{\frac{d\underline{v}}{dt} = \frac{\Delta v}{\Delta t} = \frac{v_D}{\tau}}$$

- It is possible to equate the *differential* quotient with the *difference* quotient, because the velocity change is constant. From this we obtain

$$\frac{v_D}{\tau} = -\frac{E \cdot e}{m}$$

$$\Rightarrow \quad v_D = -\frac{E \cdot e \cdot \tau}{m}$$

Inserting this equation for v_D in the old definition of the current density $j = -n \cdot e \cdot v_D$ and invoking the general version of Ohms law, $j = \sigma \cdot E$, ***yields***

$$j = \frac{n \cdot e^2 \cdot \tau}{m} \cdot E \; := \; \sigma \cdot E$$

- This gives us the final result

$$\sigma = \frac{n \cdot e^2 \cdot \tau}{m}$$

This is the *classical* formula for the conductivity of a classical "electron gas" material; i.e. metals. The conductivity contains the density n of the free electrons and their mean scattering time τ as material parameters.

- We have a good idea about n, but we do not yet know τ_{class}, the mean *classical* scattering time for classical electrons. However, since we know the order of magnitude for the conductivity of metals, we may turn the equation around and use it to calculate the order of magnitude of τ_{class}. If you do the exercise farther down, you will see that the result is:

$$\tau_{class} = \frac{\sigma \cdot m}{n \cdot e^2} \approx (10^{-13} \ldots\ldots 10^{-15})\ \text{sec}$$

- "Obviously" (as stated in many text books), this is a value that is *far too small* and thus the classical approach must be *wrong*. But is it really too small? How can *you* tell without knowing a lot more about electrons in metals?
- Let's face it: *you can't !!*. So let's look at the ***mean free path l*** instead. We have

$$l = 2 \cdot \tau \cdot (v_0 + v_D)$$

and

$$(v_0)^2 = \frac{3kT}{m}$$

- The last equation gives us a value $v_0 \approx 10^4$ **m/s** at room temperature! Now we need v_D, and this we can estimate from the equation given above to $v_D = -E \cdot \tau \cdot e/m \approx 1$ **mm/sec**, *if* we use the value for τ dictated by the measured conductivities. It is much smaller than v_0 and can be safely neglected in calculating l.

We thus can rewrite the equation for the conductivity and obtain

$$\sigma = \frac{n \cdot e^2 \cdot l}{2 \cdot m \cdot (v_0 + v_D)}$$

- Knowing σ from experiments, but not *l*, allows to determine *l*. The smallest possible mean free path l_{min} between collisions (for $v_D = 0$) thus is

$$l_{min} = \frac{2 \cdot m \cdot v_0 \cdot \sigma}{n \cdot e^2} = 2 \cdot v_0 \cdot \tau \approx (10^{-1} - 10^{1})\ \text{nm}$$

  - *And this is certainly too small!*.

- But before we discuss these results, let's see if they are actually true by doing an exercise:

Exercise 2.1-2

Derive numbers for v_0, τ, v_D, and *l*

- Now to the important question: *Why* is a mean free path in the order of the size of an atom too small?
  - Well, think about the scattering mechanisms. The distance between lattice defects is certainly much larger, and a phonon itself is "larger", too.
  - Moreover, consider what happens at temperatures below room temperatures: *l* would become even smaller since v_0 decreases - somehow this makes *no sense*.

- It does not pay to spend more time on this. Whichever way you look at it, whatever tricky devices you introduce to make the approximations better (and physicists have tried very hard!), you will *not* be able to solve the problem: The mean free paths are never even coming close to what they need to be, and the conclusion which we will reach - maybe reluctantly, but unavoidably - must be:

**There is no way to describe conductivity (in metals)
with *classical* physics!**

Scattering and Mobility

- Somewhere on the way, we have also indirectly found that the **mobility μ** as defined before is just another way to look at scattering mechanisms. Let's see why.
  - All we have to do is to compare the equation for the conductivity from above with the master equation $\sigma = q \cdot n \cdot \mu$.

- This gives us immediately

$$\mu = \frac{e \cdot \tau}{m}$$

$$\mu \approx \frac{e \cdot l}{2 \cdot m \cdot v_0}$$

  - In other words:

The *decisive* material property determining the mobility μ is the average time between scattering events or the mean free path between those events.

- The mobility **μ** thus is a basic material property, well-defined even without electrical fields, and just another way to express the scattering processes taken place by a number.

In the equations above slumbers an extremely important aspect of semicoductor technology.

- In all electronic devices carriers have to travel some distance before a signal can be produced. A **MOS** transistor, for example, switches currents on or off between its "Source" and "Drain" terminal depending on what voltage is applied to its "Gate". Source and drain are separated by some distance l_{SD}, and the "Drain" only "feels" the "on" state after the time it takes the carriers to run the distacne l_{SD}.
- How long does that take if the voltage between Source and Drain is U_{SD}?
- Easy. If we know the mobility **μ** of the carriers, we now their (average) velocity v_{SD} in the source-drain region, which by definition is $v_{SD} = \mu \cdot U_{SD}/l_{SD}$.
- The traveling time t_{SD} between source and drain for obvious reasons defines roughly the maximum frequency f_{max} the transistor can hand, we have $t_{SD} = l_{SD} / v_{SD}$ or

$$t_{SD} = \frac{l_{SD}^2}{\mu \cdot U_{SD}} \approx \frac{1}{f_{max}}$$

- The maximum frequency of a **MOS** transistor thus is directly proportional to the mobiltiy of the carriers in the material it is made from (always provided there are no other limiting factors). And since we used a rather general argument, we should not be surprised that pretty much the same relation is also true for most electronic devices, not just MOS transistors.
- This is a momentous statement: We linked a prime material parameter, the *material constant* **μ**, to one of the most important parameters of electronic circuits. We would like **μ** to be as large as possible, of course, and now we know what to do about it!

A simple exercise is in order to see the power of this knowlegde:

Exercise 2.1-3

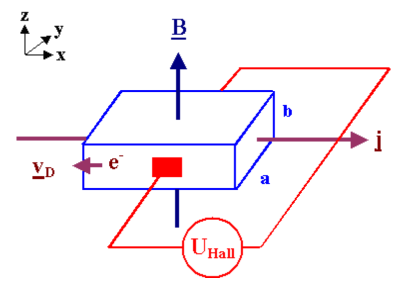
What does it take to build a **4 GHz** microprocessor?

Questionaire

Multiple Choice Questions to 2.1.2

2.1.3 The Hall Effect

- This subchapter introduces *two* important topics: The **Hall effect** as an important observation in materials science and at the same time another irrefutable proof that classical physics just can't hack it when it comes to electrons in crystals.
  - The Hall effect describes what happens to current flowing through a conducting material - a metal, a semiconductor - if it is exposed to a magnetic field $\underline{B}$.
  - We will look at this in *classical* terms; again we will encounter a fundamental problem.
- The standard geometry for doing an experiment in its most simple form is as follows:



Relevant vectors; the current is carried by electrons

  - A magnetic field $\underline{B}$ is employed perpendicular to the current direction $\underline{j}$, as a consequence a *potential difference* (i.e. a *voltage*) develops at right angles to both vectors.
  - In other words: A **Hall voltage** U_{Hall} will be measured perpendicular to $\underline{B}$ and $\underline{j}$.
  - In yet other words: An electrical field $\underline{E}_{Hall}$ develops in y-direction
  - That is already the essence of the Hall effect.

- It is relatively easy to calculate the magnitude of the *Hall voltage* U_{Hall} that is induced by the magnetic field $\underline{B}$.
  - First we note that we must also have an electrical field $\underline{E}$ parallel to $\underline{j}$ because it is the driving force for the current.
  - Second, we know that a magnetic field at right angles to a current causes a force on the moving carriers, the so-called **Lorentz force** F_L, that is given by

$$\underline{F}_L = q \cdot (\underline{v}_D \times \underline{B})$$

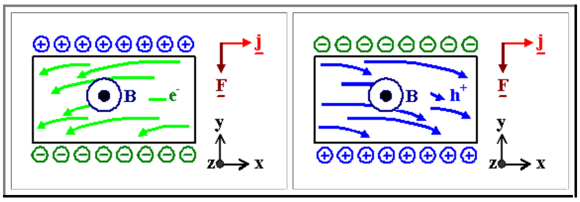
  - We have to take the drift velocity $\underline{v}_D$ of the carriers, because the other velocities (and the forces caused by these components) cancel to zero on average. The vector product assures that $\underline{F}_L$ is perpendicular to $\underline{v}_D$ and $\underline{B}$.
  - Note that instead the usual word "electron" the neutral term *carrier* is used, because in principle an electrical current could also be carried by charged particles other then electrons, e.g. positively charged ions. Remember a simple but important picture given before!
- For the geometry above, the Lorentz force $\underline{F}_L$ has only a component in y - direction and we can use a scalar equation for it. F_y is given by

$$F_y = -q \cdot v_D \cdot B_z$$

  - We have to be a bit careful, however: We know that the force is in y-direction, but we do longer know the sign. It changes if either q, v_D, or B_z changes direction and we have to be aware of that.
- With $\underline{v}_D = \mu \cdot \underline{E}$ and μ = mobility of the carriers, we obtain a rather simple equation for the force

$$F_y = -q \cdot \mu \cdot E_x \cdot B_z$$

  - It is important to note that for a fixed current density j_x the direction of the Lorentz force is independent of the sign of the charge carriers (the sign of the charge and the sign of the drift velocity just cancel each other).
- This means that the current of carriers will be deflected from a straight line in y-direction. In other words, there is a component of the velocity in y-direction and the surfaces perpendicular to the y-direction will become charged as soon as the current (or the magnetic field) is switched on. The flow-lines of the carriers will look like this:



  - The charging of the surfaces is unavoidable, because some of the carriers eventually will end up at the surface where they are "stuck".
  - Notice that the sign of the charge for a given surface depends on the sign of the charge of the carriers. Negatively charged electrons (**e^-** in the picture) end up on the surface opposite to positively charged carriers (called **h^+** in the picture).
  - Notice, too, that the direction of the force $\underline{F}_y$ is the same for both types of carriers, simply because both q and $\underline{v}_D$ change signs in the force formula
- The surface charge then induces an electrical field E_y in **y**-direction which opposes the Lorentz force; it tries to move the carriers back.
  - In *equilibrium*, the Lorentz force $\underline{F}_y$ and the force from the electrical field E_y in **y**-direction (which is of course simply $q \cdot E_y$) must be equal with opposite signs. We obtain

$$q \cdot E_y = -q \cdot \mu \cdot E_x \cdot B_z$$

$$E_y = -\mu \cdot E_x \cdot B_z$$

- The Hall voltage U_{Hall} now is simply the field in **y**-direction multiplied by the dimension d_y in **y**-direction.
  - It is clear then that the (easily measured) Hall voltage is a *direct measure* of the mobility **μ** of the carriers involved, and that its **sign** or polarity will change if the sign of the charges flowing changes.
- It is customary to define a **Hall coefficient** R_{Hall} for a given material.
  - This can be done in different, but equivalent ways. In the link we look at a definition that is particularly suited for measurements. Here we use the following definition:

$$R_{Hall} = \frac{E_y}{B_z \cdot j_x}$$

- In other words, we expect that the Hall voltage $E_y \cdot d_y$ (with d_y = dimension in **y**-direction) is proportional to the current(density) j and the magnetic field strength B, which are, after all, the main experimental parameters (besides the trivial dimensions of the specimen):

$$E_y = R_{Hall} \cdot B_z \cdot j_x$$

- The Hall coefficient is a material parameter, indeed, because we will get different numbers for R_{Hall} if we do experiments with identical magnetic fields and current densities, but different materials. The Hall coefficient, as mentioned before, has interesting properties:
  - R_{Hall} will change its sign, if the sign of the carriers is changed because then E_y changes its sign, too. It thus indicates in the most unambiguous way imaginable if positive or negative charges carry the current.
  - R_{Hall} allows to obtain the mobility **μ** of the carriers, too, as we will see immediately
- R_{Hall} is easily calculated: Using the equation for E_y from above, and the basic equation $j_x = \sigma \cdot E_x$, we obtain for *negatively* charged carriers:

$$R_{Hall} = -\frac{\mu \cdot E_x \cdot B_z}{\sigma \cdot E_x \cdot B_z} = -\frac{\mu}{\sigma}$$

- Measurements of the Hall coefficient of materials with a *known* conductivity thus give us *directly* the mobility of the carriers responsible for the conductance.
  - The **–** sign above is obtained for *electrons*, i.e. *negative* charges.
  - If positively charged carriers would be involved, the Hall constant would be positive.

- Note that while it is not always easy to measure the numerical value of the Hall voltage and thus of ***R*** with good precision, it is the easiest thing in the world to measure the *polarity* of a voltage.

- Lets look at a few experimental data:

Material	**Li**	**Cu**	**Ag**	**Au**	**Al**	**Be**	**In**	Semiconductors (e.g. **Si**, Ge, **GaAs**, **InP**,...)
R $(\times 10^{-24})$ cgs units	–1,89	–0,6	–1,0	–0,8	+1,136	+2,7	+1,774	*positive* or *negative* values, depending on "doping"

Comments:
1. the *positive* values for the metals were measured under somewhat special conditions (low temperatures; single crystals with special orientations), for other conditions negative values can be obtained, too.
2. The units are not important in the case, but multiplying with $9 \cdot 10^{13}$ yields the value in **m^3/Coulomb**

- Whichever way we look at this, one conclusion is unavoidable:
  - In certain materials including *metals*, the particles carrying the electrical current are *positively charged* under certain conditions. And this is *positively not possible* in a classical model that knows only *negatively charged electrons* as carriers of electrical current in solids!
- Again we are forced to conclude:

There is no way to describe conductivity in metals and semiconductors with *classical* physics!

Questionaire

Multiple Choice Fragen zu 2.1.3

2.1.4 Summary to: Conductors - Definitions and General Properties

What counts are the *specific* quantities:

- Conductivity σ (or the specific resistivity $\rho = 1/\sigma$.
- current density $\underline{j}$.
- (Electrical) field strength · $\underline{E}$.

$$\sigma = |q| \cdot n \cdot \mu$$

$$\underline{j} = \sigma \cdot \underline{E}$$

σ (of conductors / metals) obeys (more or less) several rules; all understandable by looking at n and particularly μ.

- Matthiesen rule:
  Reason: Scattering of electrons at defects (including phonons) decreases μ.

$$\rho = \rho_{\text{Lattice}}(T) + \rho_{\text{defect}}(N)$$

- "$\rho(T)$ rule":
  about **0,04 %** increase in resistivity per **K**
  Reason: Scattering of electrons at phonons decreases μ.

$$\Delta\rho = \alpha_\rho \cdot \rho \cdot \Delta T \approx \frac{0{,}4\%}{^\circ\mathrm{C}}$$

- Nordheim's rule:
  Reason: Scattering of electrons at **B** atoms decreases μ.

$$\rho \approx \rho_A + \text{const.} \cdot [B]$$

Major consequence: You can't beat the conductivity of pure **Ag** by "tricks" like alloying or by using other materials
(Not considering superconductors).

Non-metallic conductors are *extremely* important.

- Transparent conductors (TCO's)
  ("ITO", typically oxides).

No flat panels displays =
no notebooks etc. without ITO!

- Ionic conductors (liquid and solid).

Batteries, fuel cells, sensors, ...

- Conductors for high temperature applications; corrosive environments, ..
  (Graphite, Silicides, Nitrides, ...).

Example: $MoSi_2$ for heating elements in corrosive environments (dishwasher!).

- Organic conductors (and semiconductors).

The future High-Tech key materials?

Numbers to know (order of magnitude accuracy sufficient)

ρ(decent metals) about 2 μΩcm.
ρ(technical semiconductors) around 1 Ωcm.
ρ(insulators) > 1 GΩcm.

Questionaire

All Multiple Choice questions to 2.1

2.2 Materials and Conductivity

2.2.1 Metals

- *A few words before you start:*
    - Conductors *in general* are a bit boring, whereas conductors *in particular applications* are often hot topics (take the recent switch from **Al** to **Cu** in chip technology, for example).
    - There is a large number of highly optimized materials which are used as conductors nowadays. Just enumerating them is tiresome and not very interesting. Still, some knowledge of this issue is a must for materials scientists in the context of electronic materials.
    - As far as "theory" goes, there is either not much that goes beyond a basic knowledge of solid state physics (which, it is assumed, you already have), or very involved special theory (e.g. for superconductors or conducting polymers) - for which there is no time.
- In conclusion, we will only touch the issue, trying to present all major facets of the topic. In particular, the long list of applications for conductors (much longer than you probably would imagine) will be covered. This chapter, however, will be brief and mostly lists topics and key words.

The Basics

- The essential parameters of interest for conductors are:
- **1. Specific resistivity** ρ or **specific conductivity** σ **= 1/**ρ.
    - The defining "master" equation is

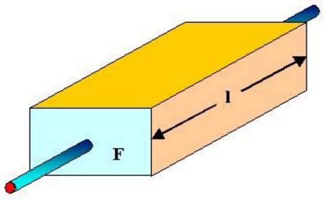
$$\sigma \; = \; |q| \cdot n \cdot \mu$$

-
    - With q **=** magnitude of the *charge* of the current carrying particles; n = *concentration* of current carrying particles (usually electrons in conductors) ; μ **= mobility** of the current carrying particles.
    - The units are

$$[\rho] \; = \; \Omega\text{m}$$

$$[\sigma] \; = \; (\Omega\text{m})^{-1} \; = \text{S/m}$$

-
    - Note that S = "**Siemens**" = **1/**Ω **= A/V** is a bit old fashioned, but still in use. Note, too, that while the **SI** standard units call for the *meter* (**m**), you will find many values given in Ω***cm***.
    - A homogeneous material with a constant cross-sectional area F and a length l thus has a resistance of $R = (\rho \cdot l)/F$



$$R \; = \; \frac{\rho \cdot l}{F}$$

-
    - Or, in other words, a cube with **1 cm** length has a resistance R given in Ω that is numerically equal to ist specific resistance ρ given in Ω**cm**.
- If *electrons* are carrying the current, we have $q = -e$ = elementary charge $= 1.602 \cdot 10^{-19}$ **C**.
    - For units, conversions, and so on consult the link!
- **2. Ohm's law**. Ohm's law (which was not a "law", but an empirical observation) formulated for the specific quantities writes

$$\underline{j} = \sigma \cdot \underline{E}$$

- With $\underline{j}$ = current density (a *vector*); $\underline{E}$ = electrical field strength (a vector); σ = specific conductivity, in general a *tensor* of **2nd** rank and, most important, *not* a function of the field strength $\underline{E}$ if not specifically noted. In other words, if the specific conductivity of a material is a constant, i.e. a fixed number with respect to $\underline{E}$, the material obeys Ohm's law.
- Ohm's law thus means that the $\underline{E}$ - $\underline{j}$ characteristics or the easily measured voltage - current characteristics are always *straight lines* through the origin! Within reasonable values of $\underline{E}$, or U, of course.

If you have *any* problem with these equations, perhaps because you feel Ohm's law should read $R = U/I$, or if you are not sure about the the meaning of the basic quantities, as e.g., *mobility*, *you have a problem*. Turn to the required reading module and other modules accessible from there.

- More about Ohm's law and the failure of classical physics in explaining the conductivity of metals can be found in a second required reading module. Add to this the required reading module for averaging vector quantities and you are ready for this chapter and others to come.

A remark to the ***mathematical notation***: HTML possibilities are limited and it is difficult to adhere to all rules of notation. In case of doubt, clarity and easy reading will have preference to formal correctness. This means:

- Whenever sensible, *cursive* symbols will be used for variables. It is *not* sensible, e.g., to use cursive letters for the velocity **v**, because the cursive ***v*** is easily mixed up with the Greek nu ν.
- All equations and the quantities used in equations are always **bold** - this greatly improves readability. However, it leaves little room for symbolizing vectors by **bold** lettering, and since underlining is cumbersome and not particularly helpful, we simply will mostly not use special notation for vectors. If you are able to understand this lecture course at all, you will know the vector (or tensor) quantities anyway.
- There are not enough letters in the alphabet to give every physical quantity an unambiguous symbol. One and the same symbol thus traditionally has several meanings, usually quite clear from the context. Occasionally, however, the danger of mix-up occurs. An example in case is the traditional use of the letter E for electrical field strength and for energies (and for Young's modulus in German). While in conventional texts one must give a different letter to these quantities, we will use the advantage of HTML and use *color coding* whenever the possibility of negative symbol interference raises its ugly head.

The *density* and *mobility* of *mobile charged* carriers thus determines the *conductivity*.

- The *carrier density* is a function of bonding (metallic, covalent in semiconductor, etc.), defects (doping in semiconductors) and temperature in general. In metals, however, n_e is nearly constant.
- The *mobility* is a function of collisions between carriers (e.g. electrons and holes) and/or between carriers and obstacles (e.g. **phonons** and crystal lattice defects).

Carrier concentration and mobility are, in general, hard to calculate from first principles. In semiconductors, the carrier density is easy to obtain, mobility is somewhat harder. In metals, the carrier density is rather fixed, but mobility is quite difficult to calculate, especially for "real" i.e. rather imperfect crystals. There are however, empirical rules or "laws".

- *Ohm's "law"* asserting that σ is *not* a function of E but only of some material property that can be expressed as a number.
- **Matthiesen's rule**, stating that

$$\rho = \rho_{\text{Lattice}}(T) + \rho_{\text{defect}}(N)$$

- With N = some measure of defect density.
- A "*rule of thumb*": ρ is proportional to T for $T >$ some T_{crit}

$$\Delta\rho = \alpha_\rho \cdot \rho \cdot \Delta T \approx \frac{0{,}4\%}{^\circ C}$$

- With **Temperature coefficient** $\alpha_\rho = 1/\rho \cdot d\rho / dT$.
- Then we have the **Wiedemann-Franz** "law", linking electrical conductivity to thermal conductivity, and so on.

The links give some graphs and numbers for representative metals.

- **Table of some metal properties**
- **$\rho(T)$ for different defect densities in Na**
- **$\rho(T)$ for different metals**

Some Values and Comments

- The range of resistivity values (at room temperature) for metals is rather limited; here are some values as well as a first and last reminder that σ and ρ, while closely related, are quite different parameters with a numerical value that depends on the choice of the units! Do not mix up **cm** and **m**!

Metal	Ag	Cu	Au	Al	Na	Zn	Ni	Fe	Sn	Pb	Hg
ρ [μΩcm]	**1,6**	**1,7**	**2,2**	**2,7**	**4,2**	**5,9**	**6,8**	**9,7**	**12**	**21**	**97**
σ = 1/ρ [$10^6 \cdot \Omega^{-1}\mathrm{cm}^{-1}$]	**0.625**	**0.588**	**0,455**	**0.37**	**0.238**	**0.169**	**0.147**	**0.103**	**0.083**	**0.046**	**0.01**
σ = 1/ρ [$10^6 \cdot \Omega^{-1}\mathrm{m}^{-1}$]	**62,5**	**58.8**	**45.5**	**37**	**23.8**	**16.9**	**14.7**	**10.3**	**8.3**	**4.6**	**1**

  - The temperature dependence, expressed e.g. in **ρ(300K)/ρ(100K)** may be a factor of **5 ...10**, so it is *not* a small factor. It may be used and is used, for measuring temperatures, e.g. with well-known **Pt** resistivity thermometers.
  - This is something you should be aware of; cf. the anecdote in the link.
- The specific resistivity, however, is not the only property that counts. In selecting a metal, important design parameters might also be:
  - *Weight*, *mechanical strength*, *corrosion resistance*, *prize*, *compatibility* with other materials, ...
- Sometimes it is advisable to look at "**figures of merit**", i.e. the numerical value coming out of a self-made formula that contains your important criteria in a suitable way.
  - One very simple example: Lets say, *weight* is important. Define a *figure of merit* = $F = \rho / d$, with d = density. The bigger F, the better.
  - You now get the following ranking (normalized to $F_{Na} = 1$):

Metal	Na	K	Ca	Al	Mg	Cu	Ag
F	**1**	**0,77**	**0,69**	**0,56**	**0,52**	**0,28**	**0,25**

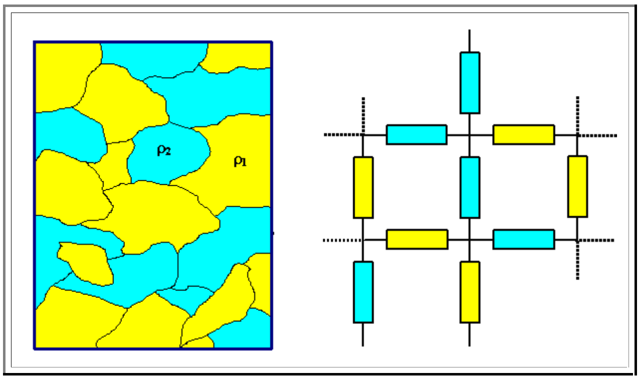
- The winner sodium! So you are going to use *Sodium* - **Na** for wiring?
  - Certainly not. Because now you will either include chemical stability C in your figure of merit (just multiply with C and assign values $C = 1$ for great stability (e.g. **Au**, **Al**,), $C = 2$ for medium stability (**Cu**, **Mg**) and $C = 5$ for unstable stuff (**Na**, **K**, **Ca**). Or any other number reflecting the importance you put on this parameter. There is no ready made recipe - if precise numbers are not existing, you take your personal grading scale.
  - And while you are at it, divide by some price index P. Use the price per **kg**, or just a rough grade scale.
  - In this simple example, you will get a surprising result: No matter what you do, the winner will be **Al**. It is the (base) material of choice for heavy duty applications when weight matters. In not so simple questions, you really may benefit from using the figure of merit approach.

Questionaire

Multiple Choice questions to 2.2.1

2.2.2 Alloys

- Pure metals are rarely used - in the real world you use **alloys**.
  - In principle, the specific resistivity ρ of an alloy can be obtained from the phase diagram and the ρ - values of the phases involved. Lets look at the extremes:
- **1.** Complete ***immiscibility***, e.g. in the case of **Au/Si**, or **Cu/W**. We may treat the resulting mix of metal particles as a *network of resistors* being linked in series and parallel. The volume fractions of the phases would constitute the weights - the treatment is not unlike the elastic modulus of compounds.



  - But no matter what kind of volume fraction you use and how you treat the resistor network - the resulting resistivity will *never* be smaller than that of the ingredient with the smallest resistivity.
- **2.** Complete *miscibility* (e.g. **Au/Ag**, **Cu/Ni**). Experimentally we find for small amounts (some **%**) of **B** in **A** (with **[B]** = concentration of **B**)

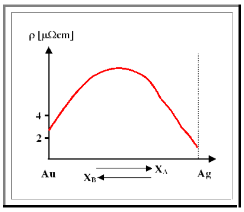
$$\rho \approx \rho_A + \text{const.} \cdot [B]$$

  - This formula is a special case of **Nordheims rule** which states .

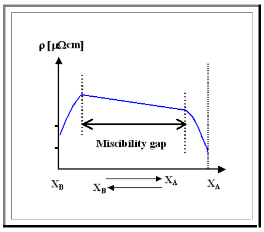
$$\rho \approx X_A \cdot \rho_A + X_B \cdot \rho_B + \text{const.} \cdot X_A \cdot X_B$$

  - This is pretty much an empirical law, it does not pay to justify it theoretically. Again, it is not possible to produce an alloy with a *resistivity smaller than one of its components*.
- If you have *intermetallic compounds* in your phase diagram, use Nordheim's rule with the intermetallic phases as X_A and X_B.
  - This leaves open the possibility that some intermetallic phase, i.e. a defined compound with its own crystal lattice, might have a lower resistivity than its constituents. While this is unlikely (if not outright impossible?) on theoretical grounds, no such intermetallics have been found so far.
- The sad fact then is that unleashing the full power of metallurgy and chemistry on mixing conductors (i.e. metals), will not give you a conductor with a specific conductivity better than **Ag**.
  - You will have to turn to **superconductors** (forgetting about cost considerations), if you can't live with **Ag**.
- Lets look at some examples:

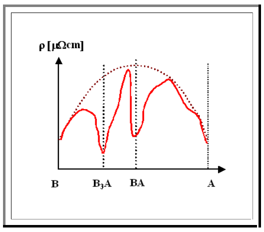
Complete miscibility
(e.g. **Au/Ag**)



Miscibility gap
(e.g. **Ag/Cu**);
phase diagram in the link.



Intermetallic phases
(e.g. **Cu/Au**)



What do we learn? Something simple:

- Again: The resistivity *always* goes up in alloys or mixed materials as compared to the pure materials.
- Things are generally complicated, but *full of potential* for custom tailored properties!

Questionaire
Multiple Choice questions to 2.1.2

2.2.3 Non-Metalic Conductors

We will just give a brief look at some especially important or useful **non-metallic conductors**:

Conducting Polymers

That polymers, usually associated with insulators, can be very good conductors was a quite unexpected discovery some **20** years ago (Noble prize **2001**). They always need some "**doping**" with ionic components, however.

- The resistivity can be ***exceedingly*** low. e.g. for *Iodine (I) doped poly-acethylene* ***(pAc)*** we may have. $\rho \leq$ **6,7 µΩcm**.
- Or in other words: If you divide by the density for some figure of merit, it beats everything else, since **{ρ/density} (pAc) > {ρ/density} (Na)**!
- More typical, however, are resistivities around (**10 .... 1000) µΩcm**.

The conduction mechanism is along **–C=C–C=C–C=** chains, it is not yet totally clear. In fact, the first question is why this kind of chain is *not* generally highly conducting. Use the link for the answer.

- The conductivity is strongly dependent on "doping" (in the **%** range!) with ions, and on many other parameters, the link gives an example.
- So do not confuse this with the doping of semiconductors, where we typically add far less than **1 %** of a dopant!

A new object of hot contemporary research are now ***semi*****conducting polymers** which have been discovered about **10** years ago.

Transparent conductors

Indium Tin Oxide (*ITO*) (including some variations) is the only really usable *transparent* conductor with reasonable conductivity (around **1 Ωcm**)! It consists of **SnO_2** doped with **In_2O_3**.

- **ITO** is technically *very important*, especially for:
  - flat panel displays, e.g. ***LCD***s .
  - solar cells.
  - research (e.g. for the electrical measurements of light-induced phenomena).
- **ITO** is one example of conducting oxides, others are **TiO**, **NiO**, or **ZnO**. The field is growing rapidly and known as "***TCO***" = Transparent Conducting Oxides

If you can find a transparent conductor much better than **ITO** (which leaves a lot to be desired), you may not get the Nobel prize, but you will become a rich person rather quickly.

- Since **In** is rare, and the demand is exploding since the advent of **LCD**s, you also would be a rich person of you invested in **In** some years ago.

Ionic conductors

Solid *Ionic conductors* are the materials behind "**Ionics**", including key technologies and products like

- Primary batteries.
- Rechargeable (secondary) batteries.
- Fuel cells.
- Sensors.
- High temperature processes, especially smelting, refining, reduction of raw materials (e.g. **Al**-production).

- There is an extra module devoted to the **Li** ion battery. This is important for you if you are interested in driving an affordable a car in **20** years or so.
- They are also on occasion the unwanted materials causing problems, e.g. in corrosion or in the degradation of dielectrics.
- See Chapter 2.4 for some details about ionic conductors.

Specialities: Intermetallics, Silicides, Nitrides etc.

- **Silicides**, i.e. metal - silicon compounds, are important for microelectronics (***ME***) technology, but also in some more mundane applications, e.g. in heating elements. Some resistivity examples for silicides:

Silicide	$MoSi_2$	$TaSi_2$	$TiSi_2$	$CoSi_2$	$NiSi_2$	PtSi	Pd_2Si
ρ (µΩcm)	40 ...100	38...50	13..16	10...18	≈ 50	28...35	30...35

  - It looks like the winner is **$CoSi_2$**. Yes, but it is difficult to handle and was only introduced more recently, like **$NiSi_2$**. In the earlier days (and at present) the other silicides given above were (and still are) used.

- Some more examples of **special conductors** which find uses out there:

Material	HfN	TiN	TiC	TiB_2	C (Graphite)
ρ (µΩcm)	30...100	40...150	ca. 100	6 ...10	1000

Superconductors

- Superconductors are in a class of their own. All kinds of materials may become superconducting at low temperatures, and there are neither general rules telling you *if* a material will become superconducting, nor at which *temperature*.
  - There will be an advanced module some time in the future.

- *Why do we need those "exotic" materials?*. There are two general reasons:
  - **1.** Because, if just *one* specific requirement exists for your application that is not met by common materials, you simply have *no choice*. For example, if you need a conductor usable at **3000 K -** you take *graphite*. No other choice. It's as simple as that.
  - **2.** Because *many* requirements must be met *simultaneously*. Consider e.g. **Al** for integrated circuits - there are plenty of important requirements; **see the link**. Since *no* material meets *all* of many requirements, an optimization process for finding an optimum material is needed.
  - **Al** won the race for chip metallization for many years, but now is crowded out by **Cu**, because in some figure of merit the importance of low resistivity in the list of requirements is much larger now than it was in the past. It essentially overwhelms *almost* all other concerns (if there would not be an almost, we would have **Ag**!).

Questionaire
Multiple Choice questions to 2.1.3

2.2.4 Summary to: Conductors - Definitions and General Properties

- What counts are the *specific* quantities:
  - Conductivity σ (or the specific resistivity $\rho = 1/\sigma$).
  - current density ***j***.
  - (Electrical) field strength · ***E***.

$[\rho] = \Omega\mathrm{m}$
$[\sigma] = (\Omega\mathrm{m})^{-1} = \mathrm{S/m}$; S = "Siemens"

- The basic equation for σ is:
  n = concentration of carriers,
  μ = mobility of carriers.

$$\sigma = |q| \cdot n \cdot \mu$$

- Ohm's law states:
  It is valid for metals, but not for all materials.

$$\underline{j} = \sigma \cdot \underline{E}$$

- σ (of conductors / metals) obeys (more or less) several rules; all understandable by looking at ***n*** and particularly ***μ***.
  - Matthiesen rule:
    Reason: Scattering of electrons at defects (including phonons) decreases ***μ***.

$$\rho = \rho_{\mathrm{Lattice}}(T) + \rho_{\mathrm{defect}}(N)$$

  - "ρ(*T*) rule":
    about **0,04 %** increase in resistivity per **K**
    Reason: Scattering of electrons at phonons decreases ***μ***.

$$\Delta\rho = \alpha_\rho \cdot \rho \cdot \Delta T \approx \frac{0{,}4\%}{^\circ\mathrm{C}}$$

  - Nordheim's rule:
    Reason: Scattering of electrons at **B** atoms decreases ***μ***.

$$\rho \approx \rho_A + \mathrm{const.} \cdot [B]$$

- Major consequence: You can't beat the conductivity of pure **Ag** by "tricks" like alloying or by using other materials
  (Not considering superconductors).

- Non-metallic conductors are *extremely* important.
  - Transparent conductors (TCO's)
    ("ITO", typically oxides).

    No flat panels displays = no notebooks etc. without **ITO**!

  - Ionic conductors (liquid and solid).

    Batteries, fuel cells, sensors, ...

  - Conductors for high temperature applications; corrosive environments, ..
    (Graphite, Silicides, Nitrides, ...).

    Example: **$MoSi_2$** for heating elements in corrosive environments (dishwasher!).

  - Organic conductors (and semiconductors).

    The future High-Tech key materials?

- Numbers to know (order of magnitude accuracy sufficient)

ρ(decent metals) about **2 μΩcm**.
ρ(technical semiconductors) around **1 Ωcm**.
ρ(insulators) > 1 GΩcm.

Questionnaire

All Multiple Choice questions to 2.1

2.3. General Applications

2.3.1 Normal Conductors

A world without *conductors* is even harder to imagine than a world without *semiconductors*. Examples for applications include

- High-voltage free-air power transmission lines.
- High voltage wires for trains (getting "scratched" all the time).
- In-house wiring.
- Low-voltage wiring (car systems).
- High current wiring (machines).
- System on-board wiring.
- Bond wires for ***IC***'s (diameter **< 30µm**).
- Metallization on chips.
- Screening electrical or magnetic fields.
- Avoidance of electrostatic charging.
- Electrodes for batteries, chemical reactors etc.
- Antennas.

Each use has special requirements which should be met by the conducting material.

Some examples for requirements

- *Money* (Use of **Au**, **Ag**, **Pt** etc. may be critical).
- *Chemistry* (general stability and reactivity; esentiall excludes **Na**, **K**, **Hg** etc. for most applications; corrosion properties, ...).
- *Mechanical* properties (Pure metals are often too soft, but alloys have higher resistivity).
- *Thermal* properties (temperature coefficient; no metal usable beyond ca. **1000 K**) .
- *Compatibility with other materials* (contact corrosion, solderability, thermoelectric and thermomechanical properties, general chip compatibility, ...).
- *Compatibility with production technologies* (e.g. thin film deposition methods, wire making (try this with a brittle superconductor!),...).

Whole families of conductors, fine-tuned for a specific applications, were developed; below are some examples.

***Cu** based conductors*

There are many precisely specified **Cu**-based conductors for all kind of specific applications, examples are given in the link.

***Al** based conductors*

This family is primarily used for high-voltage free-air cables (in combination with a steel core) because of best fitting in terms of conductivity - price - mech. ***strength*** - corrosion requirements; cf. the illustration in the link.

Others

In *one* **IC** you may find the following conductor materials:

- Poly crystalline highly doped **Si**.
- Silicides; i.e. **Si** - metal compounds like $\mathbf{NiSi_2}$.
- **Al** with ≤ **1%** of **Si** and **Cu** if the chip was made before, say, **2000**.
- **Cu** with some additions instead of **Al** if the chip was made after **2000**.
- **W**.
- **TiN**.

because one material simply does not meet the specific requirements for conductor on chips.

2.3.2 Contacts

- **Contacts**, meaning **mechanical contacts** here, are a major part of most electronic products. Even if there is no mechanical switch anymore, you still have the contact between the plug and the outlet, and/or the contact springs for the batteries.
- *Contacts* include the following items:
  - Switches, ***plugs***, relays, connections to removable parts (batteries, light bulbs, ...), ***pantographs*** (the thing on top of a locomotive), "brushes" (for motors), and so on.
  - Contacts are also the components or materials that often cause *trouble*. Contacts or switches are often the first components to break, and thus a nuisance to consumers like you and me.
- There are many specific requirements for **contact materials**:
  - Small contact resistance (it is never zero).
  - No sticking or welding under load.
  - No ***abrasion*** under load.
  - No intermixing of materials.
  - No wearing and tearing.
  - Suitable mechanical properties, e.g. good elasticity (forever) for switches.
- There are specific materials and group of materials generally favored for contacts:
  - **C** (graphite in many forms) for pantographs and whenever you want to draw a big current.
  - **Cu**, **Ag**, **Au**.
  - **Ru**, **Rh**, **Pd**, **Os**, **Ir**, **Pt**.
  - **Mo**, **W**.
  - ....
  - An example of **Ag**-based contact materials can be found in the link.
  - For contact applications we find *expensive* materials, because in many applications only small quantities are needed and the inertness of noble metals is what counts.

2.3.3 Resistors and Heating

Resistors

Basic requirements for **resistors** (still one of the most numerous component in circuits) are:

- Large region of ***R*** values (= device resistance in Ω) within *one* production technology.
- Small (ideally vanishing) temperature coefficient .
- Minimal noise.
- Small dependence of ρ on production parameters (good repeatability).
- No Ageing.
- Small thermoelectrical coefficients to **Cu** (you want a resistor, not a thermoelement).

Materials of choice include

- **Ta**, **Ta** based alloys, and in particular "**Constantan**" (**55% Cu, 44% Ni, 1% Mn**), a resistor material with an especially small temperature coefficient α_ρ, but a large thermoelectric coefficient).
- Strange mixtures of conductors and insulators including "**Cermet**" (short for Ceramics - Metals), e.g. **Cr - SiO_2**.

Details and data in the *(future)* link.

Heating

Basic requirements for **heating elements** are:

- High melting point.
- Chemical stability at high temperatures and in potentially corrosive environments.
- Mechanical strength at high temperatures.

The choice of a materials depends significantly on the range of temperatures envisioned. We have:

- **FeNiCr**, **FeNiAl** alloys.
- **Pt**, **W**, **Ta**, **Mo** - stable elements with a high melting point.
- **$MoSi_2$** Among more industrial applications also used as heaters in dish washers - this is very aggressive environment!
- Graphite (up to **3000 K** in non-oxidizing gas).

Some details and data can be found in the links.

- **Overview of resistivity and temperature range for some materials**
- **Maximum temperatures for some materials**

2.3.4 Summary to: Conductors - General Applications

- No electrical engineering without conductors!

  - Hundreds of specialized metal alloys exist just for "wires" because besides σ, other demands must be met, too:

    Money, Chemistry (try Na!), mechanical and thermal properties, compatibility with other materials, compatibility with production technologies, ...

  - Example for unexpected conductors being "best" compromise:

    Poly Si, Silicides, **TiN**, **W** in integrated circuits.

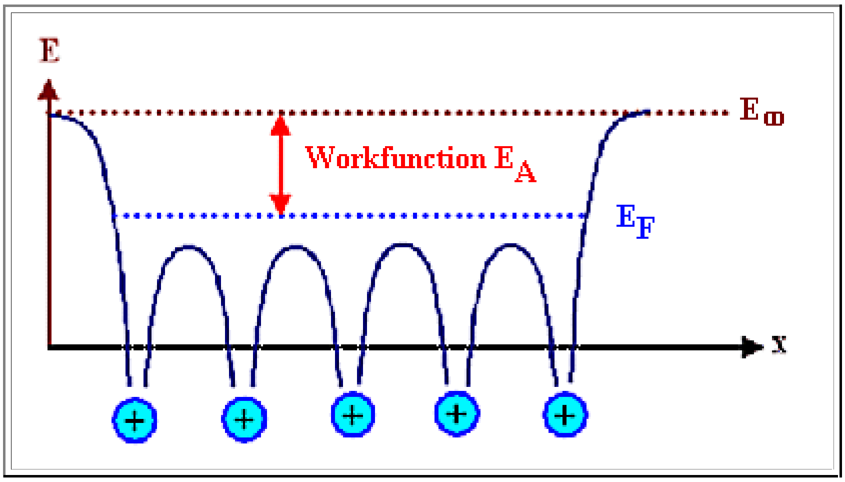
- Don't forget Special Applications:

  Contacts (switches, plugs, ...);
  Resistors;
  Heating elements; ...
  Transparent conductors

2.4. Special Applications

2.4.1 Thermionic Emission

- **Cathodes** in **cathode ray tubes** (***CRT***), in regular **electron tubes** (still used for special applications), but also in all electron beam tools as e.g. *electron microscopes* or *electron beam welding*, are one example of *special conductors*. We need to have free electrons in the material *and* we need to extract them from the material.
    - For good cathodes we wish for some specific properties: First we want to extract lots of electrons *easily* and in *large quantities* (i.e. we want high current densities for little money).
    - Second, we want to extract them from a very *small area* (for high brightness), so that we can consider the electron beam to come from a *point source* which makes (electron) optics a lot less complicated to handle!
- Lets look at the free electron gas model and see how we can extract electrons in general.



- For a metal, there are lots of electrons in the last band at all energies up to the Fermi energy, and at very low temperatures it takes at least the energy E_A to push an electron up the energy scale to E_∞, where it would be free to go wherever it wants - it is no longer "bound" to the crystal. We call that particular energy the ***work function*** of the material.

- The **work function** E_A of the material is thus the decisive quantity; it is the difference between the **Fermi energy** and the potential at infinity E_∞.

$$E_A = E_F - E_\infty$$

- If we let $E_\infty = 0$ and have the energy scale going "down", we simply have .

$$E_A = E_F$$

- The current density for thermionic emission is given by the well-known **Richardson equation**, which we obtain by calculating how many electrons will have sufficient energy (and momentum) to overcome the energy barrier to the outside world from the energy distribution of electrons in a free electron gas model.
    - The necessary calculation is not too difficult but lengthy so we will not do it here. The result, however, is simple and given below. The Richardson equation for the current density j from a hot surface states:

$$j = A \cdot T^2 \cdot \exp - \frac{E_A}{kT}$$

- From measuring $j = j(T)$ we expect (almost) **Arrhenius** behavior; E_A then follows from the slope of the plot, the constant A from its intersection with the j - axis.
- If you are unsure about what this function looks like, use the function generator and play a bit.
- The pre-exponential factor A can be calculated from the free electron gas model, but than it is only a crude approximation for real materials. Its free-electron-gas value is:
  $A_{\text{theo}} = 120\ \mathrm{A \cdot cm^{-2} \cdot K^{-2}}$.

- Lets compare that to some measured values (and bear in mind that A may depend on the Miller indices of the crystallographic plane from which we extract electrons, too - so numbers vary):

Material	**Fe**	**Ni**	**Pt**	**Ta**	**W**	**Cs**	**LaB_6**
A $[Acm^{-2}K^{-2}]$	26	30	32	55	60	162	25
E_A [eV]	4,5 - 4,8	5,15 - 5,35	5,65	4,15 - 4,8	4,2	1,8 - 2,14	2,6
T_m $[^oC]$	1 535	1 452	1 755	2 850	3 410	28,4	2 210

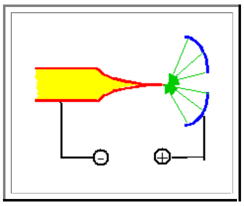
Cs has the lowest work function, but its melting point is so low that it is of no use. Optimizing everything, the winners are:

- **W**, the workhorse for cathode materials.
- **LaB_6**, a rather exotic material, because single crystals with very fine tips can be made that provide high current densities from a very small area. This is important whenever you want to focus the electron beam on a "point", e.g. in scanning electron microscopes. The focal point cannot be smaller than the area from where the electron beam was extracted from - *and you like it to be in the **nm** region*. The price one has to pay for this (besides for the **LaB_6** cathode, which is not cheap), is that the cathode has to be run in ultra high vacuum (***UHV***), because the fine tip would otherwise soon be damaged by ion bombardment resulting from ions extracted out of the residual gas atmosphere.

Questionnaire
Multiple Choice questions to 2.3.1 (and 2.3.2)

2.4.2 Field Enhanced Emission and Tunnelling Effects

- If you run a **cathode**, emitting an electron beam, with *large* electrical fields between the cathode and the anode, you will find that your workfunction E_A seems to change to smaller values as the field strength increases.
  - This is called **Schottky effect**; it is observed at large field values of **$(10^5 - 10^8)$V/cm**.
- If you apply even higher field strengths (and remember: $E = U/d$; you do not need high voltages U, only small dimensions d), E_A seems to vanish altogether.

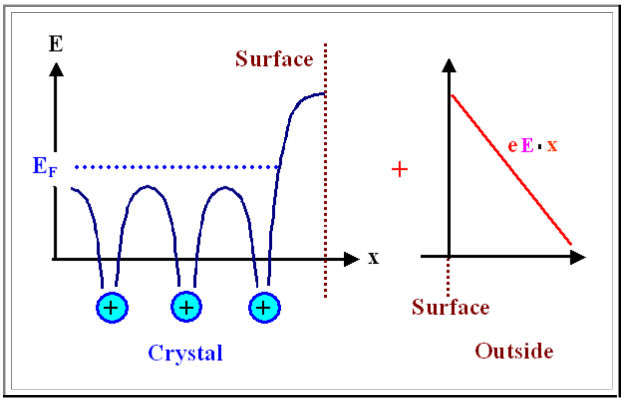


  - This effect is is called **field emission**. It works even at room temperature and is barely temperature dependent, so it can not be a temperature activated process.
  - Field emission is rather easy to obtain: all you have to do, is to make a very fine tip with a curvature of the tip in the **nm** - range as shown on the left.
  - Field emission might then occur with a few Volts between the anode and the tip, because all the field lines will have to converge on the small tip.

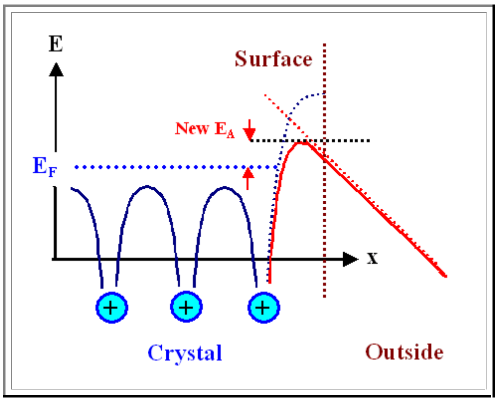
- How can we understand these effects? Whereas the *Schottky effect* is relatively straight forward, *field emission* is a manifestation of the *tunnelling effect*, a purely quantum mechanical phenomenon.
- Lets look at how the **free electron gas model** must be modified at high field strengths - and we will be able to account for *both* effects.
  - The potential energy E outside of the material is such that electrons are to be extracted - it is not constant, but varies with the field strength E simply as

$$E = e \cdot E \cdot x$$

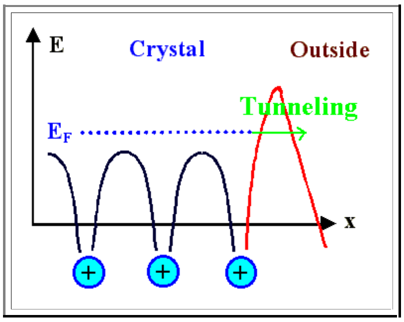
  - E, the (constant) applied field strength (written in mauve to make sure that we do not mix it up with the energy E). We have the following situation:



  - Simply summing up the energies graphically yields the qualitative energy curve for an electron at the edge of a crystal as shown below.



  - Whichever way you superimpose the potential energies, the potential barrier to the outside world will always be reduced. This explains qualitatively the *Schottky effect*.
- The *field emission effect* requires a somewhat different consideration.
  - Lets look at the *extremes* of the Schottky effect. For really high field strengths the potential barrier gets even lower and thinner, it may look somewhat like this:



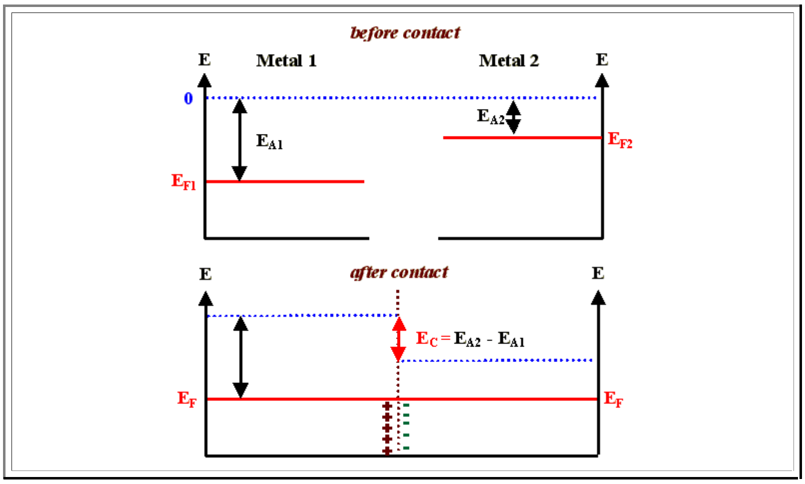
- Now the **tunneling effect** may occur. It is a phenomenon inherent in quantum mechanics and allows electron "waves" to "*tunnel*" through a potential barrier.
  - In other words, the value of the **wave function** ψ for an electron does not got to zero abruptly at a potential barrier, but decays exponentially. There is then a finite amplitude for ψ *on the other side* of the potential barrier, an effect that is felt if the barrier is "thin" and low - as in the picture above. If the field strength is high enough, large quantities of electrons can directly tunnel to the outside world. More about tunnelling in the link.
- Field emission thus is a purely quantum mechanical effect; there is no classical counterpart whatsoever. It is used in a growing number of applications:
  - **Electron microscopes** for special purposes (e.g. scanning electron microscopes with high resolution at low beam voltage, a must for the chip industry) are usually equipped with **field emission "guns"**.
  - "**Scanning Tunnelling Microscopes**" (***STM***) which are used to view surfaces with atomic resolution, directly employ tunnelling effects.
  - Large efforts are being made to construct flat panel displays with millions of miniature field emission cathodes - at least one per pixel.
  - Some semiconductor devices (e.g. the "**tunnel diode**") depend on tunnelling effects through space charge regions.
- In other contexts, tunnelling is not useful, but may *limit* what you can do. Most notorious, perhaps, is the effect that *very thin* insulators - say **5 nm** and below - are insulating no more, a growing problem for the chip industry.

Questionaire
Multiple Choice questions to 2.3.1 *and* 2.3.2

2.4.3 Thermoelectric Effects

General Consideration

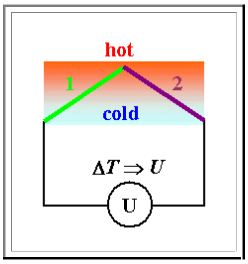
- So far we have only considered *one* conducting material; the unavoidable **contacts** between conductors, implicitly always required, were seemingly devoid of special properties.
  - We know that this *is not true* for many other contacts; e.g. combinations of
    - semiconductor - semiconductor.
    - semiconductor - conductor.
    - ionic conductor - conductor.
  - What about *metal - metal contacts*?
- We routinely solder wires of different conductors together or join them in any way, and do not worry about the contacts. Besides, maybe, a certain (usually small) **contact resistance** which is a property of the interface and must be added to the resistance of the two materials, there seems to be no other specific property of the contact.
  - But that is *only true* as long as *the temperature is constant* in the whole system of at least two conductors.
  - The reason for this is that we always get a **contact voltage**, as in the case of semiconductors, but the extension of the charged layers at the interface (the Debye lengths) is so short that no specific phenomena result from this.
  - Consider the band diagrams before and after joining two metals



- We have a dipole layer of charges at the interface which, owing to the large carrier density, is extremely thin and does not hinder current flow (it is easy for electrons to tunnel through the potential barrier).
  - We also have a *contact potential*, which is called the **Volta potential**. Since in any closed circuit (containing, e.g., the wires to the voltmeter), the *sum of the Volta potentials must be zero* in **thermal equilibrium**, it therefore can not be measured directly.
  - *If*, however, *one* of the at least two contacts needed for a closed circuit is at a temperature T_2 that is different from the temperature T_1 of the first contact, we have *non-equilibrium* and now a voltage may be measured. We observe the **Seebeck effect**, one of several **thermoelectric effects**.
- We will not go into details here (consult the link for this) but will only mention some applications and related effects.

Seebeck Effect

- The Seebeck effect is the base for **thermoelements** or **thermocouples**, the standard device for measuring temperatures (the good old mercury thermometer is virtually nonexistent in technical applications, especially at high temperatures).



  - Lets look at a typical situation: We have a thermocouple mader with a material **1** and a material **2**. It's "contacted" by whatever (material **3**, black lines). The junction of material**1** and material **2** is hot, the rest is cold (and has the same temperature).
  - The voltmeter will show a **thermovoltage** that depends on ΔT and the two materials forming the thermocouple.

  - Generally, the thermovoltage should be larger for couples of conductors with very different Fermi energies or carrier densities, since then the Volta potential is larger.
  - Being more specific, the Volta potential should follow Nernsts law. But here we are only interested in the practical aspects of thermocouples.

- For technically important materials, it is convenient to construct a voltage scale for thermocouples given in **mV/100K**.
  - The voltage measured for a temperature difference of **100 K** is then the difference of the two values given on that scale for the two materials joined in a thermocouple. The zero point was arbitrarily chosen for **Pt**.

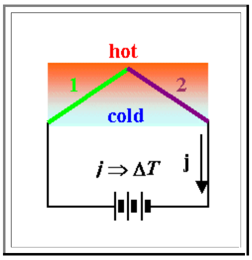
Bi	Ni	Pd	Pt	Hg	PtRh	Cu	Mo	Fe	NiCr	Sb
-7,7	-1,5	-0,3	0	0	0,7	0,77	1,2	1,92	2,6	4,8

  - Useful couples are, e.g. **Ni/NiCr**, with a thermovoltage of ca. **4 mV/100K** and a usable temperature range up to **1000 K**.

- The Seebeck effect, for many years extensively used for measuring temperatures, can also be used to convert heat energy directly into electrical energy. **Thermoelectric generators** are becoming an exciting field of materials science, because optimized materials, based on a thorough understanding of the requirements for power generation and the concomitant requirements for the materials, are becoming available.

Other Thermoelectric Effects

- There are several thermoelectrical effects which are deeply routed in **non-equilibrium thermodynamics**. Essentially, there is a "*reciprocating*" coupling of *gradients in driving forces* and *currents of any kind* (not just electrical currents but also, e.g. particle currents, heat currents, or even entropy currents).
  - *Reciprocating* means, that if a *gradient* - e.g. in the *temperature* - induces an electric *current* across a junction (the Seebeck effect), than an electric current induced by *some other means* must produce a temperature gradient. And this does *not* address the heating simply due to ohmic heating!
  - The "reversed" Seebeck effect does indeed exist, it is called the **Peltier effect**. In our schematic diagram it looks like this:



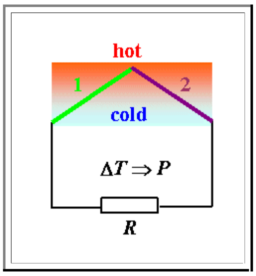
  - An electrical current, in other words, that is driven through the system by a battery, would lead to a "heat" current, transporting thermal energy from one junction to the other one. One junction then goes down in temperature, the other one goes up.
  - This effect would also occur in hypothetical materials with zero resistivities (we do not mean superconductors here). If there is some resistance R , the current will always lead to some heating of the wires everywhere which is superimposed on the Peltier effect.
  - The temperature difference ΔT between the two junctions due to the external current density j induced by the battery and the Peltier effect then is approximately given by

$$\Delta T \approx \textbf{const.} \cdot j$$

  - The removal of heat or thermal energy thus is linear with the current density
  - But there is always heating due to by ohmic losses, too. This is proportional to j^2, so it may easily overwhelm the Peltier effect and no net cooling is observed in this case.

- The Peltier effect is not useful for *heating* - that is much easier done with resistance heating - but for *cooling*!
  - With optimized materials, you can lower the temperature considerably at one junction by simply passing current through the device! The Peltier effect actually has been used for refrigerators, but now is mainly applied for controlling the temperature of specimens (e.g. chips) while measurements are being made.

- One can do a third thing with thermoelements: Generate power. You have a voltage coupled to a temperaturr difference, and that can drive a current through a load in the form of a resistor.



- Invariably the question of the efficiency η of power generation comes up. How much of the thermal energy in the system is converted to electrical energy?
- This is not easy to calculate in detail. It is, however, very easy to guess to what parameters η will be proportional:

  - η ∝ **1/κ; κ = thermal conductivity.**.
  - η ∝ σ; **σ = electrical conductivity**
  - η ∝ **S; S = Seebeck coefficient**

  If the thermal conductivity κ is large, you cannot maintain a thermal gradient for very long. If the electrical conductivity σ is small, you loose energy by heating the resistor. It is thus clear that the efficiency is proportional to **σ/κ**.
  The Seebeck coefficient **S**, finally, simply characterizes the materials to be used. Unfortunately, for most materials, a large σ implies a large κ **-** think of any metal, for example!

There is one more effect worthwhile to mention: If you have an external current *and* an external temperature gradient at the same time, you have the **Thomson effect**. But we mention that only for completeness; so far the Thomson effect does not seem to be of technical importance. Again, more information is contained in the link.

Questionaire
Multiple Choice questions to 2.3.3

2.4.4 Summary to: Conductors - Special Applications

- Thermionic emission provides electron beams.
  The electron beam current (density) is given by the *Richardson equation*:
  - A_{theo} **= 120 A · cm^{-2} · K^{-2}** for free electron gas model
    A_{exp} ≈ **(20 - 160) A · cm^{-2} · K^{-2}**
  - E_A **=** work function ≈ **(2 - >6) eV**
  - Materials of choice: **W**, **LaB$_6$** single crystal

$$j = A \cdot T^2 \cdot \exp - \frac{E_A}{kT}$$

- High field effects (tunneling, barrier lowering) allow large currents at low T from small (nm) size emitter

Needs **UHV**!

- There are several thermoelectric effects for metal junctions; always encountered in non-equilibrium.
  - *Seebeck effect:*
    Thermovoltage develops if a metal A-metal B junction is at a temperature different form the "rest", i.e. if there is a temperature gradeient

Essential for measuring (high) temperatures with a "thermoelement"
Future use for efficient conversion of heat to electricity ???

  - *Peltier effect:*
    Electrical current I through a metal - metal (or metal - semiconductor) junction induces a temperature gradient ∝ I, i.e. one of the junction may "cool down".

Used for electrical cooling of (relatively small) devices. Only big effect if electrical heating (∝ I^2) is small.

Questionaire

All Multiple Choice questions to 2.4

2.5 Ionic Conductors

2.5.1 General Remarks

- In **ionic conductors**, the current is transported by *ions* moving around (and possibly electrons and holes, too). Electrical current transport via **ions**, or *ions and electrons/holes*, is found in:
  - Conducting *liquids* called **electrolytes**.
  - **Ion conducting solids**, also called **solid electrolytes**.
- **Ionic conductivity** is important for many products:
  - Type **I** and type **II batteries** (i.e. regular and rechargeable).
  - **Fuel cells**.
  - **Electrochromic** windows and displays.
  - Solid state **sensors**, especially for reactive gases.
- In contrast to purely electronic current transport, there is *always* a chemical reaction tied to the *current flow* that takes place wherever the ionic current is converted to an electronic current - i.e. at the contacts or electrodes. There may be, however, a measurable potential difference *without* current flow in ionic systems, and therefore applications *not* involving chemical reactions.
  - This is a big difference to current flow with electrons (or holes), where no chemical reaction is needed for current flow across contacts since "chemical reactions " simply means that the system changes with time.
- If we look at the conductivity of solid ionic conductors, we look at the movement of ions in the crstal lattice - e.g. the movement (= diffusion) of **O^-** or **H^+** ions either as interstitials or as lattice ions.
  - In other words, we look at the diffusion of (ionized) atoms in some crstal lattice, described by a diffusion coefficient ***D***.
  - Since a diffusion coefficient ***D*** and a mobility **μ** describe essentially the same thing, it is small wonder that they are closely correlated - by the **Einstein-Smoluchowski relation** (the link leads you to the semiconductor Hyperscript with a derivation of the equation).

$$\boxed{\mu = \frac{e \cdot D}{kT}}$$

  - The conductivity of a solid-state ionic conductor thus becomes

$$\boxed{\sigma = e \cdot c \cdot \mu = \frac{e^2 \cdot c \cdot D}{kT} = \frac{e^2 \cdot c \cdot D_0}{kT} \cdot \exp - \frac{H^m}{kT}}$$

  - with the normal Arrhenius behaviour of the diffusion coefficient and ***H^m*** = migration enthalpy of an ion, carrying one elementary charge. In other words: we must expect complex and strongly temperature dependent behaviour; in particular if ***c*** is also a strong function of ***T***.

- **Ionics** is the topic of dedicated lecture courses, here we will only deal with two of the fundamental properties and equations - the **Debye length** and the **Nernst equation** - in a very simplified way.
  - The most general and most simple situation that we have to consider is a contact between two materials, at least one of which is a *solid ionic conductor* or solid electrolyte. Junctions with liquid electrolytes, while somewhat more complicated, essentially follow the same line of reasoning.
- Since this involves that some kind of ion can move, or, in other words, *diffuse* in the solid electrolyte, the *local concentration* ***c*** of the mobile ion can respond to two types of driving forces:
  - **1.** *Concentration gradients*, leading to particle currents j_{diff} (and, for particles with charge ***q***, automatically to an electrical current $j_{elect} = q \cdot j_{diff}$) given by **Ficks laws**

$$\boxed{j_{diff} = -D \cdot \mathrm{grad}(c)}$$

  - With ***D* = diffusion coefficient** of the diffusing particle.

- **2.** *Electrical fields* ***E***, inducing electrical current according to *Ohms law* (or whatever current - voltage - characteristics applies to the particular case), e.g.

$$j_{\text{field}} = \sigma \cdot E = q \cdot c \cdot \mu \cdot E$$

- With **µ** = **mobility** of the particle.

Both driving forces may be present *simultaneously*; the total current flow or voltage drop then results from the combined action of the two driving forces.

- Note that in one equation the current is proportional to the *gradient* of the concentration whereas in the other equation the proportionality is to the concentration *directly*. This has immediate and far reaching consequences for all cases where in equilibrium the two components must cancel each other as we will see in the next sub-chapter.

In general, the two partial currents will not be zero and some *net* current flow is observed. Under equilibrium conditions, however, there is no net current, this requires that the partial currents either are all zero, or that they must have the same magnitude (and opposite signs), so that they *cancel each other*.

- The equilibrium condition thus is

$$q \cdot j_{\text{diff}} = j_{\text{field}}$$

- The importance of this equation cannot be over emphasized. It imposes some general conditions on the *steady state concentration profile* of the diffusing ion and thus the charge density. Knowing the charge density distribution, the potential distribution can be obtained with the **Poisson equation**, and this leads to the *Debye length* and *Nernsts law* which we will discuss in the next paragraphs.

Questionaire
Multiple Choice questions to all of 2.4

2.5.2 Debye Length

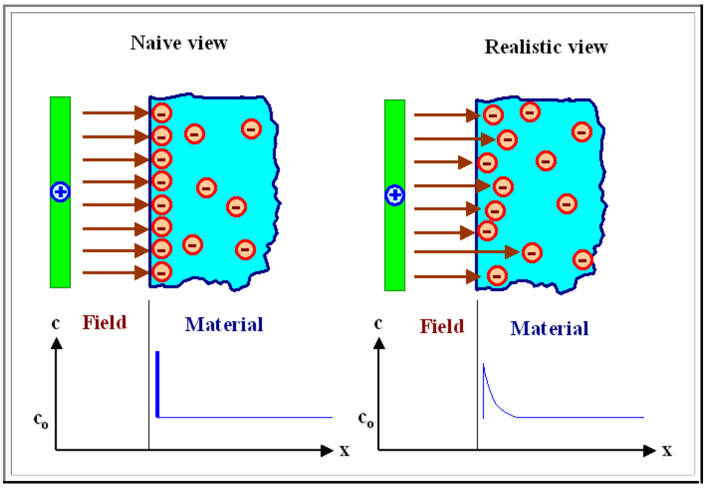
Equilibrium of Diffusion and Field Currents

- *Nernst law* is a special answer to the general and important question:

> How do *charged* and *mobile* particles redistribute themselves in an *electrical potential*
> if there are some restrictions to the obvious solution
> that they all move to one or the other pole of the field?

  - It is the answer to this question that governs not only **pn**-junctions, but also batteries, fuel cells, or gas sensors, and, if you like, simply *all junctions*.

- Let us consider a material that essentially contains mobile carriers of only *one* kind, i.e. a metal (electrons) , a (doped) semiconductor (electrons *or* holes, depending on doping), or a suitable ionic conductor (one kind of mobile ion).
  - We imagine that we hold a positively charged plate at some (small) distance to the surface of a material having mobile negative charges (a metal, a suitable ionic conductor, a **n**-doped semiconductor, ...). In other words, the positively charged plate and the material are *insulated*, and no currents of any kind can flow between the two. However, there will be an electrical field, with field lines starting at the positive charges on the plate and ending on the negative charges inside the material. We have the following situation:



- In a *naive* (and *wrong*) view, enough negatively charged carriers in the material would move to the surface to screen the field completely, i.e. prevent its penetration into the material. "Enough", to be more precise, means just the right number so that every field line originating from some charge in the positively charged plate ends on a negatively charged carrier inside the material.
  - But that would mean that the concentration of carriers at the surface would be pretty much a δ- function, or at least a function with a very steep slope. That does not seem to be physically sensible. We certainly would expect that the concentration varies smoothly within a certain distance, and this distance we call **Debye length** right away.
  - As you might know, the Debye length is a crucial material parameter not only in all questions concerning ionic conducitvity (the field of "**Ionics**"), but whenever the carrier concentration is not extremely large (i.e. comparable to the concenetration of atoms, i.e in metals).

- We will now derive a simple formula for the *Debye length*. We start from the "naive" view given above and consider its ***ramifications***:
  - If all (necessarily mobile) carriers would pile up at the interface, we would have a large concentration gradient and **Ficks law** would induce a very large *particle* current *away* from the interface, and, since the particles are charged, an *electrical* current at the same time! Since this **electrical *diffusion* current** $j_{\text{el, Diff}}$ is proportional to the concentration *gradient* **–grad ($c(x)$)**, we have:

$$\boxed{j_{\text{el, Diff}}(x) \;=\; -\,q \cdot D \cdot \text{grad}\,(c(x))}$$

  - With D = diffusion coefficient. Be clear about the fact that whenever you have a concentration gradient of mobile carriers, you will always have an electrical current by necessity. You may not notice that current because it might be cancelled by some other current, but it exists nevertheless.

- The *electrical field* $\boldsymbol{E}(x)$, that caused the concentration gradient in the first place, however, will also induce an electrical ***field* current** (also called **drift current**) $j_{\text{field}}(x)$, obeying Ohms law in the most simple case, which flows in the *opposite* direction of the electrical diffusion current. We have:

$$j_{\text{field}}(x) \;=\; q\cdot c\cdot\mu\cdot E(x)$$

- With μ = mobility, q = charge of the particle (usually a multiple of the elementary charge **e** of either sign); $q\cdot c\cdot\mu$, of course, is just the conductivity σ
- The *total* electrical current will then be the *sum* of the electrical field and diffusion current.

- In *equilibrium*, both electrical currents obviously must be *identical* in magnitude and *opposite* in sign for every x, leading for one dimension to

$$q\cdot c(x)\cdot\mu\cdot E(x) \;=\; q\cdot D\cdot\frac{dc(x)}{dx}$$

- Great, but too many unknowns. But, as we know (????), there is a relation between the diffusion coefficient D and the mobility μ that we can use; it is the **Einstein-Smoluchowski relation** (the link leads you to the semiconductor Hyperscript).

$$\mu = e\cdot D/kT$$

  - We also can substitute the electrical Field $E(x)$ by $-\,dU(x)/dx$, with $U(x)$ = potential (or, if you like, voltage) across the system. After some reshuffling we obtain

$$-\,e\,\frac{dU(x)}{dx} \;=\; \frac{kT}{c(x)}\cdot\frac{dc(x)}{dx} \;=\; kT\cdot\frac{d\,[\ln c(x)]}{dx}$$

  - We used the simple relation that $d\,(\ln c(x))\,/\,dx = 1/c(x)\cdot dc(x)/dx$. This little trick makes clear, why we always find relations between a voltage and the *logarithm* of a concentration.
  - This is a kind of basic property of ionic devices. It results from the difference of the driving forces for the two opposing currents as noted before: The diffusion current is proportional to the *gradient* of the concentration whereas the field current is directly proportional to the concentration.

- Integrating this simple differential equation once gives

$$U(x) \;+\; \frac{kT}{e}\cdot\ln c(x) \;=\; \text{const.}$$

  - Quite interesting: the sum of two functions of x must be constant for any x and for any functions conceivable; the above sum is obviously a kind of *conserved quantity*.
  - That's why we give it a name and call it the **electrochemical potential** V_{ec} (after muliplying with **e** so we have energy dimensions). While its two factors will be functions of the coordinates, its total value for any (x,y,z) coordinate in equilibrium is a *constant* (the three dimensional generalization is trivial). In other words we have

$$V_{ec} \;=\; V(x) \;+\; kT\cdot\ln c(x)$$

  - with $V(x) = e\cdot U(x)$ = electrostatic potential energy.
  - The electrochemical potential thus is a real energy like the potential energy or kinetic energy.

- Obviously, *in equilibrium* (which means that nowhere in the material do we have a *net* current flow) the *electrochemical potential must have the same value anywhere in the material*.
  - This reminds us of the **Fermi energy**. In fact, the electrochemical potential is nothing *but* the Fermi energy and the Fermi distribution in disguise.
  - However, since we are considering *classical* particles here, we get the classical approximation to the Fermi distribution which is, of course, the **Boltzmann distribution** for E_F or V_{ec}, respectively, defining the zero point of the energy scale.

- This is easy to see: Just rewriting the equation from above for $c(x)$ yields

$$c(x) = \exp - \frac{(Vx) - V_{ec}}{kT}$$

- What we have is the simple *Boltzmann distribution* for classical particles with the energy **$(Vx) - V_{electrochem}$**.

Calculating the Debye Length

- First we realize that the voltage or potential distribution (voltage times **e**) in the interior of a material *in equilibrium* can only be caused by *concentration distributions of carriers that obey equilibrium statistics*, i.e. the Boltzmann or the Fermi distribution.
  - This is simply what the equation above tells us.
- What we still need in order to calculate the Debye length is a linkage between potentials **$e \cdot U(x) = V(x)$** and concentrations **$c(x)$**.
  - This is of course what the **Poisson equation**, the main equation for electrostatics, is all about. We will only look at the one-dimensional case here. The Poisson equation than states

$$-\frac{d^2U}{dx^2} = \frac{dE}{dx} = \frac{e \cdot c(x)}{\epsilon\epsilon_0}$$

  - Now, for *good conductors* (i.e. **c(carriers) ≈ density of atoms ≈ 10^{22} cm^{-3}**), only a few of the carriers (a very small percentage) are needed to screen *any* reasonable electrical field. If you do not see this, do the exercise!

Exercise 2.5.1

Field Screening

- We may thus assume within a very good approximation that the carrier density at any point is given by the constant volume density **c_0** of the field free material, *plus* a rather small space dependent addition **$c_1(x)$**; i.e.

$$c(x) = c_0 + c_1(x)$$

  - Obviously, only **$c_1(x)$** is important for Poissons equation.
- From Boltzmanns distribution we know that

$$\frac{c(x)}{c_0} = 1 + \frac{c_1(x)}{c_0} = \exp\left(-\frac{\Delta(energy)}{kT}\right) = \exp\left(-\frac{V(x)}{kT}\right)$$

  - because the difference in energy of a carrier in the field free volume (i.e. where we have **c_0**) is simply the electrostatic energy associated with the electrical field.
  - Since we assumed **$c_1 \ll c_0$**, we may with ***impunity*** express the exponential function as a *Taylor series* of which we only retain the first term, obtaining:

$$1 + \frac{c_1(x)}{c_0} \approx 1 + \frac{V(x)}{kT}$$

$$c_1(x) = c_0 \cdot \frac{V(x)}{kT}$$

- This is a simple trick, but important. Feeding the result back into Poissons equation yields:

$$\frac{d^2\,[c_1(x)]}{dx^2} = \frac{e^2 \cdot c_0 \cdot c_1(x)}{\epsilon \cdot \epsilon_0 \cdot kT}$$

- For a simple one-dimensional case with a surface at ***x* = 0** we obtain the final solution

$$c_1(x) = c_1(x = 0) \cdot \exp - \frac{x}{d}$$

- The quantity ***d*** is the Debye length we were after, it is obviously given by

$$d = \text{Debye length} = \left(\frac{\epsilon \cdot \epsilon_0 \cdot kT}{e^2 \cdot c_0}\right)^{1/2}$$

- The *Debye length* is sometimes also called Debye-**Hückel** length (which is historically correct and just).
  - $c_1(x = 0)$, of course, is given by the boundary condition, which for our simple case is:

$$c_1\,(x = 0) = c_0 \cdot \frac{V\,(x = 0)}{kT}$$

- What is the meaning of the Debye length? Well, generalizing a bit, we look at the general case of a material having some *surplus charge* at a definite position somewhere in a material
  - Consider for example the phase boundary of a (charged) precipitate, a charged grain boundary in some crystal, or simply a (point) charge somehow held at a fixed position *somewhere* in *some* material. The treatment would be quite similar to the one-dimensional case given here.
- What we know now is quite important:
  - If you are some Debye lengths away from these fixed charges, you will not "see" them anymore; their effect on the equilibrium carrier distribution then is vanishingly small.
  - The Debye length resulting in *any* one of these situations thus is nothing but the *typical distance* needed for **screening** the surplus charge by the mobile carriers present in the material.
  - In other words, after you moved about one Debye length away from the surplus charge, its effects on the mobile charges of the material are no longer felt.
- More about the Debye length can be found in the Hyperscript "Semiconductors".

Questionaire
Multiple Choice questions to all of 2.5

2.5.3 Nernst's Equation

- **Nernst's equation** gives the *voltage* between two materials in close contact, i.e. the *potential difference* between the two materials. From the foregoing discussion, we know already two important facts about this potential:
  - It will change from one value to the other over a distance across the junction that is given by the (two) **Debye lengths** of the system.
  - The corresponding carrier concentrations are equilibrium concentrations and thus governed by the **Boltzmann distribution** (considering only classical particles at this point).
- If the potential difference is ΔU, we thus, using the Boltzmann distribution, obtain for the concentration of the carriers c_1 in material **1**, and c_2 in material **2**:

$$\frac{c_1}{c_2} = \exp - \frac{e \cdot \Delta U}{kT}$$

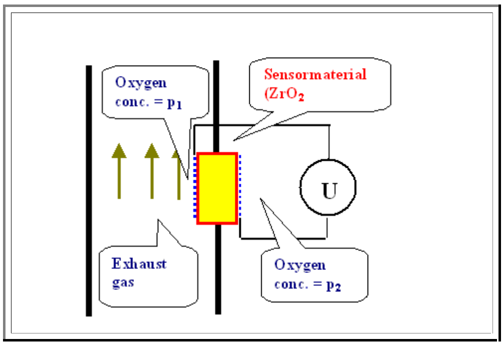
  - This is already *Nernst's equation* (or law) - *in a somewhat unusual way of writing*.
- Usually we (and everybody else) use the Boltzmann distribution to compute *concentrations* as a function of some other *known* parameters - the energy in this case. But this is *not* the only way for using a general equation!
  - Like any equation, it also works in *in reverse*: If we *know* the concentrations, we can calculate the energy difference that must go with them!
  - The important point now is that the concentrations of electrons in metals, but also of ions in ionic conductors, or holes in semiconductors, or any mobile carrier *a few Debye lengths away from the junction*, are fixed - there is no need to compute them!
  - What is *not fixed* is the *potential difference* $e \cdot \Delta U$ *a few Debye lengths away from the junction*, and that is what we now can obtain from the above equation by rewriting it for ΔU:

$$\Delta U = - \frac{kT}{e} \cdot \ln \frac{c_1}{c_2}$$

- This is **Nernst's equation** in its usual, but somewhat simplified form. We may briefly consider two complications:
  - **1.** If the particles carry z elementary charges, the first factor will now obviously write $kT/(z \cdot e)$.
  - **2.** If the *interaction* between particles is *not* negligible (which would mean, e.g., that Ficks law in its simple form would not be usable), the concentrations have to be replaced by the **activities** a of the particles.
  - If you want to know in detail what activities are - use the link. But all you have to know at this point is that activities are the particle concentrations *corrected for interaction effects*. To give an example: If a particle concentration is 10^{19} cm^{-3}, the activity might only be $5 \cdot 10^{18}$ cm^{-3}. If you use this factually wrong number, simple equations like the Boltzmann distribution that do not take into account particle interactions can still be used.
  - If the activity numbers are very different from the real concentration numbers, you are no longer doing Materials Science, but chemistry.
  - Using this, we obtain the *general* version of Nernst's law

$$\Delta U = - \frac{kT}{z \cdot e} \cdot \ln \frac{a_1}{a_2}$$

- Nernsts law, being the Boltzmann distribution in disguise, is of course extremely general. It gives the potential difference and thus the voltage of *any* contact between two materials that have sufficiently large concentrations of mobile carriers so that an equilibrium distribution can develop. It describes, among other things
  - The **contact voltage** (**Volta potential**) between two metals (i.e. **thermocouples**).
  - The built-in potential in **pn-junctions**
  - The voltage of any battery or accumulator.
  - The voltage of fuel cells.
  - The voltage produced by certain kinds of sensors.
- The last issue ***merits*** some brief explanation. Let's assume a material with a sufficiently large concentration of mobile **O^-** ions at interstitial sites (in other word, mobile interstitial point defects) at the working temperature - take **Y_2O_3** stabilized **ZrO_2** as an example (whatever that may be).
  - Use it to measure the amount of oxygen in a given gas mixture with the following **oxygen sensor** device:



- The sensor material experiences two different oxygen concentrations on its two surfaces, one of which is known (oxygen in air, a constant for all practical purposes), the other one is the concentration in the exhaust gas of a car which is supposed to be measured by the voltmeter
  - Two gas-permeable electrodes have been supplied which allow oxygen on both sides to react with the sensor material.
- In equilibrium, we will have some reaction between the oxygen in the gas and the oxygen in the crystal in the sense that oxygen will either come out, or diffuse into the material.
  - What we might expect is that the concentration of interstitial oxygen in the crystal will be larger near to the surface with the large oxygen gas concentration (air) compared to the surface exposed to a lower oxygen concentration (exhaust).
  - The gradient in the (negatively charged) oxygen concentration inside the material then will be determined by the **Debye length** of the system (in the real thing, which is **ZrO_2**, it will be just a few **nm**).
  - In total, the concentration **$[O]_S$** of mobile **O**-interstitials right at the surface will be *somehow* tied to the partial pressure **p_O** of the oxygen on both sides; lets say we have a general relation like

$$[O]_S = \left(\text{const.} \cdot p_O\right)^n$$

  - But any other (reasonable) relation you can think of will be just as good.
- Nernst's law then tells us immediately, how the voltage between the two electrodes depends on the oxygen concentration or partial pressure in the exhaust: For the assumed relation we have

$$\Delta U = -\frac{kT}{e} \cdot \ln \frac{c_1}{c_2}$$

$$\Delta U = -\frac{kT}{e} \cdot \ln \frac{(p_1)^n}{(p_2)^n}$$

$$\Delta U = -\frac{n \cdot kT}{e} \cdot \ln \frac{p_1}{p_2}$$

- This is quite remarkable: We have an equation for the voltage that develops across *some* sensor as a function of the difference of the oxygen concentration on two sides of the sensor *without knowing much* about the details of the sensor! All we have to assume is that there is *some* mobile **O^{-}**, *no* other free carriers, and that establishing equilibrium does not take forever.
- Only if you want to know the *precise* value of ***n*** do you have to delve into the detailed reactions at the interfaces.

This is essentially the working principle of not only the oxygen sensor in the exhaust system of any modern car ("**λ - Sonde**"), but of most, if not all, solid state sensors.

Questionaire
Multiple Choice questions to all of 2.4

2.5.4 Summary to: Ionic Conductors

- Electrical current can conducted by *ions* in
  - Liquid electrolytes (like $\mathbf{H_2SO_4}$ in your "lead - acid" car battery); including gels
  - Solid electrolytes (= ion-conducting crystals). Mandatory for fuel cells and sensors
  - Ion beams. Used in (expensive) machinery for "nanoprocessing".

Challenge: Find / design a material with a "good" ion conductivity at room temperature

- Basic principle
  - *Diffusion current* j_{diff} driven by concentration gradients **grad(*c*)** of the charged particles (= ions here) equilibrates with the
  - *Field current* j_{field} caused by the internal field always associated to concentration gradients of charged particles plus the field coming from the outside
  - Diffusion coefficient ***D*** and mobility **μ** are linked via theEinstein relation; concentration $c(x)$ and potential $U(x)$ or field $E(x) = -\mathrm{d}U/\mathrm{d}x$by the Poisson equation.

$$j_{\text{diff}} = -D \cdot \text{grad}(c)$$

$$j_{\text{field}} = \sigma \cdot E = q \cdot c \cdot \mu \cdot E$$

$$\mu = eD/kT$$

$$-\frac{\mathrm{d}^2U}{\mathrm{d}x^2} = \frac{\mathrm{d}E}{\mathrm{d}x} = \frac{e \cdot c(x)}{\epsilon\epsilon_0}$$

- Immediate results of the equations from above are:
  - In equilibrium we find a preserved quantity, i.e. a quantity independent of ***x*** - the electrochemical potential V_{ec}:

$$V_{ec} = \text{const.} = e \cdot U(x) + kT \cdot \ln c(x)$$

  - If you rewrite the equaiton for $c(x)$, it simply asserts that the particles are distributed on the energy scale according to the Boltzmann distrubution:

$$c(x) = \exp - \frac{(Vx) - V_{ec}}{kT}$$

  - Electrical field *gradients* and concentration *gradients* at "contacts" are coupled and non-zero on a length scale given by the **Debye length** d_{Debye}
  - The Debye length is an extremely important material parameter in "**ionics**" (akin to the space charge region width in semiconductors); it depends on temperature ***T*** and in particular on the (bulk) concentration c_0 of the (ionic) carriers.
  - The Debye length is not an important material parameter in metals since it is so small that it doesn't matter much.

$$d_{\text{Debye}} = \left(\frac{\epsilon \cdot \epsilon_0 \cdot kT}{e^2 \cdot c_0}\right)^{1/2}$$

The potential difference between two materials (her ionic conductors) in close contact thus...

- ... extends over a length given (approximately) by :

$$d_{\text{Debye}}(1) + d_{\text{Debye}}(2)$$

- ... is directly given by the Boltzmann distribution written for the energy:
  (with the c_i =equilibrium conc. far away from the contact.

- The famous *Nernst equation*, fundamental to ionics, is thus just the Boltzmann distribution in disguise!

$\dfrac{c_1}{c_2} = \exp - \dfrac{e \cdot \Delta U}{kT}$	**Boltz-mann**
$\Delta U = - \dfrac{kT}{e} \cdot \ln \dfrac{c_1}{c_2}$	**Nernst's equation**

"Ionic" sensors (most famous the **ZrO_2** - based **O_2** sensor in your car exhaust system) produce a voltage according to the Nernst equation because the concentration of ions on the exposed side depends somehow on the concentration of the species to be measured.

Questionaire

Multiple Choice questions to all of 2.5

2.6 Summary: Conductors

What counts are the *specific* quantities:

- Conductivity σ (or the specific resistivity $\rho = 1/\sigma$
- current density $\underline{j}$
- (Electrical) field strength · $\underline{E}$

$$[\sigma] = (\Omega\mathrm{m})^{-1} = \mathrm{S/m};\ \mathrm{S} = 1/\Omega = \text{"Siemens"}$$
$$[\rho] = \Omega\mathrm{m}$$

- The basic equation for σ is:
  n = concentration of carriers
  μ = mobility of carriers

$$\sigma = |q| \cdot n \cdot \mu$$

- Ohm's law states:
  It is valid for metals, but not for all materials

$$\underline{j} = \sigma \cdot \underline{E}$$

σ (of conductors / metals) obeys (more or less) several rules; all understandable by looking at n and particularly μ.

- Matthiesen rule
  Reason: Scattering of electrons at defects (including phonons) decreases μ.

$$\rho = \rho_{\mathrm{Lattice}}(T) + \rho_{\mathrm{defect}}(N)$$

- "$\rho(T)$ rule":
  about **0,04 %** increase in resistivity per **K**
  Reason: Scattering of electrons at phonons decreases μ

$$\Delta\rho = \alpha_\rho \cdot \rho \cdot \Delta T \approx \frac{0{,}4\%}{^\circ\mathrm{C}}$$

- Nordheim's rule:
  Reason: Scattering of electrons at **B** atoms decreases μ

$$\rho \approx \rho_A + \mathrm{const.} \cdot [B]$$

Major consequence: You can't beat the conductivity of pure **Ag** by "tricks" like alloying or by using other materials. (Not considering superconductors).

Non-metallic conductors are *extremely* important.

- Transparent conductors (TCO's)
  ("ITO", typically oxides)

No flat panels displays = no notebooks etc. without **ITO**!

- Ionic conductors (liquid and solid)

Batteries, fuel cells, sensors, ...

- Conductors for high temperature applications; corrosive environments, ..
  (Graphite, Silicides, Nitrides, ...)

Example: **$MoSi_2$** for heating elements in corrosive environments (dishwasher!).

- Organic conductors (and semiconductors)

The future High-Tech key materials?

Numbers to know (order of magnitude accuracy sufficient)

ρ(decent metals) about **2 μΩcm**
ρ(technical semiconductors) around **1 Ωcm**
ρ(insulators) > 1 GΩcm

- No electrical engineering without conductors!
  - Hundreds of specialized metal alloys exist just for "wires" because besides σ, other demands must be met, too:

Money, Chemistry (try **Na**!), Mechanical and Thermal properties, Compatibility with other materials, Compatibility with production technologies, ...

  - Example for unexpected conductors being "best" compromise:

Poly Si, Silicides, **TiN**, **W** in integrated circuits

- Don't forget Special Applications:

Contacts (switches, plugs, ...);
Resistors;
Heating elements; ...

- Thermionic emission provides electron beams.
The electron beam current (density) is given by the *Richardson equation*:
  - $\boldsymbol{A_{theo} = 120\ A \cdot cm^{-2} \cdot K^{-2}}$ for free electron gas model
    $\boldsymbol{A_{exp} \approx (20 - 160)\ A \cdot cm^{-2} \cdot K^{-2}}$
  - $\boldsymbol{E_A}$ **=** work function $\approx$ **(2 - >6) eV**
  - Materials of choice: **W**, **LaB$_6$** single crystal

$$j = A \cdot T^2 \cdot \exp - \frac{E_A}{kT}$$

- High field effects (tunneling, barrier lowering) allow large currents at low ***T*** from small (nm) size emitter

Needs **UHV**!

- There are several thermoelectric effects for metal junctions; always encountered in non-equilibrium.
  - *Seebeck effect:*
    Thermovoltage develops if a metal A-metal B junction is at a temperature different form the "rest", i.e. if there is a temperature gradeient

Essential for measuring (high) temperatures with a "thermoelement"
Future use for efficient conversion of heat to electricity ???

  - *Peltier effect:*
    Electrical current I through a metal - metal (or metal - semiconductor) junction induces a temperature gradient $\propto I$, i.e. one of the junction may "cool down".

Used for electrical cooling of (relatively small) devices. Only big effect if electrical heating ($\propto I^2$) is small.

- Electrical current can conducted by *ions* in
  - Liquid electrolytes (like **H$_2$SO$_4$** in your "lead - acid" car battery); including gels
  - Solid electrolytes (= ion-conducting crystals). Mandatory for fuel cells and sensors
  - Ion beams. Used in (expensive) machinery for "nanoprocessing".

Challenge: Find / design a material with a "good" ion conductivity at room temperature

- Basic principle
  - *Diffusion current* $\boldsymbol{j_{diff}}$ driven by concentration gradients **grad(*c*)** of the charged particles (= ions here) equilibrates with the
  - *Field current* $\boldsymbol{j_{field}}$ caused by the internal field always associated to concentration gradients of charged particles plus the field coming from the outside

$$j_{diff} = -D \cdot \text{grad}(c)$$

$$j_{field} = \sigma \cdot E = q \cdot c \cdot \mu \cdot E$$

- Diffusion coefficient D and mobility μ are linked via theEinstein relation;
  concentration $c(x)$ and potential $U(x)$ or field $E(x) = -dU/dx$by the Poisson equation.

$$\mu = eD/kT$$

$$-\frac{d^2U}{dx^2} = \frac{dE}{dx} = \frac{e \cdot c(x)}{\epsilon\epsilon_0}$$

- Immediate results of the equations from above are:
  - In equilibrium we find a preserved quantity, i.e. a quantity independent of x - the electrochemical potential V_{ec}:

$$V_{ec} = \text{const.} = e \cdot U(x) + kT \cdot \ln c(x)$$

  - If you rewrite the equaiton for $c(x)$, it simply asserts that the particles are distributed on the energy scale according to the Boltzmann distrubution:

$$c(x) = \exp -\frac{(Vx) - V_{ec}}{kT}$$

  - Electrical field *gradients* and concentration *gradients* at "contacts" are coupled and non-zero on a length scale given by the **Debye length** d_{Debye}

$$d_{Debye} = \left(\frac{\epsilon \cdot \epsilon_0 \cdot kT}{e^2 \cdot c_0}\right)^{1/2}$$

  - The Debye length is an extremely important material parameter in "**ionics**" (akin to the space charge region width in semiconductors); it depends on temperature T and in particular on the (bulk) concentration c_0 of the (ionic) carriers.
  - The Debye length is not an important material parameter in metals since it is so small that it doesn't matter much.

The potential difference between two materials (her ionic conductors) in close contact thus...

- ... extends over a length given (approximately) by :

$$d_{Debye}(1) + d_{Debye}(2)$$

- ... is directly given by the Boltzmann distribution written for the energy:
  (with the c_i =equilibrium conc. far away from the contact.

$$\frac{c_1}{c_2} = \exp -\frac{e \cdot \Delta U}{kT}$$

Boltzmann

- The famous *Nernst equation*, fundamental to ionics, is thus just the Boltzmann distribution in disguise!

$$\Delta U = -\frac{kT}{e} \cdot \ln \frac{c_1}{c_2}$$

Nernst's equation

- "Ionic" sensors (most famous the ZrO_2 - based O_2 sensor in your car exhaust system) produce a voltage according to the Nernst equation because the concentration of ions on the exposed side depends somehow on the concentration of the species to be measured.

Questionaire

All multiple choice questions zu 2. Conductors

3. Dielectrics

3.6.3 Summary to: Special Dielectrics

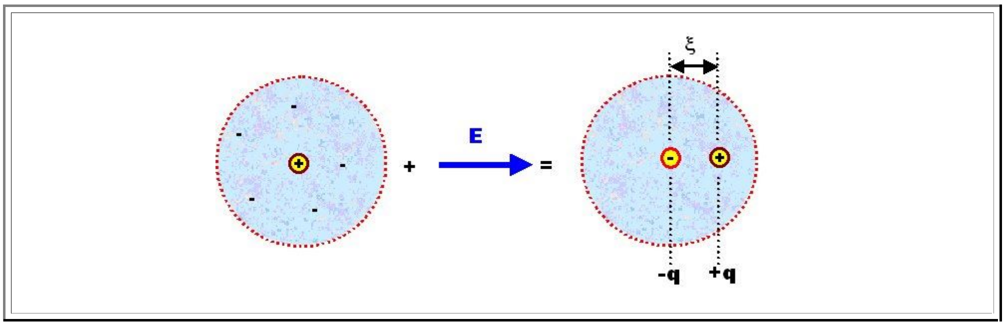
3.7 Summary: Dielectrics

3. Dielectrics

3.1 Definitions and General Relations

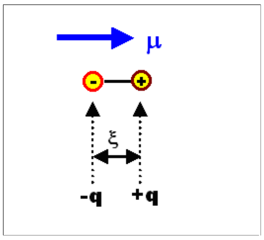
3.1.1 Polarization and Dielectric Constant

- For the sake of simplicity we assume that **dielectric materials** are *perfect* ***insulators***. In other words, there are no mobile charged particles.
- We want now to be able to answer *three questions*:
  - **1.** Given the atomic structure of the material - What is its **dielectric constant** (***DK***)?
  - **2.** How does the **DK** depend on the *frequency* of the external field?
  - **3.** How large is the *maximum field strength* a ***dielectric*** can take? Remember, no material can take arbitrarily large loads - mechanical, electical, whatever.
- For starters, we look at some general descriptions, definitions and general relations of the quantities necessary in this context.
- The dielectric constant of solids is an interesting material parameter only if the material is exposed to an electrical field (and this includes the electrical field of an **electromagnetic wave**). The effect of the *electrical field* (or just *field* for short from now on) can be twofold:
  - **1.** It *induces electrical dipoles* in the material and tries to align them in the field direction. In other words, *with* a field, dipoles come into being that do not exist *without* a field.
  - **2.** It tries to *align dipoles* that are already present in the material. In other words, the material *contains electric dipoles even without a field*.
- Of course we also may have a combination of both effects: The electrical field may change the distribution of existing dipoles while trying to align them, *and* it may generate new dipoles in addition.
- The *total* effect of an electrical field on a dielectric material is called the **polarization** of the material.
- To understand that better, lets look at the most simple object we have: A single *atom* (we do not even consider molecules at this point).
  - We have a positively charged nucleus and the electron "cloud". The smeared-out negative charge associated with the electron cloud can be averaged in space and time, and its charge center of gravity than will be at a point in space that coincides *exactly* with the location of the nucleus, because we must have spherical symmetry for atoms.
  - If we now apply an electrical field, the centers of charge will be separated. The electron cloud will be pulled in the direction of the positive pole of the field, the nucleus to the negative one. We may visualize that (ridiculously exaggerated) as follows:



  - The center of the positive and negative charges ***q* (= *z* · e)** are now separated by a distance ξ, and we thus induced a **dipole moment** μ which is defined by

$$\underline{\mu} = q \cdot \underline{\xi}$$



  - It is important to understand that $\underline{\mu}$ is a *vector* because $\underline{\xi}$ is a vector. The way we define it, its *tip will always point towards the positive charge*. For schematic drawings we simply draw a little arrow for $\underline{\mu}$.
- The magnitude of this *induced* ***dipole moment*** is a property of our particular atom, or, if we generalize somewhat, of the "particles" or building blocks of the material we are studying.

- In order to describe the ***bulk*** material - the sum of the particles - we *sum up all individual dipole moments* contained in the given volume of the material and divide this sum by the volume ***V***. This gives us the (volume independent) **polarization *P*** of the material. *Notice that we have a vector sum!*

$$\underline{P} = \frac{\Sigma\,\underline{\mu}}{V} = \langle\underline{\mu}\rangle \cdot N_V$$

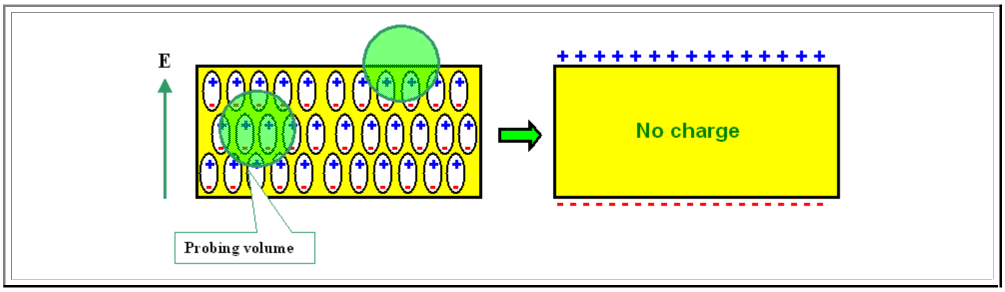
- With $\langle\underline{\mu}\rangle$ = average vector dipole moment; N_V = density of dipoles (per $\mathbf{m^3}$).
- $\underline{P}$ thus points from the *negative* to the *positive* charge, too - a convention *opposite* to that used for the electrical field.
- The physical dimension of the polarization thus is ***C*/m²**; (Coulomb per square meter). i.e. the polarization has the dimension of an *area charge*, and since $\underline{\mu}$ is a vector, $\underline{P}$ is a vector, too.

It is important to realize that a polarization $\underline{P}$ = **0** does *not* mean that the material does *not* contain dipole moments, but only that the vector sum of all dipole moments is zero.

- This will always be the case if the dipole moment vectors are *randomly distributed* with respect to their directions. Look at the picture in one of the next subchapters if you have problems visualizing this. But it will also happen if there is an ordered distribution with pairs of opposing dipole moments; again contemplate a picture in one of the following subchapters if that statment is not directly obvious.

That $\underline{P}$ has the dimension of **C/cm²**, i.e. that of an area charge, is *not* accidental but has an immediate interpretation.

- To see this, let us consider a simple plate capacitor or condenser with a *homogeneously polarized material* inside its plates. More generally, this describes an isotropic dielectric slab of material in a homogeneous electrical field. We have the following idealized situation:



- For sake of simplicity, all dipole moments have the same direction, but the subsequent reasoning will not change if there is only an average component of $\underline{P}$ in field direction. If we want to know the *charge density* ρ inside a small probing volume, it is clearly zero in the volume of the material (if averaged over probing volumes slightly larger than the atomic size), because there are just as many positive as negative charges.
- We are thus left with the *surfaces*, where there is indeed some charge as indicated in the schematic drawing. At one surface, the charges have effectively moved out a distance ξ , at the other surface they moved in by the same amount. We thus have a **surface charge**, called a *surface polarization charge*, that we can calculate.
- The number N_c of charges appearing at a surface with area ***A*** is equal to the number of dipoles contained in the surface "volume" $V_S = A \cdot \xi$ times the relevant charge **q** of the dipole. Using ξ to define the volume makes sure that we only have *one* layer of dipoles in the volume considered.
- Since we assume a homogeneous dipole distribution, we have the same polarization in any volume and thus $\underline{P} = \Sigma\underline{\mu} / V = \Sigma\underline{\mu}_S / V_S$ obtains. Therefore we can write

$$\underline{P} = \frac{\Sigma_V\ \underline{\mu}}{V} = \frac{\Sigma_S\ \underline{\mu}}{V_S} = \frac{\xi \cdot \Sigma_S\, q}{V_S} = \frac{\xi \cdot \Sigma_S\, q}{\xi \cdot A} = \frac{\Sigma_S\, q}{A}$$

$$\Sigma_S\, q = N_c = P \cdot A$$

- Σ_V or Σ_S denotes that the summation covers the total volume or the "surface" volume. Somewhere we "lost" the vector property of $\underline{P}$, but that only happens because we automatically recognize ξ as being perpendicular to the surface in question.
- *While this is a certain degree of **sloppiness**, it makes life much easier and we will start to drop the underlining of the vectors from now on whenever it is sufficiently clear what is meant.*
- The **area density** σ_{pol} of the charge on the surface is then

$$\sigma_{pol} = \frac{N_c}{A} = \frac{\underline{P} \cdot \underline{A}}{A} = |\underline{P}|$$

  - Of course, σ_{pol} is a *scalar*, which we obtain if we consider $\underline{P} \cdot \underline{A}$ to be the scalar product of the vector $\underline{P}$ and the vector $\underline{A}$; the ***latter*** being perpendicular to the surface A with magnitude $|\underline{A}| = A$.
- In purely electrical terms we thus can always replace a material with a *homogeneous* polarization $\underline{P}$ by two surfaces perpendicular to some direction, - lets say z - with a surface charge density of P_z (with, of course, different signs on the two different surfaces).
  - If the polarization vector is *not* perpendicular to the surface we chose, we must take the *component* of the polarization vector *parallel to the normal vector* of the surface considered . This is automatically taken care of if we use the vector formulation for $\underline{A}$.
- A dielectric material now quite generally reacts to the presence of an electrical field by becoming polarized and this is expressed by the direction and magnitude of $\underline{P}$.
  - $\underline{P}$ is a *measurable quantity* tied to the specific material under investigation. We now need a *material law* that connects cause and effect, i.e. a relation between the the electrical field causing the polarization and the amount of polarization produced.
  - Finding this law is of course the task of basic theory. But long before the proper theory was found, experiments supplied a simple and rather (but not quite) *empirical* "law":
- If we *measure* the polarization of a material, we usually find a linear relationship between the applied field E and P, i.e.

$$\underline{P} = \epsilon_0 \cdot \chi \cdot \underline{E}$$

  - With the proportionality constant chosen to contain ϵ_0, the **permittivity constant** (of vacuum), times a material parameter χ ("kee"), the **dielectric susceptibility**.
  - Note that including ϵ_0 in the relation is a convention which is useful in the **SI** system, where charges are always coupled to electrical fields via ϵ_0. There are other systems, however, (usually variants of the **cgs** system), which are still used by many and provide an unending source of confusion and error.
  - This equation is to dielectric material what **Ohms** law is to conductors. It is no more a real "*law of nature*" than Ohms law, but a description of many experimental observations for which we will find deeper reasons forthwith.
- Our task thus is to *calculate* χ from basic material parameters.

Connection between the Polarization $\underline{P}$ and the Electrical Displacement $\underline{D}$

- Next we need the connection between the polarization P, or the dielectric susceptibility χ, with some older quantities often used in connection with **Maxwells equations**.
  - Historically, *inside materials*, the electrical field strength E was (and still is) replaced by a vector $\underline{D}$ called the **electrical displacement** or **electrical flux density**, which is defined as

$$\underline{D} = \epsilon_r \cdot \epsilon_0 \cdot \underline{E}$$

  - and ϵ_r was (and still is) called the **(relative) dielectric constant** (**DK**) of the material (the product $\epsilon_r \cdot \epsilon_0$ is called the **permittivity**).
  - Note that in the English literature often the abbreviation κ ("Kappa") is used; in proper microelectronics slang one than talks of "low **k** materials" (pronounced "low khe" as in (O)K) when one actually means "low kappa" or "low epsilon relative".
  - D is supposed to give the "acting" *flux* inside the material.
- While this was a smart thing to do for Maxwell and his contemporaries, who couldn't know anything about materials (atoms had not been "invented" then); it is a bit unfortunate in retrospect because the *basic quantity* is the *polarization*, based on the elementary dipoles in the material, and the *material parameter* χ describing this - and not some changed "electrical flux density" and the relative dielectric constant of the material.
  - It is, however, easy (if slightly confusing) to make the necessary connections. This is most easily done by looking at a simple plate capacitor. A full treatise is found in a basic module, here we just give the results.
- The ***electric displacement*** $\underline{D}$ in a dielectric caused by some external field $\underline{E}_{ex}$ is the displacement D_0 in vacuum plus the *polarization* $\underline{P}$ of the material, i.e.

$$\underline{D} = \underline{D_0} + \underline{P} = \epsilon_0 \cdot \underline{E} + \underline{P}$$

- Inserting everything we see that the relative dielectric constant ϵ_r is simply the *dielectric susceptibility* χ plus **1**.

$$\epsilon_r = 1 + \chi$$

- For this "translations" we have used the relation $P = \epsilon_0 \cdot \chi \cdot E$, which is *not* an ***a priori*** law of nature, but an empirical relation. However, we are going to prove this relation for specific, but very general cases forthwith and thus justify the equations above.
  - We have also simply implied that $\underline{P}$ is parallel to $\underline{E}$, which is only reasonable for *isotropic materials*.
  - In *anisotropic* media, e.g. non-cubic crystals, $\underline{P}$ does not have to be parallel to $\underline{E}$, the scalar quantities ϵ_r and χ then are **tensors**. As before in the case of the electrical conductivity, the polarization and the dielectric displacement then are no longer parallel to the electrical field
  - Note right now that this is not an exotic curiosty but a major fact of life for Material Scientists and Engineers. Major industries growing rapidly at present, are based on the anisotropy of the dielctric behavior.
- The basic task in the materials science of dielectrics is now to calculate (the tensor) χ from "first principles", i.e. from basic structural knowledge of the material considered. This we will do in the following paragraphs.

Questionaire

Multiple Choice questions to 3.1.1

3.1.2 Summary to: Polarization and Dielectric Constant

The dielectric constant ϵ_r "somehow" describes the interaction of dielectric (i.e. more or less insulating) materials and electrical fields; e.g. via the equations ⇒

- $\underline{D}$ is the **electrical displacement** or **electrical flux density**, sort of replacing $\underline{E}$ in the Maxwell equations whenever materials are encountered.
- C is the capacity of a parallel plate capacitor (plate area A, distance d) that is "filled" with a dielectric with ϵ_r
- n is the index of refraction; a quantity that "somehow" describes how electromagnetic fields with extremely high frequency interact with matter.
  in this equaiton it is assumed that the material has no magnetic properties at the frequency of light.

$$\underline{D} = \epsilon_0 \cdot \epsilon_r \cdot \underline{E}$$

$$C = \frac{\epsilon_0 \cdot \epsilon_r \cdot A}{d}$$

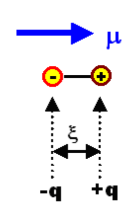
$$n = (\epsilon_r)^{1/2}$$

Electrical fields inside dielectrics polarize the material, meaning that the vector sum of electrical dipoles inside the material is no longer zero.

- The decisive quantities are the dipole moment $\underline{\mu}$, a vector, and the Polarization P, a vector, too.
- Note: The dipole moment vector points from the negative to the positive charge - contrary to the electrical field vector!
- The dipoles to be polarized are either already present in the material (e.g. in H_2O or in ionic crystals) or are induced by the electrical field (e.g. in single atoms or covalently bonded crystals like **Si**)
- The dimension of the polarization $\underline{P}$ is **[C/cm²]** and is indeed identical to the net charge found on unit area ion the surface of a polarized dielectric.

$$\underline{\mu} = q \cdot \underline{\xi}$$

$$\underline{P} = \frac{\Sigma \underline{\mu}}{V}$$



The equivalent of "Ohm's law", linking current density to field strength in conductors is the Polarization law:

- The decisive material parameter is χ ("kee"), the **dielectric susceptibility**
- The "classical" flux density D and the Polarization are linked as shown. In essence, $\underline{P}$ only considers what happens in the material, while D looks at the total effect: material plus the field that induces the polarization.

$$\underline{P} = \epsilon_0 \cdot \chi \cdot \underline{E}$$

$$\epsilon_r = 1 + \chi$$

$$\underline{D} = \underline{D}_0 + \underline{P} = \epsilon_0 \cdot \underline{E} + \underline{P}$$

Polarization by necessity moves masses (electrons and / or atoms) around, this will not happen arbitrarily fast.

- ϵ_r or χ thus must be functions of the frequency of the applied electrical field, and we want to consider the whole frequency range from **RF** via **HF** to light and beyond.

$\epsilon_r(\omega)$ is called the "**dielectric function**" of the material.

The tasks are:

- Identify and (quantitatively) describe the major mechanisms of polarization.
- Justify the assumed linear relationship between P and χ.
- Derive the dielectric function for a given material.

Questionnaire

Multiple Choice questions to all of 3.1

3.2 Mechanisms of Polarization

3.2.1 General Remarks

- We have a material and we want to know its dielectric constant ϵ_r or dielectric susceptibility χ. We would want to have those quantities as functions of various variables for the same basic materials, e.g.
  - $\chi = \chi(\omega)$,
    i.e. χ as a function of the ***angular frequency*** ω of the electrical field.
  - $\chi = \chi(T)$;
    i.e. the dependence on the temperature ***T***.
  - **$\chi = \chi$(structure)**,
    i.e. the dependence of χ on the structure of a material including the kind and density of defects in the material. As an example we may ask how χ differs from *amorphous* to *crystalline* quartz (**SiO_2**).
- The answers to all of these questions must be contained in the *mechanisms* with which atoms and molecules respond to an electrical field, i.e. in the mechanisms leading to the formation and/or orientation of dipoles. These mechanisms are called **polarization mechanisms**.
  - We want a general *theory of polarization*. This is a complex task as well it must be, given the ***plethora*** of dielectric phenomena. However, the *basic principles* are rather simple, and we are only going to look at these.
- There are essentially *four basic kinds of polarization mechanisms*:
  - **Interface polarization**.
    Surfaces, grain boundaries, interphase boundaries (including the surface of precipitates) may be *charged*, i.e. they contain dipoles which may become oriented to some degree in an external field and thus contribute to the polarization of the material.
  - ***Electronic polarization,***
    also called atom or **atomic polarization**. An electrical field will always displace the center of charge of the electrons with respect to the nucleus and thus induce a dipole moment as discussed before. The ***paradigmatic*** materials for the simple case of *atoms with a spherical symmetry* are the noble gases in all aggregate forms.
  - **Ionic polarization**.
    In this case a (solid) material must have some ionic character. It then automatically has internal dipoles, but these built-in dipoles exactly cancel each other and are unable to rotate. The external field then *induces net dipoles* by slightly displacing the ions from their rest position. The paradigmatic materials are all simple ionic crystals like **NaCl**.
  - **Orientation polarization**.
    Here the (usually liquid or gaseous) material must have *natural dipoles* which can rotate freely. In thermal equilibrium, the dipoles will be randomly oriented and thus carry no net polarization. The external field aligns these dipoles to some extent and thus induces a polarization of the material. The paradigmatic material is water, i.e. **H_2O** in its liquid form.
- Some or all of these mechanisms may act simultaneously. Atomic polarization, e.g., is always present in any material and thus becomes superimposed on whatever other mechanism there might be.
  - Real materials thus can be very complicated in their dielectric behavior. In particular, *non-spherical atoms* (as, e.g., **Si** in a crystal with its four **sp^3** orbitals) may show complex electronic polarization, and mixtures of ionic and covalent bonding (e.g. in **SiO_2**, which has about equal ionic and covalent bonding contributions) makes calculations even more difficult. But the basic mechanisms are still the ones described above.
- The last three mechanisms are ***amenable*** to basic considerations and calculations; ***interface polarization***, however, defies basic treatment. There is simply no *general* way to calculate the charges on ***interfaces*** nor their contribution to the total polarization of a material.
  - Interface polarization is therefore often omitted from the discussion of dielectric properties. Here, too, we will not pursue this matter much further.
  - It would be totally wrong, however, to conclude that *interface polarization is technically not important* because, on the one hand, many dielectrics in real capacitors rely on interface polarization while, on the other hand, interface polarization, if present, may "kill" many electronic devices, e.g. the **MOS** transistor!
  - Let's look at this in an exercise:

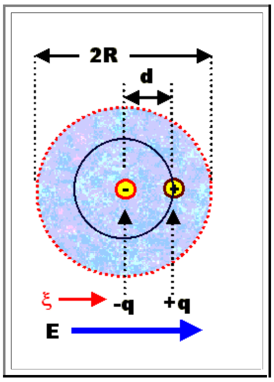
Exercise 3.2-2
Maximum polarization of water

Questionaire

Multiple Choice questions to 3.2.1

3.2.2 Electronic Polarization

- For calculating the effect of **electronic polarization**, We consider an idealized atom with *perfect* ***spherical symmetry***.



  - It has a point like charge **+ *ze*** in the nucleus, and
  - The exact opposite charge **– *ze*** homogeneously distributed in the volume of the atom, which is

$$V = \frac{4}{3}\pi R^3$$

  - With ***R* =** radius of the atom

- The charge density ρ of the electrons then is

$$\rho = -\frac{3z \cdot e}{4\pi \cdot R^3}$$

- In an electrical field ***E*** a force ***F*₁** acts on charges given by

$$F_1 = z \cdot e \cdot E$$

  - We will now drop the underlining for vectors and the mauve color for the electrical field strength ***E*** for easier readability.

- The positive charge in the nucleus and the center of the negative charges from the electron "cloud" will thus experience forces in different direction and will become separated. We have the idealized situation shown in the image above.
- The separation distance ***d*** will have a finite value because the separating force of the external field is exactly balanced by the attractive force between the centers of charge at the distance ***d***.
  - How large is his attractive force? It is not obvious because we have to take into account the attraction between a *point* charge and *homogeneously distributed* charge.
  - The problem is exactly analogous to the classical mechanical problem of a body with mass ***m*** falling through a hypothetical hole going all the way from one side of the globe to the other.
  - We know the solution to that problem: The attractive force between the point mass and the earth is equal to the attractive force between two point masses if one takes *only* the mass of the volume inside the sphere given by the distance between the center of the spread-out mass and the position of the point mass.
  - Knowing electrostatics, it is even easier to see why this is so. We may divide the force on a charged particles on any place inside a homogeneously charged sphere into the force from the "*inside*" sphere and the force from the hollow "*outside*" sphere. Electrostatics teaches us, that a sphere charged on the outside has *no field* in the *inside*, and therefore *no force* (the principle behind a Faraday cage). Thus we indeed only have to consider the "charge *inside* the sphere.
- For our problem, the attractive force ***F*₂** thus is given by

$$F_2 = \frac{\text{q(Nucleus)} \cdot \text{q(e in } d)}{4\pi\, \epsilon_0 \cdot d^{\,2}}$$

  - with **q(Nucleus) = ze** and **q(e in *d*)** = the fraction of the charge of the electrons contained in the sphere with radius ***d***, which is just the relation of the volume of the sphere with radius ***d*** to the *total* volume . We have

$$q(e \text{ in } d) = ze \cdot \frac{(4/3)\,\pi \cdot d^3}{(4/3)\,\pi \cdot R^3} = \frac{ze \cdot d^3}{R^3}$$

- and obtain for **F_2**:

$$F_2 = \left(\frac{(ze)^2}{4\,\pi\epsilon_0 \cdot R^3}\right) \cdot d$$

We have a linear force law akin to a spring; the expression in brackets is the "spring constant". Equating F_1 with F_2 gives the equilibrium distance d_E.

$$d_E = \frac{4\,\pi\epsilon_0 \cdot R^3 \cdot E}{ze}$$

Now we can calculate the **induced dipole moment** μ, it is

$$\mu = ze \cdot d_E = 4\,\pi\epsilon_0 \cdot R^3 \cdot E$$

The polarization P finally is given by multiplying with N, the density of the dipoles; we obtain

$$P = 4\,\pi \cdot N \cdot \epsilon_0 \cdot R^3 \cdot E$$

- Using the definition $P = \epsilon_0 \cdot \chi \cdot E$ we obtain the *dielectric susceptibility resulting from atomic polarization*, χ_{atom}

$$\chi_{atom} = 4\,\pi \cdot N \cdot R^3$$

Let's get an idea about the numbers involved by doing a simple exercise:

Exercise 3.2-3

Some numbers for atomic polarization

This is our first basic result concerning the polarization of a material and its resulting susceptibility. There are a number of interesting points:

- We justified the "law" of a linear relationship between E and P for the electronic polarization mechanism (sometimes also called atomic polarization).
- We can easily extend the result to a mixture of different atoms: All we have to do is to sum over the relative densities of each kind of atom.
- We can easily get an order of magnitude for χ. Taking a typical density of $N \approx 3 \cdot 10^{19}\ \text{cm}^{-3}$ and $R \approx 6 \cdot 10^{-9}\ \text{cm}$, we obtain

$$\chi \approx 8{,}14 \cdot 10^{-5}, \quad \text{or}$$

$$\epsilon_r = 1{,}000\,0814$$

In words: the electronic polarization of *spherical* atoms, while existent, is *extremely weak*. The difference to vacuum is at best in the promille range.

Concluding now that electronic polarization is totally unimportant, would be ***premature***, however. Atoms in crystals or in any solids do not generally have *spherical* symmetry. Consider **the sp^3** orbital of **Si**, **Ge** or diamond.

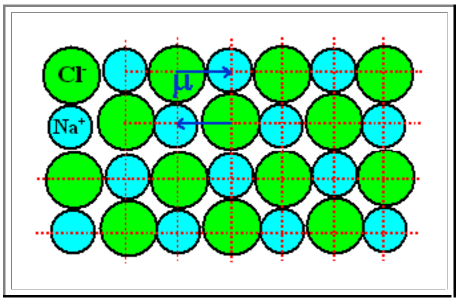
- Without a field, the center of the negative charge of the electron orbitals will still coincide with the core, but an external field breaks that symmetry, producing a dipole momentum.
- The effect can be *large* compared to spherical **s**-orbitals: **Si** has a dielectric constant (**DK**) of **12**, which comes exclusively from electronic polarization! Some values for semiconductors are given in the link.

Questionaire

Multiple Choice questions to 3.2.2

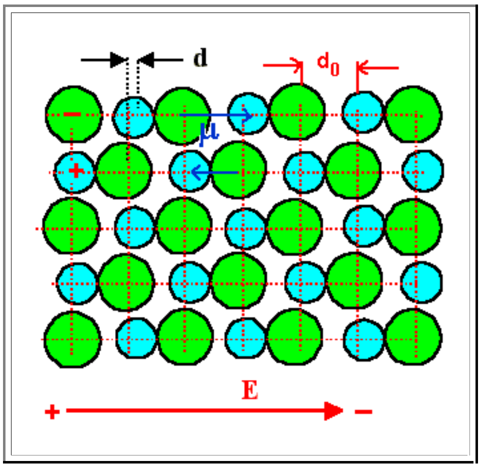
3.2.3 Ionic Polarization

- Consider a simple ionic crystal, e.g. **NaCl**.
  - The lattice can be considered to consist of **Na^+ - Cl^-** dipoles as shown below.



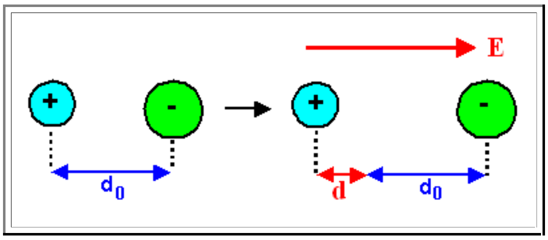
  - Each **Na^+ - Cl^-** pair is a *natural dipole*, no matter how you pair up two atoms.
  - The polarization of a given volume, however, is *exactly zero* because for every dipole moment there is a neighboring one with exactly the same magnitude, but opposite sign.
  - Note that the dipoles *can not rotate*; their direction is fixed.

- In an electric field, the ions feel forces in opposite directions. For a field acting as shown, the lattice distorts a little bit (hugely exaggerated in the drawing)



  - The **Na^+** ions moved a bit to the right, the **Cl^-** ions to the left.
  - The dipole moments between adjacent **NaCl** - pairs in field direction are now different and there is a *net dipole moment* in a finite volume now.

- From the picture it can be seen that it is sufficient to consider *one* dipole in field direction. We have the following situation:



  - Shown is the situation where the distance between the ions *increases* by ***d***; the symmetrical situation, where the distance *decreases* by ***d***, is obvious.

- How large is ***d***? That is easy to calculate:
  - The force F_1 increasing the distance is given by

$$F_1 \;=\; q \cdot E$$

  - With ***q*** **=** net charge of the ion.
  - The restoring force F_2 comes from the binding force, it is given as the derivative of the binding potential. Assuming a *linear relation* between binding force and deviation from the equilibrium distance d_0, which is a good approximation for $d \ll d_0$, we can write

$$F_2 \;=\; k_{IP} \cdot d$$

- With k_{IP} being the "*spring constant*" of the bond. k_{IP} can be calculated from the bond structure, it may also be expressed in terms of other constants that are directly related to the shape of the interatomic potential, e.g. the *modulus of elasticity* or *Youngs modulus*.
- If we do that we simply find

$$\boxed{k_{IP} = Y \cdot d_0}$$

- With Y **=** Youngs Modulus, and d_0 = equilibrium distance between atoms.

From force equilibrium. i.e. $F_1 - F_2 = 0$, we immediately obtain the following relations:

- *Equilibrium distance* d

$$\boxed{d = \frac{q \cdot E}{Y \cdot d_0}}$$

- *Induced dipole moment* μ (on top of the existing one)

$$\boxed{\mu = \frac{q^2 \cdot E}{Y \cdot d_0}}$$

- *Polarization* P

$$\boxed{P = \frac{N \cdot q^2 \cdot E}{Y \cdot d_0}}$$

Of course, this is only a very rough approximation for an *idealized* material and just for the case of increasing the distance. Adding up the various moments - some larger, some smaller - will introduce a factor **2** or so; but here we only go for the principle.

For *real* ionic crystals we also may have to consider:

- More complicated geometries (e.g. CaF_2, with ions carrying different amount of charge).
- This example was deliberately chosen: The dielectric constant of CaF_2 is of paramount interest to the semiconductor industry of the **21**st century, because CaF_2 is pretty much the only usable material with an index of refraction n (which is directly tied to the **DK** via $\epsilon_r = n^2$) that can be used for making lenses for lithography machines enabling dimensions of about **0,1 μm**.
- If the field is not parallel to a major axis of the crystal (this is automatically the case in polycrystals), you have to look at the components of μ in the field direction and average over the ensemble.

Still, the basic effects is the same and ionic polarization can lead to respectable dielectric constants ϵ_r or susceptibilities χ.

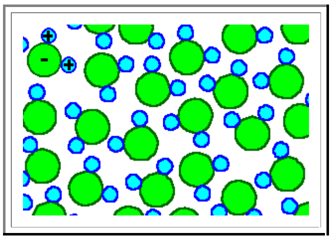
- Some values are given in the link.

Questionaire

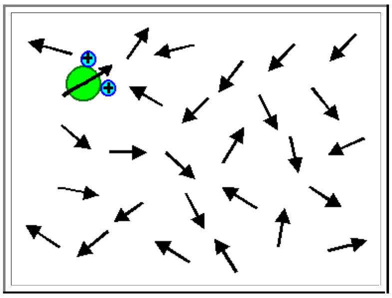
Multiple Choice questions to 3.2.3

3.2.4 Orientation Polarization

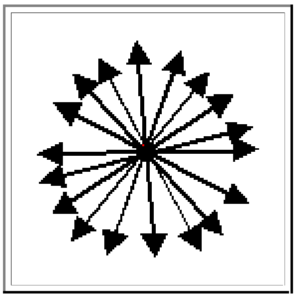
- In the case of **orientation polarization** we have a material with *built-in dipoles* that are *independent of each other*, i.e. they can *rotate freely* - in sharp contrast to ionic polarization.
  - The prime example is *liquid water*, where every water molecule is a little dipole that can have any orientation with respect to the other molecules. Moreover, the orientation changes *all the time* because the *molecules moves*! Orientation polarization for dielectric dipoles thus is pretty much limited to liquids - but we will encounter it in a major way again for *magnetic dipoles*.
  - A two-dimensional "piece of water" may - very graphically - look somewhat like the picture below that captures *one particular moment in time*. It is like a snapshot with a very, very short exposure time. A few nanoseconds later the same piece of water may look totally different in detail, but pretty much the same in general.
  - In a three-dimensional piece of water the blue and red circles would not have to be in the same plane; but that is easy to imagine and difficult to draw.



  - Shown is a bunch of water molecules that form natural dipoles because the negatively charged oxygen atom and the two positively charged **H** - atoms have different centers of charge. Each molecule carries a dipole moment which can be drawn as a vector of constant length. If we only draw a vector denoting the dipole moment, we get - in two dimensions - a picture like this:



  - Again, remember that both pictures are "*snap shots*" that only appear unblurred for very small exposure times, say picoseconds, because the dipoles wiggle, rotate, and move around rather fast, and that in *three* dimensions the vectors would also point out of the drawing plane.
- The total dipole moment is the *vector sum* of the individual dipole moments.



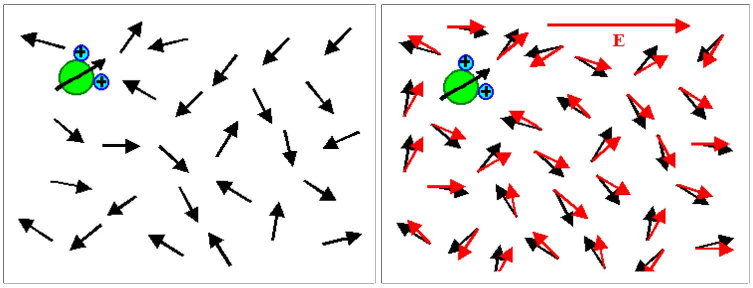
  - For dipoles oriented *at random*, at any given moment this looks like the picture below if we draw all vectors from a common origin: The sum of all dipole moments will be zero, if the dipoles are randomly oriented.
  - We can see this most easily if we have all dipoles start at the same origin. The picture, of course, is two-dimensional and crossly simplified. There would be a lot more (like **10^{20}**) dipoles for any appreciable amount of water - you really will average them to zero pretty well.

- If we now introduce a field ***E***, the dipoles would have a tendency to turn into the field because that would lower their energy.
  - If you have problems with this statement, just imagine the electrostatic interaction, which will always try to move the positive pole of the dipole towards the negative pole of the field, and vice versa for the negative pole - the dipole would align itself exactly along a field line of the external field for minimum energy.
  - Naively, we would then expect a *perfect orientation into the field* and a ***concomitantly*** large polarization because that would lead to the *minimum of the dipole energy*.
  - Well, water does have a pretty large **DK** of **81**, so there is obviously *some* orientation into the field, but it is easy (not really) to show (in an exercise) that this **DK** is several orders of magnitude too small for *fully* oriented dipole moments at some normal field strengths.

Exercise 3.2-1

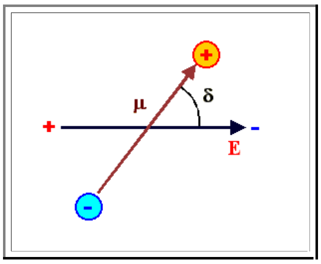
Maximum polarization of water

- In reality, the orientation into the field direction will be *counteracted by random collisions* with other dipoles, and this process is energized by the *thermal energy* "**k*T***" contained in the water.
  - Again, the dipoles are not sitting still, but moving around and rotating all the time - because they contain *thermal energy* and thus also some **entropy**.
  - Whenever two molecules collide, their new orientation is *random* - all memory of an orientation that they might have had in the electrical field is lost. This is analogous to what happens to electrons carrying an electrical current in an electrical field.
  - The electrical field only induces a little bit of *average* orientation in field direction - most of the time an individual dipole points in all kinds of directions. This is the simple truth even so some (undergraduate) text books show pictures to the contrary. The "real" picture (in the sense of a snapshot with a very short exposure time) looks like this:



Without field **With field**

  - The orientation of all dipoles is just a little bit shifted so that an average orientation in field direction results. In the picture, the effect is even exaggerated!
- In fact, the state of being *liquid* by necessity implies quite a bit of **entropy**, and entropy means *disorder*.
  - Perfectly aligned dipoles would be in *perfect order* without any entropy - this is only possible at extremely low temperatures (and even there quantum theory would not allow it) where we will not have liquids any more, or more generally, dipoles that are able to rotate freely.
  - In other words, we must look for the minimum of the **free enthalpy *G*** and not for the minimum of the **internal energy *U***. At finite temperatures the minimum of the free enthalpy requires some **entropy *S***, i.e. randomness in the dipole orientation, so we should not expect perfect orientation.
- If you are not familiar with the basics of thermodynamics, you have a problem at this point. If you do know your thermodynamics, but are a bit insecure, turn to the basic module "Thermodynamics" (in the "Defects" Hyperscript) to refresh your memory.
- We obviously need to calculate the free enthalpy ***G=U – TS*** to see what kind of average orientation will result in a given field. Note that we use ***U***, the common symbol for the (internal) energy instead of ***H***, the common symbol for the enthalpy, because ***U*** and ***H*** are practically identical for solids and liquids anyway.
  - Moreover, a mix up with the magnetic field strength usually designated by ***H***, too, would be unavoidable otherwise. (The possible mix-up between internal energy ***U*** and voltage ***U*** is not quite so dangerous in this context).
- The internal energy od a dipole is clearly a function of its orientation with respect to the field. It must be minimal, when the dipole is aligned with the field and thedipole moment has the same direction as the electrical field, and maximal if the direction is reversed.
  - This is the easy part: The energy ***U*(δ)** of a dipole with dipole moment μ in a field ***E*** as a function of the angle δ ("delta") between the dipole moment direction and the field direction.



  - From basic electrostatics we have have

$$\boxed{U(\delta) = -\ \underline{\mu} \cdot \underline{E} = -\ |\underline{\mu}| \cdot |\underline{E}| \cdot \cos\delta}$$

  - The *minimum energy* ***U*** thus would occur for **δ=0°**, i.e. for perfect alignment in proper field direction (*note the minus sign!*); the maximum energy for **δ=180°**, i.e. for alignment the wrong way around.
  - That was for two dimensions - now we must look at this in *three* dimensions.

  - In **3D** we see that all dipoles with the same angle δ between their axis and the field still have the same energy - and this means now all dipoles on a *cone* with opening angle **2**δ around the field axis if we consider possible orientations out of the plane of drawing.
  - In order to obtain the *total internal energy* $\boldsymbol{U}_{\textbf{total}}$ of a *bunch* of dipoles having *all kinds* of angles δ with the field axis, we will have to sum up all cones.
  - This means we take the number of dipoles ***N*(δ)** having a particular orientation δ times the energy belonging to that δ, and integrate the resulting function over δ from **0$^\mathrm{o}$** to **180$^\mathrm{o}$**. This is something that we could do - if we would know ***N*(δ)**.

- However, just calcuating $\boldsymbol{U}_{\textbf{total}}$ will not be of much use. We also *must* consider the entropy term **– *TS***, because we do not want to calculate the total *internal energy* $\boldsymbol{U}_{\textbf{total}}$, but the total *free enthalpy* $\boldsymbol{G=U}_{\textbf{total}}\boldsymbol{-TS}$.
  - We need to consider that term as a function of all possible angle distributions and then see for which distribution we can minimize ***G***.

- *But what is the entropy* ***S*(*N*(δ))** of an ensemble of dipoles containing ***N*(δ)** members at the angle δ as a function of the many possible distribution ***N*(δ)**? Not an easy question to answer from just looking at the dipoles.

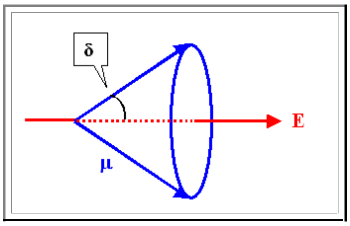
Fortunately, we do not have to calculate ***S*** *explicitly!*

- We *know* a formula for the distribution of (classical) particles on available energy levels that *automatically* gives the minimum of the free enthalpy!
  - We have a *classical* system where a number of independent particles (the dipoles) can occupy a number of energy levels (between $\boldsymbol{U}_{\textbf{min}}$ and $\boldsymbol{U}_{\textbf{max}}$) as defined by **δ=0$^\mathrm{o}$** or **δ=180$^\mathrm{o}$**, respectively.
  - Basic thermodynamics asserts that *in equilibrium*, the distribution of the particles on the available energy levels is given by the proper *distribution function* which is defined in such a way that it *always* gives the minimum of the free enthalpy.
  - Since we deal with classical particles in this approach, we have to use the **Boltzmann distribution**. We obtain for ***N*(*U*)=** number of dipoles with the energy ***U***

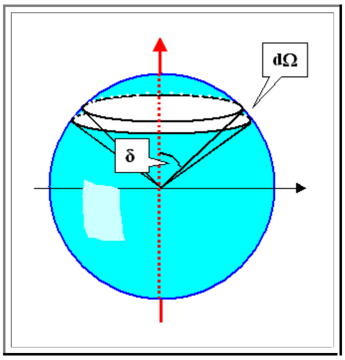
$$\boxed{N(U) = A \cdot \exp - \frac{U(\delta)}{kT}}$$

  - With a constant ***A*** that has yet to be determined.

- This Boltzmann distribution equation gives us the number of dipoles with a certain angle relative to the field direction, i.e. the number of dipoles that have their tips on a circle with an opening angle **2**δ relative to the field directions as shown below.



  - We are, however, only interested in the *component of the dipole moment parallel to the field*. For this we look at the ***solid angle*** increment **d**Ω defined on the unit sphere as the segment between δ and δ **+ d**δ.



  - The number of dipoles lying in the cone angle increment defined by δ and δ**+** Δδ is the same as the number of dipoles with tips ending on the surface of the unit sphere in the incremental angle **d**Ω. It is given by ***N*(*U*(δ)) · d**Ω.
  - Note that **d**Ω is a measure of an incremental *area*; a kind of ribbon once around the unit sphere.
  - The sum of the components μ_F of the dipole moments *in field direction* is then .

$$\boxed{\mu_\mathrm{F} = (N \cdot \mathrm{d}\Omega) \cdot (\mu \cdot \cos \delta)}$$

  - If you are not familiar with spherical coordinates, this (and what we will do with it), looks a bit like magic. Since we do not want to learn Math in this lecture, the ***essentials to spherical coordinates*** are explained in detail in a basic module.

- The *average dipole moment*, which is what we want to calculate, will now be obtained by summing up the contributions from all the **d**Ωs

$$\langle \mu_F \rangle = \frac{\displaystyle\int_0^{\pi} N(U(\delta)) \cdot \mu \cdot \cos\delta \cdot d\Omega}{\displaystyle\int_0^{\pi} N(U(\delta)) \cdot d\Omega}$$

  - And the integrals have to be taken from the "top" of the sphere to the "bottom" , i.e. from **0** to π.

- **d**Ω and δ are of course closely related, we simply have

$$d\Omega = 2\pi \cdot \sin\delta \cdot d\delta$$

- Putting everything together, we obtain a pretty horrifying integral for μ_F that runs from **0** to π

$$\langle \mu_F \rangle = \frac{\mu \cdot \displaystyle\int_0^{\pi} \sin\delta \cdot \cos\delta \cdot \exp\frac{\mu \cdot E \cdot \cos\delta}{kT} \cdot d\delta}{\displaystyle\int_0^{\pi} \sin\delta \cdot \exp\frac{\mu \cdot E \cdot \cos\delta}{kT} \cdot d\delta}$$

- One advantage is that we got rid of the undetermined constant ***A***. The integral, being a determined integral, is now simply a *number* depending on the parameters of the system, i.e. the temperature ***T***, the dipole moment **μ** and the field strength ***E***.
  - *The problem has been reduced to a mathematical exercise in solving integrals*.

- Since we are not interested at doing math, we just show the general direction toward a solution:
  - Use the substitutions

$$\beta = \frac{\mu \cdot E}{kT}$$

$$x = \cos\delta$$

  - The integral reduces to

$$<\mu_F> = \frac{\mu \cdot \int_{+1}^{-1} x \cdot \exp(\beta \cdot x) \cdot dx}{\int_{+1}^{-1} \exp(\beta \cdot x) \cdot dx}$$

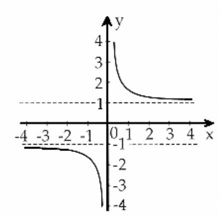
The final result after quite a bit of fiddling around is

$$<\mu_F> = \mu \cdot L(\beta)$$

- With **L(β)=Langevin function**, named after Paul Langevin, and defined as
.

$$L(\beta) = \coth(\beta) - \frac{1}{\beta}$$

$$\beta = \frac{\mu \cdot E}{kT}$$



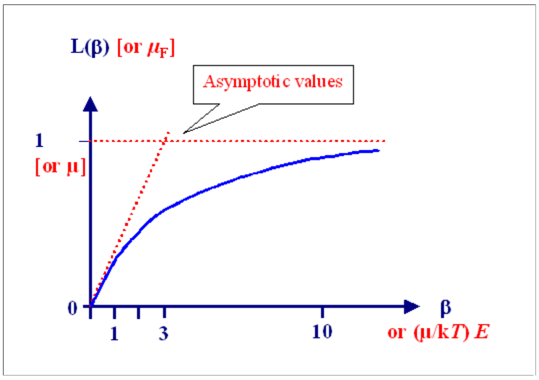
y=coth *x*

- The "**coth**" is the *hyperbolic cotangent*, defined as **coth $x=(e^x + e^{-x})/(e^x - e^{-x})=1/\tanh x$**.
- **L(β)** is a tricky function, because the **coth *x*** part looks pretty much like a hyperbola, from which the real hyperbola **1/*x*** is subtracted. What's left is almost nothing - **L(*x*)** values are between **0** and **1**

The polarization (always on average, too) is accordingly
.

$$P = N \cdot <\mu>$$

This is a definite result, but it does not help much. We need to discuss the mathematical construct "*Langevin function* **L(β)**" to get some idea of what we obtained. We look at the graph in general units and in units of the dipole moment and electrical field (in red).



- Since β is proportional to the field strength ***E***, we see that the dipole moment and the polarization increases monotonically with ***E***, eventually saturating and giving **<μ_F>=μ** which is what we must expect.
- The question is, what *range* of β values is accessible for real materials. i.e. how close to the saturation limit can we get?

For that we look at some simple approximations.

- If we develop **L(β)** into a series (consult a math textbook), we get

$$L(\beta) = \frac{\beta}{3} - \frac{\beta^3}{45} + \frac{2\,\beta^5}{945} - \;.....$$

- For large values of β we have **L(β)** ≈ **1**, while for small values of β (**β < 1**), the Langevin function can be approximated by .

$$L(\beta) \approx 1/3 \cdot \beta$$

- The slope thus is **1/3** for β → **0**.
- For "normal" circumstances, we always have β **<< 1** (see below), and we obtain as final result for the **induced dipole moment** the **Langevin - Debye** equation

$$\langle \mu_F \rangle = \frac{\mu^2 \cdot E}{3kT}$$

$$\langle P \rangle = \frac{N \cdot \mu^2 \cdot E}{3kT}$$

- These equations will be rather good approximation for small values of μ and **E** and/or large values of ***T***. For very large fields and very small temperatures the average dipole moment would be equal to the built in dipole moment, i.e. all dipoles would be strictly parallel to the field. This is, however, not observed in "normal" ranges of fields and temperatures.

- Lets see that in an example. We take
  - $E = 10^8$ **V/cm** which is about the highest field strength imaginable before we have electrical breakdown,
    $\mu = 10^{-29}$ **Asm**, which is a large dipole moment for a strongly polarized molecule, e.g. for **HCl**, and
    $T = 300$ **K**.
  - This gives us
    $\beta = 0{,}24$ - the approximation is still valid. You may want to consult exercise 3.2-1 again (or for the first time) at this point and look at the same question from a different angle.
- At $T = 30$ **K**, however, we have $\beta = 2{,}4$ and now we must think twice:
  - **1.** The approximation would no longer be good. *But*
  - **2.** We no longer would have *liquid* **HCl** (or H_2O, or liquid whatever with a dipole moment), but solid **HCl** (or whatever) , and we now look at ionic polarization and no longer at orientation polarization!
- You may now feel that this was a rather useless exercise - after all, who is interested in the **DK** of liquids? But consider: This treatment is *not* restricted to electric dipoles. It is valid for all kinds of dipoles that can rotate freely, in particular for the *magnetic dipoles* in paramagnetic materials responding to a magnetic field.
  - Again, you may react with stating "Who is interested in paramagnets? Not an electrical engineer!" Right - but the path to *ferromagnets*, which definitely are of interest, starts exactly where orientation polarization ends; you cannot avoid it.
- It is important to be aware of the basic condition that we made at the beginning: *there is no interaction between the dipoles*! This will not be true in general.
  - Two water molecules coming in close contact will of course "feel" each other and they may have preferred orientations of their dipole moments relative to each other. In this case we will have to modify the calculations; the above equations may no longer be a good approximation.

- On the other hand, if there is a *strong* interaction, we automatically have some bonding and obtain a solid - ice in the case of water. The dipoles most likely cannot orientate themselves freely; we have a different situation (usually ionic polarization). There are, however, some solids where dipoles exist that can rotate to some extent - we will get very special effects, e.g. "ferroelectricity".

Questionaire
Multiple Choice questions to 3.2.4

3.2.5 Summary and Generalization

For all three cases of polarization mechanisms, we had a *linear* relationship between the electrical field and the dipole moment (for fields that are not excessively large):

Electronic polarization	Ionic polarization	Orientation polarization
$\mu_{EP} = 4\pi \cdot \epsilon_0 \cdot R^3 \cdot E$	$\mu_{IP} = \frac{q^2}{k_{IP}} \cdot E$	$\mu_{op} = \frac{\mu^2}{3kT} \cdot E$

It seems on a first glance that we have justified the "*law*" $P = \chi \cdot E$.

- However, that is not quite true at this point. In the "law" given by equation above, ***E*** refers to the *external* field, i.e. to the field that would be present in our capacitor *without* a material inside.
- We have $E_{ex} = U / d$ for our plate capacitor held at a voltage ***U*** and a spacing between the plates of ***d***.
- On the other hand, the induced dipole moment that we calculated, always referred to the *field at the place of the dipole*, i.e. the *local* field E_{loc}. And if you think about it, you should at least feel a bit uneasy in assuming that the two fields are identical. We will see about this in the next paragraph.

Here we can only define a factor that relates μ and E_{loc}; it is called the **polarizability** α. It is rarely used with a number attached, but if you run across it, be careful if ϵ_0 is included or not; in other words what kind of unit system is used.

- We now can reformulate the three equations on top of this paragraph into one equation

$$\boxed{\underline{\mu} = \alpha \cdot E_{loc}}$$

- The ***polarizability*** α is a material parameter which depends on the polarization mechanism: For our three paradigmatic cases they are are given by

$$\alpha_{EP} = 4\pi \cdot \epsilon_0 \cdot R^3$$

$$\alpha_{IP} = \frac{q^2}{k_{IP}}$$

$$\alpha_{OP} = \frac{\mu^2}{3kT}$$

- This does not add anything new but emphasizes the proportionality to ***E***.

So we *almost* answered our first basic question about dielectrics - but for a full answer we need a relation between the *local* field and the *external* field. This, unfortunately, is *not a particularly easy problem*

- One reason for this is: Whenever we talk about electrical fields, we always have a certain scale in mind - without necessarily being aware of this. Consider: In a metal, as we learn from electrostatics, there is *no field at all*, but that is *only true* if we do not look too closely. If we look on an *atomic scale*, there are tremendous fields between the nucleus and the electrons. At a somewhat larger scale, however, they disappear or perfectly balance each other (e.g. in ionic crystals) to give no field on somewhat larger dimensions.
- The scale we need here, however, is the *atomic scale*. In the electronic polarization mechanism, we actually "looked" *inside* the atom - so we shouldn't just stay on a "rough" scale and neglect the fine details.

Nevertheless, that is what we are going to do in the next paragraph: *Neglect the details*. The approach may not be beyond reproach, but it works and gives simple relations.

Questionaire

Multiple Choice questions to all of 3.2

3.2.6 Local Field and Clausius - Mosotti Equation

- "Particles", i.e. atoms or molecules in a liquid or solid are basking in electrical fields - the external field that we apply from the outside is not necessarily all they "see" in terms of fields.
  - First, of course, there is a tremendous electrical field *inside* any atom. We have after all, positive charges and negative charges separated by a distance roughly given by the diameter of the atom.
  - Second, we have fields *between* atoms, quite evident for ionic crystals, but also possible for other cases of bonding.
- However, if you look at the materials at a scale somewhat larger than the atomic scale, all these fields must *average to zero*. Only then do we have a field-free interior as we always assume in ***electrical engineering*** ("no electrical field can penetrate a metal").
- Here, however, we are looking at the effect an external field has on *atoms* and *molecules*, and it would be ***preposterous*** to assume that what an atom "sees" as *local* electrical field is identical to what we apply from the outside.
  - Since all our equations obtained so far always concerned the *local* electrical field - even if we did not point that out in detail before - we now must find a relation between the external field and the local field, if we want to use the insights we gained for understanding the behavior of dielectrics on a macroscopic scale.
- We define the **local field** $\boldsymbol{E}_{loc}$ to be the field felt by *one* particle (mostly an atom) of the material at its position (x,y,z).
  - Since the superposition principal for fields always holds, we may express $\boldsymbol{E}_{loc}$ as a superposition of the external field $\boldsymbol{E}_{ex}$ and some field $\boldsymbol{E}_{mat}$ introduced by the surrounding material. We thus have

$$\boxed{\underline{E}_{loc} = \underline{E}_{ex} + \underline{E}_{mat}}$$

- All electrical fields can, in principle, be calculated from looking at the charge distribution $\rho(x, y, z)$ in the material, and then solving the Poisson equation (which you should know). The Poisson equation couples the charge distribution and the potential $V(x, y, z)$ as follows:

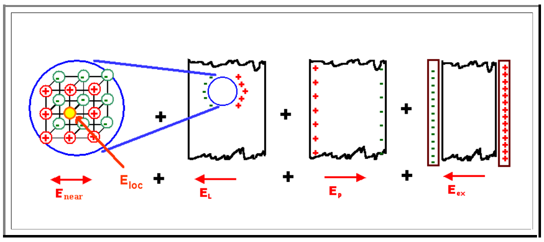
$$\Delta V = -\frac{\rho(x, y, z)}{\epsilon \cdot \epsilon_0}$$

$$\Delta = \text{Delta operator} = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}$$

  - The electrical field then is just the (negative) gradient of the potential; $\underline{E} = -\nabla V$.
- Doing this is pretty tricky, however. We can obtain usable results in a good approximation in a much simpler way, by using the time-honored **Lorentz** approach or the **Lorentz model**.
  - In this approach we decompose the total field into *four* components.
  - For doing this, we *imagine* that we remove a small sphere containing a few **10** atoms from the material. We want to know the local field in the center of this sphere while it is still in the material; this is the local field $\underline{E}_{mat}$ we are after. We define that field by the force it exerts on a charge at the center of the sphere that acts as a "probe".
- The essential trick is to calculate the field produced from the atoms inside the sphere and the field inside the now empty sphere in the material. The total local field then is simple the sum of both.
  - Like always, we do not consider the charge of the "probe" in computing the field that it probes. *The cut-out sphere thus must not contain the charge we use as the field probe!*
  - The cut-out material, in general, could produce an electrical field at its center since it is composed of charges. This is the ***1st*** component of the field, $\underline{E}_{near}$ which takes into account the contributions of the atoms or ions inside the sphere. We will consider that field in an approximation where we average over the volume of the small sphere. To make things clear, we look at an ionic crystal where we definitely have charges in our sphere.
- $\underline{E}_{near}$, however, is not the *only* field that acts on our probe. We must include the field that all the other atoms of the crystal produce *in* the hollow sphere left after we cut out some material. This field now fills the "empty" void left by taking out our sphere.
  - This field is called $\underline{E}_L$ (the "**L**" stands for Lorentz); it compensates for the cut-out part - and that provides our ***2nd*** component.
- Now we only have to add the "macroscopic" fields from **1.** the polarization of the material and **2.** the external field that causes everything:
  - The field $\underline{E}_{pol}$ is induced by the macroscopic polarization (i.e. by area charges equal to the polarization); it is the ***3rd*** component.

- The external field $\boldsymbol{E_{ex} = U/d}$ from the applied voltage at our capacitor which supplies the ***4th*** component.

In a visualization, this looks like this:



The blue "sphere" cuts through the lattice (this is hard to draw). The yellow "point" is where we consider the local field; we have to omit the contribution of the charged atom there. We now have

$$\boxed{E_{loc} = E_{ex} + E_{pol} + E_L + E_{near}}$$

How large are those fields? We know the external field and also the field from the polarization (always assuming that the material completely fills the space inside the capacitor).

$$E_{ex} = \frac{U}{d} \qquad\qquad E_{pol} = -\frac{P}{\epsilon_0}$$

We do not know the other two fields, and it is not all that easy to find out how large they are. The results one obtains, however, are quite simple:

- Lorentz showed that $\boldsymbol{E_{near} = 0}$ for *isotropic materials*, which is easy to imagine. Thus for cubic crystals (or polycrystals, or amorphous materials), we only have to calculate $\boldsymbol{E_L}$.

$\boldsymbol{E_L}$ needs some thought. It is, however, a standard problem from electrostatics in a slightly different form.

- In the standard problem one calculates the field in a materials with a **DK** given by ϵ_r that does *not* fill a rectangular capacitor totally, but is in the shape of an ellipsoid including the extreme cases of a *pure sphere*, a *thin plate* or a *thin needle*. The result is always

$$E_{ellipse} = N_P \cdot \frac{P}{\epsilon_r \cdot \epsilon_o}$$

In words: The field inside a dielectric in the shape of an ellipsoid (of any shape whatsoever) that is put between the parallel plates of a typical capacitor arrangement, is whatever it would be if the dielectric fills the space between the plates completely times a *number* $\boldsymbol{N_P}$, the value of which depends on the geometry.

- $\boldsymbol{N_P}$ is the so-called **depolarization factor**, a *pure number*, that only depends on the shape of the ellipsoid. For the extreme cases of the ellipsoid we have fixed and well-known ***depolarization*** factors:

  - **Thin plate: $N = 1$**
  - **Needle: $N = 0$**
  - **Sphere: $N = 1/3$**

.

Our case consists of having a sphere with $\boldsymbol{\epsilon_r = 1}$. We thus obtain

$$E_L = \frac{P}{3\epsilon_o}$$

- We have now all components and obtain

$$E_{loc} = \frac{U}{d} - \frac{P}{\epsilon_0} + \frac{P}{3\epsilon_o}$$

  - ***U*/*d* – *P*/ϵ$_0$** is just the field we would use in the Maxwell equations, we call it ***E*$_0$**. It is the *homogeneous field* averaged over the whole volume of the homogeneous material
  - The *local field* finally becomes

$$E_{loc} = E_0 + \frac{P}{3\epsilon_o}$$

- This seems a bit odd? How can the *local* field be different from the *average* field?
  - This is one of the tougher questions one can ask. The answer, not extremely satisfying, comes from the basic fact that *all* dipoles contribute to ***E*$_0$**, whereas for the *local field* you discount the effect of *one* charge - the charge you use for probing the field (the field of which must not be added to the rest!).
  - If you feel somewhat uneasy about this, you are perfectly right. What we are excluding here is the action of a charge on itself. While we may do that because that was *one* way of defining electrical fields (the other one is Maxwells equation defining a field as directly resulting from charges), we can not so easily do away with the *energy* contained in the field of a single charge. And if we look at this, the whole theory of electromagnetism blows up! If the charge is a point charge, we get infinite energy, and if it is not a point charge, we get other major contradictions.
  - Not that it matters in everyday aspects - it is more like a philosophical aspect. If you want to know more about this, read chapter **28** in the "Feynman lectures, Vol. 2"
- *But do not get confused now*! The relation given above is perfectly valid for everyday circumstances and ordinary matter. Don't worry - be happy that a relatively complex issue has such a simple final formula!
- We now can relate the *macroscopic* and *microscopic* parameters. With the old relations and the new equation we have a grand total of:

$$\mu = \alpha \cdot E_{loc}$$

$$P = N \cdot \alpha \cdot E_{loc}$$

$$E_{loc} = E_o + \frac{P}{3\epsilon_o}$$

- From this we obtain quite easily

$$P = N \cdot \alpha \cdot \left(E_o + \frac{P}{3\epsilon_o} \right)$$

$$P = \frac{N \cdot \alpha \cdot E_o}{1 - N \cdot \alpha / 3\epsilon_o}$$

  - With ***N* =** density of dipoles
- Using the definition of ***P***

$$P = \epsilon_o \cdot \chi \cdot E = \epsilon_o \cdot (\epsilon_r - 1) \cdot E$$

- and inserting it into the equations above gives as *final result* the connection between the polarizability α (the microscopic quantity) and the relative dielectric constant ϵ_r (the macroscopic quantity):

$$\frac{N \cdot \alpha}{3\,\epsilon_0} = \frac{\epsilon_r - 1}{\epsilon_r + 2} = \frac{\chi}{\chi + 3}$$

This is the **Clausius - Mosotti equation**, it relates the *microscopic* quantity α on the left hand side to the *macroscopic* quantity ϵ_r (or, if you like that better, $\chi = \epsilon_r - 1$) on the right hand side of the equation. This has two far reaching consequences

- We now can *calculate* (at least in principle) the dielectric constants of all materials, because we know how to calculate α.
- We have an *instrument* to measure *microscopic* properties like the polarizability α, by measuring *macroscopic* properties like the dielectric constant and converting the numbers with the Clausius-Mosotti equation.

You must also see this in an historical context: With the Clausius-Mosotti equation the dielectric properties of materials were essentially reduced to known electrical properties. There was nothing mysterious anymore about the relative dielectric constant. The next logical step now would be to apply quantum theory to dielectric properties.

3.2.7 Summary to: Polarization Mechanisms

- (Dielectric) polarization mechanisms in dielectrics are all mechanisms that

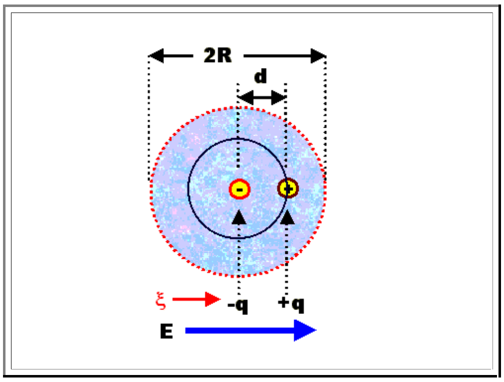
  1. Induce dipoles at all (always with $\boldsymbol{\mu}$ in field direction)
     ⇒ Electronic polarization.
  2. Induce dipoles already present in the material to "point" to some extent in field direction.
     ⇒ Interface polarization.
     ⇒ Ionic polarization.
     ⇒ Orientation polarization.

Quantitative considerations of polarization mechanisms yield

- **Justification (and limits) to the $\underline{P} \propto \underline{E}$ "law"**
- **Values for χ**
- **$\chi = \chi(\omega)$**
- **$\chi = \chi$(structure)**

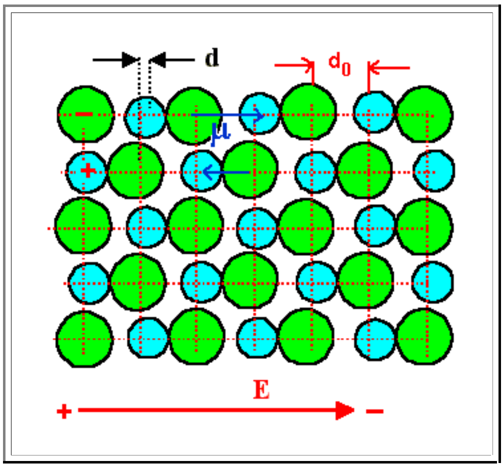
- Electronic polarization describes the separation of the centers of "gravity" of the electron charges in orbitals and the positive charge in the nucleus and the dipoles formed this way. it is always present

  - It is a very weak effect in (more or less isolated) atoms or ions with spherical symmetry (and easily calculated).
  - It can be a strong effect in e.g. covalently bonded materials like **Si** (and not so easily calculated) or generally, in solids.



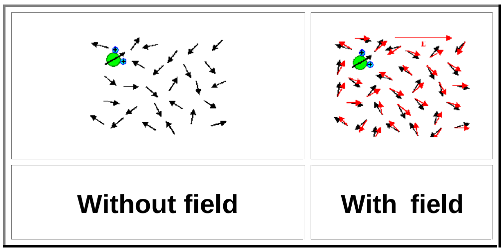
- Ionic polarization describes the net effect of changing the distance between neighboring ions in an ionic crystal like **NaCl** (or in crystals with some ionic component like **SiO_2**) by the electric field

  - Polarization is linked to bonding strength, i.e. Young's modulus **Y**. The effect is smaller for "stiff" materials, i.e. **$P \propto 1/Y$**



- Orientation polarization results from minimizing the free enthalpy of an ensemble of (molecular) dipoles that can move and rotate freely, i.e. polar liquids.

  - It is possible to calculate the effect, the result invokes the Langevin function



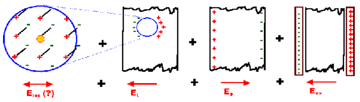
$$\boxed{L(\beta) \;=\; \coth(\beta) \;-\; \frac{1}{\beta}}$$

  - In a good approximation the polarization is given by ⇒

$$\boxed{\langle P \rangle \;=\; \frac{N \cdot \mu^2 \cdot E}{3kT}}$$

- The induced dipole moment **μ** in all mechanisms is proportional to the field (for reasonable field strengths) at the location of the atoms / molecules considered.
  - The proportionality constant is called polarizability α; it is a microscopic quantity describing what atoms or molecules "do" in a field.
  - The local field, however, is not identical to the macroscopic or external field, but can be obtained from this by the Lorentz approach
  - For isotropic materials (e.g. cubic crystals) one obtains

$$\underline{\mu} = \alpha \cdot \boldsymbol{E}_{\text{loc}}$$



$$\boldsymbol{E}_{\text{loc}} = \boldsymbol{E}_{\text{ex}} + \boldsymbol{E}_{\text{pol}} + \boldsymbol{E}_{\text{L}} + \boldsymbol{E}_{\text{near}}$$

$$\boldsymbol{E}_{\text{L}} = \frac{\boldsymbol{P}}{3\epsilon_0}$$

- Knowing the local field, it is now possible to relate the microscopic quantity α to the macroscopic quantity ∊ or ϵ_r via the Clausius - Mosotti equations ⇒
  - While this is not overly important in the engineering practice, it is a momentous achievement. With the Clausius - Mosotti equations and what went into them, it was possible for the first time to understand most electronic and optical properties of dielectrics in terms of their constituents (= atoms) and their structure (bonding, crystal lattices etc.)
  - Quite a bit of the formalism used can be carried over to other systems with dipoles involved, in particular magnetism = behavior of magnetic dipoles in magnetic fields.

$$\frac{N \cdot \alpha}{3\,\epsilon_0} = \frac{\epsilon_r - 1}{\epsilon_r + 2}$$

$$= \frac{\chi}{\chi + 3}$$

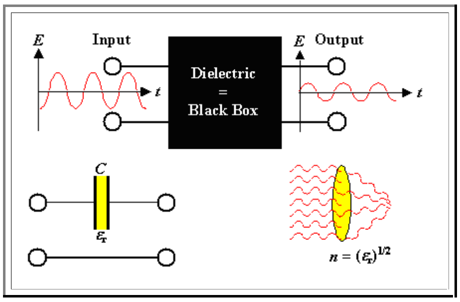
Questionaire

Multiple Choice questions to all of 3.1

3.3 Frequency Dependence of the Dielectric Constant

3.3.1 General Remarks

- All polarization mechanisms respond to an electrical field by shifting masses around. This means that masses must be accelerated and de-accelerated, and this will always take some time. So we *must* expect that the (mechanical) response to a field will depend on the *frequency* ν of the electrical field; on how often per second it changes its sign.
    - If the frequency is very large, no mechanical system will be able to follow. We thus expect that at *very large frequencies all polarization mechanisms will "die out"*, i.e. there is no response to an extremely high frequency field. This means that the dielectric constant ϵ_r will approach **1** for $\nu \Rightarrow \infty$.
- It is best to consider our dielectric now as a "**black box**". A signal in the form of an alternating electrical field ***E*** goes in at the input, and something comes out at the output, as shown below. Besides the Black Box scheme, two possible real expressions of such an abstract system are shown: A parallel-plate capacitor containing a dielectric, and an optical lens with an index of refraction $n = \epsilon_r$. The input would be a simple alternating voltage in the capacitor case, and a light wave in the lens case.



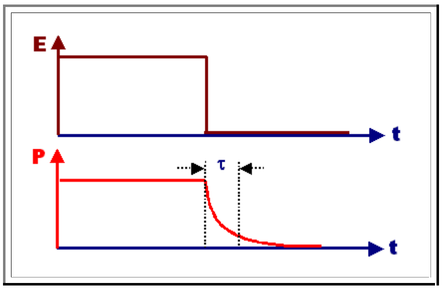
- As long as our system is linear ("twice the input ⇒ twice the output), a sinewave going in will produce a sinewave coming out, i.e. the frequency does not change.
    - If a sinewave goes in, the output then can only be a sinewave with an amplitude and a phase different from the input, as schematically shown above.
    - If a complicated signal goes in, we decompose it into its Fourier components, consider the output for all frequencies separately, and then do a Fourier synthesis.
- With complex notation, the input will be something like $E_{in} = E_{in} \cdot \exp(i\omega t)$; the output then will be $E_{out} = E_{out} \cdot \exp i(\omega t + \phi)$.
    - We just as well could write $E_{out} = f(\omega) \cdot E_{in}$ with $f(\omega) =$ complex number for a given ω or complex function of ω.
    - **f(ω)** is what we are after. We call this function that relates the output of a dielectric material to its input the **dielectric function** of the material. As we will see, the dielectric function is a well-defined and very powerful entity for any material - even if we cannot calculate it from scratch. We can however, calculate dielectric functions for some model materials, and that will give us a very good idea of what it is all about.
- Since the **index of refraction *n*** is directly given by $\epsilon_r^{1/2}$ (assuming that the material has no magnetic properties), we have a *first very general statement*:
    - There exist no microscopes with "optical" lenses for very high frequencies of electrical fields, which means electromagnetic radiation in the *deep ultraviolet* or soft **X-rays**. And indeed, there are no **X-ray microscopes** with lenses1) (however, we still have mirrors!) because there are no materials with $\epsilon_r > 1$ for the frequencies of **X**-rays.
- Looking at the polarization mechanisms discussed, we see that there is a fundamental difference in the *dynamics* of the mechanisms with regard to the response to changing forces:
    - In *two* cases (*electron and ionic polarization*), the electrical field will try to change the distance between the charges involved. In response, there is a restoring force that is (in our approximation) directly proportional to the separation distance of the dipole charges. We have, in mechanical terms, an *oscillator*.
    - The characteristic property of *any* such oscillating system is the phenomena of **resonance** at a specific frequency.
    - In the case of the *orientation polarization*, there is no direct mechanical force that "pulls" the dipoles back to random orientation. Instead we have many statistical events, that respond in their *average results* to the driving forces of electrical fields.
    - In other words, if a driving force is present, there is an equilibrium state with an (average) net dipole moment. If the driving force were to disappear suddenly, the ensemble of dipoles will assume a new equilibrium state (random distribution of the dipoles) within some characteristic time called **relaxation time**. The process knows no resonance phenomena, it is characterized by its **relaxation time** instead of a resonance frequency.
- We thus have to consider just the two basic situations: **Dipole relaxation** and **dipole resonance**. Every specific mechanism in real materials will fit one of the two cases.

[1] Well, never say never. Lenses for X-rays *do* exist for a few years by now. However, if you would see the contraption, you most likely wouldn't recognize it as a *lens*. If you want to know more, turn to the resarch of Prof. Lengeler and his group: http://2b.physik.rtwh-aachen.de

3.3.2 Dipole Relaxation and Dielectric Function

From Time Dependence to Frequency Dependence

- The easiest way to look at *relaxation phenomena* is to consider what happens if the driving force - the electrical field in our case - is suddenly switched off, after it has been constant for a sufficiently long time so that an equilibrium distribution of dipoles could be obtained.
  - We expect then that the dipoles will randomize, i.e. their dipole moment or their polarization will go to zero.
  - However, that cannot happen instantaneously. A specific dipole will have a certain orientation at the time the field will be switched off, and it will change that orientation only by some interaction with other dipoles (or, in a solid, with phonons), in other words upon collisions or other "violent" encounters. It will take a characteristic time, roughly the time between collisions, before the dipole moment will have disappeared.
  - Since we are discussing statistical events in this case, the individual characteristic time for a given dipole will be small for some, and large for others. But there will be an *average* value which we will call the **relaxation time** τ of the system. We thus expect a smooth change over from the polarization with field to zero within the relaxation time τ, or a behavior as shown below



- In formulas, we expect that ***P*** decays starting at the time of the switch-off according to

$$P(t) \;=\; P_0 \cdot \exp - \frac{t}{\tau}$$

- This simple equation describes the behavior of a simple system like our "ideal" dipoles very well. It is, however, not easy to derive from first principles, because we would have to look at the development of an ensemble of interacting particles in time, a classical task of non-classical, i.e. statistical mechanics, but beyond our ***ken*** at this point.
  - Nevertheless, we know that a relation like that comes up whenever we look at the decay of some ensemble of particles or objects, where some have more (or less) energy than required by equilibrium conditions, and the change-over from the excited state to the base state needs "help", i.e. has to overcome some energy barrier.
  - All we have to assume is that the number of particles or objects decaying from the excited to the base state is proportional to the number of excited objects. In other words, we have a relation as follows:

$$\frac{dn}{dt} \propto n \;=\; -\,\frac{1}{\tau} \cdot n$$

$$n \;=\; n_0 \cdot \exp - \frac{t}{\tau}$$

  - This covers for example radioactive decay, cooling of any material, and the decay of the foam or froth on top of your **beer**: Bubbles are an energetically excited state of beer because of the additional surface energy as compared to a droplet. If you measure the height of the head on your beer as a function of time, you will find the exponential law.
- When we turn on an electrical field, our dipole system with random distribution of orientations has too much energy relative to what it could have for a better orientation distribution.
  - The "decay" to the lower (free) energy state and the concomitant built-up of polarization when we switch on the field, will follow our universal law from above, and so will the decay of the polarization when we turn it off.

- We are, however, not so interested in the *time* dependence ***P*(*t*)** of the polarization when we apply some disturbance or *input* to the system (the switching on or off of the electrical field). We rather would like to know its *frequency* dependence ***P*(ω)** with **ω = 2πν =** angular frequency, i.e. the output to a periodic harmonic input, i.e. to a field like $E = E_0 \cdot \sin\omega t$.
  - Since *any* signal can be expressed as a **Fourier series or Fourier integral** of sin functions as the one above, by knowing ***P*(ω)** we can express the response to *any* signal just as well.
- In other words: We can switch back and forth between ***P*(τ)** and ***P*(ω)** via a **Fourier transformation**.
  - We already know the time dependence ***P*(τ)** for a switch-on / switch-off signal, and from that we can - in principle - derive ***P*(ω)**.
  - We thus have to consider the **Fourier transform** of ***P*(*t*)**. However, while clear in principle, details can become nasty. While some details are given in an advanced module, here it must suffice to say that our Fouriertransform is given by

$$\boxed{P(\omega) \;=\; \int_0^\infty P_0 \cdot \exp - \frac{t}{\tau} \cdot \exp - (i\omega t) \cdot dt}$$

  - P_0 is the static polarization, i.e. the value of ***P*(ω)** for **ω = 0 Hz** , and $i = (-1)^{1/2}$ is the imaginary unit (note that in electrical engineering usually the symbol **j** is used instead of **i**).
- This is an easy integral, we obtain

$$P(\omega) \;=\; \frac{P_0}{\omega_0 + i \cdot \omega}$$

$$\omega_0 \;=\; \frac{1}{\tau}$$

  - Note that ω_0 is *not* **2π/τ**, as usual, but just **1/τ**. That does not mean anything except that it makes writing the formulas somewhat easier.
- The ***P*(ω)** then are the *Fourier coefficients* if you describe the ***P*(*t*)** curve by a Fourier integral (or series, if you like that better, with infinitesimally closely spaced frequency intervals).
  - ***P*(ω)** thus is the polarization response of the system if you jiggle it with an electrical field given by $E = E_0 \cdot \exp(i\omega t)$ that contains just one frequency ω.
  - However, our Fourier coefficients are *complex numbers*, and we have to discuss what that means now.

Usimg Complex Numbers and Functions

- Using the powerful math of complex numbers, we end up with a *complex* polarization. That need not bother us since by convention we would only consider the *real part* of ***P*** when we are in need of real numbers.
  - Essentially, we are done. If we know the Amplitude (= E_0) and (circle) frequency ω of the electrical field in the material (taking into account possible "local field" effects), we know the polarization.
- However, there is a smarter way to describe that relationship than the equation above, with the added benefit that this "smart" way can be generalized to all frequency dependent polarization phenomena. Let's see how it is done:
- What we want to do, is to keep our basic equation that couples polarization and field strength for alternating fields, too. This requires that the susceptibility χ becomes frequency dependent. We then have

$$\boxed{P(\omega) = \epsilon_0 \cdot \chi(\omega) \cdot E(\omega)}$$

  - and the decisive factor, giving the *amplitude* of ***P*(ω)**, is **χ(ω)**.
  - The time dependence of ***P*(ω)** is trivial. It is either given by $\exp i(\omega t - \phi)$, with ϕ accounting for a possible phase shift, or simply by $\exp i(\omega t)$ if we include the phase shift in **χ(ω)**, which means it must be complex.
  - The second possibility is more powerful, so that is what we will do. If we then move from the polarization ***P*** to the more conventional electrical displacement ***D***; the relation between ***D*(ω)** and ***E*(ω)** will require a *complex dielectric function* instead of a complex susceptibiltiy, and that is the quantity we will be after from now on.

  - It goes without saying that for more complex time dependencies of the electrical field, the equation above holds for every for every **sin** component of the Fourier series of an arbitrary periodic function.

- Extracting a *frequency dependent susceptibility* $\chi(\omega)$ from our equation for the polarization is fairly easy: Using the basic equation we have

$$\epsilon_0 \cdot \chi(\omega) = \frac{P(\omega)}{E(\omega)} = \frac{P_0}{E_0} \cdot \frac{1}{\omega_0 + i \cdot \omega} = \chi_s \cdot \frac{1}{1 + i \cdot \omega/\omega_0}$$

  - $\chi_s = P_0 / E_0$ is the *static susceptibility*, i.e. the value for zero frequency.

- Presently, we are only interested in the *real* part of the complex susceptibility thus obtained. As any complex number, we can decompose $\chi(\omega)$ in a real and a imaginary part, i.e. write it as

$$\chi(\omega) = \chi'(\omega) + i \cdot \chi''(\omega)$$

  - with χ' and χ'' being the real and the imaginary part of the complex susceptibility χ . *We drop the* (ω) *by now, because whenever we discuss real and imaginary parts it is clear that we discuss frequency dependence).*
  - All we have to do in order to obtain χ' and χ'' is to expand the fraction by $1 - i \cdot \omega/\omega_0$ which gives us

$$\epsilon_0 \cdot \chi(\omega) = \frac{\chi_s}{1 + (\omega/\omega_0)^2} - i \cdot \frac{\chi_s \cdot (\omega/\omega_0)}{1 + (\omega/\omega_0)^2}$$

  - We thus have for the real and imaginary part of $\epsilon_0 \cdot \chi(\omega)$, which is almost, but not yet quite the *dielectric function* that we are trying to establish:

$$\epsilon_0 \cdot \chi' = \frac{\chi_s}{1 + (\omega/\omega_0)^2}$$

$$-\epsilon_0 \cdot \chi'' = \frac{\chi_s \cdot (\omega/\omega_0)}{1 + (\omega/\omega_0)^2}$$

- This is pretty good, because, as we will see, the real *and* imaginary part of the complex susceptibility contain an unexpected wealth of material properties. Not only the dielectric behavior, but also (almost) all optical properties and essentially also the conductivity of non-perfect dielectrics.

- Before we proceed to the *dielectric function* which is what we really want to obtain, we have to makes things a tiny bit more complicated - in three easy steps.
  - **1.** People in general like the dielectric constant ϵ_r as a material parameter far better than the susceptibility χ - history just cannot be ignored, even in physics. Everything we did above for the polarization ***P***, we could also have done for the dielectric flux density ***D*** - just replace the letter "***P***" by "***D***" and "χ" by "ϵ_r" and we obtain a complex frequency dependent dielectric constant $\epsilon_r(\omega) = \chi(\omega) + 1$ with, of course, ϵ_s instead of χ_s as the zero frequency static case.
  - **2.** So far we assumed that at very large frequencies the polarization is essentially zero - the dipole cannot follow and $\chi(\omega \to \infty) = 0$That is not necessarily true in the most general case - there might be, after all, other mechanisms that still "work" at frequencies far larger than what orientation polarization can take. If we take that into account, we have to change our consideration of relaxation somewhat and introduce the new, but simple parameter $\chi(\omega >> \omega_0) = \chi_\infty$ or, as we prefer, the same thing for the dielectric "constant", i.e. we introduce $\epsilon_r(\omega >> \omega_0) = \epsilon_\infty$.
  - **3.** Since we always have either $\epsilon_0 \cdot \chi(\omega)$ or $\epsilon_0 \cdot \epsilon(\omega)$, and the ϵ_0 is becoming cumbersome, we may just include it in what we now call the **dielectric function** $\epsilon(\omega)$ of the material. This simply means that all the ϵ_i are what they are as the relative dielectric "constant" and multiplied with ϵ_0

- This reasoning follows **Debye**, who by doing this expanded our knowledge of materials in a major way. Going through the points **1. - 3.** (which we will not do here), produces the final result for the frequency dependence of the *orientation polarization*, the so-called **Debye equations**:
  - In general notation we have pretty much the same equation as for the susceptibility χ; the only real difference is the introduction of ϵ_∞ for the high frequency limit:

$$D(\omega) \; = \; \epsilon(\omega) \cdot E(\omega) = \left(\frac{\epsilon_s - \epsilon_\infty}{1 + i\,(\omega/\omega_0)} + \epsilon_\infty \right) \cdot E(\omega)$$

- The complex function ϵ (ω) is the **dielectric function**. In the equation above it is given in a closed form for the dipole relaxation mechanism.

- Again, we write the complex function as a sum of a real part and a complex part, i.e. as $\epsilon(\omega) = \epsilon'(\omega) - i \cdot \epsilon''(\omega)$. We use a "**–**" sign, as a matter of taste; it makes some follow-up equations easier. But you may just as well define it with a **+** sign and in some books that is what you will find. For the dielectric function from above we now obtain

$$\epsilon' \; = \; \epsilon_\infty \; + \; \frac{\epsilon_s - \epsilon_\infty}{1 + (\omega/\omega_0)^2}$$

$$\epsilon'' \; = \; \frac{(\omega/\omega_0)(\epsilon_s - \epsilon_\infty)}{1 + (\omega/\ \omega_0)^2}$$

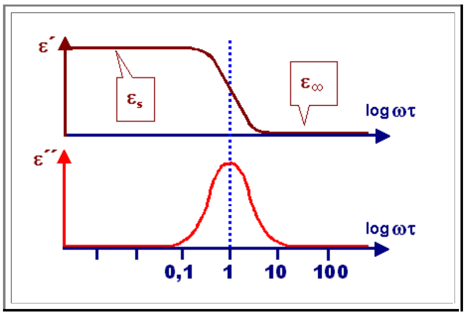
  - As it *must* be, we have

$$\epsilon'(\omega = 0) \; = \; \epsilon_s \qquad \epsilon'(\omega = 0) = \; 0$$

$$\epsilon'(\omega \rightarrow \infty) \; = \; \epsilon_\infty$$

- From working with the complex notation for sin- and cosin-functions we also know that
  - **ϵ'**, the *real part* of a complex amplitude, gives the amplitude of the response that is *in phase* with the driving force, **ϵ''**, the *imaginary part*, gives the amplitude of the response that is *phase-shifted by 90°*.

- Finally, we can ask ourselves: What does it look like? What are the graphs of **ϵ'** and **ϵ''**?
  - Relatively simple curves, actually, They always look like the graphs shown below, the three numbers that define a particular material (ϵ_s, ϵ_∞, and $\tau = 2\pi/\omega_0$) only change the numbers on the scales.



  - Note that ω for curves like this one is always on a *logarithmic scale!*

- What the dielectric function for orientation polarization looks like for real systems can be tried out with the JAVA applet below - compare that with the measured curves for water. We have a theory for the frequency dependence which is *pretty good*!

- Since ϵ_∞ must be **= 1** (or some value determined by some *other* mechanism that also exists) if we go to frequencies high enough, the essential parameters that characterize a material with orientation polarization are ϵ_s and τ (or ω_0).
  - ϵ_s we can get from the polarization mechanism for the materials being considered. If we know the dipole moments of the particles and their density, the Langevin function gives the (static) polarization and thus ϵ_s.
  - We will *not*, however, obtain τ from the theory of the polarization considered so far. Here we have to know more about the system; for liquids, e.g., the mean time before two dipoles collide and "loose" all their memory about their previous orientation. This will be expressed in some kind of diffusion terminology, and we have to know something about the random walk of the dipoles in the liquid. This, however, will go far beyond the scope of this course.

- Suffice it to say that typical relaxation times are around **10^{-11} s**; this corresponds to *frequencies in the* ***GHz*** *range* , i.e. "**cm** -waves". We must therefore expect that typical materials exhibiting orientation polarization (e.g. water), will show some peculiar behavior in the microwave range of the electromagnetic spectrum.
  - In mixtures of materials, or in complicated materials with several different dipoles and several different relaxation times, things get more complicated. The smooth curves shown above may be no longer smooth, because they now result from a superposition of several smooth curves.
  - Finally, it is also clear that τ may vary quite a bit, depending on the material and the temperature. If heavy atoms are involved, τ tends to be larger and ***vice versa***. If movements speed up because of temperature, τ will get smaller.

3.3.3 Resonance for Ionic and Atomic Polarization

- The frequency dependence of the electronic and ionic polarization mechanisms are mathematically identical - we have a driven oscillating system with a linear force law and some damping. In the simple classical approximation used so far, we may use the universal equation describing an oscillating system driven by a force with a **sin(ω*t*)** time dependence

$$m \cdot \frac{d^2x}{dt^2} + k_F \cdot m \cdot \frac{dx}{dt} + k_S \cdot x = q \cdot E_0 \cdot \exp(i\,\omega\,t)$$

  - With m = mass, k_F = **friction** coefficient; describing damping, k_S = "spring" coefficient or constant; describing the restoring force, $q \cdot E_0$ = amplitude times charge to give a force , $E = E_0 \cdot \exp(i\omega t)$ is the time dependence of electrical field in complex notation.
  - This is of course a gross simplification: In the equation above we look at *one* mass ***m*** hooked up to *one* spring, whereas a crystal consists of a hell of a lot of masses (= atoms), all coupled by plenty of springs (= bonds). Nevertheless, the analysis of just one oscillating mass provides the basically correct answer to our quest for the frequency dependence of the ionic and atomic polarization. More to that in link.
- We know the *"spring" coefficient* for the electronic and ionic polarization mechanism; however, we do not know from our simple consideration of these two mechanisms the *"friction" term*.
  - So lets just consider the general solution to the differential equation given above in terms of the general constants k_S and k_F and see what kind of general conclusions we can draw.
- From classical mechanics we know that the system has a resonance frequency ω_0, the frequency with the maximum amplitude of the oscillation, that is (for undamped oscillators) always given by

$$\omega_0 = \left(\frac{k_S}{m}\right)^{1/2}$$

  - The general solution of the differential equation is

$$x(\omega, t) = x(\omega) \cdot \exp(i\omega t + \phi)$$

  - The angle ϕ is necessary because there might be some phase shift. This phase shift, however, is automatically taken care of if we use a complex amplitude. The complex $x(\omega)$ is given by

$$x(\omega) = \frac{q \cdot E_0}{m}\left(\left(\frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + k_F^2\,\omega^2}\right) - i \cdot \left(\frac{k_F\,\omega}{(\omega_0^2 - \omega^2)^2 + k_F^2\,\omega^2}\right)\right)$$

  - $x(\omega)$ indeed is a *complex function*, which means that the amplitude is not in phase with the driving force if the imaginary part is not zero.
- Again, we are interested in a relation between the **sin** components of the polarization $P(\omega)$ and the **sin** components of the driving field $E = E_0 \cdot \exp(i\omega t)$ or the dielectric flux $D(\omega)$ and the field. We have

$$P = N \cdot q \cdot x(\omega)$$

$$D = \epsilon_0 \cdot \epsilon_r \cdot E = \epsilon_0 \cdot E + P = \epsilon_0 \cdot E + N \cdot q \cdot x(\omega)$$

- If we insert $x(\omega)$ from the solution given above, we obtain a complex relationship between D and E

$$D = \left(\epsilon_0 + \frac{N \cdot q^2}{m}\left(\left(\frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + k_F^2\,\omega^2}\right) - i\left(\frac{k_F\,\omega}{(\omega_0^2 - \omega^2)^2 + k_F^2\,\omega^2}\right)\right)\right) \cdot E$$

  - This looks pretty awful, but it encodes basic everyday knowledge!

- This equation can be rewritten using the **dielectric function** defined before with the added generalization that we now define it for the permittivity, i.e, for

$$\epsilon(\omega) = \epsilon_r(\omega) \cdot \epsilon_0 = \epsilon'(\omega) - i \cdot \epsilon''(\omega)$$

  - For the dielectric flux ***D***, which we prefer in this case to the polarization ***P***, we have as always

$$D(\omega, t) = [\epsilon'(\omega) - i \cdot \epsilon''(\omega)] \cdot E_0 \cdot \exp(i\omega t)$$

  - The time dependence of ***D*** is simple given by **exp (i**ω ***t*)**, so the interesting part is only the ω - dependent factor.

- Rewriting the equations for the real and imaginary part of ε we obtain the *general dielectric function for resonant polarization mechanisms*:

$$\epsilon' = \epsilon_0 + \frac{N \cdot q^2}{m} \left(\frac{\omega_0^2 - \omega^2}{(\omega_0^2 - \omega^2)^2 + k_F^2 \cdot \omega^2} \right)$$

$$\epsilon'' = \frac{N \cdot q^2}{m} \left(\frac{k_F \cdot \omega}{(\omega_0^2 - \omega^2)^2 + k_F^2 \cdot \omega^2} \right)$$

- These formula describe the frequency dependence of the dielectric constant of *any* material where the polarization mechanism is given by separating charges with mass ***m*** by an electrical field against a *linear* restoring force.
  - For the *limiting cases* we obtain for the real and imaginary part

$$\epsilon'(\omega = 0) = \left(\epsilon_0 + \frac{N \cdot q^2}{m} \right) \frac{1}{\omega_0^2} = \left(\epsilon_0 + \frac{N \cdot q^2}{m} \right) \frac{m}{k_S}$$

$$\epsilon'(\omega = \infty) = \epsilon_0$$

  - For **ε'(**ω **=** ∞**)** we thus have $\epsilon_r = \epsilon'/\epsilon_0 = 1$ as must be.

- The most important material parameters for dielectric constants at the low frequency limit, i.e. ω ⇒ **0**, are therefore the *masses* ***m*** of the oscillating charges, their *"spring" constants* k_S, their *density* ***N***, and the *charge* ***q*** on the ion considered.
  - We have no big problem with these parameters, and that includes the *"spring" constants*. It is a direct property of the bonding situation and in principle we know how to calculate its value.
  - The friction constant k_F does not appear in the limit values of ε. As we will see below, it is only important for frequencies around the resonance frequency.

- For this intermediate case k_F is the difficult parameter. On the atomic level, "friction" in a classical sense is *not defined*, instead we have to resort to *energy dispersion mechanisms*. While it is easy to see how this works, it is difficult to calculate numbers for k_F.
  - Imagine a single oscillating dipole in an ionic crystal. Since the vibrating ions are coupled to their neighbours via binding forces, they will induce this atoms to vibrate, too - in time the whole crystal vibrates. The ordered energy originally contained in the vibration of *one* dipole (ordered, because it vibrated in field direction) is now dispersed as *unordered* thermal energy throughout the crystal.
  - Since the energy contained in the original vibration is constant, the net effect on the single oscillating dipole is that of *damping* because its original energy is now spread out over many atoms. Formally, damping or energy dispersion can be described by some fictional "friction" force.
  - Keeping that in mind it is easy to see that all mechanisms, especially interaction with phonons, that convert the energy in an *ordered vibration in field direction* to *unordered thermal energy* always appears as a kind of friction force on a particular oscillator. Putting a number on this fictional friction force, however, is clearly a different (and difficult) business.
  - However, as soon as you realize that the dimension of k_F is **1/s** and that $1/k_F$ simply is about the time that it takes for an oscillation to "die", you can start to have some ideas - or you check the link.

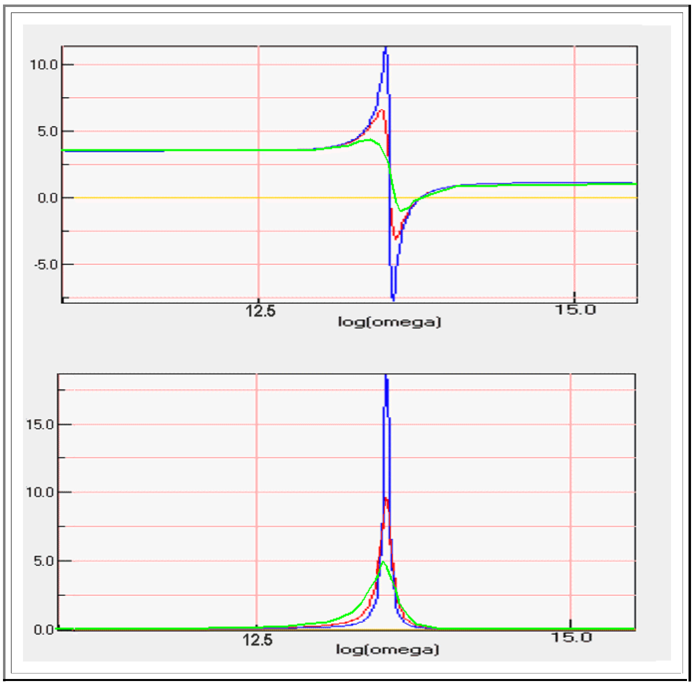
- Now lets look at some characteristic behavior and some numbers as far as we can derive them in full generality.
  - For the electronic polarization mechanism, we know the force constant, it is

$$\boxed{k_S = \frac{(ze)^2}{4\pi \cdot \epsilon_0 \cdot R^3}}$$

  - With the proper numbers for a hydrogen atom we obtain

$$\boxed{\omega_0 \approx 5 \cdot 10^{16}\ \mathrm{Hz}}$$

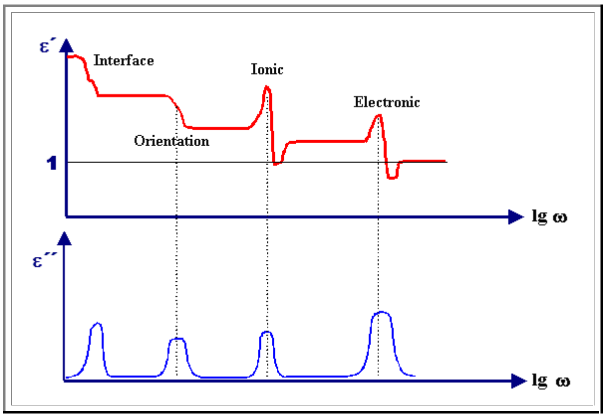
  - This is in the *ultraviolet region of electromagnetic radiation*. For all other materials we would expect similar values because the larger force constants (**$(ze)^2$** overcompensates the increasing size ***R***) is balanced to some extent by the larger mass.
  - For the *ionic polarization mechanism*, the masses are several thousand times higher, the resonance frequency thus will be considerably lower. It is, of course simply the frequency of the general lattice vibrations which, as we know, is in the **10^{13} Hz** range
- This has an important consequence:
  - The dielectric constant at frequencies higher than about the frequency corresponding to the **UV** part of the spectrum is always **1**. And since the optical index of refraction ***n*** is directly given by the DK (**$n = \epsilon^{1/2}$**), there are no optical lenses beyond the **UV** part of the spectrum.
  - In other words: You can not built a deep-**UV** or **X**-ray microscope with lenses, nor - unfortunately - lithography machines for chips with smallest dimension below about **0,2 µm**. For the exception to this rule see the footnote from before.
- If we now look at the characteristic behavior of ω **'** and ω **''** we obtain quantitatively the following curves (by using the JAVA module provided for in the link):



  - Note that ω is again on a *logarithmic scale!*
  - Note also that it is perfectly possible that **ϵ'** and therefore **ϵ_r** becomes **negative**. We won't go into what that means. however.
- The colors denote different friction coefficients ***k_F***. If ***k_F*** would be zero, the amplitude and therefore **ϵ'** would be ∞ at the resonance point, and **ϵ''** would be zero everywhere, and infinity at the resonance frequency; i.e. **ϵ''** is the Delta function.
  - While this can never happen in reality, we still may expect significantly larger ϵ values around the resonance frequency than in any other frequency region.
- That the maximum value of **ϵ''** *decreases* with increasing damping might be a bit counter-intuitive at first (in fact it was shown the wrong way in earlier versions of this Hyperscript), but for that it extends over ever increasing regions in frequency.

3.3.4 Complete Frequency Dependence of a Model Material

- The frequency dependence of a given material is superposition of the various mechanisms at work in this material. In the *idealized* case of a model material containing *all four basic mechanisms in their pure form* (a non-existent material in the real world), we would expect the following curve.



  - Note that ω is once more on a *logarithmic scale!*

- This is *highly idealized* - there is no material that comes even close! Still, there is a clear structure. Especially there seems to be a correlation between the real and imaginary part of the curve. That is indeed the case; *one* curve contains *all the information* about the other.
  - *Real* dielectric functions usually are only interesting for a small part of the spectrum. They may contain fine structures that reflect the fact that there may be *more than one* mechanism working at the same time, that the oscillating or relaxing particles may have to be treated by *quantum mechanical* methods, that the material is a *mix* of several components, and so on.
  - In the link a real dielectric function for a more complicated molecule is shown. While there is a lot of fine structure, the basic resonance function and the accompanying peak for **ε''** is still clearly visible.

- It is a general property of complex functions describing physical reality that under certain very general conditions, the real and imaginary part are directly related. The relation is called **Kramers-Kronig relation**; it is a *mathematical*, not a *physical* property, that only demands two very general conditions to be met:
  - Since two functions with a time or frequency dependence are to be correlated, one of the requirements is *causality*, the other one *linearity*.
  - The Kramers-Kronig relation can be most easily thought of as a *transformation* from one function to another by a black box; the functions being inputs and outputs. *Causality* means that there is no output before an input; *linearity* means that twice the input produces twice the output. Otherwise, the transformation can be anything.

- The Kramers-Kronig relation can be written as follows: For any complex function, e.g. $\epsilon(\omega) = \epsilon'(\omega) + i\epsilon''(\omega)$, we have the relations

$$\epsilon'(\omega) = \frac{-2\,\omega}{\pi} \int_0^\infty \frac{\omega^* \cdot \epsilon''(\omega^*)}{\omega^{*2} - \omega^2} \cdot d\omega^*$$

$$\epsilon''(\omega) = \frac{2\,\omega}{\pi} \int_0^\infty \frac{\epsilon'(\omega^*)}{\omega^{*2} - \omega^2} \cdot d\omega^*$$

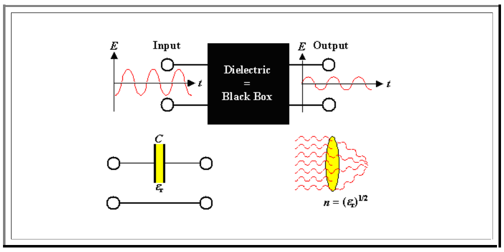
- The Kramers-Kronig relation can be very useful for experimental work. If you want to have the dielectric function of some materials, you only have to measure one component, the other one can be calculated.

Questionaire

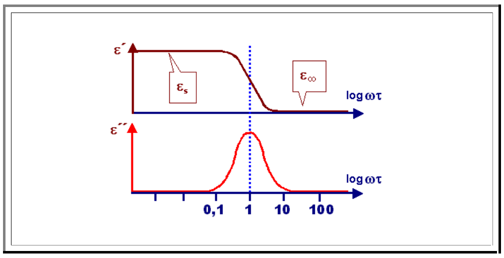
Multiple Choice questions to all of 3.3

3.3.5 Summary to: Frequency Dependence of the Dielectric Constant

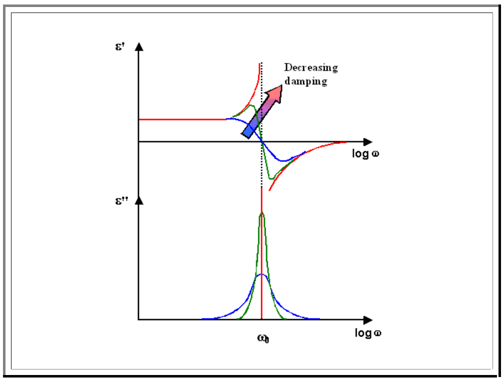
- Alternating electrical fields induce alternating forces for dielectric dipoles. Since in all polarization mechanisms the dipole response to a field involves the movement of masses, inertia will prevent arbitrarily fast movements.
  - Above certain limiting frequencies of the electrical field, the polarization mechanisms will "die out", i.e. not respond to the fields anymore.
  - This might happen at rather high (= optical) frequencies, limiting the index of refraction $n = (\epsilon_r)^{1/2}$



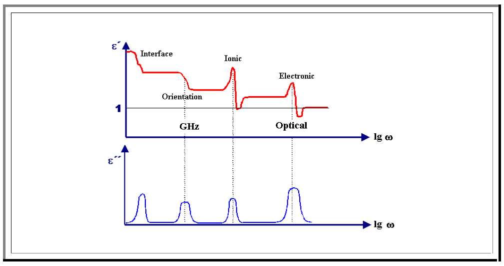
- The (only) two physical mechanisms governing the movement of charged masses experiencing alternating fields are relaxation and resonance.
- *Relaxation* describes the decay of excited states to the ground state; it describes, e.g., what happens for orientation polarization after the field has been switched off.
  - From the "easy to conceive" time behavior we deduce the frequency behavior by a Fourier transformation
  - The dielectric function describing relaxation has a typical frequency dependence in its real and imaginary part ⇒



- Resonance describes anything that can be modeled as a mass on a spring - i.e. electronic polarization and ionic polarization.
  - The decisive quantity is the (undamped) resonance frequency $\omega_0 = (k_S / m)^{½}$ and the "friction" or damping constant k_F
  - The "spring" constant is directly given by the restoring forces between charges, i.e. Coulombs law, or (same thing) the bonding. In the case of bonding (ionic polarization) the spring constant is also easily expressed in terms of Young's modulus Y. The masses are electron or atom masses for electronic or ionic polarization, respectively.
  - The damping constant describes the time for funneling off ("dispersing") the energy contained in one oscillating mass to the whole crystal lattice. Since this will only take a few oscillations, damping is generally large.
  - The dielectric function describing relaxation has a typical frequency dependence in its real and imaginary part ⇒
    The green curve would be about right for crystals.



- The complete frequency dependence of the dielectric behavior of a material, i.e. its dielectric function, contains all mechanisms "operating" in that material.
  - As a rule of thumb, the critical frequencies for relaxation mechanisms are in the**GHz** region, electronic polarization still "works" at optical (**10^{15} Hz**) frequencies (and thus is mainly responsible for the index of refraction).
  - Ionic polarization has resonance frequencies in between.
  - Interface polarization may "die out" already a low frequencies.
- A widely used diagram with all mechanisms shows this, but keep in mind that there is no real material with all 4 major mechanisms strongly present!
  ⇒



- A general mathematical theorem asserts that the real and imaginary part of the dielectric function cannot be completely independent
  - If you know the complete frequency dependence of either the real or the imaginary part, you can calculate the complete frequency dependence of the other.
  - This is done via the Kramers-Kronig relations; very useful and important equations in material practice.
    ⇒

$$\epsilon'(\omega) = \frac{-2\omega}{\pi}\int_0^\infty \frac{\omega^* \cdot \epsilon''(\omega^*)}{\omega^{*2} - \omega^2} \cdot d\omega^*$$

$$\epsilon''(\omega) = \frac{2\omega}{\pi}\int_0^\infty \frac{\epsilon'(\omega^*)}{\omega^{*2} - \omega^2} \cdot d\omega^*$$

Questionaire
Multiple Choice questions to all of 3.3

3.4. Dynamic Properties

3.4.1 Dielectric Losses

- The electric power (density) ***L*** lost per volume unit in any material as heat is *always* given by

$$\boxed{\boldsymbol{L} = \boldsymbol{j} \cdot \boldsymbol{E}}$$

  - With ***j*** **=** current density, and ***E*** **=** electrical field strength.

- In our ideal dielectrics there is *no direct current*, only **displacement currents *j*(ω) = d*D*/d*t*** may occur for alternating voltages or electrical fields. We thus have

$$\boldsymbol{j}(\omega) = \frac{\mathrm{d}\boldsymbol{D}}{\mathrm{d}t} = \epsilon(\omega)\cdot\frac{\mathrm{d}\boldsymbol{E}}{\mathrm{d}t} = \epsilon(\omega)\cdot\frac{\mathrm{d}[\boldsymbol{E}_0\exp(i\omega t)]}{\mathrm{d}t} = \epsilon(\omega)\cdot i\cdot\omega\cdot\boldsymbol{E}_0\cdot\exp(i\omega t) = \epsilon(\omega)\cdot i\cdot\omega\cdot\boldsymbol{E}(\omega)$$

  - (Remember that the dielectric function **ϵ(ω)** includes ϵ_0).

- With the dielectric function written out as **ϵ(ω) = ϵ'(ω) – i · ϵ''(ω)** we obtain

$$\boldsymbol{j}(\omega) = \underbrace{\omega\cdot\epsilon''\cdot\boldsymbol{E}(\omega)}_{\substack{\text{real part}\\ \text{of } \boldsymbol{j}(\omega);\\ \text{in phase}}} + \underbrace{i\cdot\omega\cdot\epsilon'\cdot\boldsymbol{E}(\omega)}_{\substack{\text{imaginary part}\\ \text{of } \boldsymbol{j}(\omega)\\ 90^{\circ}\text{ out of phase}}}$$

- That part of the displacement current that is *in phase* with the electrical field is given by **ϵ''**, the *imaginary* part of the dielectric function; that part that is **90°** *out of phase* is given by the *real* part of **ϵ(ω)**. The power losses thus have two components

Active power 1)

$$\boxed{L_A = \textbf{power really lost, turned into heat} = \omega\cdot|\epsilon''|\cdot E^2}$$

Reactive power

$$\boxed{L_R = \textbf{power extended and recovered each cycle} = \omega\cdot|\epsilon'|\cdot E^2}$$

1) Other possible expressions are:
actual power, effective power, real power, true power

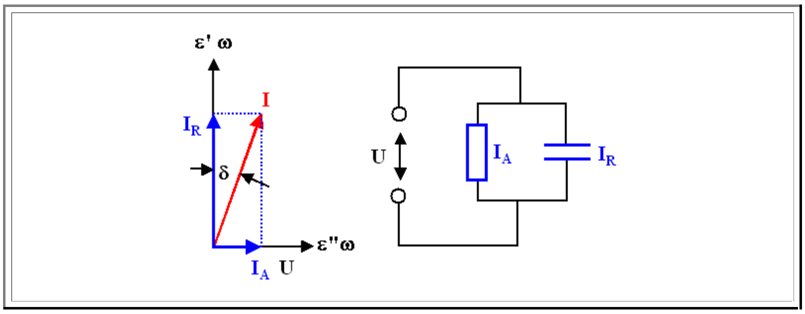
  - Remember that active, or effective, or true power is energy deposited in your system, or, in other words, it is the power that heats *up your material*! The reactive power is just cycling back and forth, so it is not heating up anything or otherwise leaving direct traces of its existence.

- The first important consequence is clear:
  - We can heat up even a "perfect" (= perfectly none **DC**-conducting material) by an **AC** voltage; most effectively at frequencies around its resonance or relaxation frequency, when **ϵ''** is always maximal.
  - Since **ϵ''** for the resonance mechanisms is directly proportional to the friction coefficient k_R, the amount of power lost in these cases thus is directly given by the amount of "friction", or power dissipation, which is as it should be.

- It is conventional, for reason we will see immediately, to use the quotient of L_A / L_R as a measure of the "quality" of a dielectric: this quotient is called "**tangens delta**" (**tg** δ) and we have

$$\frac{L_A}{L_R} := \mathrm{tg}\,\delta = \frac{I_A}{I_R} = \frac{\epsilon''}{\epsilon'}$$

  - Why this somewhat peculiar name was chosen will become clear when we look at a pointer representation of the voltages and currents and its corresponding equivalent circuit. This is a perfectly legal thing to do: We always can represent the current from above this way; in other words we can always model the behaviour of a real dielectric onto an **equivalent circuit diagram** consisting of an *ideal* capacitor with **C(ω)** and an *ideal* resistor with **R(ω)**.



- The current I_A flowing through the *ohmic resistor* of the equivalent circuit diagram is in phase with the voltage U; it corresponds to the imaginary part **ε"** of the dielectric function times ω.
  - The **90°** out-of-phase current I_R flowing through the *"perfect" capacitor* is given by the real part **ε'** of the dielectric function times ω.
  - The numerical values of both elements must depend on the frequency, of course - for ω **= 0**, ***R*** would be infinite for an ideal (non-conducting) dielectric.
  - The smaller the angle δ or **tg** δ, the better with respect to power losses.
- Using such an **equivalent circuit diagram** (with always "ideal" elements), we see that a *real* dielectric may be modeled by a fictitious "ideal" dielectric having no losses (something that does not exist!) with an ohmic resistor in parallel that represents the losses. The value of the ohmic resistor (and of the capacitor) must depend on the frequency; but we can easily derive the necessary relations.
  - How large is ***R***, the more interesting quantity, or better, the *conductivity* σ of the material that corresponds to ***R***? Easy, we just have to look at the equation for the current from above.
  - For the in-phase component we simply have

$$j(\omega) \;=\; \omega \cdot \epsilon'' \cdot E(\omega)$$

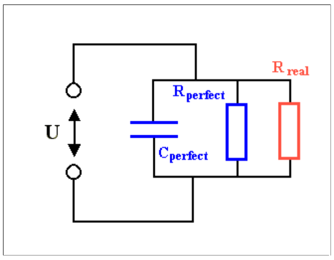
  - Since we *always* can express an in-phase current by the conductivity σ via

$$j(\omega) \;:=\; \sigma(\omega) \cdot E(\omega)$$

  - we have

$$\sigma_{DK}(\omega) \;=\; \omega \cdot \epsilon''(\omega)$$

- In other words: The dielectric losses occuring in a perfect dielectric are completely contained in the imaginary part of the dielectric function and express themselves as if the material would have a frequency dependent conductivity σ_{DK} as given by the formula above.
  - This applies to the case where our dielectric is still a *perfect* insulator at **DC** (ω **= 0 Hz)**, or, a bit more general, at low frequencies; i.e. for $\sigma_{DK}(\omega \to 0) = 0$.
- However, nobody is perfect! There is no *perfect* insulator, at best we have *good* insulators. But now it is easy to see what we have to do if a *real* dielectric is *not* a perfect insulator at low frequencies, but has some finite conductivity σ_0 even for ω **= 0**. Take water with some dissolved salt for a simple and relevant example.
  - In this case we simple *add* σ_0 to σ_{DK} to obtain the total conductivity responsible for power loss



$$\sigma_{total} \;=\; \sigma_{perfect} \;+\; \sigma_{real}$$

$$=\; \sigma_{DK} \;+\; \sigma_0$$

  - Remember: For resistors in parallel, you *add* the conductivities (or**1/R**'s) ; it is with *resistivities* that you do the $1/R_{total} = 1/R_1 + 1/R_2$ procedure.

- Since it is often difficult to separate σ_{DK} and σ_0, it is convenient (if somewhat confusing the issue), to use σ_{total} in the imaginary part of the dielectric function. We have

$$\epsilon'' = \frac{\sigma_{total}}{\omega}$$

  - We also have a completely general way now, to describe the response of *any* material to an electrical field, because we now can combine dielectric behavior and conductivity in the complete dielectric function of the material.
  - Powerful, but only important at high frequencies; as soon as the imaginary part of the "perfect" dielectric becomes noticeable. But high frequencies is where the action is. As soon as we hit the high **THz** region and beyond, we start to call what we do "*Optics*", or "*Photonics*", but the material roots of those disciplines we have right here.
- In classical electrical engineering at not too large frequencies, we are particularily interested in the relative magnitude of both current contributions, i.e in **tg**δ. From the pointer diagram we see directly that we have

$$\frac{I_A}{I_R} = \mathbf{tg}\ \delta$$

- We may get an expression for **tg** δ by using for example the Debye equations for ϵ' and ϵ'' derived for the dipole relaxation mechanism:

$$\mathbf{tg}\ \delta = \frac{\epsilon''}{\epsilon'} = \frac{(\epsilon_s - \epsilon_\infty) \cdot \omega / \omega_0}{\epsilon_s + \epsilon_\infty \cdot \omega^2 / \omega_0^2}$$

  - or, for the normal case of $\epsilon_\infty = 1$ (or , more correctly ϵ_0)

$$\mathbf{tg}\ \delta = \frac{(\epsilon_s - 1) \cdot \omega / \omega_0}{\epsilon_s + \omega^2 / \omega_0^2}$$

  - This is, of course, only applicable to real *perfect* dielectrics, i.e. for real dielectrics with $\sigma_0 = 0$.
- The total power loss, the *really interesting quantity*, then becomes (using $\epsilon'' = \epsilon' \cdot \mathbf{tg}\delta$, because **tg**$\delta$ is now seen as a *material parameter*) .

$$L_A = \omega \cdot \epsilon' \cdot E^2 \cdot \mathbf{tg}\ \delta$$

- This is a useful relation for a dielectric with a given **tg** δ (which, for the range of frequencies encountered in "normal" electrical engineering is approximately constant). It not only gives an idea of the electrical losses, but also a very rough estimate of the break-down strength of the material. If the losses are large, it will heat up and this always helps to induce immediate or (much worse) eventual breakdown.
- We also can see now what happens if the dielectric is *not ideal* (i.e. totally insulating), but slightly conducting:
  - We simply include σ_0 in the definition of **tg**δ (and then automatically in the value of ϵ'').
  - **tg** δ is then non-zero even for low frequencies - there is a constant loss of power into the dielectric. This may be of some consequence even for small **tg** δ values, as the example will show:
- The **tg** δ value for regular (cheap) insulation material as it was obtainable some **20** years ago at very low frequencies (**50 Hz**; essentially **DC**) was about **tg** δ **= 0,01**.
  - Using it for a high-voltage line (U **= 300 kV**) at moderate field strength in the dielectric (E **= 15MV/m**; corresponding to a thickness of **20 mm**), we have a loss of **14 kW/m**^{3} of dielectric, which translates into about **800 m** high voltage line. So there is little wonder that high-voltage lines were not insulated by a dielectric, but by air until rather recently!
- Finally, some examples for the **tg** δ values for commonly used materials (and low frequencies):

Material	ϵ_r	**tg δ** x 10^{-4}
Al_2O_3 (very good ceramic)	10	5....20
SiO_2	3,8	2
$BaTiO_3$	500 (!!)	150
Nylon	3,1	10...0,7
Poly...carbonate, ...ethylene ...styrol	about 3	
PVC	3	160

And now you understand how the microwave oven works and why it is essentially heating only the water contained in the food.

Questionaire

Multiple Choice questions to 3.2.1

3.4.2 Summary to: Dynamic Properties - Dielectric Losses

The frequency dependent current density *j* flowing through a dielectric is easily obtained. ⇒

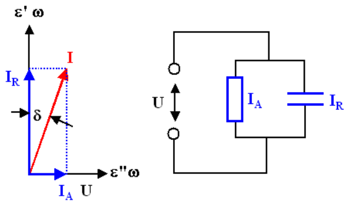
$$j(\omega) = \frac{dD}{dt} = \epsilon(\omega)\cdot\frac{dE}{dt} = \underset{\textbf{in phase}}{\omega\cdot\epsilon''\cdot E(\omega)} + \underset{\textbf{out of phase}}{i\cdot\omega\cdot\epsilon'\cdot E(\omega)}$$

- The in-phase part generates active power and thus heats up the dielectric, the out-of-phase part just produces reactive power
- The power losses caused by a dielectric are thus directly proportional to the imaginary component of the dielectric function

$$L_A = \text{power turned into heat} = \omega\cdot|\epsilon''|\cdot E^2$$

The relation between active and reactive power is called "tangens Delta" (**tg(δ)**); this is clear by looking at the usual pointer diagram of the current

$$\frac{L_A}{L_R} := \text{tg}\,\delta = \frac{I_A}{I_R} = \frac{\epsilon''}{\epsilon'}$$



- The pointer diagram for an *ideal* dielectric **σ(ω = 0) = 0**can always be obtained form an (ideal) resistor ***R***(ω) in parallel to an (ideal) capacitor ***C***(ω).
- ***R***(ω) expresses the apparent conductivity $\sigma_{DK}(\omega)$ of the dielectric, it follows that

$$\sigma_{DK}(\omega) = \omega\cdot\epsilon''(\omega)$$

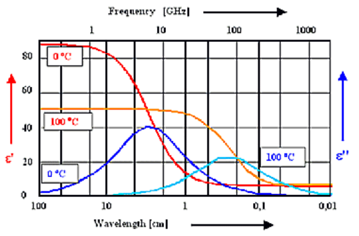
For a *real* dielectric with a non-vanishing conductivity at zero (or small) frequencies, we now just add another resistor in parallel. This allows to express *all* conductivity effects of a real dielectric in the imaginary part of its (usually measured) dielectric function via

$$\epsilon'' = \frac{\sigma_{total}}{\omega}$$

- We have no *all* materials covered with respect to their dielectric behavior - in principle even metals, but then resorting to a dielectric function would be overkill.

A good example for using the dielectric function is "dirty" water with a not-too-small (ionic) conductivity, commonly encountered in food.

- The polarization mechanism is orientation polarization, we expect large imaginary parts of the dielectric function in the **GHz** region.
- It follows that food can be heated by microwave (ovens)!



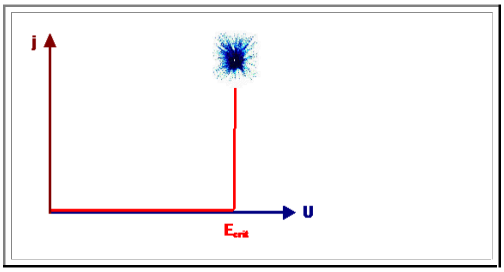
Questionaire

Multiple Choice questions to all of 3.4

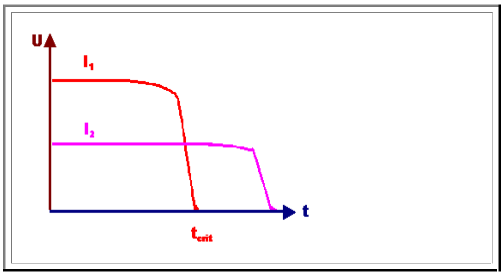
3.5 Electrical Breakdown and Failure

3.5.1 Observation of Electrical Breakdown and Failure

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



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



  - The voltage needed to press your tiny test current through the dielectric starts to decrease rapidly after some time - hours, days, weeks, ..., and this is a clear indication that you dielectric becomes increasingly leaky, and will go up in smoke soon.
  - A typical result is that breakdown of a "good" dielectric occurs after - very roughly - **1 C** of charge has been passed.
- The following table gives a rough idea of critical field strengths for certain dielectric materials

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

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

Questionnaire

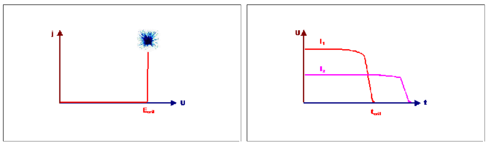
Multiple Choice questions to 3.5.1

3.5.2 Mechanisms of Electrical Breakdown

- What are the atomic mechanisms by which breakdown occurs or dielectrics fail? This is a question not easily answered because there is no general mechanism expressible in formulas. Most prominent are the following disaster scenarios:
- **Thermal breakdown**
    - A tiny little current that you can't even measure is flowing *locally* through "weak" parts of the dielectric. With increasing field strength this current increases, producing heat *locally*, which leads to the generation of point defects. Ionic conductivity sets in, more heat is produced *locally*, the temperature goes up even more.... - *boooom*!
    - This is probably the most common mechanism in run-of-the-mill materials which are usually not too perfect.
- **Avalanche breakdown**
    - Even the most perfect insulator contains a few free electron. Either because there is still a non-zero probability for electrons in the conduction band, even for large band gaps, or because defects generate some carriers, or because irradiation (natural radioactivity may be enough) produces some.
    - In large electrical field these carriers are accelerated; if the field strength is above a certain limit, they may pick up so much energy that they can rip off electrons from the atoms of the materials. A chain reaction then leads to a swift ***avalanche*** *effect*; the current rises exponentially ... *boom*!
- **Local discharge**
    - In small cavities (always present in sintered ceramic dielectrics) the field strength is even higher than the average field (ϵ is small)- a microscopic arc discharge may be initiated. Electrons and ions from the discharge bombard the inner surface and erode it. The cavity grows, the current in the arc rises, the temperature rises ... - *boooom*!
- **Electrolytic breakdown**
    - Not as esoteric as it sounds! Local electrolytical (i.e involving moving ions) current paths transport some conducting material from the electrodes into the interior of the dielectric. Humidity (especially if it is acidic) may help. In time a filigree conducting path reaches into the interior, reducing the local thickness and thus increasing the field strength. The current goes up....*booom*!
    - This is a very irreproducible mechanism because it depends on many details, especially the local environmental conditions. It may slowly built up over years before it suddenly runs away and ends in sudden break-through.

- Do the incredibly good dielectrics in integrated circuits fail eventually? After all, they are worked at very high field strength, but the field never increases much beyond its nominal value.
    - The answer is that they *do* fail. The mechanisms are far from being clear and it is one of the more demanding tasks in the field to predict the life-time of a dielectric in a chip. Empirically, however, an interesting relation has been found:
    - The dielectric will fail after a certain amount of charge has been passed through it - very roughly about **1 As**. This allows to test the chip dielectrics: A very small current is forced through the dielectric; the voltage necessary to do that is monitored. If the voltage rapidly goes down after about **1 As** of total charge has been passed, the dielectric is **OK**. Now its life time can be predicted: Since every time the voltage is on, a tiny little current flows, the life time can be roughly predicted from the leakage current and the average frequency of voltage switching. About **10 a** should be obtained.

3.5.3 Summary to: Electrical Breakdown and Failure



- The first law of materials science obtains: At field strengths larger than some critical value, dielectrics will experience (destructive) electrical breakdown
  - This might happen suddenly (then calls break-down) , with a bang and smoke, or
  - it may take time - months or years - then called failure.
  - Critical field strength may vary from **< 100 kV/cm** to **> 10 MV / cm**.

- Highest field strengths in practical applications do not necessarily occur at high voltages, but e.g. in integrated circuits for very thin (a few **nm**) dielectric layers
  - Properties of thin films may be quite different (better!) than bulk properties!

Example 1: TV set, **20 kV** cable, thickness of insulation = **2 mm.** ⇒ ***E* = 100 kV/cm**
Example 2: Gate dielectric in transistor, **3.3 nm** thick, **3.3 V** operating voltage. ⇒ ***E* = 10 MV/cm**

- Electrical breakdown is a major source for failure of electronic products (i.e. one of the reasons why things go "kaputt" (= broke)), but there is no simple mechanism following some straight-forward theory. We have:
  - *Thermal breakdown*; due to small (field dependent) currents flowing through "weak" parts of the dielectric.
  - *Avalanche breakdown* due to occasional free electrons being accelerated in the field; eventually gaining enough energy to ionize atoms, producing more free electrons in a runaway avalanche.
  - *Local discharge* producing micro-plasmas in small cavities, leading to slow erosion of the material.
  - *Electrolytic breakdown* due to some ionic micro conduction leading to structural changes by, e.g., metal deposition.

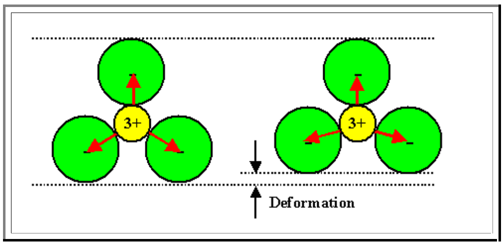
Questionaire
Multiple Choice questions to all of 3.5

3.6 Special Dielectrics

3.6.1 Piezo Electricity and Related Effects

Piezo Electricity

- The polarization of a material must not necessarily be an effect of electrical fields only; it may come about by other means, too.
  - Most prominent is the *inducement of polarization by mechanical deformation*, which is called *piezo electricity*. The reverse mechanism, the inducement of mechanical deformation by polarization, also falls under this heading.
- The principle of piezo electricity is easy to understand:



  - Let'as consider a crystal with ionic components and some arrangement of ions as shown (in parts) in the picture above. In the undeformed symmetrical arrangement, we have three dipole moments (red arrows) that exactly cancel in vector addition.
  - If we induce some elastic deformation as shown, the **symmetry is broken** and the three dipole moments no longer cancel - we have induced polarization by mechanical deformation.
  - We also realize that symmetry is somehow important. If we were to deform the "crystal" in a direction perpendicular to the drawing plane, nothing with respect to polarization would happen. This tells us:
    - Piezo electricity can be pronounced in single crystals if they are deformed in the "right" direction, while it may be absent or weak in polycrystals with randomly oriented grains.
    - Piezo electricity must be described by a **tensor** of second rank. What this means is that we must consider the full tensor properties of the susceptibility χ or the dielectric constant ϵ_r when dealing with piezoelectricity proper.
  - If one looks more closely at this, it turns out that the crystal symmetry must meet certain conditions. Most important is that it must not have an **inversion center**.
- We won't look into the tensor properties of piezoelectricity but just note that for piezo electric materials we have a general relation between polarization ***P*** and deformation ***e*** of the form

$$\boldsymbol{P} \;=\; \textbf{const.} \cdot \boldsymbol{e}$$

  - With ***e*** **=** mechanical ***strain*** **=** $\Delta l / l$ = relative change of length. (Strain is usually written as ϵ; but here we use ***e*** to avoid confusion with the dielectric constant).
- In piezo electric materials, mechanical deformation produced polarization, i.e an electrical field inside the material. The reverse then must be true, too:
  - Piezo electrical materials exposed to an electrical field will experience a force and therfore undergo mechanical deformation, i.e. get somewhat shorter or longer.
- So piezo electricity is restricted to crystals with relatively low symmetry (there must be no center of symmetry; i.e. no inversion symmetry) in single crystalline form (or at least strongly textured poly crystals). While that appears to be a rather limiting conditions, piezo electricity nevertheless has major technical uses:
  - Most prominent, perhaps, are the **quartz oscillators**, where suitable (and small) pieces of single crystals of quartz are given a very precisely and purely mechanically defined resonance frequency (as in tuning forks). Crystalline quartz happens to be strongly piezo electric; if it is polarized by an electrical field of the right frequency, it will vibrate vigorously, otherwise it will not respond. This can be used to control frequencies at a very high level of precision. More about quartz oscillators (eventually) in the link
  - Probably just as prominent by now, although a rather recent big break-through, are **fuel injectors** for advanced ("common rail") Diesel engines. Makes for more fuel efficient and clean engines and is thus a good thing. More to that in this link.The materials of choice for this mass application is ***PZT***, Lead zirconate titanate. This link gives a short description.
  - While for fuel injectors relatively large mechanical displacements are needed, the piezoelectric effect can just as well be used for precisely controlled very small movements in the order of fractions of **nm** to **µm**, as it is, e.g., needed for the scanning tunnelling microscope.

- There are many more applications (consult the links from above), e.g. for
  - Microphones.
  - Ultrasound generators.
  - Surface acoustic wave filters (**SAW**).
  - Sensors (e.g. for pressure or length).

Electrostriction

An effect that must be kept separate from the piezo electricity is **electrostriction**, where again mechanical deformation leads to polarization.

- It is an effect observed in many material, but usually much weaker than the piezo electric effect. Much simplified, the effect results if dipoles induced by electronic polarization are not exactly in field direction (e.g. in covalent bonds) and then experience a mechanical force (leading to deformation) that tries to rotate them more into the field direction.
- The deformation **e** in this case depends on the *square of the electrical field* because the field induces the dipoles *and* acts on them. We have

$$\boxed{e \;=\; \frac{\Delta l}{l} \;=\; \text{const} \cdot E^2}$$

- Because of the quadratic dependence, the sign of the field does not matter (in contrast to piezo electricity).
- There is *no* inverse effect - a deformation does not produce an electric field.

Electrostriction can be used to produce extremely small deformations in a controlled way; but it is not really much used.

Pyro Electricity

Polarization can also be induced by sudden changes in the temperature, this effect is called **pyro electricity**; it is most notably found in natural **tourmalin** crystals.

- The effect comes about because pyro electrical crystals are naturally polarized on surfaces, but this polarization is compensated by mobile ions in a "dirt" skin, so that no net polarization is observed.
- Changes in temperature change the natural polarization, but because the compensation process may take a rather long time, an outside polarization is observed for some time.

Electrets

The word "electret" is a combination of *electr*icity and magn*et* - and that tells it all:

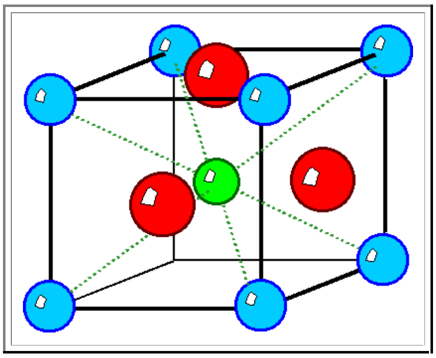
- Electrets are the electrical analog of (permanent) magnets: Materials that have a permanent macroscopic polarization or a permanent charge. Ferroelectric materials (see next sub-chapter) might be considered to be a sub-species of electrets with a permanent polarization that is "felt" if the internal domains do not cancel each other.
- Electrets that contain surplus charge that is not easily lost (like the charge on your hair after brushing it on dry days) are mostly polymers, like fluoropolymers or polyproylene.

Electrets have been a kind of scientific curiosity since the early 18th century (when people did a lot of rubbing things to generate electricity), their name was coined in 1885 by Oliver **Heaviside**

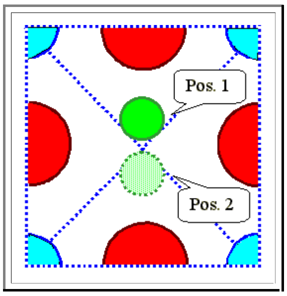
- Lately, however, they were put to work. Cheap electret microphones are now quite ubiquitous; electrostatic filters and copy machines might employ electrets, too.
- It is a safe bet that some of the "exotic" materials mentioned in this sub-chapter 3.6 (and some materials not even mentioned or maybe not yet discovered) will be turned into products within *your* career as an engineer, dear student!

3.6.2 Ferro Electricity

- The name, obviously, has nothing to do with "*Ferro*" (= Iron), but associates the analogy to ferro magnetism. It means that in some special materials, the electrical dipoles are not *randomly* distributed, but interact in such a way as to align themselves even *without* an external field.
    - We thus expect *spontaneous polarization* and a *very large* dielectric constant (**DK**).
    - This should be very useful - e.g. for making capacitors - but as in the case of ferro *magnetism*, there are not too many materials showing this behavior.
- The best known material used for many application is **$BaTiO_3$** (**Barium titanate**).
    - It has a simple lattice as far as materials with three different atoms can have a simple lattice at all. The doubly charged **Ba^{2+}** atoms sits on the corners of a cube, the **O^{2-}** ions on the face centers, and the **Ti^{4+}** ion in the center of the cube.
    - We have **8 Ba^{2+}** ions belonging to **1/8** to the elementary cell, **6 O^{2-}** ions belonging to **1/2** to the elementary cell, and one **Ti^{4+}** ion belonging in total to the cell, which gives us the **$BaTiO_3$** stoichiometry.
    - This kind of crystal structure is called a **Perovskite** structure; it is very common in nature and looks like the drawing below (only three of the six oxygen ions are shown for clarity):



- Often, the lattice is not exactly cubic, but slightly distorted. In the case of **$BaTiO_3$** this is indeed the case: The **Ti** - ion does not sit in the *exact* center of the slightly distorted cube, but slightly off to one side. It thus has *two* symmetrical positions as schematically (and much exaggeratedly) shown below

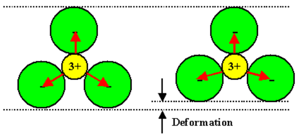


- Each elementary cell of **$BaTiO_3$** thus carries a dipole moment, and, what's more important, *the moments of neighbouring cells tend to line up*.

- The interactions between the dipoles that lead to a line-up can only be understood with quantum mechanics. It is not unlike the interactions of *spins* that lead to ferro magnetism.
    - We will not go into details of ferro electricity at this point. Suffice it to say that there are many uses. Traditionally, many capacitors use ferro-electric materials with high **DK** values. In recent years, a large interest in ferro-electrics for uses in integrated circuits has developed; we have yet to see if this will turn into new products.

3.6.3 Summary to: Special Dielectrics

- Polarization **$\underline{P}$** of a dielectric material can also be induced by mechanical deformation ***e*** or by other means.
    - *Piezo electric materials* are anisotropic crystals meeting certain symmetry conditions like crystalline quartz (**SiO_2**): the effect is linear.
    - The effect also works in reverse: Electrical fields induce mechanical deformation
    - Piezo electric materials have many uses, most prominent are quartz oscillators and, recently, fuel injectors for Diesel engines.

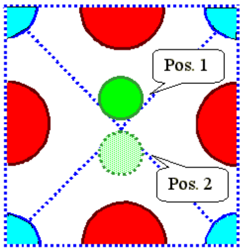


$$P = \text{const.} \cdot e$$

- *Electrostriction* also couples polarization and mechanical deformation, but in a quadratic way and only in the direction "electrical fields induce (very small) deformations".
    - The effect has little uses so far; it can be used to control very small movements, e.g. for manipulations in the **nm** region. Since it is coupled to electronic polarization, many materials show this effect.

$$e = \frac{\Delta l}{l} = \text{const} \cdot E^2$$

- *Ferro electric materials* posses a permanent dipole moment in any elementary cell that, moreover, are all aligned (below a critical temperature).
    - There are strong parallels to ferromagnetic materials (hence the strange name).
    - Ferroelectric materials have large or even very large ($\epsilon_r > 1.000$) dielectric constants and thus are to be found inside capacitors with high capacities (but not-so-good high frequency performance)



$BaTiO_3$ unit cell

- *Pyro electricity* couples polarization to temperature changes; *electrets* are materials with permanent polarization, ... There are more "curiosities" along these lines, some of which have been made useful recently, or might be made useful - as material science and engineering progresses.

Questionaire

Multiple Choice questions to all of 3.6

3.7 Summary: Dielectrics

- The dielectric constant ϵ_r "somehow" describes the interaction of dielectric (i.e. more or less insulating) materials and electrical fields; e.g. via the equations ⇒
  - $\underline{D}$ is the **electrical displacement** or **electrical flux density**, sort of replacing $\underline{E}$ in the Maxwell equations whenever materials are encountered.
  - C is the capacity of a parallel plate capacitor (plate area A, distance d) that is "filled" with a dielectric with ϵ_r
  - n is the index of refraction; a quantity that "somehow" describes how electromagnetic fields with extremely high frequency interact with matter.
    in this equaiton it is assumed that the material has no magnetic properties at the frequency of light.

$$\underline{D} = \epsilon_0 \cdot \epsilon_r \cdot \underline{E}$$

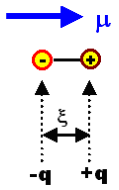
$$C = \frac{\epsilon_0 \cdot \epsilon_r \cdot A}{d}$$

$$n = (\epsilon_r)^{½}$$

- Electrical fields inside dielectrics polarize the material, meaning that the vector sum of electrical dipoles inside the material is no longer zero.
  - The decisive quantities are the dipole moment $\underline{\mu}$, a vector, and the Polarization P, a vector, too.
  - Note: The dipole moment vector points from the negative to the positive charge - contrary to the electrical field vector!
  - The dipoles to be polarized are either already present in the material (e.g. in H_2O or in ionic crystals) or are induced by the electrical field (e.g. in single atoms or covalently bonded crystals like **Si**)
  - The dimension of the polarization $\underline{P}$ is **[C/cm^2]** and is indeed identical to the net charge found on unit area ion the surface of a polarized dielectric.

$$\underline{\mu} = q \cdot \underline{\xi}$$

$$\underline{P} = \frac{\Sigma \underline{\mu}}{V}$$



- The equivalent of "Ohm's law", linking current density to field strength in conductors is the Polarization law:
  - The decisive material parameter is χ ("kee"), the **dielectric susceptibility**
  - The "classical" flux density D and the Polarization are linked as shown. In essence, $\underline{P}$ only considers what happens in the material, while D looks at the total effect: material plus the field that induces the polarization.

$$\underline{P} = \epsilon_0 \cdot \chi \cdot \underline{E}$$

$$\epsilon_r = 1 + \chi$$

$$\underline{D} = \underline{D}_0 + \underline{P} = \epsilon_0 \cdot \underline{E} + \underline{P}$$

- Polarization by necessity moves masses (electrons and / or atoms) around, this will not happen arbitrarily fast.
  - ϵ_r or χ thus must be functions of the frequency of the applied electrical field, and we want to consider the whole frequency range from **RF** via **HF** to light and beyond.

$\epsilon_r(\omega)$ is called the "**dielectric function**" of the material.

- The tasks are:
  - Identify and (quantitatively) describe the major mechanisms of polarization.
  - Justify the assumed linear relationship between P and χ.
  - Derive the dielectric function for a given material.

(Dielectric) polarization mechanisms in dielectrics are all mechanisms that

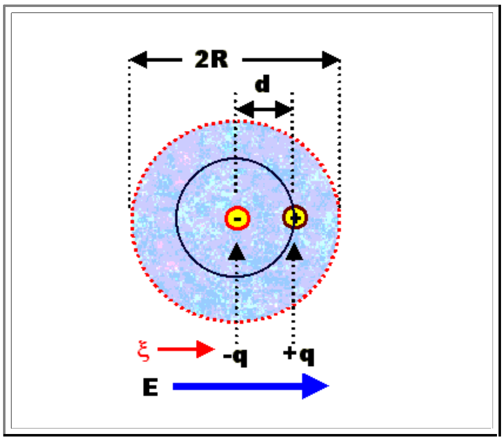
1. Induce dipoles at all (always with $\boldsymbol{\mu}$ in field direction)
   ⇒ Electronic polarization.
2. Induce dipoles already present in the material to "point" to some extent in field direction.
   ⇒ Interface polarization.
   ⇒ Ionic polarization.
   ⇒ Orientation polarization.

Quantitative considerations of polarization mechanisms yield

- **Justification (and limits) to the $\underline{P} \propto \underline{E}$ "law"**
- **Values for χ**
- **$\chi = \chi(\omega)$**
- **$\chi = \chi$(structure)**

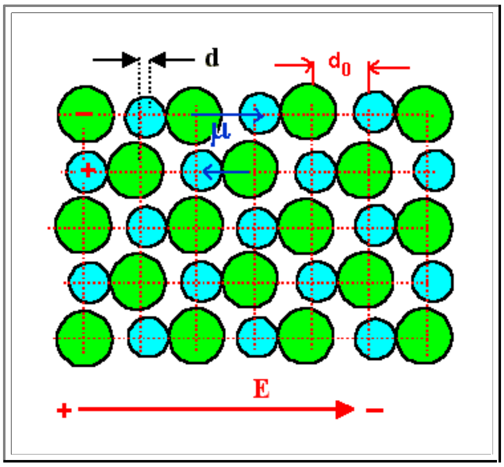
Electronic polarization describes the separation of the centers of "gravity" of the electron charges in orbitals and the positive charge in the nucleus and the dipoles formed this way. it is always present

- It is a very weak effect in (more or less isolated) atoms or ions with spherical symmetry (and easily calculated).
- It can be a strong effect in e.g. covalently bonded materials like **Si** (and not so easily calculated) or generally, in solids.



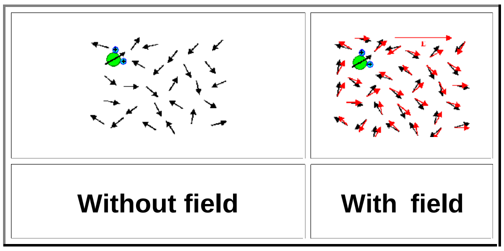
Ionic polarization describes the net effect of changing the distance between neighboring ions in an ionic crystal like **NaCl** (or in crystals with some ionic component like **SiO_2**) by the electric field

- Polarization is linked to bonding strength, i.e. Young's modulus **Y**. The effect is smaller for "stiff" materials, i.e. **P ∝ 1/Y**



Orientation polarization results from minimizing the free enthalpy of an ensemble of (molecular) dipoles that can move and rotate freely, i.e. polar liquids.

- It is possible to calculate the effect, the result invokes the Langevin function



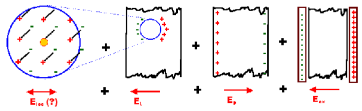
$$L(\beta) = \coth(\beta) - \frac{1}{\beta}$$

- In a good approximation the polarization is given by ⇒

$$\langle P \rangle = \frac{N \cdot \mu^2 \cdot E}{3kT}$$

- The induced dipole moment **μ** in all mechanisms is proportional to the field (for reasonable field strengths) at the location of the atoms / molecules considered.
  - The proportionality constant is called polarizability α; it is a microscopic quantity describing what atoms or molecules "do" in a field.
  - The local field, however, is not identical to the macroscopic or external field, but can be obtained from this by the Lorentz approach
  - For isotropic materials (e.g. cubic crystals) one obtains

$$\underline{\mu} = \alpha \cdot \boldsymbol{E}_{\mathrm{loc}}$$



$$\boldsymbol{E}_{\mathrm{loc}} = \boldsymbol{E}_{\mathrm{ex}} + \boldsymbol{E}_{\mathrm{pol}} + \boldsymbol{E}_{\mathrm{L}} + \boldsymbol{E}_{\mathrm{near}}$$

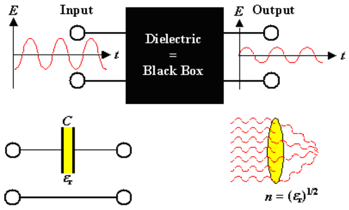
$$\boldsymbol{E}_{\mathrm{L}} = \frac{\boldsymbol{P}}{3\epsilon_0}$$

- Knowing the local field, it is now possible to relate the microscopic quantity α to the macroscopic quantity ϵ or ϵ_r via the Clausius - Mosotti equations ⇒
  - While this is not overly important in the engineering practice, it is a momentous achievement. With the Clausius - Mosotti equations and what went into them, it was possible for the first time to understand most electronic and optical properties of dielectrics in terms of their constituents (= atoms) and their structure (bonding, crystal lattices etc.)
  - Quite a bit of the formalism used can be carried over to other systems with dipoles involved, in particular magnetism = behavior of magnetic dipoles in magnetic fields.

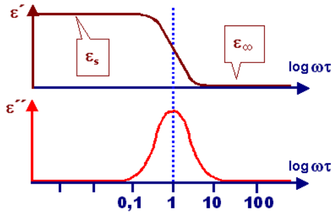
$$\frac{N \cdot \alpha}{3\,\epsilon_0} = \frac{\epsilon_r - 1}{\epsilon_r + 2}$$

$$= \frac{\chi}{\chi + 3}$$

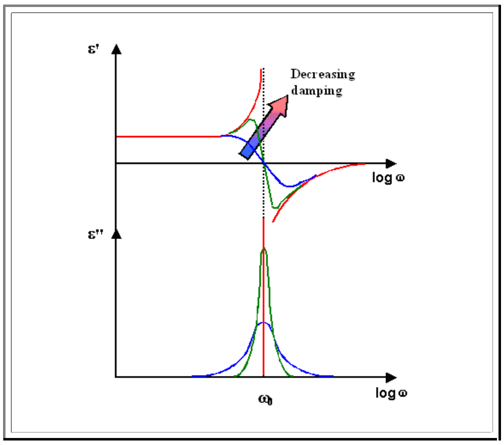
- Alternating electrical fields induce alternating forces for dielectric dipoles. Since in all polarization mechanisms the dipole response to a field involves the movement of masses, inertia will prevent arbitrarily fast movements.
  - Above certain limiting frequencies of the electrical field, the polarization mechanisms will "die out", i.e. not respond to the fields anymore.
  - This might happen at rather high (= optical) frequencies, limiting the index of refraction $n = (\epsilon_r)^{1/2}$



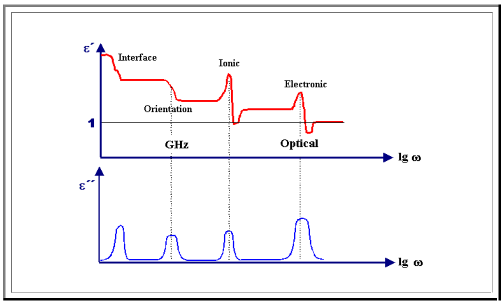
- The (only) two physical mechanisms governing the movement of charged masses experiencing alternating fields are relaxation and resonance.
- *Relaxation* describes the decay of excited states to the ground state; it describes, e.g., what happens for orientation polarization after the field has been switched off.
  - From the "easy to conceive" time behavior we deduce the frequency behavior by a Fourier transformation
  - The dielectric function describing relaxation has a typical frequency dependence in its real and imaginary part ⇒



- Resonance describes anything that can be modeled as a mass on a spring - i.e. electronic polarization and ionic polarization.
  - The decisive quantity is the (undamped) resonance frequency $\omega_0 = (k_S / m)^{1/2}$ and the "friction" or damping constant k_F
  - The "spring" constant is directly given by the restoring forces between charges, i.e. Coulombs law, or (same thing) the bonding. In the case of bonding (ionic polarization) the spring constant is also easily expressed in terms of Young's modulus Y. The masses are electron or atom masses for electronic or ionic polarization, respectively.
  - The damping constant describes the time for funneling off ("dispersing") the energy contained in one oscillating mass to the whole crystal lattice. Since this will only take a few oscillations, damping is generally large.
  - The dielectric function describing relaxation has a typical frequency dependence in its real and imaginary part ⇒ The green curve would be about right for crystals.



- The complete frequency dependence of the dielectric behavior of a material, i.e. its dielectric function, contains all mechanisms "operating" in that material.
  - As a rule of thumb, the critical frequencies for relaxation mechanisms are in the **GHz** region, electronic polarization still "works" at optical (**10^{15} Hz**) frequencies (and thus is mainly responsible for the index of refraction).
  - Ionic polarization has resonance frequencies in between.
  - Interface polarization may "die out" already a low frequencies.
- A widely used diagram with all mechanisms shows this, but keep in mind that there is no real material with all 4 major mechanisms strongly present! ⇒



- A general mathematical theorem asserts that the real and imaginary part of the dielectric function cannot be completely independent
  - If you know the complete frequency dependence of either the real or the imaginary part, you can calculate the complete frequency dependence of the other.
  - This is done via the Kramers-Kronig relations; very useful and important equations in material practice. ⇒

$$\epsilon'(\omega) = \frac{-2\,\omega}{\pi} \int_0^\infty \frac{\omega^* \cdot \epsilon''(\omega^*)}{\omega^{*2} - \omega^2} \cdot d\omega^*$$

$$\epsilon''(\omega) = \frac{2\,\omega}{\pi} \int_0^\infty \frac{\epsilon'(\omega^*)}{\omega^{*2} - \omega^2} \cdot d\omega^*$$

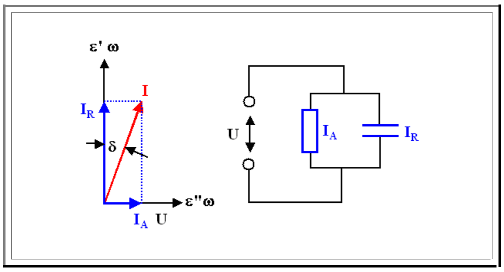
- The frequency dependent current density j flowing through a dielectric is easily obtained. ⇒
  - The in-phase part generates active power and thus heats up the dielectric, the out-of-phase part just produces reactive power
  - The power losses caused by a dielectric are thus directly proportional to the imaginary component of the dielectric function

$$j(\omega) = \frac{dD}{dt} = \epsilon(\omega)\cdot\frac{dE}{dt} = \underbrace{\omega \cdot \epsilon'' \cdot E(\omega)}_{\text{in phase}} + \underbrace{i \cdot \omega \cdot \epsilon' \cdot E(\omega)}_{\text{out of phase}}$$

$$L_A = \text{power turned into heat} = \omega \cdot |\epsilon''| \cdot E^2$$

- The relation between active and reactive power is called "tangens Delta" (**tg(δ)**); this is clear by looking at the usual pointer diagram of the current

$$\frac{L_A}{L_R} := \textbf{tg}\ \delta = \frac{I_A}{I_R} = \frac{\epsilon''}{\epsilon'}$$



  - The pointer diagram for an *ideal* dielectric **σ(ω = 0) = 0**can always be obtained form an (ideal) resistor ***R*(ω)** in parallel to an (ideal) capacitor ***C*(ω)**.
  - ***R*(ω)** expresses the apparent conductivity $\sigma_{DK}(\omega)$ of the dielectric, it follows that

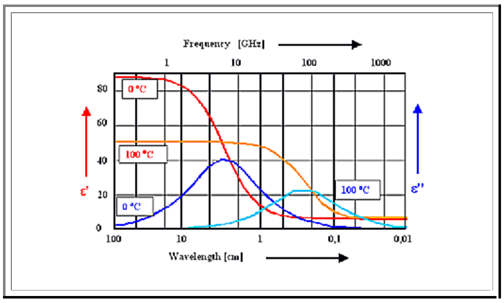
$$\sigma_{DK}(\omega) = \omega \cdot \epsilon''(\omega)$$

- For a *real* dielectric with a non-vanishing conductivity at zero (or small) frequencies, we now just add another resistor in parallel. This allows to express *all* conductivity effects of a real dielectric in the imaginary part of its (usually measured) dielectric function via

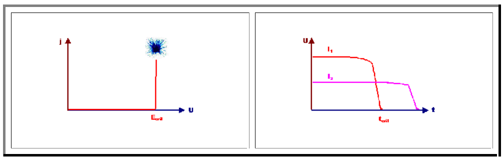
$$\epsilon'' = \frac{\sigma_{total}}{\omega}$$

  - We have no *all* materials covered with respect to their dielectric behavior - in principle even metals, but then resorting to a dielectric function would be overkill.

- A good example for using the dielectric function is "dirty" water with a not-too-small (ionic) conductivity, commonly encountered in food.
  - The polarization mechanism is orientation polarization, we expect large imaginary parts of the dielectric function in the **GHz** region.
  - It follows that food can be heated by microwave (ovens)!



- The first law of materials science obtains: At field strengths larger than some critical value, dielectrics will experience (destructive) electrical breakdown
  - This might happen suddenly (then calls break-down) , with a bang and smoke, or
  - it may take time - months or years - then called failure.
  - Critical field strength may vary from **< 100 kV/cm** to **> 10 MV / cm**.

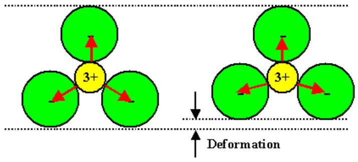


- Highest field strengths in practical applications do not necessarily occur at high voltages, but e.g. in integrated circuits for very thin (a few **nm**) dielectric layers
  - Properties of thin films may be quite different (better!) than bulk properties!

Example 1: TV set, **20 kV** cable, thickness of insulation = **2 mm.** ⇒ ***E* = 100 kV/cm**
Example 2: Gate dielectric in transistor, **3.3 nm** thick, **3.3 V** operating voltage. ⇒ ***E* = 10 MV/cm**

- Electrical breakdown is a major source for failure of electronic products (i.e. one of the reasons why things go "kaputt" (= broke)), but there is no simple mechanism following some straight-forward theory. We have:
    - *Thermal breakdown*; due to small (field dependent) currents flowing through "weak" parts of the dielectric.
    - *Avalanche breakdown* due to occasional free electrons being accelerated in the field; eventually gaining enough energy to ionize atoms, producing more free electrons in a runaway avalanche.
    - *Local discharge* producing micro-plasmas in small cavities, leading to slow erosion of the material.
    - *Electrolytic breakdown* due to some ionic micro conduction leading to structural changes by, e.g., metal deposition.

- Polarization $\underline{P}$ of a dielectric material can also be induced by mechanical deformation e or by other means.
    - *Piezo electric materials* are anisotropic crystals meeting certain symmetry conditions like crystalline quartz (**SiO_2**): the effect is linear.
    - The effect also works in reverse: Electrical fields induce mechanical deformation
    - Piezo electric materials have many uses, most prominent are quartz oscillators and, recently, fuel injectors for Diesel engines.

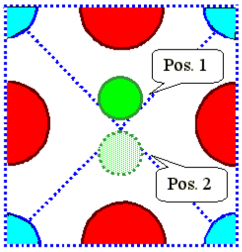


$$P = \text{const.} \cdot e$$

- *Electrostriction* also couples polarization and mechanical deformation, but in a quadratic way and only in the direction "electrical fields induce (very small) deformations".
    - The effect has little uses so far; it can be used to control very small movements, e.g. for manipulations in the **nm** region. Since it is coupled to electronic polarization, many materials show this effect.

$$e = \frac{\Delta l}{l} = \text{const} \cdot E^2$$

- *Ferro electric materials* posses a permanent dipole moment in any elementary cell that, moreover, are all aligned (below a critical temperature).
    - There are strong parallels to ferromagnetic materials (hence the strange name).
    - Ferroelectric materials have large or even very large (**$\epsilon_r > 1.000$**) dielectric constants and thus are to be found inside capacitors with high capacities (but not-so-good high frequency performance)



$BaTiO_3$ unit cell

- *Pyro electricity* couples polarization to temperature changes; *electrets* are materials with permanent polarization, ... There are more "curiosities" along these lines, some of which have been made useful recently, or might be made useful - as material science and engineering progresses.

Questionaire

Multiple Choice questions to all of 3

4. Magnetic Materials

4.1 Definitions and General Relations

4.1.1 Fields, Fluxes and Permeability

4.1.2 Origin of Magnetic Dipoles

4.1.3 Classifications of Interactions and Types of Magnetism

4.1.4 Summary to: Magnetic Materials - Definitions and General Relations

4.2 Dia- and Paramagnetism

4.2.1 Diamagnetism

4.2.2 Paramagnetism

4.2.3 Summary to: Dia- and Paramagnetism

4.3 Ferromagnetism

4.3.1 Mean Field Theory of Ferromagnetism

4.3.2 Beyond Mean Field Theory

4.3.3 Magnetic Domains

4.3.4 Domain Movement in External Fields

4.3.5 Magnetic Losses and Frequency Behavior

4.3.6 Hard and Soft Magnets

4.3.7 Summary to: Ferromagnetism

4.4 Applications of Magnetic Materials

4.4.1 Everything Except Data Storage

4.4.2 Magnetic Data Storage

4.4.3 Summary to: Technical Materials and Applications

4.5. Summary: Magnetic Materials

4. Magnetic Materials

4.1 Definitions and General Relations

4.1.1 Fields, Fluxes and Permeability

- There are many analogies between dielectric and magnetic phenomena; the big difference being that (so far) there are *no magnetic "point charges"*, so-called magnetic monopoles, but only *magnetic dipoles*.
  - The first basic relation that we need is the relation between the magnetic flux density ***B*** and the magnetic field strength ***H*** *in vacuum*. It comes straight from the Maxwell equations:

$$B = \mu_o \cdot H$$

  - The symbols are:
    - ***B*** = **magnetic flux density** or **magnetic induction**,
    - μ_o = **magnetic permeability of the vacuum = 4π · 10^{-7} Vs/Am = 1,26 · 10^{-6}Vs/Am**
    - ***H*** = **magnetic field strength**

    .
  - The units of the magnetic field ***H*** and so on are
    - **[*H*] = A/m**
    - **[*B*]** = **Vs/m^2**, with **1Vs/m^2 = 1 Tesla**.

    $\underline{B}$ and $\underline{H}$ are vectors, of course.
  - **10^3/4π A/m** used to be called **1 Oersted**, and **1 Tesla** equales **10^4 Gauss** in the old system.
  - Why the eminent mathematician and scientist *Gauss* was dropped in favor of the somewhat shady figure *Tesla* remains a mystery.
- If a material is present, the relation between magnetic field strength and magnetic flux density becomes

$$B = \mu_o \cdot \mu_r \cdot H$$

  - with μ_r = **relative permeability of the material** in complete analogy to the *electrical flux density* and the *dielectric constant*.
  - The relative permeability of the material **μ_r** is a material parameter without a dimension and thus a *pure number* (or several pure numbers if we consider it to be a *tensor* as before). It is the material property we are after.
- Again, it is useful and conventional to split ***B*** into the *flux density in the vacuum* plus the *part of the material* according to

$$B = \mu_o \cdot H + J$$

  - With ***J*** = **magnetic polarization** in analogy to the dielectric case.
- As a new thing, we now we define the **magnetization *M*** of the material as

$$M = \frac{J}{\mu_o}$$

  - That is only to avoid some labor with writing. This gives us

$$B = \mu_o \cdot (H + M)$$

- Using the independent definition of ***B*** finally yields

$$M = (\mu_r - 1) \cdot H$$

$$M := \chi_{mag} \cdot H$$

- With χ_{mag} **= (μ_r – 1) = magnetic susceptibility**.
- It is really straight along the way we looked at dielectric behavior; for a direct comparison use the link

The magnetic susceptibility χ_{mag} is the *prime material parameter* we are after; it describes the response of a material to a magnetic field in exactly the same way as the dielectric susceptibility $\chi_{dielectr}$. We even chose the same abbreviation and will drop the suffix most of the time, believing in your intellectual power to keep the two apart.

- Of course, the four *vectors* ***H***, ***B***, ***J***, ***M*** are all parallel in isotropic homogeneous media (i.e. in amorphous materials and poly-crystals).
- In anisotropic materials the situation is more complicated; χ and μ_r then must be seen as **tensors**. That should be no surprise anymore.

We are left with the question of the *origin of the magnetic susceptibility*. There are no **magnetic monopoles**that could be separated into magnetic dipoles as in the case of the dielectric susceptibility, there are only *magnetic dipoles* to start from.

- Why there are no magnetic monopoles (at least none have been discovered so far despite extensive search) is one of the tougher questions that you can ask a physicist; the ultimate answer seems not yet to be in. So just take it as a fact of life.
- In the next paragraph we will give some thought to the the origin of magnetic dipoles.

4.1.2 Origin of Magnetic Dipoles

- Where are magnetic dipoles coming from? The classical answer is simple: A **magnetic moment *m*** is generated whenever a *current flows in closed circle*.
    - *Of course, we will not mix up the letter **m** used for magnetic moments with the m^*_e, the mass of an electron, which we also need in some magnetic equations*.
    - For a current *I* flowing in a circle enclosing an area *A*, *m* is defined to be

$$m = I \cdot A$$

    - This does not only apply to "regular" current flowing in a wire, but in the extreme also to a *single electron circling around an atom*.
- In the context of **Bohrs model** for an atom, the magnetic moment of such an electron is easily understood:
    - The current *I* carried by *one* electron orbiting the nucleus at the distance *r* with the frequency $\nu = \omega/2\pi$ is .

$$I = e \cdot \frac{\omega}{2\pi}$$

    - The area *A* is πr^2, so we have for the magnetic moment m_{orb} of the electron

$$m_{orb} = e \cdot \frac{\omega}{2\pi} \cdot \pi r^2 = \tfrac{1}{2} \cdot e \cdot \omega \cdot r^2$$

- Now the *mechanical* angular momentum *L* is given by

$$L = m^*_e \cdot \omega \cdot r^2$$

    - With m^*_e = mass of electron (*the * serves to distinguish the mass m^*_e from the magnetic moment m^e of the electron)*, and we have a simple relation between the mechanical angular momentum *L* of an electron (which, if you remember, was the decisive quantity in the Bohr atom model) and its magnetic moment *m*.

$$\underline{m}_{orb} = -\frac{e}{2m^*_e} \cdot \underline{L}$$

    - The *minus sign* takes into account that mechanical angular momentum and magnetic moment are antiparallel; as before we note that this is a *vector equation* because both *m* and *L* are (polar) vectors.
    - The quantity $e/2m^*_e$ is called the **gyromagnetic relation** or quotient; *it should be a fixed constant* relating *m* and any given *L*.
    - However, in real life it often deviates from the value given by the formula. How can that be?
    - Well, try to remember: Bohr's model is a mixture of classical physics and quantum physics and *far too simple* to account for everything. It is thus small wonder that conclusions based on this model will not be valid in all situations.
- In *proper quantum mechanics* (as in Bohr's semiclassical model) *L* comes in *discrete values only*. In particular, the fundamental assumption of Bohr's model was $L = n \cdot \hbar$, with n = quantum number = **1, 2, 3, 4, ...**
    - It follows that m_{orb} *must be quantized, too*; it must come in multiples of

$$m_{orb} = \frac{h \cdot e}{4\pi \cdot m^*_e} = m_{Bohr} = 9.27 \cdot 10^{-24}\ \text{Am}^2$$

  - This relation defines a **fundamental unit for magnetic dipole moments**, it has its own name and is called a **Bohr magneton**.
  - It is for magnetism what an elementary charge is for electric effects.

- But electrons orbiting around a nucleus are not the *only* source of magnetic moments.
  - Electrons always have a **spin** ***s***, which, on the level of the Bohr model, can be seen as a built-in angular momentum with the value $\hbar \cdot s$. The spin quantum number ***s*** is ½, and this allows two directions of angular momentum and magnetic moment , always symbolically written as .

$$s = \begin{cases} +1/2 \\ -1/2 \end{cases}$$

- It is possible, of course, to compute the circular current represented by a charged ball spinning around its axis if the distribution of charge in the sphere (or on the sphere), is known, and thus to obtain the magnetic moment of the spinning ball.
  - Maybe that even helps us to understand the internal structure of the electron, because we know its magnetic moment and now can try to find out what kind of size and internal charge distribution goes with that value. Many of the best physicists have tried to do exactly that.
  - However, as it turns out, whatever assumptions you make about the internal structure of the electron that will give the right magnetic moment will always get you into *deep trouble* with *other properties* of the electron. There simply is *no internal structure* of the electron that will explain its properties!
  - We thus are forced to simply accept as a *fundamental property of an electron* that it always carries a magnetic moment of

$$m^e = \frac{2 \cdot h \cdot e \cdot s}{4\pi \cdot m^*_e} = \pm m_{Bohr}$$

  - The factor **2** is a puzzle of sorts - not only because it appears at all, but because it is actually **= 2.00231928**. But pondering this peculiar fact leads straight to quantum electrodynamics (and several Nobel prizes), so we will not go into this here.

- The total magnetic moment of an atom - *still within the Bohr model* - now is given by the (vector)sum of all the "orbital" moments and the "spin" moments of all electrons in the atom, taking into account all the quantization rules; i.e. the requirement that the angular momentums ***L*** cannot point in arbitrary directions, but only in fixed ones.
- *This is were it gets complicated* - even in the context of the simple Bohr model. A bit more to that can be found in the link. But there are few rules we can easily use:
  - All *completely filled orbitals* carry *no* magnetic moment because for every electron with spin ***s*** there is a one with spin **-*s***, and for every one going around "*clockwise*", one will circle "*counter clockwise*". This means:
  - *Forget the inner orbitals* - everything cancels!
  - Spins on not completely filled orbitals tend to *maximize their contribution*; they will first fill all available energy states with spin up, before they team up and cancel each other with respect to magnetic momentum.
  - The *chemical environment*, i.e. bonds to other atoms, incorporation into a crystal, etc., may strongly change the magnetic moments of an atom.
- The net effect for a given (isolated) atom is simple. Either it has a magnetic moment in the order of a Bohr magneton because not all contributions cancel - or it has none. And it is possible, (if not terribly easy), to calculate what will be the case. A first simple result emerges: Elements with an *even* number of electrons have generally *no magnetic moment*.
- We will not discuss the rules for getting the permanent magnetic moment of a single atom from the interaction of spin moments and orbital moments, but are going to look at the possible effects if you
  - bring atoms together to form a solid, or
  - subject solids to an external magnetic field ***H***
- A categorization will be given in the next paragraph.

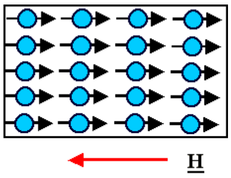
Questionnaire

Multiple Choice questions to 4.1.2

4.1.3 Classifications of Interactions and Types of Magnetism

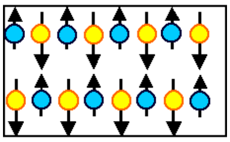
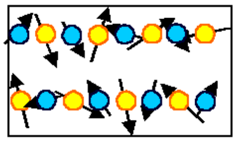
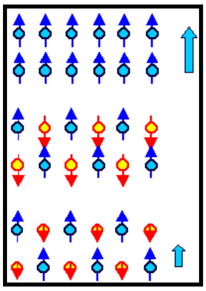
Dia-, Para-, and Ferromagnetism

- We want to get an idea of what happens to *materials* in external magnetic fields. "Material", in contrast to a single atom, means that we have plenty of (possibly different) atoms in close contact, i.e with some bonding. We can distinguish two basic cases:
  - **1.** The atoms of the material have *no magnetic moment* of their own. This is generally true for about one half of the elements; the ones with even atomic numbers and therefore an even number of electrons. The magnetic moments of the spins tends to cancel; the atoms will only have a magnetic moment if there is an orbital contribution. Of course, the situation may change if you look at *ions* in a crystal.
  - **2.** At least *some* of the atoms of the material have a magnetic moment. That covers the other half of the periodic table: All atoms with an odd number of electrons will have one spin moment left over. Again, the situation may be different if you look at ionic crystals.
- Lets see what can happen if you consider interactions of the magnetic moments with each other and with a magnetic field. First, we will treat the case of solids with *no magnetic moments* of their constituents, i.e. **diamagnetic materials**.
  - The following table lists the essentials

Diamagnetic Materials		
Magnetic moment?	No	
Internal magnetic interaction?	None	
Response to external field	Currents (and small magn. moments) are induced by turning on the field because the orbiting electrons are slightly disturbed. The induced magn. moments oppose the field. No temperature dependence Mechanism analogous to electronic polarisation in dielectrics,	 *The* black *arrows should be seen as being very short!!!!*
Value of μ_r	$\mu_r \le\approx 1$ in **diamagnetic** Small effect in "regular" materials	$\mu_r = 0$ in **superconductors** (ideal diamagnet)
Value of B	$B \le\approx \mu_0 \cdot H$	$B = 0$ in superconductors
Typical materials	All elements with filled shells (always even atomic number)	all noble gases, H_2, Cu, H_2O, NaCl, Bi, ... Alkali or halogene *ions*

- Since you cannot expose material to a magnetic field without encountering a changing field strength **d*H*/d*t*** (either by turning on the field on or by moving the specimen into a field), currents will be induced that produce a magnetic field of their own.
  - According to **Lenz's law**, the direction of the current and thus the field is always such as to oppose the generating forces. Accordingly, the **induced magnetic moment** will be *antiparallel* to the external field.
  - This is called **diamagnetism** and it is a weak effect in normal materials.
- There is an exception, however: **Superconductors**, i.e. materials with a **resistivity = 0** at low temperatures, will have their mobile charges responding without "resistance" to the external field and the induced magnetic moments will *exactly cancel* the external field.
  - Superconductors (at least the "normal" ones (or "type **I**" as they are called) therefore are always *perfectly field free* - a magnetic field cannot penetrate the superconducting material.
  - That is just as amazing as the zero resistance; in fact the magnetic properties of superconductors are just as characteristic for the superconducting state of matter as the resistive properties.

- There will be a backbone II module for superconductors in due time
- If we now look at materials where at least *some of the atoms* carry a permanent magnetic moment, we have to look first at the possible *internal interactions* of the magnetic moments in the material and then at their interaction with an *external field*. Two limiting cases can be distinguished.
    - **1.** Strong internal interaction (i.e. interaction energies **»** **k*T***, the thermal energy). **Ferromagnetism** results
    - **2.** No or weak interaction. We have **paramagnetic materials**.
- The first case of strong interaction will more or less turn into the second case at temperatures high enough so that **k*T* >>** interaction energy, so we expect a temperature dependence of possible effects. A first classification looks like this:

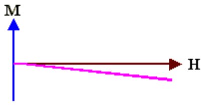
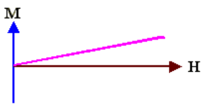
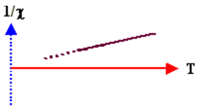
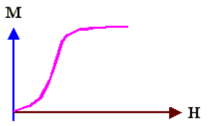
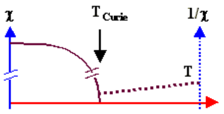
Paramagnetic and Ferromagnetic Materials		
Magnetic moment?	Yes	
Internal agnetic interaction?	Strong	Weak
Ordered regions?	Yes  This example shows a **ferrimagnetic material** Ordered magnetic structures that are stable in time. Permanent magnetization is obtained by the (vector) sum over the individual magnetic moments.	No  Example for a **paramagnetic material** Unordered magnetic structure, fluctuating in time. Averaging over time yields no permanent magnetization
Response to external field	A large component of the magnetic moment may be in field direction	Small average orientation in field direction. Mechanism *fully* analogous to orientation polarization for dielectrics
Kinds of ordering	Many possibilities. Most common are *ferro*-, *antiferro*-, and *ferri*magnetism as in the self-explaining sequence below: 	
Value of μ_r	**μ_r >> 1** for ferromagnets **$\mu_r \approx 1$** for anti-ferromagnets **μ_r > 1** for ferrimagnets	**$\mu_r \geq\approx 1$**
***T*-dependence**	Paramagnetic above **Curie Temperature**	Weak ***T***-dependence
Paramagnetic materials (at room temperature)		Mn, Al, Pt, O_2(gas and liquid), rare earth ions, ...

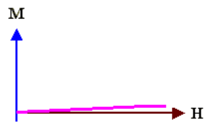
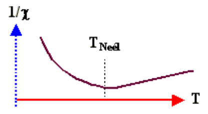
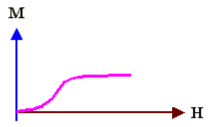
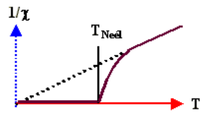
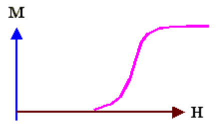
Ferromagnetic materials (with Curie- (or Neél) ***T***)	Ferro elements: Ferro technical: Ferri: Antiferro: (no technical uses)	Fe (770 ^{0}C), Co (1121 ^{0}C), Ni (358 ^{0}C), Gd (16 ^{0}C) "AlNiCo", Co_5Sm, $Co_{17}Sm_2$, "NdFeB" Fe_3O_4, MnO (116 ^{0}C), NiO (525 ^{0}C), Cr (308 ^{0}C)

- This table generated a lot of new names, definitions and question. It sets the stage for the dealing with the various aspects of *ferromagnetism* (including *ferri*- and *anti-ferro* magnetism as well as some more kinds of internal magnetic ordering. A few examples of ferromagnetic materials are given in the link.
  - There might be *many more types* of ordering: Any fixed relation between two vectors qualify. As an example, moment **2** might not be parallel to moment **1** but off by ***x*** degrees; and the succession of many moments might form a spiral pattern.
  - If you can think of some possible ordering (and it is not forbidden by some overruling law of nature), it is a safe bet that mother nature has already made it in some exotic substance. But, to quote Richard **Feynman**:
  - *"It is interesting to try to analyze what happens when a field is applied to such a spiral* (of magnetic ordering) *- all the twistings and turnings that must go on in all those atomic magnets. (Some people like to amuse themselves with the theory of these things!)"* (Lectures on Physics, Vol **II**, **37-13**; Feynmans emphasizes).
- Well, we don't, and just take notice of the fact that there is *some* kind of magnetic ordering for *some* materials.
  - As far as the element are concerned, the only ferromagnets are: **Fe**, **Ni**, and **Co**. (**Mn** almost is one, but not quite).
  - Examples for antiferromagnets include **Cr**, ...
  - And there are many, many *compounds*, often quite strange mixtures (e.g. **NdFeB** or **Sm_2Co_{17}**), with *remarkable and often useful* ferro-, ferri, antiferro, or,..., properties.

Temperature Dependence of Magnetic Behavior

- How do we distinguish an *antiferromagnetic* material from a *paramagnet* or a *diamagnet*? They all appear not to be very "magnetic" if you probe them with a magnetic field.
  - We have to look at their behavior in a magnetic field *and* at the *temperature dependence* of that behavior. Ordering the atomic magnetic moments is, after all, a *thermodynamical effect* - it always has to compete with entropy - and thus should show some specific temperature dependence.
  - There are indeed quite characteristic curves of major properties with temperature as shown below.

Magnetization ***M* = *M*(*H*)**	**Magnetic susceptibility** $\chi_{mag} = \chi_{mag}(T)$	**Remarks**
		For **diamagnets** the susceptibility is negative and close to zero; and there is no temperature dependence.
		For **paramagnets**, the susceptibility is (barely) larger than zero and decreases with ***T***. Plotted as **1/χ(*T*)** we find a *linear* relationship.
		For **ferromagnets** the susceptibility is large; the magnetization increases massively with **H**. Above a critical temperature T_{Cu}, the *Curie temperature*, paramagnetic behavior is observed.

		**Antiferromagnets** are like paramagnets *above* a critical temperature T_{Ne} called **Neél temperature**. Below T_{Ne} the susceptibility is small, but with a ***T***-dependence quite different from paramagnets.
		**Ferrimagnets** behave pretty much like *ferro*magnets, except that the effect tends to be smaller. The **$1/\chi(T)$** curve is very close to zero *below* a critical temperature, also called Neél temperature.
	Just for good measure, the behaviour of one of the more exotic magnetic materials. Shown is a **metamagnet**, behaving like a *ferro* magnet, but only *above* a critical magnetic field strength.	

The question now will be if we can understand at least *some* of these observations within the framework of some simple theory, similar to what we did for dielectric materials

- The answer is: *Yes*, we can - but only for the rather uninteresting (for engineering or applications) *dia*- and *para*magnets.
- Ferro magnets, however, while *extremely interesting electronic materials* (try to imagine a world without them), are a different matter. A real understanding would need plenty of quantum theory (and has not even been fully achieved yet); it is far outside the scope of this lecture course. But a phenomenological theory, based on some assumptions that we do not try to justify, will come straight out from the theory of the orientation polarization for dielectrics, and that is what we are going to look at in the next subchapters.

Questionaire

Multiple Choice questions to 4.1.3

4.1.4 Summary to: Magnetic Materials - Definitions and General Relations

- The **relative permeability** μ_r of a material "somehow" describes the interaction of magnetic (i.e. more or less all) materials and magnetic fields ***H***, e.g. vial the equations ⇒
  - ***B*** is the **magnetic flux density** or **magnetic induction**, sort of replacing ***H*** in the Maxwell equations whenever materials are encountered.
  - ***L*** is the inductivity of a linear solenoid (also called coil or inductor) with length ***l***, cross-sectional area ***A***, and number of turns ***t***, that is "filled" with a magnetic material with μ_r.
  - ***n*** is *still* the index of refraction; a quantity that "somehow" describes how electromagnetic fields with extremely high frequency interact with matter.
    For all practical purposes, however, $\mu_r = 1$ for optical frequencies

$$B = \mu_0 \cdot \mu_r \cdot H$$

$$L = \frac{\mu_0 \cdot \mu_r \cdot A \cdot w^2}{l}$$

$$n = (\epsilon_r \cdot \mu_r)^{1/2}$$

- Magnetic fields inside magnetic materials polarize the material, meaning that the vector sum of magnetic dipoles inside the material is no longer zero.
  - The decisive quantities are the *magnetic* dipole moment ***m***, a vector, and the *magnetic* Polarization ***J***, a vector, too.
  - Note: In contrast to dielectrics, we define an additional quantity, the **magnetization *M*** by simply including dividing ***J*** by μ_o.
  - The magnetic dipoles to be polarized are either already present in the material (e.g. in **Fe, Ni or Co**, or more generally, in all **paramagnetic** materials, or are induced by the magnetic fields (e.g. in **diamagnetic** materials).
  - The dimension of the magnetization ***M*** is **[A/m]**; i.e. the same as that of the magnetic field.

$$B = \mu_o \cdot H + J$$

$$\underline{J} = \mu_o \cdot \frac{\Sigma \underline{m}}{V}$$

$$M = \frac{J}{\mu_o}$$

- The magnetic polarization ***J*** or the magnetization ***M*** are *not* given by some magnetic surface charge, because ⇒.

There is no such thing as a *magnetic monopole*, the (conceivable) counterpart of a negative or positive electric charge

- The equivalent of "Ohm's law", linking current density to field strength in conductors is the *magnetic* Polarization law:
  - The decisive material parameter is $\chi_{mag} = (\mu_r - 1)$ **= magnetic susceptibility**.
  - The "classical" induction ***B*** and the magnetization ***M*** are linked as shown. In essence, ***M*** only considers what happens in the material, while ***B*** looks at the total effect: material plus the field that induces the polarization.

$$M = (\mu_r - 1) \cdot H$$

$$M := \chi_{mag} \cdot H$$

$$B = \mu_o \cdot (H + M)$$

- Magnetic polarization mechanisms are formally similar to dielectric polarization mechanisms, but the physics can be entirely different.

Atomic mechanisms of magnetization are not directly analogous to the dielectric case

Magnetic moments originate from:

- The intrinsic magnetic dipole moments ***m*** of elementary particles with spin is measured in units of the Bohr magneton $\mathbf{m}_{\mathbf{Bohr}}$.
- The magnetic moment m^e of the electron is ⇒
- Electrons "orbiting" in an atom can be described as a current running in a circle thus causing a magnetic dipole moment; too

$$m_{\mathrm{Bohr}} = \frac{h \cdot e}{4\pi \cdot m^*_e} = 9.27 \cdot 10^{-24}\ \mathrm{Am}^2$$

$$m^e = \frac{2 \cdot h \cdot e \cdot s}{4\pi \cdot m^*_e} = 2 \cdot s \cdot m_{\mathrm{Bohr}} = \pm m_{\mathrm{Bohr}}$$

The total magnetic moment of an atom in a crystal (or just solid) is a (tricky to obtain) sum of all contributions from the electrons, and their orbits (including bonding orbitals etc.), it is either:

- **Zero** - we then have a **diamagnetic material**.

Magnetic field induces dipoles, somewhat analogous to elctronic polarization in dielectrics.
Always very weak effect (except for superconductors)
Unimportant for technical purposes

- In the order of a few Bohr magnetons - we have a essentially a **paramagnetic material**.

Magnetic field induces some order to dipoles; strictly analogous to "orientation polarization" of dielectrics.
Always very weak effect
Unimportant for technical purposes

In some **ferromagnetic** materials spontaneous ordering of magnetic moments occurs below the Curie (or Neél) temperature. The important families are

- Ferromagnetic materials ⇑⇑⇑⇑⇑⇑⇑
  large $\boldsymbol{\mu_r}$, **extremely important**.
- Ferrimagnetic materials ⇑⇓⇑⇓⇑⇓⇑
  still large $\boldsymbol{\mu_r}$, **very important**.
- Antiferromagnetic materials ⇑⇓⇑⇓⇑⇓⇑
  $\boldsymbol{\mu_r} \approx \mathbf{1}$, unimportant

Ferromagnetic materials:
Fe, Ni, Co, their alloys
"AlNiCo", $\mathbf{Co_5Sm}$, $\mathbf{Co_{17}Sm_2}$,
"NdFeB"

There is characteristic temperature dependence of $\boldsymbol{\mu_r}$ for all cases

Questionnaire

Multiple Choice questions to all of 4.1

4.2 Dia- and Paramagnetism

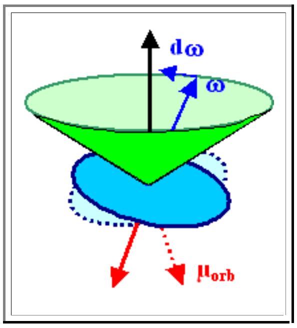
4.2.1 Diamagnetism

What is it Used for?

- It is customary in textbooks of electronic materials to treat dia- and paramagnetism in considerable detail. Considering that there is not a *single* practical case in electrical engineering where it is of any interest if a material is dia- or paramagnetic, there are only two justifications for doing this:
    - Dia- and paramagnetism lend themselves to *calculations* (and engineers like to calculate things).
    - It helps to *understand* the phenomena of magnetism in general, especially the quantum mechanical aspects of it.
- In this script we are going to keep the treatment of dia- and paramagnetism at a minimal level.

Diamagnetism - the Essentials

- The first thing to note about diamagnetism is that *all* atoms and therefore *all* materials show diamagnetic behavior.
    - Diamagnetism thus is always superimposed on all other forms of magnetism. Since it is a small effect, it is hardly noticed, however.
    - Diamagnetism results because all matter contains electrons - either "orbiting" the nuclei as in insulators or in the valence band (and lower bands) of semiconductors, or being "free", e.g. in metals or in the conduction band of semiconductors. All these electrons can respond to a (changing) magnetic field. Here we will only look at the (much simplified) case of a *bound electron orbiting a nucleus in a circular orbit*.
- The basic response of an *orbiting* electron to a changing magnetic field is a **precession** of the orbit, i.e. the polar vector describing the orbit now moves in a circle around the magnetic field vector ***H***:



- The angular vector ω characterizing the blue orbit of the electron will experience a force from the (changing) magnetic field that forces it into a circular movement on the green cone.
- Why do we emphasize "changing" magnetic fields? Because there is no way to bring matter into a magnetic field without changing it - either be switching it on or by moving the material into the field.

- What exactly happens to the *orbiting* electron? The reasoning given below follows the semi-classical approach contained within Bohr's atomic model. It gives essentially the right results *(in cgs units!)*.
    - The changing magnetic field, **d*H*/d*t***, generates a force ***F*** on the orbiting electron via inducing a voltage and thus an electrical field ***E***. We can always express this as

$$F = m^*_e \cdot a = m^*_e \cdot \frac{dv}{dt} := e \cdot E$$

- With ***a*** = acceleration = **dv/d*t* = e · *E*/m*$_e$**.

- Since **d*H*/d*t*** primarily induces a voltage ***V***, we have to express the field strength ***E*** in terms of the induced voltage ***V***. Since the electron is orbiting and experiences the voltage during *one* orbit, we can write:

$$E = \frac{V}{L}$$

- With ***L*** **=** length of orbit = **2π · *r***, and ***r*** **=** radius of orbit.
- ***V*** is given by the basic equations of induction, it is

$$V = -\frac{d\Phi}{dt}$$

  - With Φ = magnetic flux $= H \cdot A$; and A = area of orbit $= \pi \cdot r^2$. The *minus sign* is important, it says that the *effect* of a changing magnetic fields will be opposing the *cause* in accordance with **Lenz's law**.

- Putting everything together we obtain

$$\frac{dv}{dt} = \frac{e \cdot E}{m^*_e} = \frac{V \cdot e}{L \cdot m^*_e} = -\frac{e \cdot r}{2\, m^*_e} \cdot \frac{dH}{dt}$$

- The total change in **v** will be given by integrating:

$$\Delta v = \int_{v_1}^{v_2} dv = -\frac{e \cdot r}{2m^*_e} \cdot \int_{0}^{H} dH = -\frac{e \cdot r \cdot H}{2\, m^*_e}$$

- The magnetic moment m_{orb} of the undisturbed electron was $m_{orb} = ½ \cdot e \cdot v \cdot r$
  - By changing **v** by Δ**v**, we change m_{orb} by Δm_{orb}, and obtain

$$\Delta m_{orb} = \frac{e \cdot r \cdot \Delta v}{2} = -\frac{e^2 \cdot r^2 \cdot H}{4m^*_e}$$

- That is more or less the *equation for diamagnetism* in the primitive electron orbit model.
  - What comes next is to take into account that the magnetic field does not have to be perpendicular to the orbit plane and that there are many electrons. We have to add up the single electrons and average the various effects.
  - Averaging over all possible directions of H (taking into account that a field in the plane of the orbit produces zero effect) yields for the *average* induced magnetic moment almost the same formula:

$$\Delta m_{orb} = \langle \Delta m_{orb} \rangle = -\frac{e^2 \cdot \langle r \rangle^2 \cdot H}{6m^*_e}$$

  - $\langle r \rangle$ denotes that we average over the orbit radii at the same time

- Considering that not just *one*, but all z electrons of an atom participate, we get the final formula:

$$\Delta m = \langle \Delta m_{orb} \rangle = -\frac{e^2 \cdot z \cdot r^2 \cdot H}{6\, m^*_e}$$

- The additional magnetization M caused by Δm is *all the magnetization there is for diamagnets*; we thus we can drop the Δ and get

$$M_{Dia} = \frac{\langle \Delta m \rangle}{V}$$

- With the definition for the magnetic susceptibility $\chi = M/H$ we finally obtain for the relevant material parameter for diamagnetism

$$\boxed{\chi_{dia} = -\frac{e^2 \cdot z \cdot \langle r \rangle^2}{6\, m^*_e \cdot V} = -\frac{e^2 \cdot z \cdot \langle r \rangle^2}{6\, m^*_e} \cdot \rho_{atom}}$$

- With ρ_{atom} = number of atoms per unit volume

Plugging in numbers will yield χ values around **$-(10^{-5} - 10^{-7})$** in good agreement with experimental values.

4.2.2 Paramagnetism

The treatment of paramagnetism in the most simple way is exactly identical to the treatment of orientation polarization. All you have to do is to replace the *electric dipoles* by magnetic dipoles, which we call *magnetic moments*.

- We have permanent dipole moments in the material, they have no or negligible interaction between them, and they are free to point in any direction *even in solids*.
- This is a major difference to electrical dipole moments which can *only* rotate if the whole atom or molecule rotates; i.e. only in liquids. This is why the treatment of magnetic materials focusses on ferromagnetic materials and why the underlying symmetry of the math is not so obvious in real materials.
- In an external magnetic field the magnetic dipole moments have a tendency to orient themselves into the field direction, but this tendency is opposed by the thermal energy, or better entropy of the system.

Using exactly the same line of argument as in the case of orientation polarization, we have for the potential energy ***W*** of a magnetic moment (or dipole) ***m*** in a magnetic field ***H***

$$W(\varphi) = -\ \mu_0 \cdot \underline{m} \cdot \underline{H} = -\ \mu_0 \cdot m \cdot H \cdot \cos\varphi$$

- With φ = angle between $\underline{H}$ and $\underline{m}$.

In *thermal equilibrium*, the number of magnetic moments with the energy ***W*** will be $N[W(\varphi)]$, and that number is once more given by the Boltzmann factor:

$$N[W(\varphi)] = c \cdot \exp -(W/kT) = c \cdot \exp \frac{m \cdot \mu_0 \cdot H \cdot \cos\varphi}{kT} = N(\varphi)$$

- As befoere, ***c*** is some as yet undetermined constant.

As before, we have to take the component in field direction of all the moments having the same angle with ***H*** and integrate that over the unit sphere. The result for the induced magnetization m_{ind} and the total magnetization ***M*** is the same as before for the induced dielectric dipole moment:

$$m_{\text{ind}} = m \cdot \left(\coth\beta - \frac{1}{\beta}\right)$$

$$M = N \cdot m \cdot L(\beta)$$

$$\beta = \frac{\mu_0 \cdot m \cdot H}{kT}$$

- With **L(β) = Langevin function = cothβ – 1/β**

The only interesting point is the *magnitude* of β. In the case of the orientation polarization it was ≤ **1** and we could use a simple approximation for the Langevin function.

- We know that ***m*** will be of the order of magnitude of **1 Bohr magneton**. For a rather large magnetic field strength of $5 \cdot 10^6$ **A/m**, we obtain as an estimate for an upper limit $\beta = 1.4 \cdot 10^{-2}$, meaning that the range of β is even smaller as in the case of the electrical dipoles.
- We are thus justified to use the simple approximation **L(β) = β/3** and obtain

$$M = N \cdot m \cdot (\beta/3) = \frac{N \cdot m^2 \cdot \mu_0 \cdot H}{3kT}$$

- The paramagnetic susceptibility $\chi = M/H$, finally, is

$$\chi_{para} = \frac{N \cdot m^2 \cdot \mu_0}{3kT}$$

- Plugging in some typical numbers (A Bohr magneton for ***m*** and typical densities), we obtain $\chi_{para} \approx +10^{-3}$; i.e. an *exceedingly small* effect, but with certain characteristics that will carry over to ferromagnetic materials:
    - There is a strong temperature dependence and it follows the "**Curie law**":

$$\chi_{para} = \frac{\text{const}}{T}$$

    - Since ferromagnets of all types turn into paramagnets above the Curie temperature T_C, we may simply expand Curies law for this case to

$$\chi_{ferro}(T > T_C) = \frac{\text{const*}}{T - T_C}$$

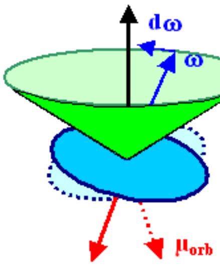
- In summary, paramagnetism, stemming from some (small) average alignement up of permanent magnetic dipoles associated with the atoms of the material, *is of no (electro)technical consequence*. It is, however, important for analytical purposes called "**Electron spin resonance**" (***ESR***) techniques.
- There are other types of paramagnetism, too. Most important is, e.g., the **paramagnetism of the free electron gas**. Here we have magnetic moments associated with spins of electrons, but in a *mobile* way **-** they are not fixed at the location of the atoms
    - But as it turns out, other kinds of paramagnetism (or more precisely: calculations taking into account that magnetic moments of atoms can not assume any orientation but only sone quantized ones) do not change the general picture: *Paramagnetism is a weak effect*.

4.2.3 Summary to: Dia- and Paramagnetism

- Dia- and Paramagentic propertis of materials are of no consequence whatsoever for products of electrical engineering (or anything else!)
  - Only their common denominator of being essentially "non-magnetic" is of interest (for a submarine, e.g., you want a non-magnetic steel)
  - For research tools, however, these forms of magnitc behavious can be highly interesting ("paramagentic resonance")

Normal diamagnetic materials: $\chi_{\text{dia}} \approx -(10^{-5} - 10^{-7})$
Superconductors (= ideal diamagnets): $\chi_{\text{SC}} = -1$
Paramagnetic materials: $\chi_{\text{para}} \approx +10^{-3}$

- Diamagnetism can be understood in a semiclassical (Bohr) model of the atoms as the response of the current ascribed to "circling" electrons to a changing magnetic field via classical induction ($\propto$ **dH/dt**).
  - The net effect is a precession of the circling electron, i.e. the normal vector of its orbit plane circles around on the green cone. ⇒
  - The "Lenz rule" ascertains that inductive effects oppose their source; diamagnetism thus weakens the magnetic field, $\chi_{\text{dia}} < 0$ must apply.



- Running through the equations gives a result that predicts a very small effect. ⇒ A proper quantum mechanical treatment does not change this very much.

$$\chi_{\text{dia}} = -\frac{e^2 \cdot z \cdot \langle r \rangle^2}{6\, m^*_e} \cdot \rho_{\text{atom}} \approx -(10^{-5} - 10^{-7})$$

- The formal treatment of paramagnetic materuials is mathematically completely identical to the case of orientation polarization
  - The range of realistc β values (given by largest ***H*** technically possible) is even smaller than in the case of orientation polarization. This allows tp approximate **L(β)** by **β/3**; we obtain:

$$\chi_{\text{para}} = \frac{N \cdot m^2 \cdot \mu_0}{3kT}$$

  - Insertig numbers we find that χ_{para} is indeed a number just slightly larger than **0**.

$$W(\varphi) = -\ \mu_0 \cdot \underline{m} \cdot \underline{H} = -\ \mu_0 \cdot m \cdot H \cdot \cos\varphi$$

Energy of magetic dipole in magnetic field

$$N[W(\varphi)] = c \cdot \exp -(W/kT) = c \cdot \exp \frac{m \cdot \mu_0 \cdot H \cdot \cos\varphi}{kT} = N(\varphi)$$

(Boltzmann) Distribution of dipoles on energy states

$$M = N \cdot m \cdot L(\beta)$$

$$\beta = \frac{\mu_0 \cdot m \cdot H}{kT}$$

Resulting Magnetization with Langevin function **L(β)** and argument β

4.3 Ferromagnetism

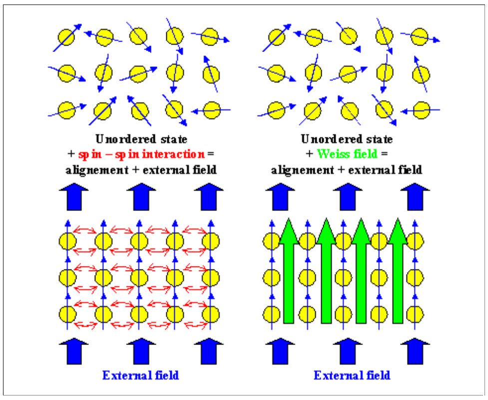
4.3.1 Mean Field Theory of Ferromagnetism

The Mean Field Approach

- In contrast to dia- and paramagnetism, *ferromagnetism* is of *prime importance* for electrical engineering. It is, however, one of the most difficult material properties to understand.
  - It is not unlike "ferro"electricity, in relying on strong interactions between neighbouring atoms having a permanent magnetic moment ***m*** stemming from the *spins* of electrons.
  - But while the interaction between electric dipoles can, at least in principle, be understood in classical and semi-classical ways, *the interaction between spins of electrons is an exclusively quantum mechanical effect with no classical analogon*. Moreover, a theoretical treatment of the three-dimensional case giving reliable results still eludes the theoretical physicists.
  - Here we must accept the fact that only **Fe**, **Co**, **Ni** (and some rare earth metals) show strong interactions between spins and thus ferromagnetism in elemental crystals.
  - In *compounds*, however, many more substances exist with spontaneous magnetization coming from the coupling of spins.
- There is, however, a relatively *simple theory of ferromagnetism*, that gives the proper relations, temperature dependences etc., - with one major drawback: It starts with an *unphysical assumption*.
  - This is the **mean field theory** or the **Weiss theory** of ferromagnetism. It is a phenomenological theory based on a central (wrong) assumption:

Substitute the elusive spin - spin interaction between electrons
by the interaction of the spins with a very strong magnetic field.

  - In other words, *pretend*, that in addition to your external field there is a *built-in magnetic field* which we will call the **Weiss field**. The Weiss field will tend to line up the magnetic moments - you are now treating ferromagnetism as an *extreme* case of paramagnetism. The scetch below illustrates this



- Of course, if the material you are looking at *is* a real ferromagnet, you don't have to *pretend* that there is a built-in magnetic field, because there *is* a large magnetic field, indeed. But this looks like mixing up cause and effect! What you want to result from a calculation is what you start the calculation with!
  - This is called a self-consistent approach. You may view it as a closed circle, where cause and effect loose their meaning to some extent, and where a calculation produces some results that are fed back to the beginning and repeated until some parameter doesn't change anymore.
  - Why are we doing this, considering that this approach is rather questionable? Well - it works! It gives the right relations, in particular the temperature dependence of the magnetization.
- The local magnetic field H_{loc} for an external field H_{ext} then will be

$$H_{loc} = H_{ext} + H_{Weiss}$$

  - Note that this has not much to do with the local electrical field in the Lorentz treatment. We call it "local" field, too, because it is supposed to contain everything that acts *locally*, including the modifications we ought to make to account for effects as in the case of electrical fields. But since our fictitious "Weiss field" is so much larger than everything coming from real fields, we simply can forget about that.

- Since we treat this *fictive* field H_{Weiss} as an internal field, we write it as a superposition of the external field ***H*** and a field stemming from the internal magnetic polarization ***J***:

$$\boxed{H_{loc} = H_{ext} + w \cdot J}$$

  - With ***J*** = magnetic polarization and ***w*** **= Weiss's factor**; a constant that now *contains the physics of the problem*.

- This is the decisive step. We now identify the Weiss field with the magnetic polarization that is caused by it. And, yes, as stated above, we now do mix up cause and effect to some degree: the fictitiuos Weiss field causes the alignments of the individual magnetic moments which than produce a magnetic polarization that causes the local field that we identify with the Weiss field and so on.
  - But that, after all, *is* what happens: the (magnetic moments of the) spins interact causing a field that causes the interaction, that ....and so on . If your mind boggles a bit, that is as it should be. The magnetic polarization caused by spin-spin interactions and mediating spin-spin interaction just *is* - asking for cause and effect is a futile question.
  - The Weiss factor ***w*** now contains *all the local effects* lumped together - in analogy to the Lorentz treatment of local fields, μ_0, and the interaction between the spins that leads to ferromagnetism as a result of some fictive field.
  - But lets be very clear: *There is no internal magnetic field* H_{Weiss} *in the material* before the spins become aligned. This completely fictive field just leads - within limits - to the same interactions you would get from a proper quantum mechanical treatment. Its big advantage is that it makes calculations possible if you determine the parameter ***w*** experimentally.

- All we have to do now is to repeat the calculations done for paramagnetism, substituting H_{loc} wherever we had ***H***. Lets see where this gets us.

Orientation Polarization Math with the Weiss Field

- The potential energy ***W*** of a magnetic moment (or dipole) **m** in an external magnetic field **H** now becomes

$$\boxed{\begin{aligned} W &= -\, m \cdot \mu_0 \cdot (H + H_{Weiss}) \cdot \cos\varphi \\ &= -\, m \cdot \mu_0 \cdot (H + w \cdot J) \cdot \cos\varphi \end{aligned}}$$

- The Boltzmann distribution of the energies now reads

$$\boxed{N(W) = c \cdot \exp - \frac{W}{kT} = c \cdot \exp \frac{m \cdot \mu_0 \cdot (H + w \cdot J) \cdot \cos\varphi}{kT}}$$

- The Magnetization becomes

$$\boxed{\begin{aligned} M &= N \cdot m \cdot L(\beta) \\ &= N \cdot m \cdot L\left(\frac{m \cdot \mu_0 \cdot (H + w \cdot J)}{kT}\right) \end{aligned}}$$

  - In the last equation the argument of **L(β)** is spelled out; it is quite significant that β contains ***w · J***.

- The total polarization is $J = \mu_0 \cdot M$, so we obtain the final equation

$$\boxed{J = N \cdot m \cdot \mu_0 \cdot L(\beta) = N \cdot m \cdot \mu_0 \cdot L\left(\frac{m \cdot \mu_0 \cdot (H + w \cdot J)}{kT}\right)}$$

- Written out in full splendor this is

$$J = N \cdot m \cdot \mu_0 \cdot \coth\left(\frac{m \cdot \mu_0 \cdot (H + w \cdot J)}{kT}\right) - \frac{N \cdot kT}{(H + w \cdot J)}$$

- What we really want is the magnetic polarization ***J*** as a function of the external field ***H***. Unfortunately we have a *transcendental* equation for ***J*** which can not be written down directly without a "***J***" on the right-hand side.
  - What we also like to have is the value of the spontaneous magnetization ***J*** for no external field, i.e. for **H = 0**. Again, there is no analytical solution for this case.
  - There is an easy graphical solution, however: We actually have *two* equations for which must hold *at the same time*:
  - The argument β of the Langevin function is

$$\beta = \frac{m \cdot \mu_0 \cdot (H + w \cdot J)}{kT}$$

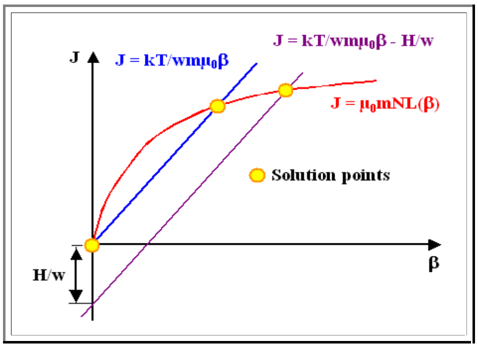
  - Rewritten for ***J***, we get our first equation:

$$J = \frac{kT \cdot \beta}{w \cdot m \cdot \mu_0} - \frac{H}{w}$$

  - This is simply a straight line with a slope and intercept value determined by the interesting variables ***H*, *w***, and ***T***.
- On the other hand we have the equation for ***J***, and this is our second independent equation
.

$$J = N \cdot m \cdot \mu_0 \cdot L(\beta) = N \cdot m \cdot \mu_0 \cdot L\left(\frac{m \cdot \mu_0 \cdot (H + w \cdot J)}{kT}\right)$$

  - This is simply the Langevin function which we know for any numerical value for β
- All we have to do is to draw *both* functions in a ***J* -** β diagram
  - We can do that by simply putting in some number for β and calculating the results. The intersection of the two curves gives the solutions of the equation for ***J***.
  - This looks like this



  - Without knowing anything about β, we can draw a definite conclusion:
  - For ***H* = 0** we have *two* solutions (or none at all, if the straight line is too steep): One for ***J* = 0** and one for a rather large ***J***.
  - It can be shown that the solution for ***J* = 0** is unstable (it disappears for an arbitrarily small field ***H***) so we are left with a *spontaneous large magnetic polarization* without an external magnetic field as the first big result of the mean field theory.
- We can do much more with the mean field theory, however.

- *First*, we note that switching on an *external magnetic field* does not have a large effect. $\boldsymbol{J}$ increases somewhat, but for realistic values of $\boldsymbol{H/w}$ the change remains small.
- *Second*, we can look at the *temperature dependence* of $\boldsymbol{J}$ by looking at the straight lines. For $\boldsymbol{T \to 0}$, the intersection point moves all the way out to infinity. This means that all dipoles are now lined up in the field and **L(β)** becomes **1**. We obtain the *saturation value* $\boldsymbol{J_{sat}}$

$$J_{sat} = N \cdot m \cdot \mu_0$$

- *Third*, we look at the effect of increasing *temperatures*. Raising $\boldsymbol{T}$ increases the slope of the straight line, and the two points of intersection move together. When the slope is equal to the slope of the Langevin function (which, as we know, is **1/3**), the two points of solution merge at $\boldsymbol{J = 0}$; if we increase the slope for the straight line even more by increasing the temperature by an incremental amount, solutions do no longer exist and the spontaneous magnetization disappears.

- This means, there is a *critical temperature* above which ferromagnetism disappears. This is, of course, the **Curie temperature** $\boldsymbol{T_C}$.
  - At the Curie temperature $\boldsymbol{T_C}$, the slope of the straight line and the slope of the Langevin function for **β = 0** must be identical. In formulas we obtain:

$$\frac{dJ}{d\beta} = \frac{kT_C}{w \cdot m \cdot \mu_0} = \text{slope of the straight line}$$

$$\left.\frac{dJ}{d\beta}\right|_{\beta = 0} = N \cdot m \cdot \mu_0 \cdot \frac{dL(\beta)}{d\beta} = \frac{N \cdot m \cdot \mu_0}{3}$$

  - We made use of our old insight that the slope of the Langevin function for **β → 0** is **1/3**.
- Equating both slopes yields for $\boldsymbol{T_C}$

$$T_C = \frac{N \cdot m^2 \cdot \mu_0{}^2 \cdot w}{3k}$$

- This is pretty cool. We did not solve an transcendental equation nor go into deep quantum physical calculations, but still could produce rather simple equations for prime material parameters like the Curie temerature.
  - If we only would know $\boldsymbol{w}$, the Weiss factor! Well, we do *not* know $\boldsymbol{w}$, but now we can turn the equation around: If we know $\boldsymbol{T_C}$, we can *calculate* the Weiss factor $\boldsymbol{w}$ and thus the *fictive magnetic field* that we need to keep the spins in line.
  - In **Fe**, for example, we have $\boldsymbol{T_C = 1043}$ **K**, $\boldsymbol{m = 2{,}2 \cdot m_{Bohr}}$. It follows that

$$H_{Weiss} = w \cdot J = 1.7 \cdot 10^9 \text{ A/m}$$

  - This is a truly *gigantic* field strength telling us that quantum mechanical spin interactions, if existent, are not to be laughed at.
  - If you do not have a feeling of what this number means, consider the unit of $\boldsymbol{H}$: A field of $\boldsymbol{1{,}7 \cdot 10^9}$ **A/m** is produced if a current of $\boldsymbol{1{,}7 \cdot 10^9}$ **A** flows through a loop (= coil) with $\boldsymbol{1 \text{ m}^2}$ area. Even if you make the loop to cover only $\boldsymbol{1 \text{ cm}^2}$, you still need $\boldsymbol{1{,}7 \cdot 10^5}$ **A**.
- We can go one step further and approximate the Langevin function again for temperatures $\boldsymbol{> T_C}$, i.e. for **β < 1** by

$$L(\beta) \approx \frac{\beta}{3}$$

  - This yields

$$J(T > T_C) \approx \frac{N \cdot m^2 \cdot \mu_0{}^2}{3kT} \cdot (H + w \cdot J)$$

- From the equation for T_C we can extract w and insert it, arriving at

$$J(T > T_C) \approx \frac{N \cdot m^2 \cdot \mu_0{}^2}{3k(T - T_C)} \cdot H$$

- Dividing by H gives the susceptibility χ for $T > T_C$ and the final formula

$$\chi = \frac{J}{H} = \frac{N \cdot m^2 \cdot \mu_0{}^2}{3k \cdot (T - T_C)} = \frac{\text{const.}}{T - T_C}$$

- This is the famous Curie law for the paramagnetic regime at high temperatures which was a phenomenological thing so far. Now we derived it with a theory and will therefore call it **Curie - Weiss law**.

In summary, the mean field approach ain't that bad! It can be used for attacking many more problems of ferromagnetism, but you have to keep in mind that it is only a description, and not based on sound principles.

Questionaire
Multiple Choice questions to 4.3.1

4.3.2 Beyond Mean Field Theory

Some General Considerations

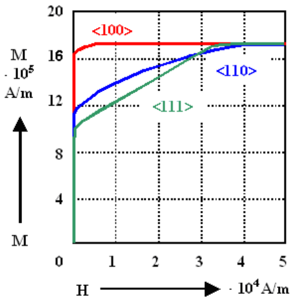
- According to the mean field theory, if a material is ferromagnetic, *all* magnetic moments of the atoms would be coupled and point in the same direction. We now ask a few questions:
  - **1.** *Which direction* is that going to be for a material just sitting there? Is there some preferred internal direction or are all directions equal? In other words: Do we have to make the fictitious Weiss field H_{Weiss} larger in some directions compared to other ones? Of course, we wonder if some crystallographic directions have "special status".
  - **2.** What happens if an external field is superimposed in some direction that does *not* coincide with a preferred internal direction?
  - **3.** What happens if it *does*? Or if the external field is *parallel* to the internal one, but pointing in the *opposite* direction?
- The (simple) mean field theory remains rather silent on those questions. With respect to the first one, the internal alignment direction would be determined by the direction of the fictive field H_{Weiss}, but since this field does not really exist, each direction seems equally likely.
  - In real materials, however, we might expect that the direction of the magnetization is not totally random, but has some specific preferences. This is certainly what we must expect for *crystals*.
  - A specific direction in real ferromagnetic materials could be the result of crystal anisotropies, inhomogeneities, or external influences - none of which are contained within the mean field theory (which essentially treats perfectly isotropic and infinitely large materials).
- *Real* ferromagnetic materials thus are more complicated than suggested by the mean field theory - for a very general reason:
- Even if we can lower the internal *energy* U of a crystal by aligning magnetic moments, we still must keep in mind that the aim is always to minimize the *free enthalpy* $G = U - TS$ of the *total* system.
  - While the *entropy* part coming from the degree of orderliness in the system of magnetic moments has been taken care of by the general treatment in the frame work of the orientation polarization, we must consider the enthalpy (or energy) U of the *system* in more detail. So far we only minimized U with respect to *single* magnetic moments in the Weiss field.
  - This is so because the mean field approach essentially relied on the fact that by aligning the spins relative to the (fictitious) Weiss field, we lower the energy of the individual spin or magnetic moments as treated before by some energy W_{align}. We have

$$\boxed{U_{align} = U_{random} - W_{align}}$$

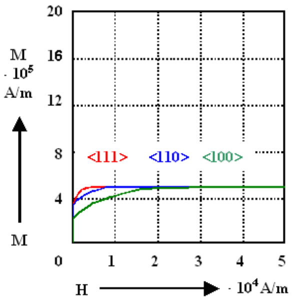
  - But, as discussed above, real materials are mostly (poly)crystals and we must expect that the real (quantum-mechanical) interaction between the magnetic moments of the atoms are different for different directions in the crystal. There is some anisotropy that must be considered in the U_{align} part of the free enthalpy.
  - Moreover, there are other contributions to U not contained in the mean field approach. Taken everything together makes quantitative answers to the questions above exceedingly difficult.
- There are, however, a few relatively simple general rules and experimental facts that help to understand what really happens if a ferromagnetic material is put into a magnetic field. Let's start by looking at the **crystal anisotropy**.

Crystal Anisotropy

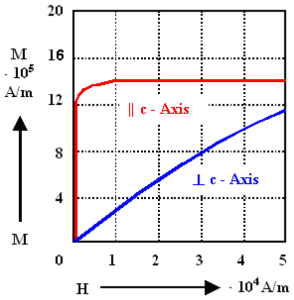
- Generally, we must expect that there is a preferred crystallographic direction for the spontaneous magnetization, the so-called "**easy directions**". If so, it would need some energy to change the magnetization direction into some other orientations; the "**hard directions**".
  - That effect, if existent, is easy to measure: Put a single crystal of the ferromagnetic material in a magnetic field H that is oriented in a certain crystal direction, and measure the magnetization of the material in that direction:
  - If it happens to be an *easy* direction, you should see a strong magnetization that reaches a **saturation** value - obtained when all magnetic moments point in the desired direction - already at low field strength H. If, on the other hand, H happens to be in a *hard* direction, we would expect that the magnetization only turns into the H direction ***reluctantly***, i.e. only for large values of H will we find saturation.
- This is indeed what is observed, classical data for the elemental ferromagnets **Fe**, **Ni**, **Co** are shown below:



Anisotropy of magnetization in **Fe**.



Anisotropy of magnetization in **Ni**.



Anisotropy of magnetization in **Co**.

Anisotropy of the magnetization in **Fe** (***bcc*** lattice type), **Ni** (***fcc*** lattice type), and **Co** (***hcp*** lattice type).

- The curves are easy to interpret qualitatively along the lines stated above; consider, e.g., the **Fe** case:
  - For field directions not in **<100>**, the spins become aligned in the **<100>** directions pointing as closely as possible in the external field direction.
  - The magnetization thus is just the component of the **<100>** part in the field direction; it is obtained for arbitrarily small external fields.
  - Increasing the magnetization, however, means turning spins into a "hard" directions, and this will proceed reluctantly for large magnetic fields.
  - At sufficiently large fields, however, all spins are now aligned into the external field directions and we have the same magnetization as in the easy direction.
- The curves above contain the material for a simple little exercise:

Exercise 4.3-2

Magnetic moments of Fe, Ni, Co

Questionnaire

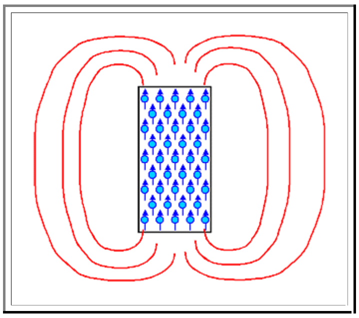
Multiple Choice questions to 4.3.2

4.3.3 Magnetic Domains

Reducing the External Magnetic Field

If we now turn back to the question of what you would observe for the magnetization of a single crystal of ferromagnetic material just sitting on your desk, you would now expect to find it completely magnetized in its *easy direction* - even in the presence of a not overly strong magnetic field.

- This would look somewhat like this:



- There would be a large *internal* magnetization *and* a large *external* magnetic field ***H*** - we would have an ideal **permanent magnet**.
- And we also would have a high energy situation, because the external magnetic field around the material contains magnetic field energy W_{field}.
- In order to make life easy, we do not care how large this energy is, even so we could of course calculate it. We only care about the general situation: We have a rather large energy outside the material caused by the perfect line up of the magnetic dipoles in the material
- How do we know that the field energy is rather large? Think about what will happen if you put a material as shown in the picture next to a piece of iron for example.
- What we have is obviously a strong **permanent magnet**, and as we know it will violently attract a piece of iron or just any ferromagnetic material. That means that the external magnetic field is strong enough to line up all the dipoles in an other ferromagnetic material, and that, as we have seen, takes a considerable amount of energy.

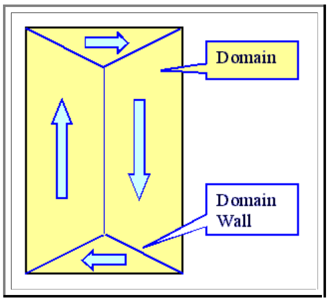
The **internal energy *U*** of the system thus must be written

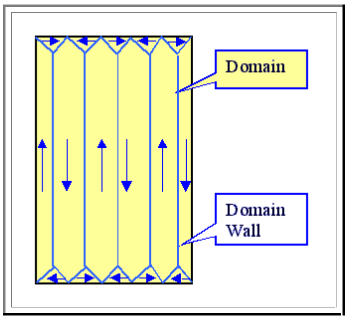
$$U_{align} = U_{random} - W_{align} + W_{field}$$

- The question is if we can somehow lower W_{field} substantially - possibly by spending some smaller amount of energy elsewhere. Our only choice is to *not* exploit the maximum alignment energy W_{align} as it comes from perfect alignment in *one* direction.
- In other words, are there non-perfect alignments patterns that only cost *a little bit* W_{align} energy, but safe *a lot* of W_{field} energy? Not to mention that we always gain a bit in entropy by not being perfect?

The answer is *yes* - we simply have to introduce **magnetic domains**.

- Magnetic domains are regions in a crystal with different directions of the magnetizations (but still pointing in one of the easy directions); they must by necessity be separated by **domain walls**. The following figures show some possible configurations





Both domain structures *decrease* the external field and thus W_{field} because the flux lines now can close inside the material. And we kept the alignment of the magnetic moments in most of the material, it only is disturbed in the domain walls. Now which one of the two configurations shown above is the better one?

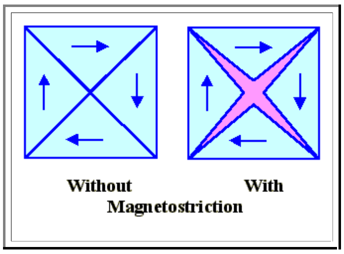
- Not so easy to tell. With *many* domains, the magnetic flux can be confined better to the inside of the material, but the total domain wall area goes up - we loose plenty of W_{align}.
- The energy lost by non-perfect alignment in the domains walls can be expressed as a property of the domain wall, as a **domain wall energy**. A magnetic domain wall, by definition a two-dimensional defect in the otherwise perfect order, thus carries an energy (per $\mathbf{cm^2}$) like any other two-dimensional defect.

  - There must be an optimum balance between the energy gained by *reducing* the external field, and the energy lost in the domain wall energy. And all the time we must remember that the magnetization in a domain is always in an easy direction (without strong external fields).

- We are now at the ***end of our tether***. While the ingredients for minimizing the system energy are perfectly clear, nobody can calculate exactly what kind of stew you will get for a given case.
  - *Calculating* domain wall energies from first principles is already nearly hopeless, but even with experimental values and for perfect single crystals, it is not simple to deduce the *domain structure* taking into account the anisotropy of the crystal and the external field energy.
- And, too make things even worse (for theoreticians) there are even *more* energetic effects that influence the domain structure. Some are important and we will give them a quick look.

Magnetostriction and Interaction with Crystal Lattice Defects

- The interaction between the magnetic moments of the atoms that produces alignment of the moments - ferromagnetism, ferrimagnetism and so on - necessarily acts as a force between the atoms, i.e. the interaction energy can be seen as a potential and the (negative) derivative of this potential is a *force*.
  - This interaction force must be added to the general binding forces between the atoms.
  - In general, we must expect it to be *anisotropic* - but not necessarily in the same way that the binding energy could be anisotropic, e.g. for covalent bonding forces.
  - The total effect thus usually will be that the lattice constant is *slightly* different in the direction of the magnetic moment. A *cubic* crystal may become orthorhombic upon magnetization, and the crystal *changes dimension* if the direction of the magnetization changes.
- A crystal "just lying there" will be magnetized in several directions because of its magnetic domains and the anisotropy of the lattice constants averages out: A cubic crystal is still - *on average* - cubic, but with a slightly changed lattice constant.
  - However, if a large external field $\boldsymbol{H}_{\mathbf{ex}}$ forces the internal magnetization to become oriented in field directions, the material now (usually) responds by some *contraction in field direction* (no more averaging out); this effect is called **magnetostriction**. This word is generally used for the description of the effect that the interatomic distances are different if magnetic moments are aligned.
  - The *amount* of magnetostriction is different for different magnetic materials, again there are no straight forward calculations and experimental values are used. It is a complex phenomena.
  - Magnetostriction is a useful property; especially since recently "giant magnetostriction" has been discovered. Technical uses seem to be just around the corner at present.
- Magnetostriction also means that a piece of crystal that contains a magnetic domain would have a somewhat *different* dimension as compared to the same piece without magnetization.
  - Lets illustrate that graphically with an (oversimplified, but essentially correct) picture:



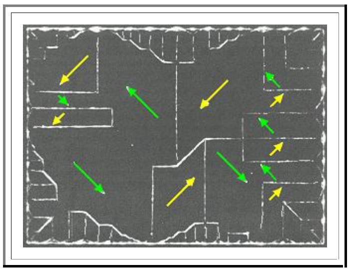
  - In this case the magnetostriction is perpendicular to the magnetization. The four domains given would assume the shape shown on the right hand side.
- Since the crystal does not come apart, there is now some *mechanical strain and stress* in the system. This has two far reaching consequences
  - **1.** We have to add the *mechanical energy* to the energy balance that determines the domain structure, making the whole thing even more complicated.
  - **2.** We will have an *interaction* of domain walls with structural defects that introduce mechanical stress and strain in the crystal. If a domain wall moves across a dislocation, for example, it might relieve the stress introduced by the dislocation in one position and increase it in some other position. Depending on the signs, there is an attractive or repelling force. In any case, there is some interaction: *Crystal lattice defects attract or repulse domain walls*.
- Generally speaking, both the domain structure and the movement of domain walls will be influenced by the internal structure of the material. A rather perfect single crystal may behave magnetically quite differently from a polycrystal full of dislocations.
  - This might be hateful to the *fundamentalists* among the physicists: There is not much hope of calculating the domain structure of a given material from first principles and even less hope for calculating what happens if you deform it mechanically or do something else that changes its internal structure.

- However, we have the *engineering point of view*:

This is great

- The complicated situation with respect to domain formation and movement means that there are *many ways* to influence it.
- We do not have to live with a few materials and take them as they are, we have many options to tailor the material to specific needs. Granted, there is not always a systematic way for optimizing magnetic materials, and there might be much trial and error - but progress is being made.

What a real domain structure looks like is shown in the picture below. Some more can be found in the link.



- We see the domains on the surface of a single crystalline piece of **Ni**. How domains can be made visible is a long story **-** it is not easy! We will not go into details here.

Summarizing what we have seen so far, we note:

- **1.** The **domain structure** of a given magnetic material *in equilibrium* is the result of *minimizing the free enthalpy* mostly with respect to the *energy* term.
- **2.** There are several contributions to the energy; the most important ones being magnetic stray fields, magnetic anisotropy, magnetostriction and the interaction of the internal structure with these terms.
- **3.** The domain structure can be very complicated; it is practically impossible to calculate details. Moreover, as we will see, it is not necessarily always the equilibrium structure!

But this brings us to the next subchapter, the movement of *domain walls* and the *hysteresis curve*.

Questionaire
Multiple Choice questions to 4.3.3

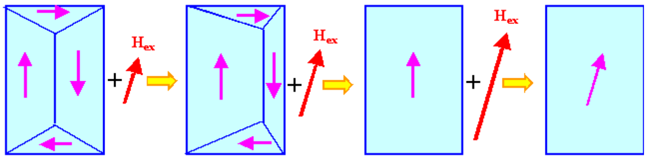
4.3.4 Domain Movement in External Fields

Domain Movement in External Fields

What happens if we apply an external field to a ferromagnet with its equilibrium domain structure?

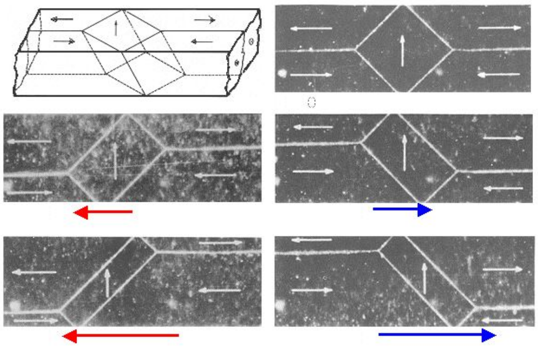
- The domains oriented most closely in the direction of the external field will gain in energy, the other ones loose; always following the basic equation for the energy of a dipole in a field.
- Minimizing the total energy of the system thus calls for increasing the size of favorably oriented domains and decreasing the size of unfavorably oriented ones. Stray field considerations still apply, but now we have an external field anyway and the stray field energy looses in importance.
- We must expect that the most favorably oriented domain will win for large external fields and all other domains will disappear.
- If we increase the external field beyond the point where we are left with only one domain, it may now even become favorable, to orient the atomic dipoles off their "easy" crystal direction and into the field.
- After that has happened, all atomic dipoles are in field direction - more we cannot do. The magnetization than reaches a saturation value that cannot be increased anymore.

Schematically, this looks like as shown below:



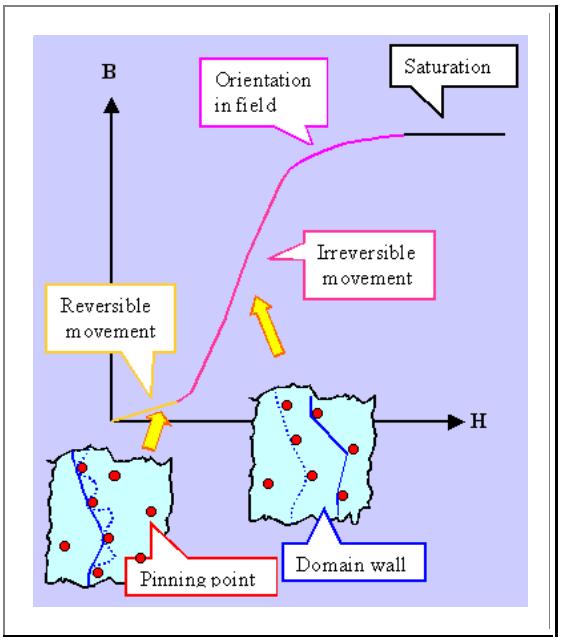
- Obviously, domain walls have to move to allow the new domain structure in an external magnetic field.

What this looks like in reality is shown below for a small single crystal of iron.



- As noted before, domain walls interact with *stress* and strain in the lattice, i.e. with *defects* of all kinds. They will become "stuck" (the proper expression for things like that is "**pinned**") to defects, and it needs some force to ***pry*** them off and move them on. This force comes from the external magnetic field.

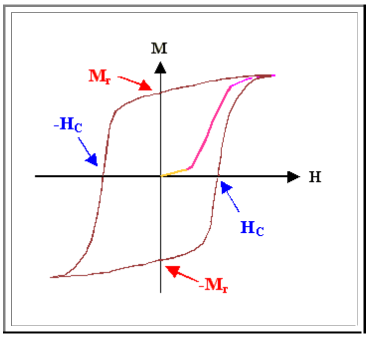
The magnetization curve that goes with this looks like this:



- For small external fields, the domain walls, being pinned at some defects, just bulge out in the proper directions to increase favorably oriented domains and decrease the others. The magnetization (or the magnetic flux ***B***) increases about linearly with ***H***
- At larger external fields, the domain walls overcome the pinning and move in the right direction where they will become pinned by other defects. Turning the field of will not drive the walls back; the movement is irreversible.
- After just one domain is left over (or one big one and some little ones), increasing the field even more will turn the atomic dipoles in field direction. Since even under most unfavorable condition they were at most **45^o** off the external direction, the increase in magnetization is at most **1/cos(45^o) = 1.41**.
- Finally, saturation is reached. All magnetic dipoles are fully oriented in field direction, no further increase is possible.

If we switch off the external field anywhere in the irreversible region, the domain walls might relax back a little, but for achieving a magnetization of zero again, we must use force to move them back, i.e. an external magnetic field pointing in the opposite direction.

- In total we obtain the well known **hysteresis** behavior as shown in the **hysteresis curve** below.



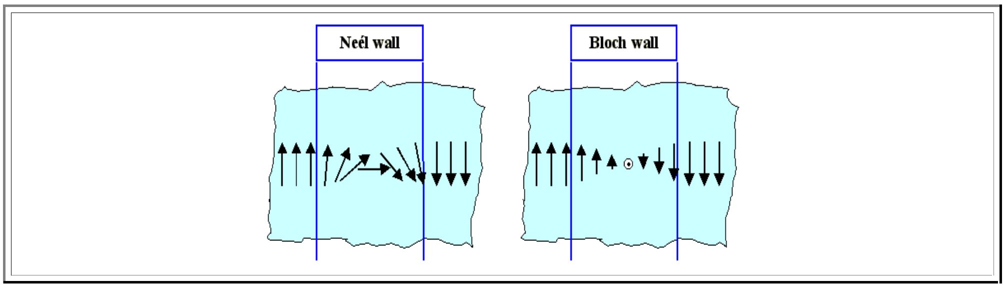
- The resulting hysteresis curve has two particular prominent features:
  - The remaining magnetization for zero external field, called the **remanence** M_R, and
  - the magnitude of the external field needed to bring the magnetization down to zero again. This is called **coercivity** or coercive field strength H_C.
- *Remanence* and *coercivity* are two numbers that describe the major properties of ferromagnets (and, of course, ferrimagnets, too). Because the exact shape of the hysteresis curve does not vary too much.
- Finally, we may also address the **saturation magnetization** M_S as a third property that is to some extent independent of the other two.

- Technical optimization of (ferro)magnetic materials first always focuses on these two numbers (plus, for reasons to become clear very soon, the resistivity).
- We now may also wonder about the dynamic behaviour, i.e. what happens if we change the external field with ever increasing frequency.

Domain Wall Structure

The properties of the domain walls, especially their interaction with defects (but also other domain walls) determine most of the magnetic properties of ferromagnets.

- What is the structure of a domain wall? How can the magnetization change from one direction to another one?
- There are two obvious geometric ways of achieving that goal - and that is also what really happens in practically all materials. This is shown below.



- What kind of wall will be found in real magnetic materials? The answer, like always is: Whichever one has the smallest (free) energy
  - In most bulk materials, we find the **Bloch wall**: the magnetization vector turns bit by bit like a screw out of the plane containing the magnetization to one side of the Bloch wall.
  - In thin layers (oft the same material), however, **Neél walls** will dominate. The reason is that Bloch walls would produce stray fields, while Neél walls can contain the magnetic flux in the material.
- Both basic types of domain walls come in many sub-types, e.g. if the magnetization changes by some defined angle other than **180°**. In thin layers of some magnetic material, special domain structures may be observed, too.
- The interaction of domain walls with magnetic fields, defects in the crystal (or structural properties in amorphous magnetic materials), or intentionally produced structures (like "scratches", localized depositions of other materials, etc., can become fantastically complicated.
  - Since it is the domain structure together with the response of domain walls to these interactions that controls the hystereses curve and therefore the basic magnetic properties of the material, things are even more complicated as described before.
  - But do keep in mind: The underlying basic principles is the minimization of the free enthalpy, and there is nothing complicated about this. The fact that we can no easily write down the relevant equations, no to mention solving them, does not mean that we cannot understand what is going on. And the material has no problem in solving equations, it just assumes the proper structure, proving that there are solutions to the problem.

Questionaire
Multiple Choice questions to 4.3.4

4.3.5 Magnetic Losses and Frequency Behavior

General Remarks

- So far we have avoided to consider the frequency behavior of the magnetization, i.e. we did not discuss what happens if the external field oscillates!
    - The experience with electrical polarization can be carried over to some magnetic behaviour, of course. In particular, the frequency response of *paramagnetic* material will be quite similar to that of electric dipole orientation, and diamagnetic materials show close parallels to the electronic polarization frequency behaviour.
    - Unfortunately, this is of (almost) no interest whatsoever. The "almost" refers to **magnetic imaging** employing **magnetic resonance imaging** (***MRI***) or *nuclear spin resonance imaging* - i.e. some kind of "*computer tomography*". However, this applies to the paramagnetic behavior of the magnetic moments of the *nuclei*, something we haven't even discussed so far.
- What is of interest, however, is what happens in a *ferromagnetic* material if you have expose it to an *changing*, i.e. oscillating magnetic field. $H = H_o \cdot \exp(i\omega t)$
    - Nothing we discussed for dielectrics corresponds to this questions. Of course, the frequency behavior of ferroelectric materials would be comparable, but we have not discussed this topic.
    - Being wise from the case of dielectric materials, we suspect that the frequency behavior and some **magnetic energy losses** go in parallel, as indeed they do.
- In contrast to dielectric materials, we will start with looking at magnetic losses *first*.

Hystereses Losses

- If we consider a ferromagnetic material with a given hysteresis curve exposed to an oscillating magnetic field at low frequencies - so we can be sure that the internal magnetization can instantaneously follow the external field - we may consider *two* completely independent mechanisms causing losses.
    - **1.** The changing magnetic field induces currents wandering around in the material - so called ***eddy currents***. This is different from dielectrics, which we always took to be insulators: ferromagnetic materials are usually conductors.
    - **2.** The movement of domain walls needs (and disperses) some energy, these are the *intrinsic* magnetic losses or hystereses losses.
- Both effects add up; the energy lost is converted into heat. Without going into details, it is clear that the losses encountered increase with
    - **1.** The frequency ***f*** in *both* cases, because every time you change the field you incur the same losses per cycle.
    - **2.** The maximum magnetic flux B_{max} in both cases.
    - **3.** The conductivity $\sigma = 1/\rho$ for the eddy currents, and
    - **4.** The magnetic field strength ***H*** for the magnetic losses.
- More involved calculations (see the advanced module) give the following relation for the total ferromagnetic loss P_{Fe} per unit volume of the material

$$\boxed{P_{Fe} \;\approx\; P_{eddy} \;+\; P_{hyst} \;\approx\; \frac{\pi \cdot d^2}{6\rho} \cdot (f \cdot B_{max})^2 \;+\; 2f \cdot H_C \cdot B_{max}}$$

-
    - With d = thickness of the material perpendicular to the field direction, H_C = coercivity.
- It is clear what you have to do to minimize the eddy current losses:
    - Pick a ferromagnetic material with a high resistivity - *if* you can find one. That is the point where *ferri***magnetic** materials come in. What you loose in terms of maximum magnetization, you may gain in reduced eddy losses, because many ferrimagnets are ceramics with a high resistivity.
    - Make d small by stacking insulated thin sheets of the (conducting) ferromagnetic material. This is, of course, what you will find in any run-of-the-mill transformer.
- We will not consider eddy current losses further, but now look at the remaining **hystereses losses** P_{hyst}
    - The term $H_C \cdot B_{max}$ is pretty much the area inside the hystereses curve. Multiply it with two times the frequency, and you have the hystereses losses in a good approximation.
    - In other words: There is nothing you can do - for a given material with its given hystereses curve.

- Your only choice is to select a material with a hystereses curve that is *just right*. That leads to several questions:
  - **1.** What kind of hystereses curve do I need for the application I have in mind?
  - **2.** What is available in terms of hystereses curves?
  - **3.** Can I change the hystereses curve of a given material in a defined way?
- The answer to these questions will occupy us in the next subchapter; here we will just finish with an extremely cursory look at the frequency behavior of ferromagnets.

Frequency Response of Ferromagnets

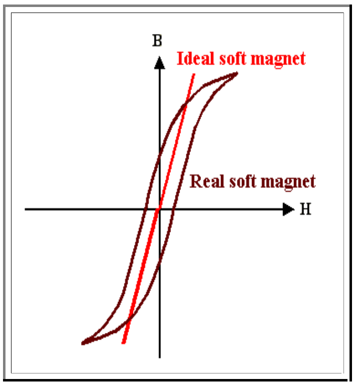
- As already mentioned, we only have to consider ferromagnetic materials - and that means the back-and-forth movement of domain walls in response to the changing magnetic field.
  - We do not have a direct feeling for how fast this process can happen; and we do not have any simplified equations, as in the case of dielectrics, for the forces acting on domain walls. Note that the atoms do *not* move if a domain wall moves - only the direction of the magnetic moment that they carry.
  - We know, however, from the bare fact that permanent magnets exist, or - in other words - that coercivities can be large, that it can take rather large forces to move domain walls - they might not shift easily.
  - This gives us at least a feeling: It will not be easy to move domain walls *fast* in materials with a large coercivity; and even for materials with low coercivity we must not expect that they can take large frequencies, e.g. in the optical region
  - There are materials, however, that still work in the **GHz** region.
- And that is where we stop. There simply is no general way to express the frequency dependence of domain wall movements.
  - That, however, does not mean that we cannot define a **complex magnetic permeability $\mu = \mu' + i\mu''$** for a particular magnetic material.
  - It can be done and it has been done. There simply is no *general* formula for it and that limits its general value.

Questionaire
Multiple Choice questions to 4.3.5

4.3.6 Hard and Soft Magnets

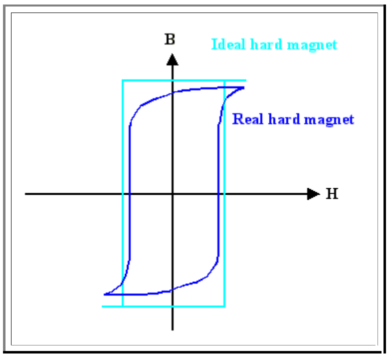
Definitions

- Lets quickly go over the three questions from the preceding sub-chapter
  - **1.** What kind of hystereses curve do I need for the application I have in mind?
- Lets look at two "paradigmatic" applications: A **transformer core** and a **magnetic memory**.
  - The transformer core is ferromagnetic in order to "transport" a large magnetic flux ***B*** produced by the primary coil to the secondary coil. What I want is that the induced flux ***B*** follows the primary field ***H*** as closely as possible.
  - In other words: There should be *no* hystereses loop - just a straight line, as shown below



  - The ideal curve, without any hystereses, does not exist. What you get is something like the curve shown for a real **soft magnet** - because that is what we call a material with a kind of slender hystereses curve and thus small values of coercivity and remanence
  - If we switch on a positive field ***H*** and then go back to zero again, a little bit of magnetization is left. For a rather small reverse field, the magnetic flux reverses, too - the flux ***B*** follows ***H*** rather closely, if not exactly.
  - Hystereses losses are small, because the area enclosed in the hystereses loop is small.
  - But *some* losses remain, and the "transformer core" industry will be very happy if you can come up with a material that is just **1 %** or **2 %** "softer" than what they have now.
  - Beside losses, you have another problem: If you vary ***H*** sinusoidally, the output will be a somewhat distorted sinus, because ***B*** does not follow ***H*** linearly. This may be a problem when transforming *signals*.

- A soft magnetic material will obviously not make a good **permanent magnet**, because its remaining magnetization (its remanence) after switching off the magnetic field ***H*** is small.
  - But a permanent magnet is what we want for a **magnetic storage material**. Here we want to induce a large permanent magnetization by some external field (produced by the "writing head" of our storage device) that stays intact for many years if needs be. Some more information about magnetic storage can be found in an extra module
  - It should be *strong enough* - even so it is contained in a tiny area of the magnetic material on the tape or the storage disc - to produce a measurable effect if the reading head moves over it. It should not be *too strong*, however, because that would make it too difficult to erase it if we want to overwrite it with something else. In short, it should look like this

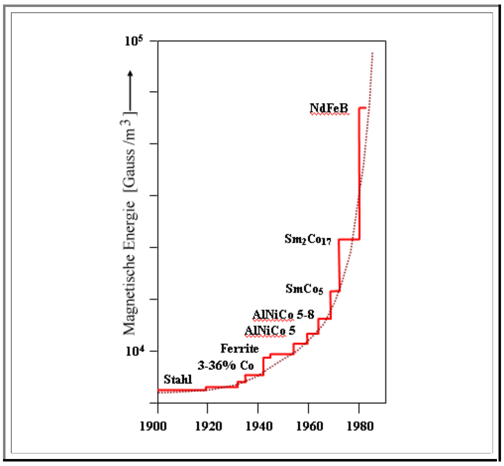


  - We can define what we want in terms of coercivity and remance. Ideally, the hystereses curve is very "square.
  - At some minimum field, the magnetization is rather large and does not change much anymore.
  - If we reverse the field direction, not much happens for a while, but as soon as we move above slightly above the coercivity value, the magnetization switches direction completely.
  - Ferromagnetic losses are unavoidable, we simply must live with them

- Pretty much all possible applications - consult the list in the next section - either calls for *soft* or for *hard* magnets; there isn't much in between.
  - So we now must turn to the second and third question:

Tailoring Hystereses Curves

- The question was: What is available in terms of *hystereses curves*? Good question; it immedately provokes another questions:
  - What is available in terms of ferromagnetic *materials*? The kind of hystereses behavior you get is first of all a property of the specific material you are looking at.
  - For arbitrary chemical compounds, there is little predictive power if they are ferromagnetic or not. In fact, the rather safe bet is that some compound *not* containing **Fe**, **Ni**, or **Co** is *not* ferromagnetic.
  - Even if we restrict ourselves to some compound or alloy containing at least one of the ferromagnetic elements **Fe**, **Ni** or **Co**, it is hard to predict if the result will be ferromagnetic and even harder to predict the kind of hystereses curve it will have. Pure **Fe** in its (high temperature) **fcc** lattice variant is not magnetic, neither are most variants of stainless steel, for example.
- But progress has been made - triggered by an increasing theoretical understanding (there are theories, after all), lots of experience and semi-theoretical guide lines - and just plain old trying out in the lab.
  - This is best demonstrated by looking at the "strength" of permanent magnets as it went up over the years:

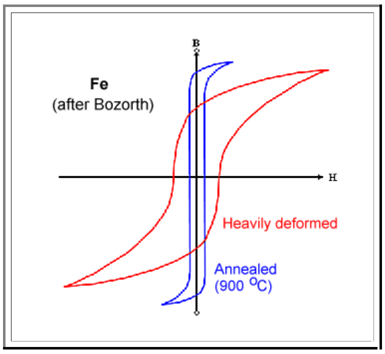


- Not bad. And pretty exotic materials emerged. Who thinks of Cobalt - Samarium compounds, or Neodymium - iron - boron?
  - What will the future bring. Well, I don't know and you shall see!
  - But we can do a little exercise to get some idea of what might be possible

Exercise 6.3.1

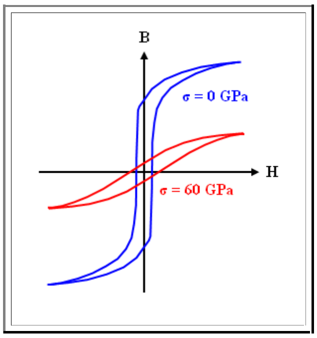
Maximum Magnetization

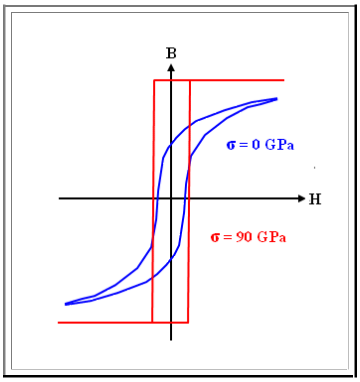
- The final question was: Can I change the hystereses curve of a given material in a defined direction?
  - The answer is: Yes, you can - within limits, of course.
  - The hystereses curve results from the relative ease or difficulty of moving domain walls in a given material. And since domain walls interact with stress and strain in a material, their movement depends on the internal structure of the material; on the kind and density of crystal lattice defects.
- This is best illustrated by looking at hystereses curves of *one and the same material* with different internal structures.



  - There is a big difference for annealed, i.e. relatively defect free iron and heavily deformed iron, i.e. iron full of dislocations, as the figure on the left nicely illustrates
  - We will find similar behavior for most ferromagnetic materials (not for all, however, because some are amorphous).
  - Instead of manipulating the defects in the materials to see what kind of effect we get, we can simply put it under mechanical stress, e.g. by pulling at it. This also may change the hystereses curve very much:

- Here we have the hystereses curves of pure **Ni** samples with and without mechanical tension. The effects are quite remarkable





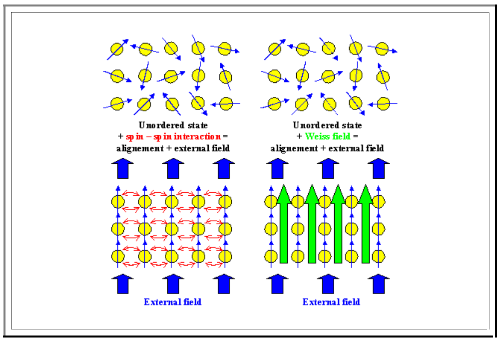
  - In this case the tension force was parallel to the external field ***H***
  - There is a big change in the remanence, but not so much difference in the coercivity.
  - In this case the tension force was at *right angles* to the external field ***H***
  - Big changes in the remanence, not so much effect in the coercivity. We have an almost box-like shape, coming close to the ideal hard magnet from above.

- The final word thus is:
  - There is a ***plethora*** of ways to design ferromagnetic properties out there. The trouble is, we are just learning now how to do it a little bit better than by pure trial and error.
  - The future of magnetism looks bright. With an increased level of understanding, new materials with better properties will result for almost sure. Time will tell.

4.3.7 Summary to: Ferromagnetism

In ferromagnetic materials the magnetic moments of the atoms are "correlated" or lined-up, i.e. they are all pointing in the same direction

- The physical reason for this is a quantum-mechanical spin-spin interaction that has no simple classical analogue.
- However, exactly the same result - complete line-up - could be obtained, if the magnetic moments would feel a strong magnetic field.
- In the "mean field" approach or the "Weiss" approach to ferromagnetism, we simply assume such a magnetic field H_{Weiss} to be the cause for the line-up of the magnetic moments. This allows to treat ferromagnetism as a "special" case of paramagnetism, or more generally, "orientation polarization".



For the magnetization we obtain ⇒

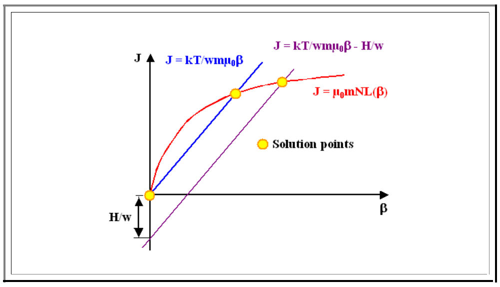
- The term $w \cdot J$ describes the Weiss field via $H_{\text{loc}} = H_{\text{ext}} + w \cdot J$; the Weiss factor w is the decisive (and unknown) parameter of this approach.
- Unfortunately the resulting equation for J, the quantity we are after, cannot be analytically solved, i.e. written down in a closed way.

$$J \;=\; N \cdot m \cdot \mu_0 \cdot \mathrm{L}(\beta) \;=\; N \cdot m \cdot \mu_0 \cdot \mathrm{L}\left(\frac{m \cdot \mu_0 \cdot (H + w \cdot J)}{\mathrm{k}T}\right)$$

Graphical solutions are easy, however ⇒

- From this, and with the usual approximation for the Langevin function for small arguments, we get all the major ferromagnetic properties, e.g.
    - Saturation field strength.
    - Curie temperature T_C.

$$T_C \;=\; \frac{N \cdot m^{\,2} \cdot \mu_0{}^2 \cdot w}{3\mathrm{k}}$$

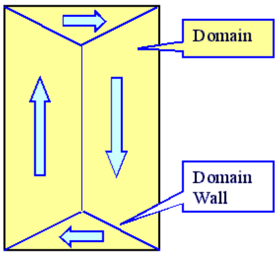


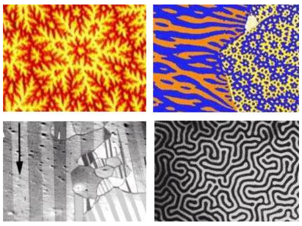
    - Paramagnetic behavior above the Curie temperature.
    - Strength of spin-spin interaction via determining w from T_C.
- As it turns out, the Weiss field would have to be far stronger than what is technically achievable - in other words, the spin-spin interaction can be exceedingly strong!

- In single crystals it must be expected that the alignments of the magnetic moments of the atom has some preferred crystallographic direction, the "easy" direction.

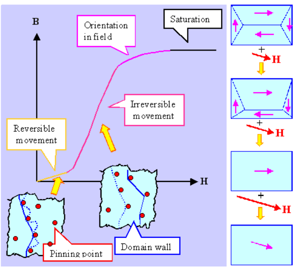
Easy directions:
Fe (bcc) <100>
Ni (fcc) <111>
Co (hcp) <001> (c-direction)

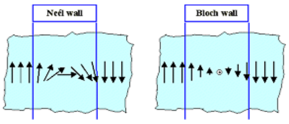
- A single crystal of a ferromagnetic material with all magnetic moments aligned in its easy direction would carry a high energy because:
    - It would have a large external magnetic field, carrying field energy.
- In order to reduce this field energy (and other energy terms not important here), magnetic domains are formed ⇒. But the energy gained has to be "payed for" by:
    - Energy of the domain walls = planar "defects" in the magnetization structure. It follows: Many small domains —> optimal field reduction —> large domain wall energy "price".
    - In polycrystals the easy direction changes from grain to grain, the domain structure has to account for this.
    - In all ferromagnetic materials the effect of magnetostriction (elastic deformation tied to direction of magnetization) induces elastic energy, which has to be minimized by producing a optimal domain structure.
- The domain structures observed thus follows simple principles but can be fantastically complicated in reality ⇒.



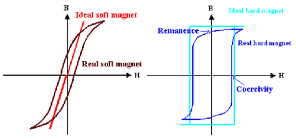


- For ferromagnetic materials in an external magnetic field, energy can be gained by increasing the total volume of domains with magnetization as parallel as possible to the external field - at the expense of unfavorably oriented domains.
    - Domain walls must move for this, but domain wall movement is hindered by defects because of the elastic interaction of magnetostriction with the strain field of defects.
    - Magnetization curves and hystereses curves result ⇒, the shape of which can be tailored by "defect engineering".
- Domain walls (mostly) come in two varieties:
    - Bloch walls, usually found in bulk materials.
    - Neél walls, usually found in thin films.





- Depending on the shape of the hystereses curve (and described by the values of the remanence M_R and the coercivity H_C, we distinguish hard and soft magnets ⇒.
- Tailoring the properties of the hystereses curve is important because magnetic losses and the frequency behavior is also tied to the hystereses and the mechanisms behind it.
    - Magnetic losses contain the (trivial) eddy current losses (proportional to the conductivity and the square of the frequency) and the (not-so-trivial) losses proportional to the area contained in the hystereses loop times the frequency.
    - The latter loss mechanism simply occurs because it needs work to move domain walls.
- It also needs time to move domain walls, the frequency response of ferromagnetic materials is therefore always rather bad - most materials will not respond anymore at frequencies far below **GHz**.



Questionaire

Multiple Choice questions to all of 4.3

4.4 Applications of Magnetic Materials

4.4.1 Everything Except Data Storage

General Overview

What are typical applications for magnetic materials? A somewhat stupid question - after all we already touched on several applications in the preceding subchapters.

- But there are most likely more applications than you (and I) are able to name. In addition, the material requirements within a specific field of application might be quite different, depending on details.
- So lets try a systematic approach and list all relevant applications together with some key requirements. We use the abbreviation **M_S**, **M_R**, and **H_C** for saturation, remanence, and coercivity, resp., and **low** ω, **medium** ω, and **high** ω with respect to the required frequency range.

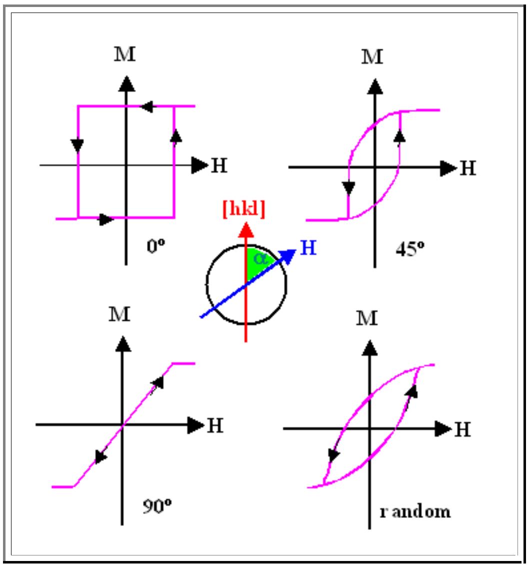
Field of application	Products	Requirements	Materials
Soft Magnets			
Power conversion electrical - mechanical	Motors Generators Electromagnets	Large **M_R** Small **H_C** Low losses **=** small conductivity low ω	**Fe** based materials, e.g. **Fe +** ≈ **(0,7 - 5)% Si** **Fe +** ≈ **(35 - 50)% Co**
Power adaption	(Power) Transformers		
Signal transfer	Transformer ("Überträger")	Linear **M - H** curve	
	LF ("low" frequency; up to ≈ **100 kHz**)	Small conductivity medium ω	**Fe +** ≈ **36 % Fe/Ni/Co** ≈ **20/40/40**
	HF ("high" frequency up to ≈ **100 kHz**)	Very small conductivity high ω	**Ni - Zn** ferrites
Magnetic field screening	"**Mu-metal**"	Large **dM/dH** for **H** ≈ **0** ideally **μ_r = 0**	**Ni/Fe/Cu/Cr** ≈ **77/16/5/2**
Hard Magnets			
Permanent magnets	Loudspeaker Small generators Small motors Sensors	Large **H_C** (and **M_R**)	**Fe/Co/Ni/Al/Cu** ≈**50/24/14/9/3** **$SmCo_5$** **Sm_2Co_{17}** **"NdFeB" (= $Nd_2Fe_{14}B$)**
Data storage analog	Video tape Audio tape	Medium **H_C**(and **M_R**), hystereses loop as rectangular as possible	**NiCo**, **CuNiFe**, **CrO_2** **Fe_2O_3**
Data storage digital	Ferrite core memory Drum		
	Hard disc, Floppy disc		
	Bubble memory	Special domain structure	Magnetic garnets **(AB_2O_4**, or **$A_3B_5O_{12}$**), e.g. with **A =** Yttrium (or mixtures of rare earth), and **B =** mixtures of **Sc**, **Ga**, **Al** Most common: **$Gd_3Ga_5O_{12}$**
Specialities			

Quantum devices	***GMR*** reading head ***MRAM***	Special spin structures in multilayered materials	

- As far as materials are concerned, we are only scratching the surface here. Some more materials are listed in the link
    - Data storage is covered in a separate module, here we just look at the other applications a bit more closely.

Soft Ferromagnets

- The general range of applications for soft magnets is clear from the table above. It is also clear that we want the hystereses loop as "flat" as possible, and as steeply inclined as possible. Moreover, quite generally we would like the material to have a high resistivity.
    - The requirements concerning the maximum frequency with which one can run through the hystereses loop are more specialized: Most power applications do not need high frequencies, but the microwave community would love to have more magnetic materials still "working" at **100 Ghz** or so.
- Besides trial and error, what are the guiding principles for designing soft magnetic materials? There are simple basic answers, but it is not so simple to turn these insights into products:
    - Essentially, *remanence* is directly related to the ease of movement of domain walls. If they can move easily in response to magnetic fields, remanence (and coercivity) will be low and the hystereses loop is flat.
- The essential quantities to control, partially mentioned before, therefore are:
    - The *density of domain walls*. The fewer domain walls you have to move around, the easier it is going to be.
    - The *density of defects* able to "pin" domain walls. These are not just the classical lattice defects encountered in neat single- or polycrystalline material, but also the cavities, inclusion of second phases, scratches, microcracks or whatever in *real* sintered or hot-pressed material mixtures.
    - The general *anisotropy* of the magnetic properties; including the anisotropy of the magnetization ("*easy*" and "hard" direction, of the magnetostriction, or even induced the shape of magnetic *particles* embedded in a non-magnetic matrix (we must expect, e.g. that elongated particles behave differently if their major axis is in the direction of the field or perpendicular to it). Large anisotropies generally tend to induce large obstacles to domain movement.
- A few general recipes are obvious:
    - Use well-annealed material with few grain boundaries and dislocations. For **Fe** this works, as already shown before.
    - Align the grains of e.g. polycrystalline **Fe**-based material into a favorable direction, i.e. use materials with a **texture**.
    - Doing this by a rather involved process engineered by **Goss** for **Fe** and **Fe-Si** alloys was a major break-through around **1934**. The specific power loss due to hystereses could be reduced to about **2.0 W/kg** for regular textured **Fe** and to **0.2 W/kg** for (very difficult to produce) textured **Fe** with **6% Si** (at **50 Hz** and **B** ≈ **1 T**)
    - Use isotropic materials, in particular **amorphous metals** also called **metallic glasses**, produced by extremely fast cooling from the melt. Stuff like $\mathbf{Fe_{78}B_{13}Si_{19}}$ is made (in very thin very long ribbons) and used.
    - Total losses of present day transformer core materials (including eddy current losses) are around **0,6 W/kg** at **50 Hz** which, on the one hand, translates into an efficiency of **99,25 %** for the transformer, and a financial loss of roughly **1 $/kg** and year - which is not to be neglected, considering that big transformer weigh many tons.



- Reduce the number of domains. One solution would be to make very small magnetic particles that can only contain one domain embedded in some matrix. This would work well if the easy direction of the particles would always be in field direction, i.e. if all particles have the same crystallographic orientation pointing in the desired direction as shown below.
- This picture, by the way, was calculated and is an example of what can be done with theory. It also shows that single domain magnets can have ideal soft or ideal hard behavior, depending on the angle between an easy direction and the magnetic field.
- Unfortunately, for randomly oriented particles, you only get a mix - neither here nor there.

Well, you get the drift. And while you start thinking about some materials of your own invention, do not forget: We have not dealt with eddy current losses yet, or with the resistivity of the material.

- The old solution was to put **Si** into **Fe**. It increases the resistivity substantially, without degrading the magnetic properties too much. However it tends to make the material brittle and very hard to process and texture.
- The old-fashioned way of stacking thin insulated sheets is still used a lot for big transformers, but has clear limits and is not very practical for smaller devices.
- Since eddy current losses increase with the square of the frequency, metallic magnetic materials are simply not possible at higher frequencies; i.e. as soon as you deal with signal transfer and processing in the **kHz**, **MHz** or even **GHz** region. We now need *ferrites*.

Questionaire
Multiple Choice questions to 4.4.1

4.4.2 Magnetic Data Storage

This topic was regularly handled in the Seminar and therfore not included here.

- *Since the Seminar has been abandoned, this page might be writtten in the near future - bear with me.*

4.4.3 Summary to: Technical Materials and Applications

- Uses of ferromagnetic materials may be sorted according to:
  - Soft magnets; e.g. **Fe** - alloys
  - Hard magnets; e.g. metal oxides or "strange" compounds.

- **Everything profiting from an "iron core": Transformers, Motors, Inductances, ...**
- **Shielding magnetic fields.**

- **Permanent magnets for loudspeakers, sensors, ...**
- **Data storage (Magnetic tape, Magnetic disc drives, ...**

- Even so we have essentially only **Fe**, **Ni** and **Co** (+ **Cr**, **O** and **Mn** in compounds) to work with, innumerable magnetic materials with optimized properties have been developed.
  - New complex materials (including "nano"materials) are needed and developed all the time.

Strongest permanent magnets:
Sm_2Co_{17}
$Nd_2Fe_{14}B$

- Data storage provides a large ***impetus*** to magnetic material development and to employing new effects like "**GMR**"; giant magneto resistance; a purely quantum mechanical effect.

4.5. Summary: Magnetic Materials

- The **relative permeability μ_r** of a material "somehow" describes the interaction of magnetic (i.e. more or less all) materials and magnetic fields ***H***, e.g. vial the equations ⇒
  - ***B*** is the **magnetic flux density** or **magnetic induction**, sort of replacing ***H*** in the Maxwell equations whenever materials are encountered.
  - ***L*** is the inductivity of a linear solenoid, or)coil or inductor) with length ***l***, cross-sectional area ***A***, and number of turns ***t***, that is "filled" with a magnetic material with μ_r.
  - ***n*** is *still* the index of refraction; a quantity that "somehow" describes how electromagnetic fields with extremely high frequency interact with matter.
    For all practical purposes, however, $\mu_r = 1$ for optical frequencies

$$B = \mu_o \cdot \mu_r \cdot H$$

$$L = \frac{\mu_0 \cdot \mu_r \cdot A \cdot w^2}{l}$$

$$n = (\epsilon_r \cdot \mu_r)^{½}$$

- Magnetic fields inside magnetic materials polarize the material, meaning that the vector sum of magnetic dipoles inside the material is no longer zero.
  - The decisive quantities are the *magnetic* dipole moment ***m***, a vector, and the *magnetic* Polarization ***J***, a vector, too.
  - Note: In contrast to dielectrics, we define an additional quantity, the **magnetization *M*** by simply including dividing ***J*** by μ_o.
  - The magnetic dipoles to be polarized are either already present in the material (e.g. in **Fe, Ni or Co**, or more generally, in all **paramagnetic** materials, or are induced by the magnetic fields (e.g. in **diamagnetic** materials).
  - The dimension of the magnetization ***M*** is **[A/m]**; i.e. the same as that of the magnetic field.

$$B = \mu_o \cdot H + J$$

$$\underline{J} = \frac{\Sigma \underline{m}}{V}$$

$$M = \frac{J}{\mu_o}$$

- The magnetic polarization ***J*** or the magnetization ***M*** are *not* given by some magnetic surface charge, because ⇒.

There is no such thing as a *magnetic monopole*, the (conceivable) counterpart of a negative or positive electric charge

- The equivalent of "Ohm's law", linking current density to field strength in conductors is the *magnetic* Polarization law:
  - The decisive material parameter is $\chi_{mag} = (\mu_r - 1) =$ **magnetic susceptibility**.
  - The "classical" induction ***B*** and the magnetization are linked as shown. In essence, ***M*** only considers what happens in the material, while ***B*** looks at the total effect: material plus the field that induces the polarization.

$$M = (\mu_r - 1) \cdot H$$

$$M := \chi_{mag} \cdot H$$

$$B = \mu_o \cdot (H + M)$$

- Magnetic polarization mechanisms are formally similar to dielectric polarization mechanisms, but the physics can be entirely different.

Atomic mechanisms of magnetization are not directly analogous to the dielectric case

Magnetic moments originate from:

- The intrinsic magnetic dipole moments ***m*** of elementary particles with spin is measured in units of the Bohr magneton$\mathbf{m_{Bohr}}$.
- The magentic moment m^e of the electron is ⇒
- Electrons "orbiting" in an atom can be described as a current running in a circle thus causing a magnetic dipole moment; too

$$m_{Bohr} = \frac{h \cdot e}{4\pi \cdot m^*_e} = 9.27 \cdot 10^{-24}\,\mathrm{Am}^2$$

$$m^e = \frac{2 \cdot h \cdot e \cdot s}{4\pi \cdot m^*_e} = 2 \cdot s \cdot m_{Bohr} = \pm m_{Bohr}$$

- The total magentic moment of an atom in a crystal (or just solid) is a (tricky to obtain) sum of all contributions from the electrons, and their orbits (including bonding orbitals etc.), it is either:
  - **Zero** - we then have a **diamagmetic material**.

Magnetic field induces dipoles, somewhat analogous to elctronic polarization in dielectrics.
Always very weak effect (except for superconductors)
Unimportant for technical purposes

  - In the order of a few Bohr magnetons - we have a essentially a **paramagnetic material**.

Magnetic field induces some order to dipoles; strictly analogous to "orientation polarizaiton" of dielectrics.
Alsways very weak effect
Unimportant for technical purposes

- In some **ferromagnetic** materials spontaneous ordering of magenetic moments occurs below the Curie (or Neél) temperature. The important familiess are
  - Ferromagnetic materials ⇑⇑⇑⇑⇑⇑⇑
    large $\mathbf{\mu_r}$, **extremely important**.
  - Ferrimagnetic materials ⇑⇓⇑⇓⇑⇓⇑
    still large $\mathbf{\mu_r}$, **very important**.
  - Antiferromagnetic materials ⇑⇓⇑⇓⇑⇓⇑
    $\mathbf{\mu_r} \approx \mathbf{1}$, unimportant

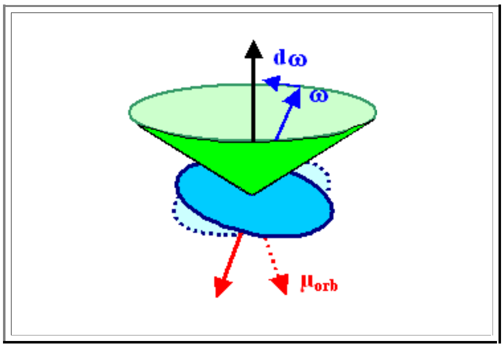
Ferromagnetic materials:
Fe, Ni, Co, their alloys
"AlNiCo", $\mathbf{Co_5Sm}$, $\mathbf{Co_{17}Sm_2}$,
"NdFeB"

- There is characteristic temperatuer dependence of $\mathbf{\mu_r}$ for all cases

- Dia- and Paramagentic propertis of materials are of no consequence whatsoever for products of electrical engineering (or anything else!)
  - Only their common denominator of being essentially "non-magnetic" is of interest (for a submarine, e.g., you want a non-magnetic steel)
  - For research tools, however, these forms of magnitc behavious can be highly interesting ("paramagentic resonance")

Normal diamagnetic materials: $\chi_{dia} \approx \mathbf{-(10^{-5} - 10^{-7})}$
Superconductors (= ideal diamagnets): $\chi_{SC} = \mathbf{-1}$
Paramagnetic materials: $\chi_{para} \approx \mathbf{+10^{-3}}$

- Diamagnetism can be understood in a semiclassical (Bohr) model of the atoms as the response of the current ascribed to "circling" electrons to a changing magnetic field via classical induction (∝ **dH/dt**).
  - The net effect is a precession of the circling electron, i.e. the normal vector of its orbit plane circles around on the green cone. ⇒
  - The "Lenz rule" ascertains that inductive effects oppose their source; diamagnetism thus weakens the magnetic field, $\chi_{dia} < 0$ must apply.



- Running through the equations gives a result that predicts a very small effect. ⇒ A proper quantum mechanical treatment does not change this very much.

$$\chi_{dia} = -\frac{e^2 \cdot z \cdot \langle r \rangle^2}{6\, m^*_e} \cdot \rho_{atom} \approx -(10^{-5} - 10^{-7})$$

- The formal treatment of paramagnetic materiuals is mathematically completely identical to the case of orientation polarization

$$W(\varphi) = -\ \mu_0 \cdot \underline{m} \cdot \underline{H} = -\ \mu_0 \cdot m \cdot H \cdot \cos\varphi$$

Energy of magetic dipole in magnetic field

$$N[W(\varphi)] = c \cdot \exp -(W/kT) = c \cdot \exp \frac{m \cdot \mu_0 \cdot H \cdot \cos\varphi}{kT} = N(\varphi)$$

(Boltzmann) Distribution of dipoles on energy states

$$M = N \cdot m \cdot L(\beta)$$

$$\beta = \frac{\mu_0 \cdot m \cdot H}{kT}$$

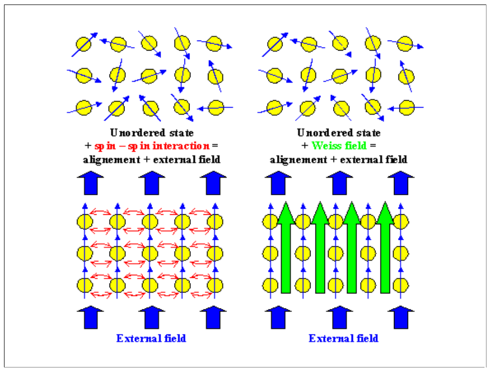
Resulitn Magnetization with Langevin function **L(β)** and argument **β**

  - The range of realistc β values (given by largest ***H*** technically possible) is even smaller than in the case of orientation polarization. This allows tp approximate **L(β)** by **β/3**; we obtain:

$$\chi_{para} = \frac{N \cdot m^2 \cdot \mu_0}{3kT}$$

  - Insertig numbers we find that χ_{para} is indeed a number just slightly larger than **0**.

- In ferromagnetic materials the magnetic moments of the atoms are "correlated" or lined-up, i.e. they are all pointing in the same direction
  - The physical reason for this is a quantum-mechanical spin-spin interaction that has no simple classical analogue.
  - However, exactly the same result - complete line-up - could be obtained, if the magnetic moments would feel a strong magnetic field.
  - In the "mean field" approach or the "Weiss" approach to ferromagnetism, we simply assume such a magnetic field H_{Weiss} to be the cause for the line-up of the magnetic moments. This allows to treat ferromagnetism as a "special" case of paramagnetism, or more generally, "orientation polarization".



For the magnetization we obtain ⇒

- The term $w \cdot J$ describes the Weiss field via $H_{loc} = H_{ext} + w \cdot J$; the Weiss factor w is the decisive (and unknown) parameter of this approach.
- Unfortunately the resulting equation for J, the quantity we are after, cannot be analytically solved, i.e. written down in a closed way.

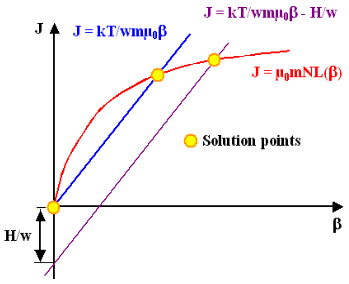
$$J = N \cdot m \cdot \mu_0 \cdot L(\beta) = N \cdot m \cdot \mu_0 \cdot L\left(\frac{m \cdot \mu_0 \cdot (H + w \cdot J)}{kT}\right)$$

Graphical solutions are easy, however ⇒

- From this, and with the usual approximation for the Langevin function for small arguments, we get all the major ferromagnetic properties, e.g.
  - Saturation field strength.
  - Curie temperature T_C.

$$T_C = \frac{N \cdot m^2 \cdot \mu_0^2 \cdot w}{3k}$$

  - Paramagnetic behavior above the Curie temperature.
  - Strength of spin-spin interaction via determining w from T_C.
- As it turns out, the Weiss field would have to be far stronger than what is technically achievable - in other words, the spin-spin interaction can be exceedingly strong!



In single crystals it must be expected that the alignments of the magnetic moments of the atom has some preferred crystallographic direction, the "easy" direction.

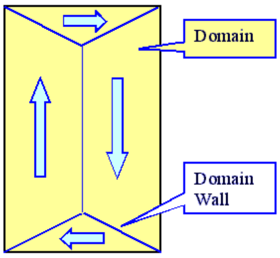
Easy directions:
Fe (bcc) <100>
Ni (fcc) <111>
Co (hcp) <001> (c-direction)

A single crystal of a ferromagnetic material with all magnetic moments aligned in its easy direction would carry a high energy because:

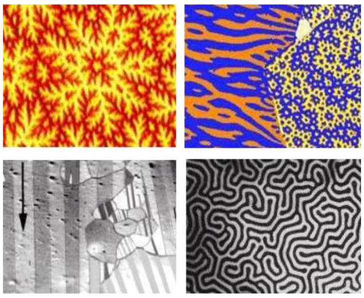
- It would have a large external magnetic field, carrying field energy.

In order to reduce this field energy (and other energy terms not important here), magnetic domains are formed ⇒. But the energy gained has to be "payed for" by:

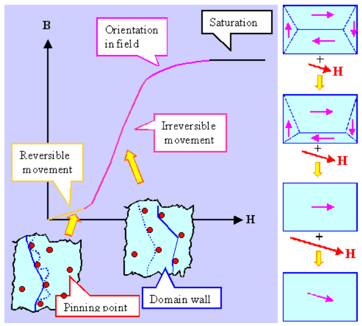
- Energy of the domain walls = planar "defects" in the magnetization structure. It follows: Many small domains —> optimal field reduction —> large domain wall energy "price".
- In polycrystals the easy direction changes from grain to grain, the domain structure has to account for this.
- In all ferromagnetic materials the effect of magnetostriction (elastic deformation tied to direction of magnetization) induces elastic energy, which has to be minimized by producing a optimal domain structure.



The domain structures observed thus follows simple principles but can be fantastically complicated in reality ⇒.

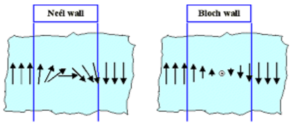


- For ferromagnetic materials in an external magnetic field, energy can be gained by increasing the total volume of domains with magnetization as parallel as possible to the external field - at the expense of unfavorably oriented domains.
  - Domain walls must move for this, but domain wall movement is hindered by defects because of the elastic interaction of magnetostriction with the strain field of defects.
  - Magnetization curves and hystereses curves result ⇛, the shape of which can be tailored by "defect engineering".

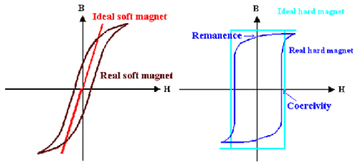


- Domain walls (mostly) come in two varieties:

  - Bloch walls, usually found in bulk materials.
  - Neél walls, usually found in thin films.



- Depending on the shape of the hystereses curve (and described by the values of the remanence M_R and the coercivity H_C, we distinguish hard and soft magnets ⇛.
- Tailoring the properties of the hystereses curve is important because magnetic losses and the frequency behavior is also tied to the hystereses and the mechanisms behind it.
  - Magnetic losses contain the (trivial) eddy current losses (proportional to the conductivity and the square of the frequency) and the (not-so-trivial) losses proportional to the area contained in the hystereses loop times the frequency.
  - The latter loss mechanism simply occurs because it needs work to move domain walls.
- It also needs time to move domain walls, the frequency response of ferromagnetic materials is therefore always rather bad - most materials will not respond anymore at frequencies far below **GHz**.



- Uses of ferromagnetic materials may be sorted according to:
  - Soft magnets; e.g. **Fe** - alloys

- **Everything profiting from an "iron core": Transformers, Motors, Inductances, ...**
- **Shielding magnetic fields.**

  - Hard magnets; e.g. metal oxides or "strange" compounds.

- **Permanent magnets for loudspeakers, sensors, ...**
- **Data storage (Magnetic tape, Magnetic disc drives, ...**

- Even so we have essentially only **Fe**, **Ni** and **Co** (+ **Cr**, **O** and **Mn** in compounds) to work with, innumerable magnetic materials with optimized properties have been developed.
  - New complex materials (including "nano"materials) are needed and developed all the time.

Strongest permanent magnets:
Sm_2Co_{17}
$Nd_2Fe_{14}B$

- Data storage provides a large ***impetus*** to magnetic material development and to employing new effects like "**GMR**"; giant magneto resistance; a purely quantum mechanical effect.

Questionaire

Multiple Choice questions to all of 4

5. Optics

5. Optics

5.1 Basic Optics

5.1.1 What is Light?

- What is **light**? You know the answer, of course. To refresh your memory, here is the definition:

> **Light** is the common name for **electromagnetic waves**
> with wavelengths just below a micrometer (**400 nm - 800 nm**).
>
> **Light** is the common name for **photons**
> with energies just above **1 eV** (**1,8 eV - 3,2 eV**).

  - So we still have the good old ***dichotomy*** between the **wave picture**, championed by **Huygens** and the **particle picture** first championed by **Newton**. As you know, Newton lost the fight but was ***redeemed*** to some extent by **Einstein** in **1905**. More to that in the link.

- It is of course **quantum theory** that reconciles the otherwise incompatible viewpoints. Either you have some ideas how this works or you don't. In the latter case you need to do some work on your own. I cannot go into this kind of "details" here.
  - In case of doubt think about what you learned about "electron waves". For example the $\psi = \psi_0 \cdot \exp(i\underline{kr})$ wavefunction for a free electron that turned a particle into a wave, and the $|\psi|^2$ that turns a wave back into a particle. It's just as easy for photons.
  - If you don't get it - tough luck! We're not doing quantum mechanics here. Just accept and follow the simple rule:

> **For the *propagation* of light:**
> **use the *wave model***
> **For the *generation* and disappearance (= *absorption*) of light:**
> **use the *photon model***

- Now we need to consider a few very basic numbers and relations.
  - The key properties and parameters that should come to mind when considering light propagating in vacuum or in some transparent material with dieelectric constant ϵ_r, magnetic permeability μ_r (always ≈ **1** for optical frequencies) and index of refraction ***n*** are:

Relations concerning light

	Propagation in vacuum	**Propagation in material with *index of refraction n***
Wavelength	λ_0	λ_0/n
Frequency	ν	ν
Energy	$h\nu = \hbar\omega$	$h\nu = \hbar\omega$
Propagation speed	c_0 = natural constant $= \nu \cdot \lambda$ $= (\epsilon_0 \cdot \mu_0)^{-1/2}$	$c(n) = \dfrac{c_0}{n} = \dfrac{1}{(\epsilon_0\epsilon_r \cdot \mu_0\mu_r)^{1/2}}$ $n = (\epsilon_r \cdot \mu_r)^{1/2} \approx \epsilon_r^{1/2}$
Wave vector	$\lvert\underline{k}\rvert = 2\pi/\lambda_0$ $= \omega/c_0$	$\lvert\underline{k}\rvert = 2\pi/\lambda$ $= 2\pi n/\lambda_0 = \omega/c$
Momentum	$\underline{p} = \hbar\underline{k} = \hbar\omega/c_0$	$\underline{p} = \hbar\underline{k} = \hbar\omega/c$
Snellius law	$n = \sin\alpha/\sin\beta$ with α, β the angle of incidence or propagation, resp.	

We can actually derive all the materials stuff and Snellius law as given in the last entry very easily by considering energy and momentum conservation. We will do that in a litle exercise.

Exercise 5.1.1

Derivation of Snellius' law

- The following table gives a few basic numbers for these quantities *that you must know* .

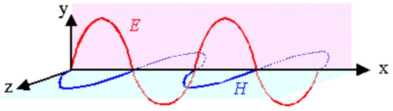
Numbers concerning light		
	Rough order of ten value	**Better value**
Wavelength	**≈ 1 µm**	**500 nm (390 to 750 nm)**
Frequency	**≈ 10^{15} Hz**	**$5 \cdot 10^{15}$ Hz**
Energy	**≈ 1 eV**	**2.5 eV**
c_0 (vacuum)		**300 000 km/s = $3 \cdot 10^8$ m/s**
Momentum ratio	$\dfrac{p(\text{2,5 eV electron})}{p(\text{2,5 eV photon})} \approx 10^3$	

- The momentum entry serves to remind you that photons have very little momentum relative to electrons (and phonons) of the same energy.

At a slightly higher level of sophistication we remember that light is an electromagnetic wave consisting of an interwoven electric and magnetic field; ***E*** and ***H***.

- In complex notation (with the understanding that we only use the real part; *in contrast to quantum theory*) a standard **plane wave** with amplitude ***E*** or ***H***, wavelength $\lambda = 2\pi/|\underline{k}|$ and circle frequency ω that propagates in ***k*** direction writes and looks as shown below.

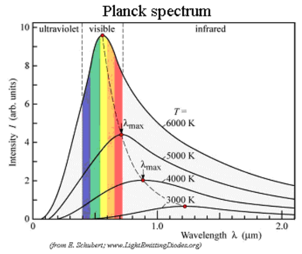
$$\underline{E}(\underline{r},t);\ \underline{H}(\underline{r},t) \quad = \underline{E}_0;\ \underline{H}_0 \cdot \exp\{i(\underline{kr} - \omega t)\}$$

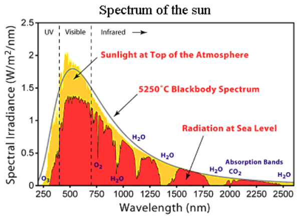


- The electric or magnetic field are *vector* quantities, always perpendicular to each other. Light thus always has a **polarization** vector associated with it, defined as the direction of the electric field vector (always perpendicular to the propagation direction).

Where does light come from? The sun, of course, is a major producer of light and so is any other hot body. **Max Planck**, as you know, first described the spectrum of light emitted by a hot "**black body**" in his famous work that was the beginning of quantum theory. The link gives a short and simple derivation

- What Planck calculated and what the sun actually does is shown in the following pictures.





- The sun comes pretty close to a black-body spectrum and the same is true for a light bulb or any other light source relying on high temperatures.
- It is clear to a Materials Scientist or Engineer that the sun is hot because nuclear fusion going on in its interior delivers the necessary energy, and that the radiation energy flooding the earth is the one and only energy on which life depends.

  - Right now (**2011**) we enter the age of massive solar energy harvesting via solar cells and wind or water power. The necessary materials science and engineering for doing this on a large scale will provide work and jobs for many years to come - but that will not concern us here.

- Besides hot bodies we also have "cold" light sources like **light emitting diodes** (***LED***s) and *Lasers*. We will come to that later in more detail.

- Most light sources and all hot bodies produce **incoherent light** (travelling in all directions with random phases) and **multi-chromatic** light (having all kinds of frequencies), which is a far cry from the $E = E_0 \cdot \exp(i\underline{kr})$ fully coherent and mono-chromatic plane wave that we like to use as mathematical representation. Sun light or artificial light sources used for illumination thus generate extremely "messy" light from a purists viewpoint. The messy light is nevertheless quite important in a general sense (imagine it missing!) but not of much *technical* interest - besides generating it.
  Laser light, by contrast, is a good approximation to the plane wave model but not of much use for illuminating your kitchen.

  - What we are interested in here is *working* with light. That means we have to consider manipulating it by running it through or off *materials*. What comes to mind in this context are optical products and components. The table below gives an incomplete list of a few catchwords that you should know in this context

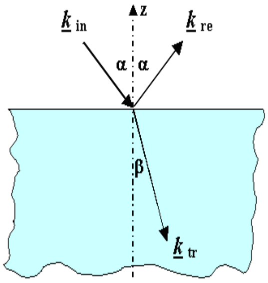
Products and components around light technology		
Components	**Products**	**Field**
Lenses (and apertures)	Microscope, Glasses, Camera, Film projector	Geometric optics
Mirrors	Reflector telescope; steppers, optical MEMS	
Prisms	Binocular; Reflectors	
Filters	Color photography etc.	
Diffraction gratings	Spectrometer	⇓
Anti reflection coatings	Solar cells, glasses, lenses	
Linear polarizers	Cameras, sun glasses, optical measurements,	
Circular polarizers	3-D cinema, advaned measurements	"Tensor" Optics Interference Optics
Interference filters; Interferometers	Optical precision measurements	
Phase shifters, High resolution optics	Lithography for Microelectronics	
"Digital" optics	Beamers, Displays, Cameras, optical MEMS	
Faraday, Kerr, Pockel, ... effects	LCD display, ultrafast modulation, advanced analytics	
Optical fibers	Optical communication, sensors	
LED, OLED, Lasers	High efficiency light sources, Displays, Processing, ..	⇓
Non-linear materials	Frequency doubling	
Photonic crystals	"Optical" semiconductors	Quantum Optics
Quantum dots	Optical computing	

- The message is loud and clear. We have to move from simple **geometric optics** to **"tensor" optics** and **interference optics**, arriving finally at **quantum optics**. We keep in mind, however, that there is only one *kind of optics* - those catchwords do not describe different optic realities but just different approaches to one and the same thing.

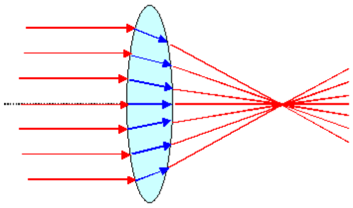
5.1.2 Basic Geometric Optics

The Very Basics of Geometric Optics

- The essence of basic high-school **geometric optics** is shown in the following pictures:



Reflection and Refraction



Focussing by an ideal convex lens

  - Optically transparent materials ("glass") have an **index of refraction *n* > 1**, and light hitting a transparent material is *reflected* and *refracted*. **Convex** (or collecting or converging) lenses and **concave** (or dispersing or diverging) lenses allow to manipulate the light path, e.g. by focusing a parallel beam of light as shown.
  - Let's clarify the terms at this point:

Reflection is, well, reflection; always with "angle in" = "angle out".
Refraction is the sudden "***bending***" or "flexing" of light beams at the interface between two different materials. The term belongs to *geometric optics*. A light beam going through a lens is refracted.
Diffraction is the continous "***bending***" of light beams around corners and all the other effects bringing about directional changes and interference effects. A light beam going through an optical grid is diffracted. The term belongs to *wave optics*.

- The decisive material quantity in geometric optics (and beyond) is the *index of refraction* together with **Snellius law**.
  - What we know about Snellius law and some other basic optics parameters like the speed of propagation ***c*** inside materials, frequency ν and wavelength λ in materials or in vacuum, is

$\frac{\sin\alpha}{\sin\beta} = n$	**Snellius law**
$n = \frac{c_0}{c} = \epsilon_r^{1/2}$	**From Maxwell equations**
$c = \nu \cdot \lambda$	**Always valid**
$\lambda_{mat} = \frac{1}{n} \cdot \lambda_{vac}$	**From the equations above**

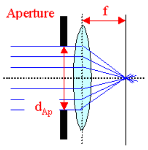
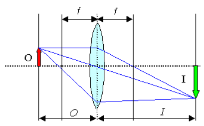
- Knowing *only* the index of refraction ***n*** makes it already possible to construct **light paths** or light rays running through optical devices like lenses or prisms.
  - Going a bit beyond that, we would also like to know the the **optical dispersion**, i.e. ***n* = *n*(ν)** so we can construct light paths for the various frequencies (= colors) of visible light. Obviously, all we need to know for this is the frequency dependence of the dielectric constant **ε(ν)**, something we have treated extensively before.

- A nice thing in geometric optics is that the direction of the light paths is always **reversible**. Change the arrow directions in the picture above (or in all other pictures like that) and they are still correct.
  - A not-so-nice thing might be that the definition of $\mathbf{n = \epsilon_r^{1/2}}$ becomes troublesome if $\mathbf{\epsilon < 0}$, which, as we know, is perfectly possible. Could there be an imaginary or even **negative index of refraction**? The answer is yes **-** as you will see later.
- If we go one step beyond simple geometric optics with *ideal* lenses, we realize that some modifications need to be made:
  1. *Real* lenses have all kinds of problems called **lens errors** or **lens aberrations**.
  2. Some light is always **reflected** at interfaces between media with different indices of refraction.
  3. The intensity of the light is always **attenuated** or damped whenever it passes through a material.
  4. **Focal "points"** have finite dimensions in in the order of the wave length of the light.
  5. All of the above may depend to some extent on the **polarization** of the light.
  - In going from point **1** to point **5** we move, of course, from geometric optics to *wave optics*.
- Let's look very briefly on the first point. The major lens aberrations are:
  - **Spherical aberration**. Following Snellius' law, and tracing the light rays for spherical lenses, it becomes clear that light rays running not close to the center of the lens are focussed to a point different from those close to the lens. The effect is small if some aperture keeps the light rays close to the optic axis. The lens then has a small **numerical aperture *NA***.
    The **NA** for a single lens is roughly the quotient of (possibly aperture defined) diameter / focal length; i.e. a crude measure of the size of the lens; see the picture below. Of course, lenses with small NA will not suffer much from spherical aberration but will also not transmit much light and thus produce "dark" pictures. The solution might be aspherical lenses but usually combinations of spherical lenses are used.
  - **Chromatic aberration**. Different wavelengths or "colors" are not focussed on the same focal point because we always have some *dispersion* and the index of refraction is a function of the wave length; $\mathbf{n = n(\lambda)}$. The "solution" is the achromatic lens, always a combination of two lenses made from different glasses with $\mathbf{n = n(\lambda)}$ or dispersion curves that compensate the effects of chromatic aberration to a sufficient extent.
  - **Astigmatism** occurs if the radius of curvature defining the surface of a lens is not exactly the same everywhere (probably true for the lens in your eye). Instead of a focal point you get a smeared out longish spot. A similar effect applies even to perfectly hemispherical lenses if the light rays coming in are inclined relative to the optical axis.
  - And so on. It is almost a miracle that we can see so well using a rather imperfect lens, and that sophisticated optical apparatus like your binocular or camera objective is not only extremely good but also dirt cheap.
- It is of considerable interest for Materials Science that the **electromagnetic lenses** used in electron microscopes have pretty much the same "aberration" problems as optical lenses, causing all kinds of trouble. Unlike optical lenses, however, there are usually no "easy" fixes except using small apertures, i.e. small **NA** values.

A few Examples

- Note: If what follows doesn't bore you to tears, you have a problem!

- Imaging through a convex lens. The lens has a focal length ***f***; always a positive number. The Object **O** is at a distance of ***O*** cm, the image **I** will occur at a distance ***I*** cm.

$NA \approx \dfrac{d_{Ap}}{f}$			Focal length ***f*** $f = \dfrac{r}{2}$ ***r* = radius of curvature** Imaging equation $\dfrac{1}{O} + \dfrac{1}{P} = \dfrac{1}{f}$

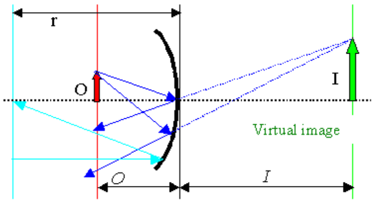
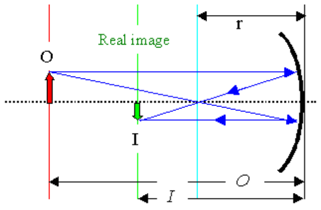
  - That's what the geometric construction looks like. We call the picture "real" if it is on the other side of the lens as seen from the object.
  - The **focal length** is the decisive number for the lens, next would be its *numerical aperture* (roughly given by its lateral size). For convex lenses the focal length is a positive number, for concave lenses it is a *negative* number!
  - Instead of the focal length ***f*** one often gives values of **dioptre *D*** (or ***diopter***), which is simply $\mathbf{D = 1/f}$. My reading glasses with **3** diopters thus have a focal length of **33 cm**.

- From a materials point of view the **dispersion properties** of the (transparent) dielectric, i.e. $\boldsymbol{n(\omega) = [\epsilon(\omega)]^{1/2}}$ in the optical wavelength region is of supreme importance.
  - As we know, at optical frequencies dispersion is always determined by resonance phenomena linked to atomic polarization. We also know, that $\boldsymbol{\epsilon, n \rightarrow 1}$ for high frequencies, i.e. for ultraviolet (**UV**) and beyond. In fact, we don't have a lot of good optical materials with high index of refraction in the visible region. Here are a few numbers.

Material	Air	Water liquid	Water Ice	Benzene	Eye lens (human)	PMMA (PC, ..)	Salt (NaCl)	Crown glass	Flint glass	Diamond	TiO_2	GaP	Silicon (Si)	GaAs
n	1,00027	1,333	1,31	1.501	1,386 - 1,406	≈ 1,57	1,50	1,5 - 1,54	1,6 - 1,62	2,419	2,496	3,5	3,96	3,93
	Transparent in visible light											Only IR; not transparent in visible light		

- Typically diamond is described as the "material" with the highest index of refraction. The semiconductors to the right of it actually beat diamond fair and square - but of course only at frequencies where $\boldsymbol{h\nu < E_g}$ obtains ($\boldsymbol{E_g}$ is the band gap). At higher frequencies semiconductors are perfectly opaque. Silicon thus is only a good optical material in the infrared (**IR**) region of the spectrum.
- So what is **crown glass** as opposed to **flint glass**? Or any of the umpteenth other varieties of glass? Look it up. As a guide: crown glass is your basic run-of-the-mill "lime" glass (i.e. $\boldsymbol{SiO_2}$ with added **Na**, **K** and so on); flint glass (also known as lead glass) is what "**crystal**" (English for fancy goblets, tumblers, etc. for wine etc.) is made from; i.e. $\boldsymbol{SiO_2}$ with added lead oxide **PbO**.

- Now let's image with a **concave mirror** of spherical shape. We can produce **real** or **virtual** images as shown.

	Focal length ***f*** ***f* = *r*** — ***r* = radius of curvature**
	Imaging equation $$\frac{1}{O} + \frac{1}{P} = \frac{1}{r}$$

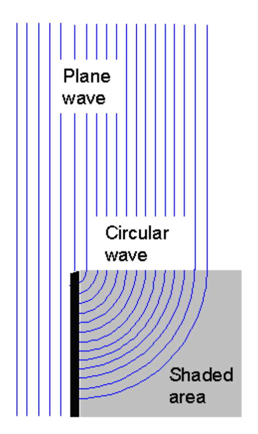
  - That's what geometric constructions look like. If the image is on the other side of the mirror as seen from the object, we call it "*virtual*" image.
  - Note that spherical mirrors have severe problems with spherical aberration. That's why you tend to use a **parabolic mirror** where all light rays coming in parallel to the optical axis are deflected to the same focal point. However, remember the first law of economics? There is no such thing as a free lunch! If the light comes in somewhat inclined to the optical axis of a parabolic mirror, its imaging properties are actually worse than that of a spherical mirror. That's why you always have a long tube on front of the mirror.

5.1.3 Basic Wave Optics

Some Basics

Wave optics starts with **Huygens** (1629 - 1695) and **Young** (1773 - 1829); see the link for some details. Wave optics proceeds in essentially two steps. *First,* the **Huygens principle** is applied, *second* interference between the resulting waves is added.

- *Step 1:* A wave hitting some edge or just about anything will produce a **circular wave** as shown below. The effect is easily visible when looking at water waves on some relatively undisturbed water surface with obstacles.
- (Plane) waves hitting an obstacle as shown below thus are detectable even in "shaded" places. Note that the circular wave would go all the way around the edge; this is not shown for simplicity. The consequences are clear:
- Even sharp edges appear always blurred if one looks closely enough. i.e. on a **µm** scale. This happens for any plane wave, be it light, radio waves (allowing you to receive the rhythmic noise of your choice even in the "shade" of, e.g., buildings), or electron waves (in electron microscopes). Just the lengths scales are different, going with the basic wave lengths of the wave considered.



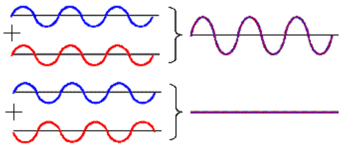
The blue lines show the maxima (or minima; whatever you prefer) of the amplitude of a traveling wave at some point in time t_0. The distance between the lines is thus the wavelength λ
Somewhat later, the whole system of lines would have moved somewhat to the right (the propagation direction here).
The distance of one wavelength is covered after a time $t_\lambda = \lambda / v$; **v** is the propagation speed. The time t_λ is the inverse frequency ν
Thus we have $v = \nu \cdot \lambda = c$ for light.

- The consequences are clear: The resolution d_{min} of a lens with the **numerical aperture** ***NA***, i.e. its capability to image two points at a distance d_{min} separately and not as some blur, is wave-length limited and given by

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

- It's easy to see in a "hand-waving" manner why the numerical aperture comes in. Imagine some lens and reduce its numerical aperture by putting a real ***opaque*** aperture with a hole in front of it. The aperture edges will induce a blur that get's worse the smaller the hole and therefore **NA**. Your resolution goes down with decreasing **NA**.
- On the other hand, your lens aberrations become worse with increasing **NA**. The resulting conflict for optimized optical apparatus is clear and encompasses a lot of intricate and very advanced topics in optics, e.g. how to make structures on microelectronic chips with lateral dimensions around **30 nm <** λ with optical lithography.

Now we consider *step 2*: two waves can *interfere* with one another. The principle as shown below is clear.

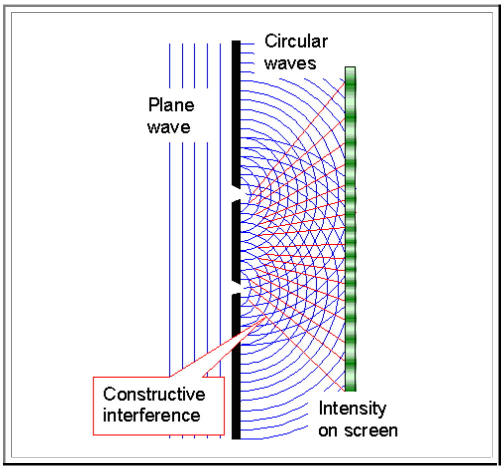


- For a **phase difference = 0** we have **constructive interference**; the amplitudes are doubled. For a phase difference **=** $180^0 = \pi$ **destructive interference** "cancels" the wave; the amplitude is zero.
- The apparent **paradox** of how you can get nothing from something (where are the two single waves and the energy they carry now?) is not trivial to solve; for details see the link.

The paradigmatic experiment for showing interference effects is, of course, the **double slit experiment**. If you consider that for electron waves, and in particular just for *one* electron (or photon), you are smack in the middle of quantum theory.

- In the **wave picture** the two spherical (or here cylindrical) waves emanating from the two slits interfere to give the pattern shown below. There is no problem at all.

- In the **photon picture**, a photon (or electron) passing through the two slits *interferes with itself*. This boggles the mind quite a bit but the result is the same: You get the interference pattern as shown, with pronounced minima and maxima of the intensity, which now correspond to the *probability* of detecting the particle.



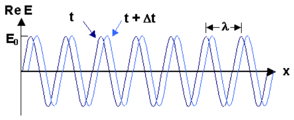
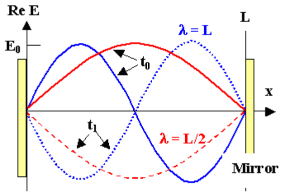
- Whenever we look at non-trivial optics, we need to consider interference effects. In the real world (as opposed to the ideal world shown in the pictures above), we need to consider the fact that our waves are almost never **mono-chromatic** (all have the same wavelength) and **coherent** (all have the same phase) plane waves extending into infinity in every direction.
  - The exception is, of course, the typical **Laser** beam, where we have an (almost) mono-chromatic and (almost) coherent beam. However, a "Laser beam" is typically "thin" and doesn't extend in all directions. So it is *not* a *simple* plane wave!
- A first important conclusion can be arrived at.
  - If we look at an ensemble of waves with the same wavelength, or better: with the same **wave vector** $\boldsymbol{k} = 2\pi/\lambda$ since it contains in addition to the wavelength also the **direction of propagation** (that's why it's a *vector*), we note that:

An ensemble of sufficiently many waves with the same $\pm\boldsymbol{k}$
and random phases interferes to exactly zero (plus some noise)

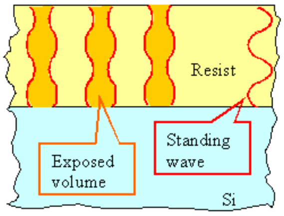
  - "*Random phases*" means that all phases are equally probable. The proof of the theorem is easy: If one of the many waves has a phase α, there will be some other wave with the phase **–**α - the two will cancel. A visual proof constructed in a different context (that should be familiar) but fits just as well here is shown in the link.
- From this you realize immediately that inside some hollow tube of length ***L*** that reflects waves at either end, only waves with $\lambda = 2L/m;\ m = 1, 2, 3, \ldots$ can exist.
  - Waves not meeting this criterion will, upon reflection at the end of the tube, produce a phase-shifted wave with some new phase, twice this phase upon the second reflection and so on. Pretty soon you have waves with random phases in there and—see above.
  - What that also means is: We now have also *all* musical instruments covered, in fact everything where the term "**resonator**" comes up. This includes also the free electron gas and to some extent electrons in a periodic potential as well as the Bragg condition for diffraction of waves at crystals. The list goes on. Quantum theory deals with wave functions ψ after all, and the big difference to classical physics comes from the simple fact that you let your wave functions interfere before you take the square, producing in terms of probabilities the classical equivalent of a particle.

Standing Waves

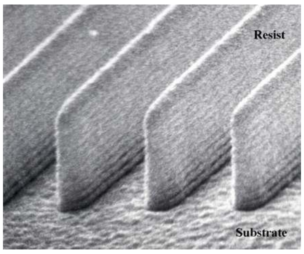
- If we look at simple waves propagating in just *one* direction, i.e. at a one-dimensional problem like sound waves inside an organ pipe, we quickly get the concept of a **standing wave**, the superposition of two plane waves with everything equal except the sign of the wave vector $\boldsymbol{k}$.
  - First let's look at the pictures and relations below; they are only meant to refresh your memory

"Running" Plane Wave	Standing Wave
	
$\underline{E}(\underline{r},t) = \underline{E}_0 \cdot \exp\{i(\underline{k}\underline{r} - \omega t)\}$ Re E $= \underline{E}_0 \cdot \cos\{2\pi/\lambda - \omega t\}$	$\underline{E}(\underline{r},t) = \underline{E}_0 \cdot \exp\{i(\underline{k}\underline{r} - \omega t)\} \pm \underline{E}_0 \cdot \exp-\{i(\underline{k}\underline{r} - \omega t)\}$ Re E $= 2\underline{E}_0 \cdot \cos(2\pi/\lambda) \cdot \cos(\omega t)$ $\lambda = 2L/m$; m = 1, 2, 3,..; (a "quantum" number) L = resonator length

- The pictures and equations are true for acoustic waves, light waves or electron "waves" - just for any wave. You should know standing waves from acoustics - it's the base of any musical instrument, after all, and you *hear* them all the time.
  - How about standing light waves? You ever *seen* some?
  - No? So put two mirrors at some distance ***L*** in the **cm** range and admit some light. Are you now going to *see* a standing light wave between the mirrors, as you should expect from all of the above? No you don't - for several reasons:
    1. The **coherence length** l_{coh} of normal light is too short. Normal light is not a infinitely extended plane wave but has some finite "length" that is far below **cm**. You're essentially missing the extended waves between the mirrors that are superimposed and thus you can't have a standing wave. Organ pipes that are **500 m** long don't work either.
    2. Fine, so let's use *coherent light*, however made. OK - you will get standing waves now but you won't notice. The minimal difference in wavelength between two allowed standing waves is $\Delta\lambda = [2L/m - 2L/(m+1)$; for large m this simplifies to $\Delta\lambda \approx 2L/m^2$. Since λ is in the **µm** region, and ***L*** in the **cm** region, m is around **10.000** and $\Delta\lambda \approx 10^{-4}\,\lambda$. In other words, the allowed wavelengths of the standing waves are so close to each other that pretty much all light wavelengths can live inside your resonator. You won't notice a difference to an arbitrary spectrum. Our lecture room here, even so it is a resonator for acoustic waves in principle, doesn't appear to produce nice musical tones because it is simply too large for acoustic wave lengths.
    3. OK, so let's make a resonator - two mirrors once more - but only *separated by a distance of a few* ***µm***. Now you did it. You could have distinct standing light waves in there. But what do you expect to *see* now? Think!
  - Why do you *hear* the standing acoustic waves in an organ pipe or in any (classical) musical instrument? Because some of the wave leaks out, eventually hitting your ear. The tone (= pressure amplitude inside the pipe) then would soon be gone if one wouldn't keep feeding acoustic waves into the resonator (by blowing into the organ pipe, for example). Same here. Some light must leak out so you can see it. If the leaking (and the feeding light into the resonator) is done in a certain way, we call the resulting instrument a **Laser**. We'll get back to this.
- The long and short is: Yes, interference effects and standing waves are of supreme importance for modern optics and you should refresh you memory about the basics of that if necessary.
  - The following picture shows standing light waves very graphically. We are looking at the unexposed section of a photo resist from microelectronics. The part that was exposed to light has been etched off. The ripples on the left-over resist (= light sensitive polymer) correspond to the extrema of the amplitude of a standing light wave.



The principle of what is happening.



- The surface of the resist and the surface of the substrate were rather flat and only a few wavelengths apart. When light (monochromatic and rather coherent) was fed to the system, a standing wave developed inside the resist. While *you* wouldn't have seen anything special, the light intensity "seen" by the resist varied periodically with depth, and that's why the light-induced chemical changes that allow to etch out the exposed part, leave "ripples" on the side walls.
- If this is all gobbledegook to you, you need to look up "lithography" within the context of semiconductor technologies.

Questionaire
Multiple Choice questions to 5.1.1 - 5.1.3

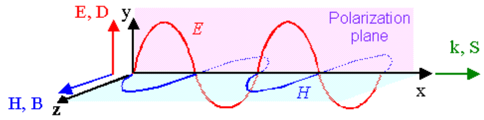
5.1.4 Energy Flow, Poynting Vector and Polarization

Energy Flow and Poynting Vector

Light as a wave, by definition, is *polarized* since an arbitrary light wave can always be decomposed into plane waves described by the electrical field vector $\underline{E}(\underline{r},t)=\underline{E}_0\exp\{i(\underline{kr}-\omega t)\}$ and the corresponding expression for the magnetic field.

- The electrical field *vector* (or the electrical displacement vector $\underline{D}$ in material) thus is contained in some plane (the **polarization plane**) that contains the direction of propagation (direction of $\underline{k}$-vector). Looking for example at a picture we had before but now augmented with the relevant *vector quantities*, we see that in this case:

  - Polarization plane=**(y,x)** plane.
  - Propagation direction=$\underline{k}$ direction=*Poynting vector* $\underline{S}$ direction=**x** direction here.
  - E-field amplitude in **y**-direction given by E_0.
  - H-field amplitude in **z**-direction given by H_0.



$$\underline{E}_y(t) = \underline{E}_{y0}\cdot\exp\{i(\underline{kx}-\omega t)\}$$

$$\underline{H}_z(t) = \underline{H}_{z0}\cdot\exp\{i(\underline{kx}-\omega t)\}$$

This all looks well-known and clear. Usually we don't bother to go much deeper at this point. In a Master study course, however, you now ponder the picture and the equations above for a while. Then a few questions should come up, for example:

- Can polarization be described by a *vector*? Could I just use a unit vector in E_0-direction?
- What happens if E_0; H_0 are *not* parallel to D_0; B_0, something to be expected in anisotropic materials?
- How do I describe polarization in the *particle picture*?
- Is the electrical and magnetic field really *in phase* as shown above?
- How *large*, in **V/cm**, is the electrical field E_0 in a typical light wave. How large (roughly) is it for *one* photon of a given energy?
- Let's assume you know the electrical field E_0: how large is the *magnetic field* H_0 associated with it?
- The energy a photon brings along is $E_{Ph}=h\nu$. How is this *light energy* transported by an electromagnetic wave?

It's not very difficult to conceive and to understand these questions. How about the answers?

- *The bad news is*: There are no easy answers for some of those questions.

- *The good news is*: For most of what follows you don't have to worry about the answers. I will go into details later whenever necessary. For now you can use this link to get some ideas about the answers and, even better, you could do the simple but illuminating exercise *below*.

Let's start with the last question. Light flow equals energy flow. All of us who once experienced a sun burn (or shot down one of those evil alien space ships with a Laser gun) know that. How does light energy "flow" exactly?

- Well, we (should) know how much energy *density* **W =energy /cm^3** is contained in an electric or magnetic field in vacuum (in materials use D and B instead of E and H): We have:

$$W_{elect} = \frac{\epsilon_0\cdot E^2}{2} \qquad W_{mag} = \frac{\mu_0\cdot H^2}{2}$$

$$[W_{elect;\ magn}] = [\mathrm{Ws\ m^{-3}}]$$

We do not yet know, how E and H are linked. All we need to know is that this follows "straight" from the Maxwell equations and is given by

$$E_0 = \left(\frac{\mu_r\mu_0}{\epsilon_r\epsilon_0}\right)^{1/2}\cdot H_0 = Z_w\cdot H_0$$

- We defined a new quantity $\boldsymbol{Z_W = E_0/H_0 = (\mu_0\mu_r/\epsilon_0\epsilon_r)^{1/2}}$ that is called the **wave impedance** of the medium. This is an apt name since Z_W does have the dimension "*Ohm*" (Ω) as required for an impedance.
- For the **vacuum impedance** ($\boldsymbol{\epsilon_r, \mu_r = 1}$) we get

$$\boxed{Z_W(\text{Vacuum}) = \mathbf{376{,}7}\ \Omega}$$

- This is a bit strange if you think about it. We just connected vacuum, the *nothing*, with a defined property that can be expressed in a simple number! You can buy a resistor with **377** Ω and wonder why it somehow reflects a property of vacuum.

Be that as it may, we now define a *new* vector ***S*** as follows:

$$\underline{S} = \underline{E} \times \underline{H}$$

$$= \text{Poynting vector}$$

- The *Poynting vector* has interesting properties:
    - $\underline{S}$ points in the direction of the propagation of the light, i.e. in wavevector $\underline{k}$ direction as shown in the figure above.
    - $|\underline{S}| = S = E_0 \cdot H_0 \cdot \cos^2(\underline{k}\underline{r} - \omega t)$, the magnitude of the Poynting vector, gives directly the **power flux** in **W/cm^2** of an electromagnetic wave
    - The average power delivered to an area is thus .

$$\langle S \rangle = \frac{E_0H_0}{2} = \frac{E_0^2}{Z_W} = E_0^2 \cdot \left(\frac{\epsilon_r\epsilon_0}{\mu_r\mu_0}\right)^{1/2}$$

$$\langle S \rangle \propto E_0^2 \cdot \epsilon_r^{1/2} \propto E_0^2 \cdot n$$

- Don't mix up the **power flux** in **W/cm^2** with the **specific energy** measured in in **Ws/cm^3**or **J/cm^3** that is contained in fields as shown above! It's time for an exercise to get some ideas about the *numbers* associated with light power and field strength!

Exercise 5.1.2

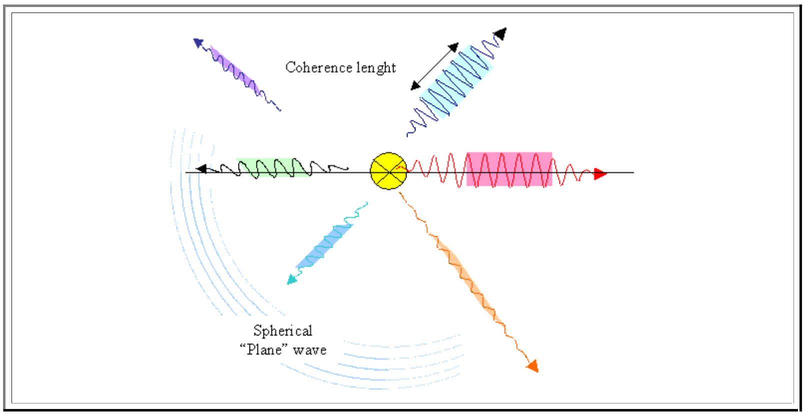
Energy, field strength and photons

We still have a bunch of open questions from above but before we tackle those we have to learn a bit about polarization.

Polarization of Light

It is clear that an idealized *plane* wave, given essentially by the **exp[i($\underline{k}\underline{r}$ – ωt)]** term, describes a *monochromatic* and fully *coherent* light wave propagating in exactly *one direction* that is also **linearly polarized** by definition.

- Some Laser beams may come close to emitting a plane wave matching that description but laser light is a far cry from "normal" light.
- Most real or normal light beams, e.g. the light "beams" emanating from light bulbs, are *not polarized* as we know. "Real" light from light bulb looks - very schematically - as shown below. We are forced to assume that the electromagnetic wave has a beginning in space and time (at the light bulb, like now) and an end. The length of the whole thing we call "**coherence length**". The color (=wavelength), and the polarization plane for one of those waves with a beginning and an end, has some fixed value but that is different from the next wave coming along. The distribution of wavelengths we call the *spectrum* of the light source, and the distribution of polarization planes will be random. Each ones of theses "wavelets" is caused by an electron changing its state from higher to lower energy.



- The picture might be interpreted as *photons* coming out of the light bulb. If one of those wavelets would be seen as a completely *abstract symbol* for a photon, this would be true. However. photons are *not* just short pieces of an electromagnetic wave, so one must not interpret this figure as a *picture* of photons.

- We have a number of *new* questions to consider (never mind that we did not answer all the old ones yet):

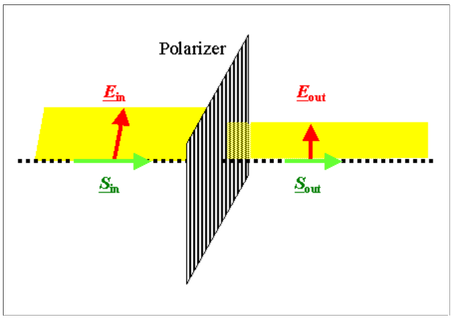
1. How do we *detect* if and in which direction a light beam is polarized?
2. How can we *polarize* an unpolarized light beam if the need arises?
3. Does is matter if and how a light beam is polarized for the *applications* we have in mind?
4. If it matters: how do we *optimize* the needed polarization?
5. How does the polarization of a wave relate to *photon properties*?
6. Are there other *polarization modes* besides *linear* polarization?

We are now at the point where serious optics starts.
It ain't exactly easy but we will only scratch the surface here.

- Let's look at these questions one by one. We will encounter rather difficult topics that we only treat in a ***perfunctory*** way here. Some topics will come up later again.

- **1. How do we *detect* if and how a light beam is polarized?**

  - By using a **polarizer**, i.e. some material that transmits *only* that ***E***-field component of an incoming light beam that is parallel to some material specific **polarization direction** as shown below.



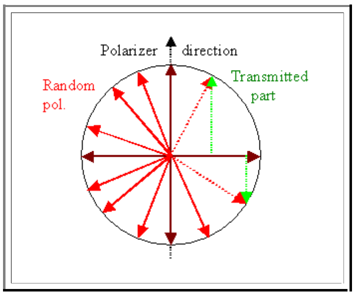
From Rogilbert; French Wikipedia

  - Turn your polarizer and the light intensity coming out stays either constant if the incoming beam is not polarized at all (but is always lower in intensity) or varies between zero and maximum (=incoming) intensity in a **cos^2** fashion if the incoming beam is **100 %** linearly polarized, as shown above. If the beam is *partially* polarized you have a mix of the two extremes.

- The second question (**2. How can we *polarize* an unpolarized light beam if the need arises?**) is answered too: Use a *polarizer*! Of course we have a new question now, the really tough one.

  - **7. What, *exactly*, gives a material polarizing properties?**
    And how can I manipulate or engineer those properties? That's indeed a tough question as we shall see in one of the next sub chapters. Before we try to answer this one, we look at a few simple things first.

What we perceive now is that a standard non-polarized light beam simply consists of many intrinsically polarized beams but with randomly distributed polarization directions. The total polarization effect then is zero.



- This is easy to perceive with the help of the figure on the left.
- Randomly distributed *polarization vectors* sum up to zero=no polarization.
- Note that we now indirectly answered one of the earlier questions from above: Yes, you can use a **polarization vector** like we do in this case and most others. But beware! Polarization in general is more tricky than you might think and a simple vector is not always enough to describe it mathematically. (Remember the problems encountered in using a polar vector for describing rotations?). You will need a matrix if you go about it systematically.
- The picture also makes clear why the polarized beam emerging from a polarizer has a lower intensity than the incoming beam, and also how much lower it will be.
- If we want to make life easier we can replace a large number of beams with random polarization directions (red vectors) by just two beams at right angles (brown vectors).
- It is time for a quick exercise:

Exercise 5.1.3

Polarization

3. Does is matter if and how a light beam is polarized for the *applications* we have in mind?

- The answer is simple: No, it does not matter for many *simple* or *standard* applications. Your glasses, binoculars, microscopes, and so on, work perfectly well with "normal" light. With a polarization filter your camera will make somewhat better pictures in certain circumstances but that is usually not decisive.
- However, for anything a bit more sophisticated, like all the optical things that a *Material Scientist and Engineer* needs for her or his work or develops for others, it will matter *very much* as we will see. Here is a table that should give you some idea:

Application	**Polarization matters *a lot***	**Polarization doesn't matter *much***
Simple optical instruments (Cameras, binoculars, ...)		Forget it
Better optical instruments (Cameras, ..	you use a "Pol filter"	
Optical communications	not yet but soon	most cases now
Optical measurements	most depend on polarization	
3-dim movies	including circular polarization	
Laser	in many cases	some cases
LCD displays	impossible without polarization	

4. If it matters: how do we *optimize* the needed polarization?

- The answer is simple again: by constantly coming up with new or improved *materials*. Note the extremely simple and important truth:

If a light beam does not interact with some *material*,
its properties will not change.

- **5. How does the polarization of a wave relate to *photon* properties?**
  - Well - the photon has a spin of **1/2**! Naively put, there is a kind of (polar) vector associated with it. So there might be a relation between polarization and spin?
  - Yes - up to a point. But it's not that simple! Whatever the case may be, my advice is: *forget it*! Whenever polarization comes up, *only* think in the wave picture. Otherwise you simply run into all kinds of unnecessary problems. However, it is perfectly possible to describe all the effects of polarization in the particle picture, too. There is no contradiction.
- We have the last questions to consider and that will open a large can of worms, as the saying goes. Let's start with No. 6: Are there other *polarization modes* besides *linear* polarization?
  - Yes there are. They are called *circular* and *elliptical* polarization. Here we go:

Other Modes of Polarization

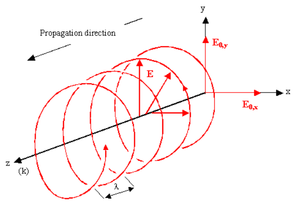
- Some idealized plane wave travelling in z direction with the electrical field vector pointing in x-direction writes as $\underline{E}_x(z,t) = E_{0,x}\exp\{i(\underline{k}z - \omega t)$ for the the electric field part. This idealized light wave is a solution of the Maxwell equations for some (idealized) conditions.
- $\underline{E}_y(z,t) = E_{0,y}\exp\{i(\underline{k}z - \omega t + \pi/2)$; i.e. the same wave but with the electrical field pointing in y-direction *and* a *phase difference* of $\pi/2$ (or any other you care to insert) is also a solution, of course.
- If those waves travel inside a material, all we have to do is to replace the electrical field $\underline{E}$ by $\underline{D}$, the dielectric displacement (and $\underline{H}$ by $\underline{B}$, of course), and to take care of the change in wavelength or the magnitude of k, respectively.
  - If the material is *isotropic* and *linear*, we have $\underline{D} = \epsilon_0\epsilon_r\underline{E}$. That means that not only the two waves from above are still a solution of the Maxwell equations for the case considered, *so is any superposition* of those (and possibly other) solutions. So whatever we are about to discuss works just as well in "simple" materials. Most optical materials are amorphous (called "glass"), and meet the requirement *isotropic* and *linear* in a fair approximation so we don't have to worry *yet*.
  - However, many crystalline materials are neither *isotropic* nor *linear* - and then we will be in trouble. Those materials are of course very interesting to a Materials Scientist but the theory of what happens is difficult and we will deal with that later.
- Now let's add up the two $\underline{E}$ (or $\underline{D}$) solutions from above and see what we get.

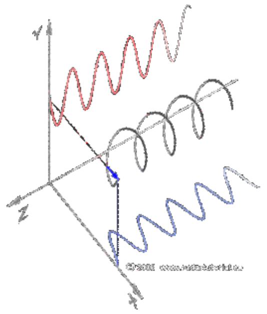
$$E_x(z,t) + E_y(z,t) \quad = \quad E_{0,x}\exp\{i(\underline{k}z - \omega t) \quad + \quad E_{0,y}\exp\{i(\underline{k}z - \omega t + \pi/2)$$

- The real part of this complex equation gives the electrical field vector, we have

$$\underline{E} \quad = \quad E_{0,x}\cdot\cos(\underline{k}z - \omega t) + E_{0,y}\cdot\sin(\underline{k}z - \omega t)$$

  - The full electrical field $\underline{E}$ thus is obtained by adding two vectors at right angles with magnitudes that change sinusoidally between $(\pm\underline{E}_0)$. The total effect is a vector with length $\underline{E}_0$ that *rotates* with a circle frequency ω around the z-axis.
  - What you get is called **circular polarization** for obvious reasons; it looks like shown below. Note that all of this appears to be far more complicated than it actually is, this is made clear by the animation below which shows the basic simplicity of what is going on.





(From wikipedia)

You realize, if you think about it, that we just opened the promised rather large can of worms. We started with simple, intuitively clear linear polarization and now progressed to circular polarization by superimposing two *special* linearly polarized waves. Of course, we can produce two basic kinds of circular polarization: **left-handed** and **right-handed** circular polarized light this way. Equally of course, we can now superimpose *all kinds of waves* with all kinds of phases. *Where will it end*?

Actually, it's not as bad as it might be. Superimposing whatever you like just gives you **elliptically polarized light** (plus, perhaps, a background of unpolarized light).

The extremes of an ellipse are straight lines (*linearly* polarized light, just superimpose the two waves without a phase difference) and the circle (phase difference **π/2**; equal amplitudes).

We will stop at this point. Far more important than going into intricate details of all this polarization stuff is to consider:

- Question 2 from above, not yet addressed: How do we *make* all these kinds of polarization? (*Answer*: with special *materials*, of course).
- What is it good for? Who needs circular or elliptically polarized light? (*Answer*: you - if you watch, e.g., **3D**-movies)

We will look very briefly at some answers in the next chapter, where we deal with *light and materials*.

Questionnaire

Multiple Choice questions to 5.1.4

5.1.5 Summary to: 5.1 Optics

- Know your numbers and relations for visible light!
  - Wavelengths: $\lambda \approx$ **400 nm - 800 nm.**

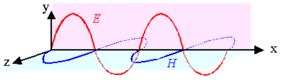
    $\lambda_{mat} = \lambda_0 / n$.
  - Frequency: $\nu \approx$ **10 ^{15} Hz**.
  - Index of refraction: $n = \epsilon_r^{1/2} \approx$ **1,5 - 2,5**
  - Energy $E \approx$ **1,8 eV - 3,2 eV**.
  - Dispersion relation: $c_0 = \nu \lambda_0 = 3 \cdot 10^8$ **m/s**
    $c_{Mat} = \nu \lambda_0 / n(\lambda)$

For the *propagation* of light:
use the *wave model*
For the *generation* and disappearance (= *absorption*) of light:
use the *photon model*

Snellius law:
$n = \sin\alpha / \sin\beta$ with α, β the angle of incidence
or propagation, resp.

- Know yout basic equations and terminology

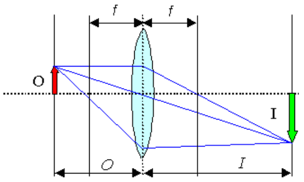
$$\begin{matrix} \underline{E}(\underline{r},t) \\ \underline{H}(\underline{r},t) \end{matrix} = \begin{matrix} \underline{E}_0 \\ \underline{H}_0 \end{matrix} \cdot \exp\{i(\underline{k}\underline{r} - \omega t)\}$$



Coherent monochromatic plane wave
$\underline{E}$ and $\underline{H}$ perpendicular and in phase

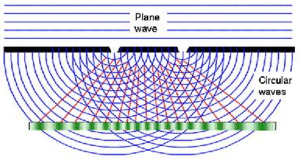
  - **Reflection** always with "angle in" = "angle out".
  - **Refraction** is the sudden "***bending***" or "flexing" of light beams at the interface
  - **Diffraction** is the continous "***bending***" of light beams around corners; interference effects.

- **Geometric optics**
  Key paramters
  - Focal length ***f*** and
    numerical aperture **NA** of lenses, mirrors.
  - Image formation by simple geometric constration
  - Various aberrations (spherical. chromatic, astigmatism, coma, ...) limit performance.



- **Wave optics**
  Huygens principle: and interference
  - Ultimate limit to resolution

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



- Know your basic types of waves:
  - (Running, coherent, monochromatic) *plane wave*.
  - *Standing waves* = superposition of plane waves.
  - Incoherent, multichromatic *real* waves

- Relation s between electrical field ***E***, magnetic field ***H*** and **Poynting vector** (energy flow vector) ***S* = *E* × *H***

$$\langle S \rangle = \frac{E_0 H_0}{2} = \frac{E_0^2}{Z_W}$$

This equation links *energy flow* (easy in photon picture) to *field strength* in wave picture.

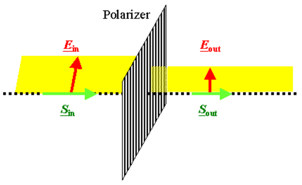
  - ***Z*W = wave impedance** of the medium.
    ***Z*W(vacuum) = 376,7 Ω**

$$W_{elect} = \frac{\epsilon_0 \cdot E^2}{2} \qquad W_{mag} = \frac{\mu_0 \cdot H^2}{2}$$

$$[W_{elect;\ magn}] = [\mathrm{Ws\ m^{-3}}]$$

$$E_0 = \left(\frac{\mu_r \mu_0}{\epsilon_r \epsilon_0}\right)^{1/2} \cdot H_0 = Z_W \cdot H_0$$

- **Polarization** = key to "advanced" optics.
  Simple case: **linear polarization**.
  - Plane of polarization contains ***E***-vector and ***S*** (***k***) vector.
  - Any (coherent) wave is polarized but *net polarization* of many waves with random polarization is zero!
  - Light *intensity* (∝ ***E*^{2}**) between polarizers at angle α scales with **(cosα)^2**.



- General case: *elliptical* polarization; important are the two extremes: *linear* and **circular** polarization.
  - For circular polarizaiton the ***E***-vector rotates on a circle while moving "forward". This results from a superposition of two plane waves with ***E***-vectors ar right angles and a *phase difference* of π**/2**.
  - Technically important (3-dim Cinema; Lab optics)

Questionaire

Multiple Choice questions to all of 5.1

5.2 Optics and Materials

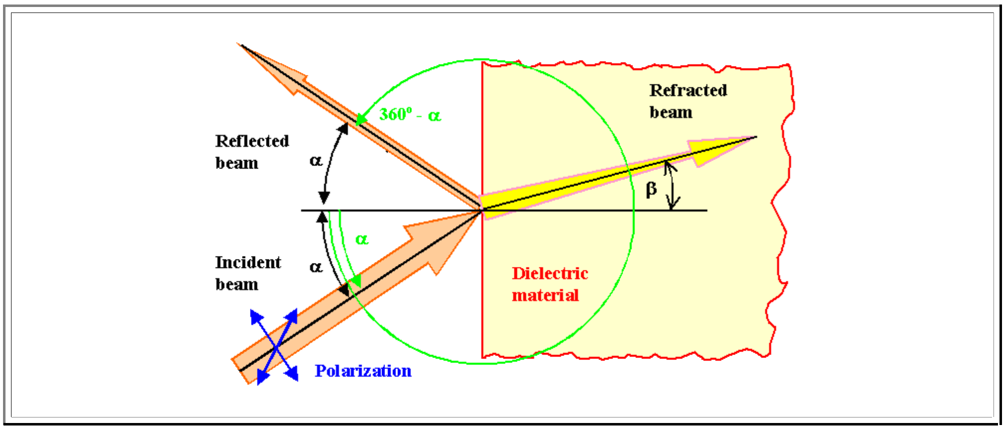
5.2.1 Interaction between Light and Matter

The Task

We have a (monochromatic, coherent, polarized) light beam (a plane wave in other words) and a piece of material. We direct our idealized "perfect" beam on the material and ask ourselves: what is going to happen?

- First we have to discuss the properties of the material a bit more. It might be:
    - Optically fully **transparent** for all visible wavelength (like diamond or glass) or optically ***opaque*** (like metals).
    - Optically partially **transparent** only for parts of the visible wavelength (like GaP or all semiconductors with bandgaps in the visible energy range) or colored glass.
    - **Opaque** and black (= fully absorbing) like ***soot*** or highly reflective (like a mirror).
    - Perfectly flat (like a polished Si wafer) or rough (like paper).
    - Uniform / homogeneous (like glass or water) or non-uniform (like milk: fat droplets in water).
    - Isotropic (like glass) or anisotropic (like all non-cubic crystals).
    - Large (like anyything you can see) or small (like the Au nanoparticles in old church window that produce the red color).
- I'm not sure I have exhausted the list. Be glad now that for the time being we look at a simple and "perfect" light beam and not at *real* light.falling on a *real* material.
- So we have a tall task before us. We first make it a bit easier by considering only flat, uniform, isotropic and transparent materials like glass or (transparent) cubic single crystals like diamond.

All that can happen for these rather ideal conditions is shown in the following picture:



In essence we have an incoming beam, a reflected beam and a diffracted beam that "goes" into the material.

- The incident "perfect" beam must have some kind of polarization. Even if it is unpolarized we should *from now on* think of it as consisting of *two* linearly polarized parallel beams with equal intensity and polarization directions differing by **90°**. Same thing for the two other beams. Consider them to be *two* beams with a **90°** difference in polarization direction. **This is important!**
- We must expect that the reflected and diffracted or **transmitted beams** might be polarized, too, but we must not assume that their polarization is the same as that of the incoming beams. We deal with that in the next sub chapter.
- The situation in the picture above is very slightly simplified because we don't consider so-called "**evanescent waves**" at the interface, and we only discuss a *linear system* **-** the frequency of the light doesn't change. There are no beams with doubled frequency, for example (can happen in some crystals).

What do we know about the three (times two) light beams shown?
I'll drop the plural from now on. But remember: think of all beams as consisting of *two* linearly polarized parallel beams with equal intensity and polarization directions differing by **90°**.

Incoming beam. We know all about the incident beam because we "make" it. In the simplest case it's a plane wave with an electrical field given by $\underline{E} = \underline{E}_0\exp(\underline{k}_{in}\cdot\underline{r} - \omega t)$ or, if you prefer, $\underline{E} = \underline{E}_0\cos(\underline{k}_{in}\cdot\underline{r} - \omega t)$. The basic parameters of the incident beam are:

- **Intensity:** We can describe the intensity I_{in} by looking at E_0^2, the square of the electrical field strength amplitude. In the particle picture it would be the number of photons per second. In regular or technical optics, we have special units, "invented" for dealing with light but we will not cover that here.
- **Frequency:** We assume monochromatic light with the circle frequency ω.
- **Polarization:** We assume some arbitrary constant polarization (= direction of the $\underline{E}$ vector). By definition,

the polarization direction is perpendicular to the direction of the wave vector ***k***. We have only *one* beam now (the intensity of the second one is zero).

- **Phase:** We can pick *any* initial phase since it's numerical value depends on the (arbitrary) zero point of the coordinate system chosen.
  Note that by just picking a direction (like this **—**), you have not yet decided where the tip of the vector is (—> or <—), so we must pick that too. Switching to the other direction then implies a **phase change** of **180^0** or π *or reversing the sign of* ***E***.
- **Coherence:** We always assume full coherence, i.e. there is only *one* phase. An **incoherent beam**, for comparison, would be a mixture of beams like "our" beam but with different (= random) phases.

Since we assume a *linear* system, we can always discuss "colored" light by discussing each frequency separately. We also can deal with arbitrary polarizations by decomposing it into the two basic polarizations considered below which we need to discuss separately. An arbitrary polarization then is just a superposition of the two basic cases; we are back to our "two beam" picture from above.

Reflected beam:. The reflected beam will essentially be identical to the incoming beam except for

- **Intensity:** The intensity I_{ref} will be different from that of the incoming beam; we have $0 < I_{re} < I_{in}$.
- **Direction:** We know that we have a mirror situation i.e. $\alpha_{out} = \alpha_{in}$. Note that its actually $\alpha_{out} = 360^o - \alpha_{in}$ if you measure the angles in *one* coordinate system. *Why* we know that we have a mirror situation is actually a tricky question!
- **Phase:** We might have to consider that the phase of the reflected beam changes at the surface of the material.

Refracted beam: The refracted or **transmitted** beam runs through the material. We know that there is always some attenuation, damping, extinction or what ever you like to call it. I will call it **attenuation**. What do we have to consider?

- **Intensity:** We know from **energy conservation** that $I_{tr}(z = 0) = I_{in} - I_{re}$.
- **Attenuation:** We expect exponential attenuation according to $I_{tr}(z) = I_{tr}(z = 0) \cdot \exp{-z/\alpha_{ab}}$ if we put the ***z***-direction in the direction of the transmitted beam for simplicity. The quantity α_{ab} obviously is an **absorption length**, giving directly the distance after which the intensity decreased to **1/e** or to about **1/3**.
- **Direction:** Snellius law applies, i.e. $\sin\alpha/\sin\beta = n$. We also know that the index of refraction ***n*** is given by ϵ_r, the "dielectric constant"; we have $n = (\epsilon_r)^{1/2}$.
- **Phase:** We might have to consider that the phase of the transmitted beam changes at the interface of the material.

While we seem to know a lot already, some tough questions remain, essentially relating to intensities, phases and attenuation as a function of the polarization, the angle of incidence, and the properties of the material.

As it will turn out, dealing with *attenuation* is easy. All we have to do is to remember that we replaced the simple "dielectric constant" ϵ_r some time ago by a complex dielectric function $\epsilon(\omega) = \epsilon'(\omega) + \epsilon''(\omega)$. Since the index of refraction is simply given by the square root of the dielectric constant, we might expect that the dielectric function not only contains the index of refraction but additional information concerning attenuation. We will look at that in sub-chapter 5.2.3 in more detail.

The questions relating to *intensities* and *phases* will not go away that easily, however. We will see how to find answers in the next sub-chapter.

Finding the Answers

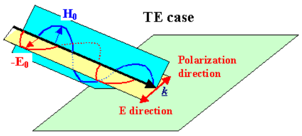
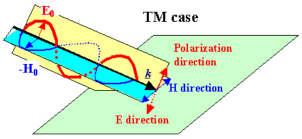
How to we have to proceed in order to find answers to the questions raised above? In other words, how can we derive the so-called **Fresnel equations** that contain the answers?

First we need to look a bit more closely at the polarization. Remembering our old convention is helpful:

Polarization (if not otherwise stated) is always taken in the *E*-field direction.

As long as we only deal with **linear polarization**, meaning that the polarization direction is always the same (i.e. not a function of time), we can describe *any* wave with some or none polarization by a superposition of the two waves discussed above with two orthogonal polarization directions.

We now need to define one of the two polarizations directions. The other one then follows automatically. We use the following convention like everybody else:

<table>
<tr>
<td> </td>
<td> </td>
</tr>
<tr>
<td>TE polarization:
("Transversal electric")</td>
<td>TM polarization:
("Transversal magnetic")</td>
</tr>
<tr>
<td><i>E</i> lies <i>perpendicular</i> to the (blue) <i>plane of incidence</i> defined by the normal vector <i>n</i> of the material surface considered and the wave vector k_{in} of the incoming wave.
<i>E</i> thus lies in the surface of the material. Also described as _|_ (perpendicular) case because the <i>E</i>-vector is perpendicular to the "plane of incidence".
The magnetic field vector <i>H</i> then lies in the "plane of incidence".</td>
<td><i>E</i> lies <i>in</i> the (yellow) <i>plane of incidence</i> defined by the normal vector <i>n</i> of the material surface considered and the wave vector k_{in} of the incoming wave.
<i>E</i> then has no components in the material surface. Also described as || (parallel) case because the <i>E</i>-vector is in the "plane of incidence".
The magnetic field vector <i>H</i> then lies in in the surface of the material.</td>
</tr>
</table>

- Once more: Arbitrary (linear) polarizations can always be described by a suitable superposition of these two basic case.

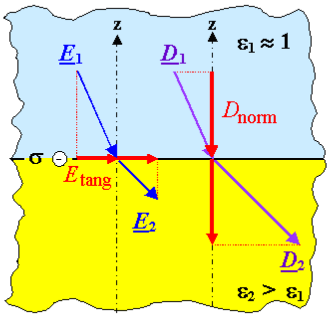
Energy conservation gives us a simple and obvious relation for the energies or intensities flowing along with the waves:

$$I_{tr}(z = 0) \;=\; I_{in} - I_{re}$$

- Note in this context that the energy of a electromagnetic wave with electrical field amplitude E_0 traveling in a medium with dielectric constant ϵ is proportional to $\epsilon^{1/2}(E_0)^2$ as we have figured out before. Note also that the transmitted beam might be attenuated so its energy is eventually transferred to the medium it's traveling in.

Going beyond that, however, needs some work. We must, in essence, start with the Maxwell equations, look at the "electromagnetic wave" case and solve them for the proper boundary conditions at the boundary of the two media.

- Or do we? Actually, we don't have to - as long as we remember (or accept) that there are simple boundary conditions for all the fields coming up in electromagnetism as illustrated below:

<table>
<tr>
<td rowspan="4"> </td>
<td>Electrical field <i>E</i></td>
<td>E_{tang} = const</td>
</tr>
<tr>
<td>Dielectric displacement
$D = \epsilon_0 \epsilon_r E$</td>
<td>D_{norm} = const.</td>
</tr>
<tr>
<td>Magnetic field <i>H</i></td>
<td>H_{tang} = const</td>
</tr>
<tr>
<td>Magnetic induction
$B = \mu_0 \mu_r E$</td>
<td>B_{norm} = const</td>
</tr>
</table>

The picture shows some interface between two (dielectric) materials. In the picture the first one is something like air or vacuum ($\epsilon_r \approx 1$) but the relations holds for all possible combinations. How do we derive the boundary conditions?

- It's easy. For an electrical field you need a gradient in some charges. For a change of an electrical field vector you need an additional charge gradient right at the place where the field vector is supposed to change. In the above case, for a change on the tangential component you would need a lateral gradient in the charge distribution on the surface.
- If you understood chapter 3, you know that some surface charge with an area charge density σ is generated by polarization. Looking long enough at Gauss' law will show that a lateral charge gradient cannot exits. The tangential components E_{tang} of ***E*** therefore must *not* change going across the boundary. The normal component, in contrast, must change because we do have an addition charge gradient perpendicular to the surface.
- Using related arguments makes clear that for the dielectric displacement ***D*** the normal component must remain constant. For the magnetic field ***H*** and the magnetic flux density or induction ***B*** corresponding relations apply. You might reason that this provides the definition of ***D*** and ***B***. It's just extremely useful to have vectors meeting those boundary conditions.

Fortified with these boundary conditions, valid for any fields including the rapidly oscillating electric and magnetic fields of light, the derivation of the *Fresnel equations* - that's what we are after - is not too difficult as we will see in the next sub-chapter.

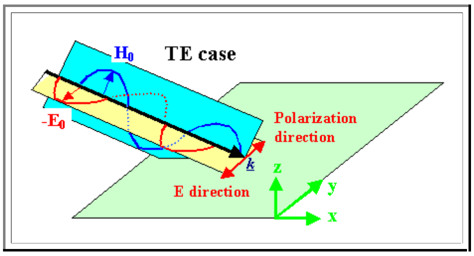
Questionnaire

Multiple Choice questions to 5.2.1

5.2.2 Fresnel Equations

Deriving the Fresnel Equations

- Let's look at the TE mode (or _|_ mode) once more but now with the **coordinate system** needed for the equations coming up.



  - The electrical field of the incoming beam thus writes as $\underline{E}_{in} = (0, E_{in}, 0)$, i.e. there is only an oscillating component in y-direction. For the y-component E_{in} we can write $E_{in} = E_{in,0}\exp[-i(\underline{k}_{in} z\cos\alpha + \underline{k}_{in} x\sin\alpha)]$, decomposing the wave in an z and x component. We omitted the ωt phase factor because it will drop out anyway as soon as we go to intensities.
  - Next we should write the corresponding equations for the reflected wave and the transmitted wave (requiring changes in the k-vector).
  - Then we need the same set of equations for the magnetic field. For that we have to know how the magnetic field of an electromagnetic wave can be derived from its electrical field. That means back to the Maxwell equations once more or for a taste of that to sub-chapter 5.1.4.
  - After *you* did that you consider the *boundary conditions* as outlined before. Now you can start to derive the Fresnel equations. *You*, not me. It's tedious but good exercise. Let's just look at the general way to proceed.
- First we write down the continuity of the tangential or here parallel component of $\underline{E}$ (and always same thing for $\underline{H}$ in principle). Since $\underline{E}$ has only components in y-directions we have for those components

$$E_{in} + E_{ref} \;=\; E_{tr}$$

  - While this looks a bit like the energy or *intensity* conservation equation from before, it is not! It is *completely different*, in fact! Our E's here are *field strengths* and *not* energy!
- So let's look at the energy flux in z-direction now, as given by the Poynting vector $\underline{S}$. It must be continuous since energy is neither generated nor taken out at the interface as noted before. With the relation for energy from before and dropping the index "$_0$" for easier writing and reading, we obtain

$$\epsilon_1^{1/2}\,[(E_{in})^2 - (E_{ref})^2]\cdot\cos\alpha \;=\; \epsilon_2^{1/2}\,(E_{tr})^2\cdot\cos\beta$$

  - This equation is simply the good old **Snellius law** in slight disguise (figure it out yourself, noting that $I_{tr} = I_{in} - I_{ref}$).
- Dividing that equation by the one above it (remember: $(E_{in})^2 - (E_{ref})^2 = (E_{in} - E_{ref})\cdot(E_{in} + E_{ref})$ gives

$$\epsilon_1^{1/2}\,[E_{in} - E_{ref}]\cdot\cos\alpha \;=\; \epsilon_2^{1/2} E_{tr}\cdot\cos\beta$$

  - With the good old relation $(\epsilon_1/\epsilon_2)^{1/2} = n_1/n_2 = \sin\alpha/\sin\beta$ (with α and β as given in the old figure for simplicity) and some shuffling of terms we finally obtain the **Fresnel equations for the TE case** .

Fresnel Equations TE case

$$E_{ref} \;=\; E_{in}\cdot\frac{\sin\beta\cos\alpha - \sin\alpha\cos\beta}{\sin\beta\cos\alpha + \sin\alpha\cos\beta} \;=\; -\,E_{in}\cdot\frac{\sin(\alpha-\beta)}{\sin(\alpha+\beta)}$$

$$E_{tr} \;=\; E_{in}\cdot\frac{2\sin\beta\cos\alpha}{\sin(\alpha+\beta)}$$

- Going through the whole thing for the **TM** case (something I will not do here) gives the **Fresnel equations for the TM case**

Fresnel equations TM case

$$E_{ref} = E_{in} \cdot \frac{\sin\beta\cos\beta - \sin\alpha\cos\alpha}{\sin\beta\cos\beta + \sin\alpha\cos\alpha} = -E_{in} \cdot \frac{\tan(\alpha - \beta)}{\tan(\alpha + \beta)}$$

$$E_{tr} = E_{in} \cdot \frac{2\sin\beta\cos\alpha}{\sin(\alpha + \beta) \cdot \cos(\alpha - \beta)}$$

- Relatively simple equations - but with a lot of power! Before we look at these equations a bit more closely, we modify them by dividing everything by E_{in}, so we get relative numbers (in % if you like) for the field strength or the relative intensities $(E/E_{in})^2$.
  - The resulting numbers for the relative field strengths we call the **Fresnel coefficients**. We have four Fresnel coefficients: one each for reflection or transmittance, and that always separately for the **TE** and **TM** case.

Using the Fresnel Equations

- A first extremely easy thing to do is to calculate the Fresnel coefficients for normal incidence ($\alpha = 0^o$). What we get for the standard case of going from air (less dense medium, $n = 1$ to some appreciable n (denser medium) for *both* the **TE** *and* **TM** case is

$$\frac{E_{ref}}{E_{in}} = -\frac{n - 1}{n + 1}$$

$$\frac{I_{ref}}{I_{in}} = \left(\frac{n - 1}{n + 1}\right)^2$$

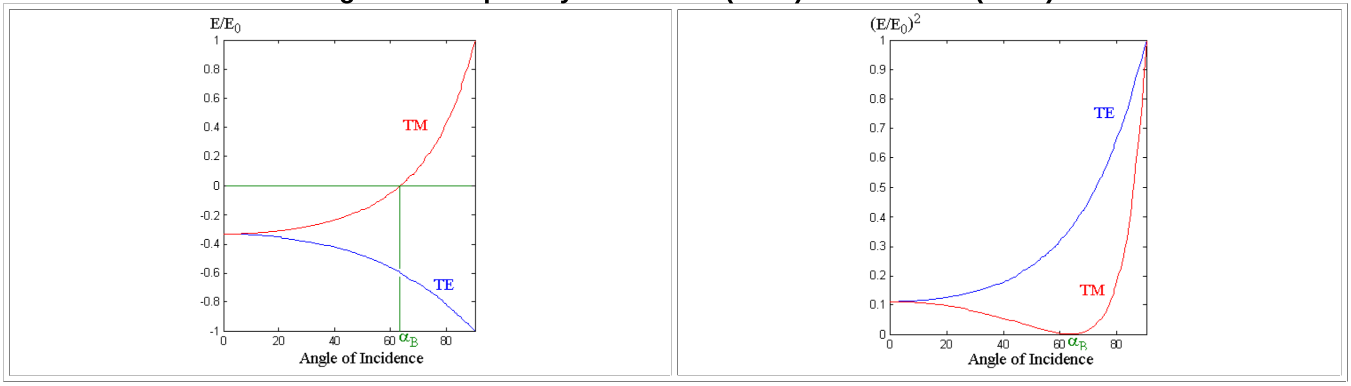
  - In other words, shining light straight on some glass with $n = 2$ means that almost **10 %** of the intensity will be reflected! This has immediate and dire consequences for optical instruments: you *must* provide some "anti-reflection" coating - otherwise your intensity gets too low after the light passed through a few lenses..
- We need to do a bit of exercise here:

Exercise 5.2.1

Fresnel coefficients

  - If we now speculate a little and consider metals as a material with *very large* dielectric constants and thus n, it is clear that they will reflect almost **100 %**.
- Next we plot the Fresnel coefficients as a function of α, the angle of incidence. We need four figures with **8** graphs to get the major points clear:
  - Two figures showing the relative field strength (E_{ref}/E_{in}), always with two separate graphs for the two basic cases **TE** and **TM**.
    1. Case 1: $n_1 > n_2$, i.e. going from the less dense to the optically denser material
    2. Case 2: $n_1 < n_2$, i.e. going from the optically denser to the less dense material
  - Two figures showing the relative *intensity* $(E_{ref}/E_{in})^2$, always with separate graphs for the two basic cases **TE** and **TM**; same cases as above
- First we look at **case 1** with $n_1 < n_2$, i.e. going from the less dense to the optically denser material
  - We take $n_1/n_2 = ½$ e.g. going from air with $n = 1$ into some glass with $n = 2$.
  - Here are the **4** graphs for this case; we look at the reflected beam.

Case 1: Going from an optically less dense (*n* = 1) into a dense (*n* = 2) material



Electrical Field
E/E_0

Reflected beam

Intensity
$(E/E_0)^2$

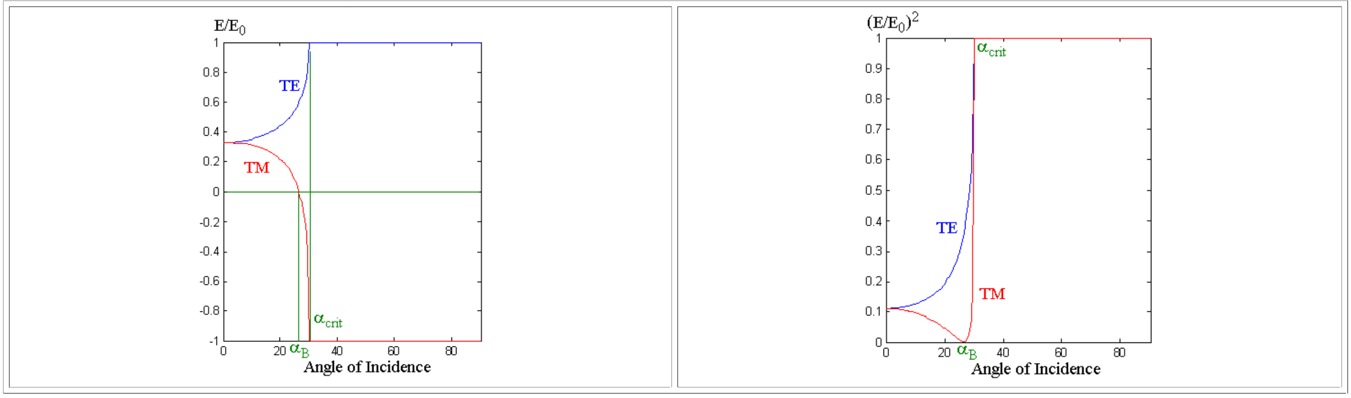
- Let's look at the field strength first What we see is

  1. The numbers are negative for small α or almost perpendicular incidence. This means that we have a phase shift of **180°** between the incident and the reflected wave as outlined before.
  2. The relative amplitude of the transmitted beam is simply **1 -** E_{ref}/E_{in}. It becomes small for large α's.
  3. In the **TM** case the field strength is exactly zero at a certain angle α_B or "**Brewster angle**". This means that there is *no reflection* for this polarization; all the light will be transmitted.
  4. So if the incident light consists of waves with arbitrary polarization, the component in **TM** direction will not be reflected and that means that whatever will be reflected must be polarized in TE direction. *We have a way to polarize light!*.
  5. For **grazing incidence** or large α's almost all of the light will be reflected in either case.

- Looking a the intensities does not show anything new; you just see the "strength" of the reflected beam more clearly.

- Now let's look at **case 2** with $n_1 > n_2$, i.e. going from the more dense to the optically less dense material
  - We take $n_1/n_2 = 2$ e.g. going from some glass with ***n* = 2** into air with ***n* = 1**.
  - Here are the **4** graphs for this case.

Case 2: Going from an optically dense (*n* = 2) into a less dense (*n* = 1) material



Electrical Field
***E*/E_0**

Reflected Beam

Intensity
$(E/E_0)^2$

- Let's look at the field strength first What we see is

  1. The numbers are positive for small α or almost perpendicular incidence. This means that we have no phase shift of **180°** between the incident and the reflected wave as outlined before. Not that the reflected wave is the one staying inside the optically dense material.
  2. The relative amplitude of the wave leaving the material is simply **1 -** E/E_0. It goes to zero rather quickly for increasing α's
  3. In the **TM** case the field strength is exactly zero at a certain angle α_B or "**Brewster angle**". This means that there is *no reflection* for this polarization, all the light will be transmitted. Note that the value for the Brewster angle here is different from the one in the case going from the less dense to the more dense material.
  4. So if the incident light consists of waves with arbitrary polarization, the component in **TM** direction will not be reflected and that means that whatever will be reflected must be polarized in **TE** direction. *We have a way to*

polarize light inside a material!

5. At some **critical angle** α_{crit} all light in either mode will be reflected. Beyond α_{crit} the Fresnel equations have only complex number solution and that means there is no field strength or energy outside the material. Light waves impinging at an angle **>**α_{crit} will be reflected right back into the material. α_{crit} is also known as the angel of **total reflection**.
6. That means that only light within a cone with opening angle **<** α_{crit} will be able to get of the material. It should be clear to you that a serious problem concerning **light emitting diodes,** (***LED***) is encountered here.

It's time for another exercise:

Exercise 5.2.2

Fresnel equations

Questionnaire

Multiple Choice questions to 5.2.2

5.2.3 The Complex Index of Refraction

Dielectric Function and the Complex Index of Refraction

- Light is an electromagnetic wave. We have an electrical field that oscillates with some frequency (around 10^{15} **Hz** as you should now know by heart). If it impinges on a dielectric material (= no free electrons), it will jiggle the charges inside (bound electrons) around a bit. We looked at this in detail in **chapter 3**.
  - An electrical field caused some *polarization* of the dielectric material. This lead straight to the *dielectric constant* ϵ_r.

Attention!	**The word "polarization" above and in chapter 3 has *nothing* to do with the "polarization" of light!**	**Attention!**

  - Since the word "light" is synonymous to "oscillating electrical field", it is no surprise that ϵ_r is linked to the *index of refraction* $n = \epsilon_r^{½}$.
  - For oscillating electrical fields we needed to look at the *frequency dependence of the polarization* and that lead straight to the complex *dielectric function* $\epsilon_r(\omega) = \epsilon'(\omega) + i\epsilon''(\omega)$ instead of the simple dielectric constant ϵ_r. Go back to chapter 3.3.2 if you don't quite recall all of this.
- The dielectric function, after some getting used to, made life much easier and provided for new insights not easily obtainable otherwise. In particular, it encompassed the "ideal" dielectric losses *and* losses resulting from non-ideality. i.e. from a finite conductivity in its imaginary part.
- So it's logical to do exactly the same thing for the index of refraction. We replace n by a **complex index of refraction** n^* defined as

$$n^* = n + i\,\kappa$$

  - We don't use n' and n'' as symbols for the real and imaginary part but denote the real part by the (old) symbol n and the imaginary part by κ. This is simply to keep up with tradition and has no special meaning.
- We use the old relation between the index of refraction and the dielectric constant but now write it as

$$(n + i\kappa)^2 = \epsilon' + i\epsilon''$$

  - With $n = n(\omega)$; $\kappa = \kappa(\omega)$, since ϵ' and ϵ'' are frequency dependent as discussed before.
- Re-arranging for n and κ yields somewhat unwieldy equations:

$$n^2 = \frac{1}{2}\left(\left(\epsilon'^{\,2} + \epsilon''^{\,2}\right)^{½} + \epsilon'\right)$$

$$\kappa^2 = \frac{1}{2}\left(\left(\epsilon'^{\,2} + \epsilon''^{\,2}\right)^{½} - \epsilon'\right)$$

- Anyway - That is all. Together with the Fresnel equations we now have a lot of optics covered. Example of a real complex indexes of refraction are shown in the link.
  - So lets see how this works, and what κ, the so far unspecified imaginary part of n^*, will give us.

The Meaning of the Imaginary Part κ

- First, lets get some easier formula. In order to do this, we remember that **ε''** was connected to the "dielectric" and static (ohmic) conductivity of the material and express **ε''** in terms of the (total) conductivity σ_{DK} as

$$\boxed{\epsilon'' = \frac{\sigma_{DK}}{\epsilon_0 \cdot \omega}}$$

  - Note that in contrast to the definition of **ε''** given before in the context of the dielectric function, we have an ϵ_0 in the **ε''** part. We had, for the sake of simplicity, made a convention that the ε in the dielectric function contain the ϵ_0, but here it is more convenient to write it out, because then $\epsilon' = \epsilon_0 \cdot \epsilon_r$ is reduced to ϵ_r and directly relates to the "simple" index of refraction ***n***
  - Using that in the expression $(n + i\kappa)^2$ gives

$$\boxed{(n + i\kappa)^2 = n^2 - \kappa^2 + i \cdot 2n\kappa = \epsilon' + i \cdot \frac{\sigma_{DK}}{\epsilon_0 \cdot \omega}}$$

  - We have a complex number on both sides of the equality sign, and this demands that the real and imaginary parts must be the same on both sides, i.e.

$$\boxed{\begin{aligned} n^2 - \kappa^2 &= \epsilon' \\ n\kappa &= \frac{\sigma_{DK}}{2\epsilon_0\omega} \end{aligned}}$$

  - Separating ***n*** and κ finally gives

$$\boxed{\begin{aligned} n^2 &= \frac{1}{2}\left(\epsilon' + \left(\epsilon'^{\,2} + \frac{\sigma_{DK}^2}{4\epsilon_0^2\omega^2}\right)^{1/2}\right) \\ \kappa^2 &= \frac{1}{2}\left(-\epsilon' + \left(\epsilon'^{\,2} + \frac{\sigma_{DK}^2}{4\epsilon_0^2\omega^2}\right)^{1/2}\right) \end{aligned}}$$

  - Similar to what we had above, but now with basic quantities like the "relative dielectric constant" since $\epsilon' = \epsilon_r$ and the total conductivity σ_{DK}.

- Now lets look at the *physical* meaning of **n** and κ, i.e. the real and complex part of the complex index of refraction, by looking at an electromagnetic wave traveling through a medium with such an index.
  - For that we simply use the general formula for the electrical field strength ***E*** of an electromagnetic wave traveling in a medium with refractive index **n***. For simplicities sake, we do it one-dimensional in the **x**-direction (and use the index "**x**" only in the first equation). In the most general terms we have

$$\boxed{E_x = E_{0,\,x} \cdot \exp i \cdot (k_x \cdot x - \omega \cdot t)}$$

  - With k_x = component of the wave vector in **x**-direction = $k = 2\pi/\lambda$, ω = circular frequency = $2\pi\nu$.

- There is no index of refraction in the formulas but you know (I hope) what to do.
  - You must introduce the velocity **v** of the electromagnetic wave in the material and use the relation between frequency, wavelength, and velocity to get rid of ***k*** or λ, respectively. In other words, we use

$$v = \frac{c}{n^*} \qquad v = \nu \cdot \lambda$$

$$\boldsymbol{k} = \frac{2\pi}{\lambda} = \frac{\omega \cdot n^*}{c}$$

- Of course, **c** is the speed of light in vacuum. Insertion yields

$$E_x = E_{0,x} \cdot \exp i \cdot \left(\frac{\omega \cdot n^*}{c} \cdot x - \omega \cdot t \right) = E_{0,x} \cdot \exp i \cdot \left(\frac{\omega \cdot (n + i \cdot \kappa)}{c} \cdot x - \omega \cdot t \right)$$

$$E_x = E_{0,x} \cdot \exp \cdot \left(\frac{i \cdot \omega \cdot n \cdot x}{c} - \frac{\omega \cdot \kappa \cdot x}{c} - i \cdot \omega \cdot t \right)$$

- The red expression is nothing but the wavevector, so we get a rather simple result:

$$E_x = \underbrace{\exp - \frac{\omega \cdot \kappa \cdot x}{c}}_{\text{Decreasing amplitude}} \cdot \underbrace{\exp[\, i \cdot (k_x \cdot x - \omega \cdot t)]}_{\text{Plane wave}}$$

- Spelt out: if we use a complex index of refraction, the propagation of electromagnetic waves in a material is whatever it would be for an ideal material with only a *real* index of refraction *times* a *attenuation factor* that decreases the amplitude exponentially as a function of depth ***x***.
  - Obviously, at a depth often called **absorption length** or **penetration depth W = c/**ω **·** κ, the intensity decreased by a factor **1/e**.
  - The imaginary part κ of the complex index of refraction thus describes rather directly the attenuation of electromagnetic waves in the material considered. It is known as **damping constant**, **attenuation index**, **extinction coefficient**, or (rather misleading) *absorption constant*. Misleading, because an absorption constant is usually the α in some exponential decay law of the form $I = I_0 \cdot \exp - \alpha \cdot x$ or what we called **W = c/**ω **·** κ above.
  - Note: Words like "constant", "index", or "coefficient" are also misleading - because κ is not constant, but depends on the frequency just as much as the real and imaginary part of the dielectric function.

Using the Complex Index of Refraction

- The equations above go beyond just describing the optical properties of (perfect) dielectrics because we can include all kinds of conduction mechanisms into σ, and all kinds of dielectric polarization mechanisms into **ε'**.
  - We can even use these equations for things like the reflectivity of metals, as we shall see.
- Keeping in mind that typical ***n***'s in the visible region are somewhere between **1.5 - 2.5** (***n*** ≈ **2.5** for diamond is one of the highest values as your girl friend knows), we can draw a few quick conclusions: From the simple but coupled equations for ***n*** and κ follows:
  - For $\sigma_{DK} = 0$ (and, as we would assume as a matter of course, $\epsilon_r > 0$ (but possibly < 1?)) we obtain immediately $n = (\epsilon_r)^{1/2}$ and $\kappa = 0$ - the old-fashioned simple relation between just ϵ_r and ***n***. Remember that $\sigma_{DK} = 0$ applies *only* if

    1. the static conductivity σ_{stat} is close to zero, and
    2. we have frequencies where $\epsilon'' \approx 0$, i.e well outside the resonance "peak" for optical frequencies.
- Generally, we would *like* κ to be rather small for "common" optical materials!
  - We also *expect* κ to be rather small for "common" optical materials, because optical materials are commonly insulators, i.e. so at least $\sigma_{static} \approx 0$ applies.
- Let's look at some numbers now. With $\omega \approx 10^{16}$ **Hz** and $c = 3 \cdot 10^{10}$ **cm/s**, we have a penetration depth $W \approx 3 \cdot 10^{-6}/\kappa$.

- If, for example, the penetration depth should be in excess of **1 km** (for optical communication, say), $\kappa < 3 \cdot 10^{-11}$ is needed. It should be clear that this is quite a tough requirement on the material. How does it translates into requiremetns for σ_{DK} or ϵ''?

Exercise 5.2.3
Attenuation and dielectric function

If we now look at the other extreme, materials with large σ_{DK} values (e.g. metals), both ***n*** and κ will become large.

- Looking at the Fresnel equations we see that for large ***n*** values the intensity of the reflected beam approaches **100 %**, and large κ values mean that the little bit of light that is not reflected will not go very deep.
- Light that hits a good conductor thus will be mostly reflected and does not penetrate. Well, that is exactly what happens when light hits a metal, as we know from everyday experience.

Questionnaire
Multiple Choice questions to 5.2.3

5.2.4 Polarization and Materials

Linear Polarization

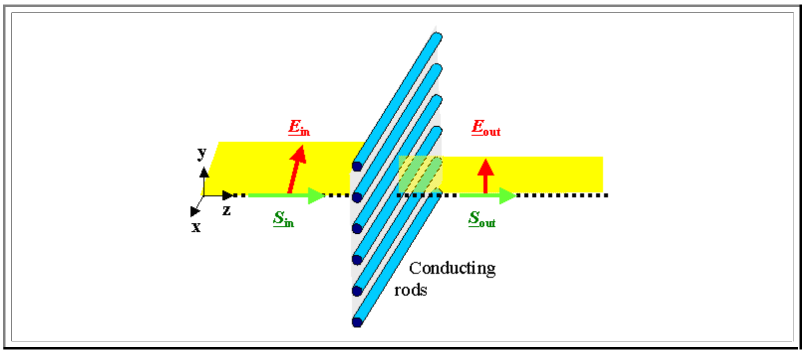
- What kind of properties need a material have to have, so it *linearly* polarizes light?
  - For starters: if you run the light *through* the material, it needs to be transparent. But who says that you can only polarize light by running it *through* a material? How about bouncing it off some material or, in other words, *reflecting* it at some mirror-like surface?
  - This should trigger a flash-back. As you already know, it is perfectly possible to polarize light by utilizing reflection; just look once more at the Fresnel equations.
- All things considered, there are several ways to polarizes light, each one with its own requirements, advantages, and disadvantages. Here we'll just take a first glimpse at some major **polarizing methods**.
- **1. Special geometries**.
  - As we have learned under the heading "*Fresnel equations*", a non-polarized light beam impinging on *any material* under some special angle ("*Brewster angle*") will produce a fully polarized *reflected* beam. Of course, some materials are more suitable than others (the index of refraction should have a decent value, not too close to $n=1$ but not too large either) but no *special* properties of the materials are required.
  - Using the "Brewster angle" approach is indeed a major way to produce polarized light. It is more or less limited to "advanced" application, however, because there are some problems. Time for a quick exercise:

Exercise 5.2.4

Polarizers based on reflection

  - For everyday applications like your "Polaroid" sun glasses or the cheap glasses used for watching **3D** movies in the cinema, you need something less special and far, far cheaper. You need:
- **2. Polarization filters or foils**
  - We have a typically thin and bendable transparent foil (or a thin unbendable glass-like sheet). Unpolarized light goes in from one side, *linearly* polarized light comes out on the other side. There are *two* basic ways or principles for that:

    1. The material contains "nanorod" conductors arranged in a grid with dimensions in the wavelength region.
    2. The material is "*birefringent*" or in neutral terms, it is *optically anisotropic*, meaning that ϵ and therefore also n are *tensors*.
  - The *bad* news are: The second principle is the more important one!
    The *good* news are: We are *not* going to look at that in great detail, since there is not enough time. Tensor optics also may cause the bulk intake of large quantities of alcohol to soothe your brain, and you are too young for this.
- The *first* principle is easy to understand:
  - Just imagine for a moment that you need to polarize *radio waves* with a wave length in the **cm** region and not light. All you need to do is to use a bunch of aligned **conducting rods** as shown below:

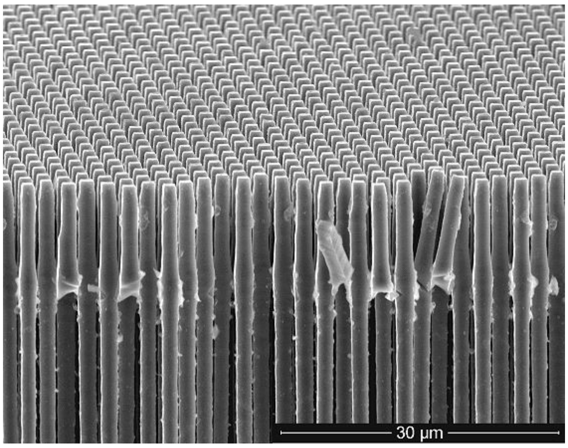


- *Surprise*! The polarization direction is *not* parallel to the rods as one would naively imagine but **perpendicular** to it.
  - It's clear what happens, though. The field components *parallel* to the conducting rods will simply be short circuited, causing currents $\underline{j} = \sigma \underline{E}$ in the rod, and thus heating. The components *perpendicular* to the rods cannot cause much current and pass mostly through. The figure above, by the way, is identical to a figure we had before except that the symbolic polarizer in this picture has now materialized into a defined device or product.
  - If you wonder about possible *diffraction* effects at the grid: There aren't any if the distance between the rods is smaller than a wavelength. Just do an Ewald sphere construction to see this.
- The priciple, of course, also works for the electromagnetic waves we call light. All that remains to do, is to produce aligned conductive rods on a **100 nm** or so scale; **nanorods** in other words. *Any ideas*?

- Well, a guy named **Edwin Herbert Land** (1909 - 1991) had an idea - in 1932. Take a polymer foil that can easily be stretched to a large extent (look at this link to get an idea), get some proper stuff inside (microscopic **herapathite crystals** in Lang's case; funny things in their own right involving **dog piss**), and then stretch the whole foil, aligning the little crystals. Use the link to get the full story.
- I don't know what Lang thought when he did this. Some people (and text books) believe that he did *align conducting rods*, indeed. Chances are that he did not, however. The herapathite crystals in Lang's case are actually very complex crystals; their structure was found out only a few years ago. Herapathite (also known as Chininbisulfatpolyiodid) forms intrinsically **polarizing crystals**, and all the stretching of the foil just insured that their polarization axes were aligned.

So back to square 1: *Any ideas* of how to align conducting *nano*rods on a **100 nm** or so scale?

- Of course you have an idea: Use the basic *micro* (or by now *nano*) technologies of semiconductor technology and produce something as shown in the picture below



Si nanorods (actually rather microrods but one could make them smaller if required)

- The rods shown consist of **Si** and they would certainly polarizes (**IR**) light besides doing a few more things (due to their 3-dim.arrangement). It's however not a practical way for cheap mass-production stuff. Moreover, while light coming out on the "other" side would certainly be linearly polarized, one can just feel that it would not be a lot of light that makes it through the structure. More about this structure (which was not made for optics but for something utterly different) in the link.
- If we look a bit beyond "normal" optics into the far **UV** and **IR**, polarizing a beam is not all that easy and structures as shown above might be the "future" in this case. Here is the link for some more information about this.
- Time for the famous last word: I could not find out in April 2011 if there are actually (cheap) polarizing foils out there that actually work on the "conducting rod" principle. So let's go for a *third* way.

3. Polarizing materials

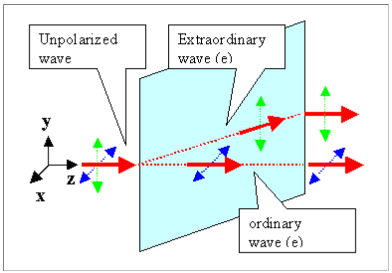
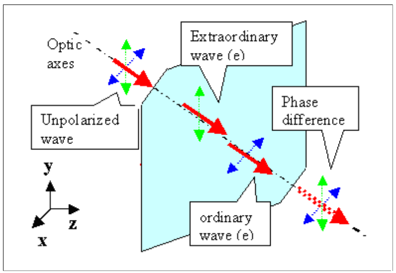
- A good polarizing material, e.g. some *crystal* and not a "foil", transmits an incoming beam without too much absorption and emits one *or even two* beams that are more or less polarized.
- We can already make one statement about the properties such a material must have: All directions must *not be equal* since the polarization direction is per definition special. Isotropic materials like most amorphous stuff and *all cubic crystals*, thus *cannot* be polarizing materials.
- A stretched polymer foil is no longer isotropic since the polymer chains now are somewhat aligned in the stretching direction. However, it will not necessarily polarize light. *Try it!*. Stretch some suitable candy wrap and see if it now polarizes light. Changes are it will not. So just being structurally *anisotropic* is not a good enough condition for being a polarizer.

That means we must now deal with not-so-simple Bravais lattices, crystals, or other structures. It also means that properties like the dielectric constant ϵ_r and thus the index of refraction ***n***, the conductivity σ_r, possible magnetic properties, the modulus of elasticity (Young's modulus) and so on are no longer simple *scalar* number but **tensors** of second, third or even fourth rank! We are now discussing crystalline *"optical tensor materials"*.

- Talk about opening a can of worms! Nevertheless, understanding and exploiting the tensor properties of materials (mostly crystals) is where the action is right now (2011), and where it will be for the foreseeable future.
  At this point I will only enumerate some of the important effects of tensor materials on polarization.

Optical Anisotropy and Tensor Materials

- The basic ideas are easy to state. First, if there is an optical anisotropy, it can take two basic forms (and then mixtures of the two, of course):
- **1. The index of refraction depends on the crystal direction**. The optical effects resulting from this are called "**birefringence**" or **double refraction** (Latin: bi=two, twice, refringere=to break up, to refract).
- **2. The absorption depends on the polarization (and the crystal direction)**. In some directions far less light of some polarization will come out after travelling through a crystal of given thickness than light polarized in the other direction. This effect is called **dichroism** (Greek for two-colored).
  - It is an unfortunate word because historically it was used first for a *different* effect: White light is split into into distinct beams of different wavelengths because of the crystal anisotropy and this is completely different from regular dispersion! Since the absorption effect may strongly depend on the wavelength, too, everything can be mixed up wonderfully. The word "**pleochroism**" ("more-colored") is occasionally used as a more general term, containing all of the above.
- You know, of course (?) that the two effects must be related because they must both be contained in the complex index of refraction, which simply (haha) happens to be a **tensor** now. All tensor components are functions of the wavelength because we still have dispersion *and*, since we also have the Fresnel equations, they must be functions of the polarization, too
- All we need to do now is to run through the derivation of the Fresnel equations once more. Just consider now a complex index of refraction that is also a tensor of of second rank with a symmetry that is somehow related to the crystal symmetry instead of a scale for boundary conditions and propagation. That would be a program that could entertain (some of) us for many weeks - so we are not going to do it here.
  - It is important to realize, however, that with our modern computers this is actually an easy task. As soon as a basic program has be written, and the functions describing the tensor components of the complex index are inputted, the rest is child's play. That's why for you, the young Materials Science and Engineer, the complex index is so important. For old guys like me it was not important, since we simply couldn't do the necessary math with a slide ruler as the only math tool. We had to resort to approximations and special case studies, and that's where all these fance names and distinctions comes from.
- We are going to give the whole topic just a very superficial look, focussing on the practical side. As a first simplification we simply ignore di- or pleochroism for the time being and only give *birefringence* a slightly closer look. We also ignore absorption and thus the imaginary part of the complex index of refraction.
  - Recalling the little bit what we know about tensors, we are certain that there is always a coordinate system or "axes", where only the diagonal elements of the refractive index tensor are different from zero. These axes we call the **principal axes** of the crystal under consideration.
  - We then can always define *three* **principal refractive indices n_1, n_2, n_3** as the tensor components in the principal axes system. Note that those ***n***'s will depend on the polarization and the wavelength. Of course, the principle axes will be related to the symmetry of the crystal in some way.
  - The rest is math. Not too difficult if you are used to tenor algebra but a bit mind boggling for normal people like us. For example, the dielectric displacement ***D*** or the Poynting vector ***S*** may no longer be parallel to the electrical field ***E*** or the wave vector ***k***, respectively. Now what would that mean?
- Be that as it may, the fact of importance here is that *in general*, an unpolarized light beam entering an anisotropic crystals splits into *two* orthogonally polarized light beams inside the birefringent material (typically a single crystal, a collection of small aligned single crystals, or an amorphous foil "somehow" made anisotropic and birefringent). *Two* different light beams thus will run through the crystal and eventually exit on the other side as shown below.

	
Incoming beam *not* along *optic axis* (see below), but at *right angle* to some surface. ⇒ No change of direction under "ordinary" circumstances and for the *ordinary* beam. Pronounced change of direction for *extraordinary* beam.	Incoming beam *along* optic axis (see beleow) and at *right angle* to surface. ⇒ No change of direction under "ordinary" circumstances *and* for both beams in birefringent material. *Both* beams run parallel but with *different* velocities.

- The two figures above illustrate the basic effects that will occur. It is best to discuss birefringences for the two special cases shown
- **1. case:** The incoming (unpolarized) beam is *not* parallel to a principal axis of the crystal but its angle of incidence is **90^o**, i.e. it is exactly perpendicular to the surface of the crystal. This is a special condition, but not so very special because it is easy to do in an experiment, All you need are some flat surfaces.
  - For an *isotropic* material with a *scalar* index of refraction ***n*** we know that there will be no refraction or bending of the beam. A little bit will be reflected (see the Fresnel equation), and the polarization does not matter. The transmitted beam will travel through the material at a velocity $c=c_0/n$.
  - Exactly the same thing happens for the **ordinary wave** or **o**-wave in our tensor material here. However, the **o**-wave now is *fully polarized* as shown
  - But in major contrast to isotropic materials with a scalar ***n***, an *extraordinary* thing happens in our birefringent material with a *tensor* ***n***. An **extraordinary wave** or **e**-wave is also generated at the interface. This **e**-wave travels under some angle with a velocity that is *different* from that of the **o**-wave. The **e**-wave is also *fully polarized*, but with a direction orthogonal to that of the **o**-wave.
- **2. case:** The incoming (unpolarized) beam is *parallel* to an *optical axis* of the crystal, and its angle of incidence is still **90^o**, i.e. it is exactly perpendicular to the surface of the crystal. This is a rather special condition now because it means that our crystal must have a planar surface perpendicular to an special crystal direciion.
  - The so-called **optic axis** of the birefringent crystal is coupled to our principal axes from above. We have something new now (or just an extreme case of the general situation).
  - The **o**-wave *and* the **e**-wave travel in the *same* direction=optic axis but with *different velocities*. Both waves are still fully polarized in orthogonal directions.
  - The two waves emerging form the crystal then have *different phases*. The exact phase difference depends on the distance covered inside the material, i.e. on the thickness of the material.
- That's just a description of what you will observe, of course. Calculating the relative intensities of the two beams, the intensities of reflected beams (??? (there is only one)), the two angles of refraction, the polarization directions, and the propagation velocities from the index of refraction tensor is possible, of course—but well beyond our scope here.
  - One question remains: How many optic axes are there? The answer is: *Two* for "fully" anisotropic material and *one* for somewhat more symmetric materials. That tells you that the optic axis and the principal axis are not the same thing because we always have *three* pricipal axes. It's all in the table below.

Dielectric tensor in *principal* axes			
Anisotropic Materials			**Isotropic Materials**
$\epsilon_r = \begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_3 \end{pmatrix}$	$\epsilon_r = \begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_3 \end{pmatrix}$		$\epsilon_r = \begin{pmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_1 \end{pmatrix}$
General case **Two** optical axes	Practical case **One** optical axis		Simple case **All** directions / axes are equal
Mica	$\epsilon_3 > \epsilon_1$ Positive uniaxial	$\epsilon_3 < \epsilon_1$ Negative uniaxial	Glass Diamond Cubic crystal (NaCl, CaF_2, Si, GaAs, ..)
	Ice quartz TiO_2 (rutile)	Calcite ($CaCO_3$) Tourmaline $LiNbO_3$	

- I can't end this chapter without a quick reference to the ubiquitous direct effect of birefringent crystals discovered long ago. There is deceptively simple effect already known to Huygens and Co: Take an "*Iceland spar*" crystal (**$CaCO_3$** or **calcite**; easy to get), put it on some paper with writing on it and you see everything *twice* (see below).
  - Here are some pictures taken in my office:

Birefringence with a calcite crystal

Circular and Elliptical Polarization

- The topic is now to produce circular and elliptical polarization. This is easy because we already did that just above where we read: "*The two waves emerging form the crystal then have different phases*"
  - That sentence addressed the ordinary and extraordinary beam traveling in *parallel* along one of the two (or just the one) optic axes of a birefringent crystal.
  - The two beams travel in parallel—but with different speeds. So the have different phases at *different depths*. If we want *circular* polarized light, we just have to make sure that the thickness of the material we use has the right value for a phase difference of π **/2** between the two beams. The beams emerging then have exactly the phase difference needed for circular polarization.
- So far - so easy. Now let's do it in reality. You need a circular polarizer, and you now have all this textbook knowledge. What's next?
  - Easy. You look into proper catalogues and buy a "**quarter lambda plate**" with a thickness ***d*** given by

$$(n_e - n_o) \cdot d = \lambda/4$$

  - Then you laminate it with a linear polarizer to make sure that only light polarized in a certain direction enters the quarter lambda plate. What you get is one part of the glasses you get for watching 3-dim. movies. Alternatively, you just take those glasses with you, and you have two circular polarizers (plus the linear ones) for free. Don't ask me, how it is possible to make those optical components for very little money. I don't know—but intend to find out.

Questionnaire

Multiple Choice questions to 5.2.4

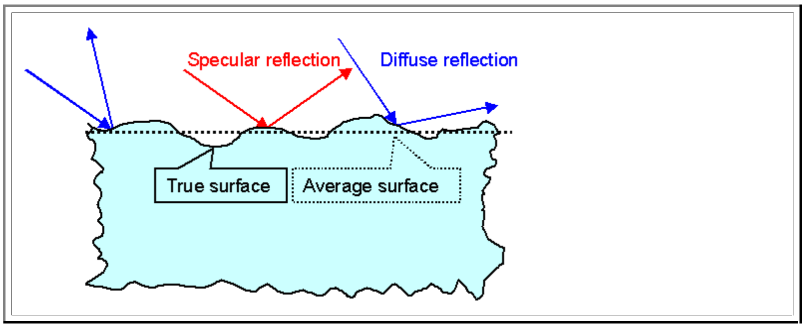
5.2.5 Not so Perfect Materials

Specular and Diffuse Reflection

- A piece of paper has a fairly flat surface but you can see it from all directions. Why? Clouds are formed by condensed water in the atmosphere and you can cleary see them radiantly white against a blue sky. Why? Before you answer, consider that you can't see the *uncondensed* water vapor even so the water concentration is about the same as in the cloud. And why, exactly, is the sky blue? Why is milk (= water with a few percent of fat) white but water with a few percent of alcohol colorless and transparent? Even if it contains far more than a few percent of e.g. alcohol?
  - The answer to the first question should be obvious. Even if the reflection law is fully valid for white paper, i.e. $\alpha_{in} = \alpha_{out}$, every pixel of it is normally hit by light coming from *all* direction so some of it will always be reflected right into your eye.
  - Sure. Now go and shine a Laser beam on a piece of paper in an otherwised completely dark room and you will *still* see the paper, no matter from which direction you look at it. Obviously some of the Laser light is still *scattered* into your eyes even so the beam reflected from its general surface goes somewhere else. Why? Because the paper is not *really* or *optically* flat. Some parts of it always reflect light in the direction you are looking.
- We are now looking at the interaction of light with matter of not-so-perfect properties, especially at matter with properties that change on a small length scale.

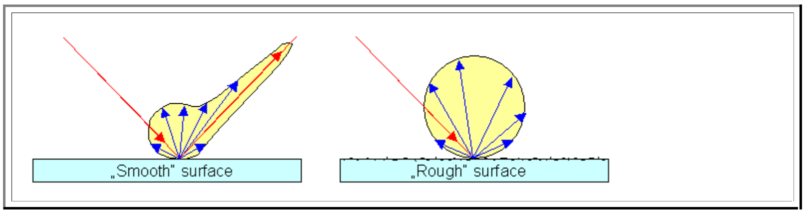
- The paper, for example, is not perfectly flat but has a roughness on some length scale far larger than the wavelength of light, but smaller than what your eye can easily resolve as shown in the picture. Some parts of it thus are always reflecting the light into your eye.
  - Condensed water vapor means you have small water droplets suspended in the atmosphere (if they're too big, they fall down and we call it rain); milk means you have some small emulgated grease particles in the water.
  - Light is *scattered* at those small things in all directions and the **scattering of light** is one topic we encounter if we look at not-so-perfect materials
- Since "small"and "large"are relative values we must use the obvious natural length scale when dealing with light as our ruler and that is of course the wave length $\lambda \approx$ **1 µm**. We then distinguish three cases depending on the length scale l_{mat} inherent to our material
  - $l_{mat} << \lambda$: The extreme case would be scattering at single atoms or molecules. Proper **nanoparticles** also belong into this group. This kind of scattering is called **Rayleigh scattering**.
  - $l_{mat} >> \lambda$: No problem, we covered that already. Just look at any part of the sample by itself, apply what we discussed before, and then add up across the sample
  - $l_{mat} \approx \lambda$: Now we have a problem. What will happen in this case is difficult to deal with and no general rules apply. This kind of scattering is called **Mie scattering**.

- Let's look at the easy stuff first and define some terms.
  - **Specular reflection** is the new word for the $\alpha_{in} = \alpha_{out}$ "normal" reflection. The only difference is that now we take the **α's** with respect to the "*average*" surface as shown below.



  - **Diffuse reflection** is reflection in all other directions. It is easily conceived as "proper" reflection from those parts of the surface that deviate from the average.

It is clear that the relative magnitudes of both reflection types depends on exactly what the surface looks like at a small scale, and that one could have all kinds of reflected intensity distributions vs. angle. It is thus convenient to characterize a surface (for a given wave length) by polar diagrams as shown below



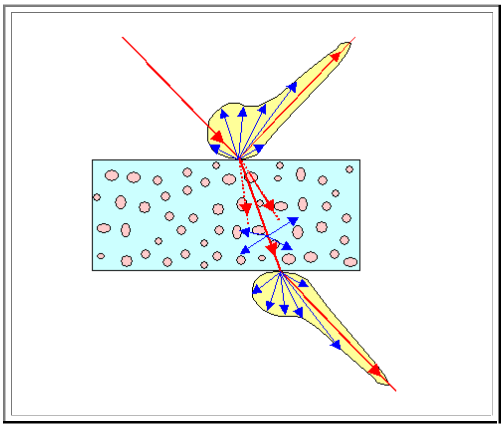
- An **"optically" smooth** surface scatters only little light outside the specular reflection direction, a perfectly rough surface scatters evenly in all directions, producing a polar diagram that can be typically described by $\boldsymbol{I = I_0 \cos\alpha}$.
- It is clear that "optically smooth" means that all deviations from the average perfect surface, i.e. all quantitative roughness parameters, must be substantially smaller than the wavelength of the light considered.
- If we take optics to the extreme, e.g. with **DUV** or **EUV** (= deep or extreme ultraviolet) lithography with wavelengths down to the **20 nm** region, we need to use mirrors for reasons discussed before. Those mirrors need to be large (numerical aperture is important) and flat to *atomic dimensions*! Not so easy to make and calibrate!

Transparency, Translucency, Opacity

Now let's look at light beams transmitted through a not-so-perfect but basically transparent material. Those materials, by definition, have a small κ value, i.e. a small imaginary part of the complex index of refraction or, same thing, of the dielectric function.

- As shown in the picture, we have

  - Specular and diffuse reflection at the surface the light is impinging on. This is descirbed by the polar diagram characterizing this surface.
  - Scattering of the transmitted light (running in different directions) at defects or imperfections contained in the material (fat droplets in milk, air bubbles in glass, ...).
  - Specular and diffuse reflection at the internal surface the light is coming out off. This is described by a (different) polar diagram characterizing this surface.



- If your material is not fully **transparent** (i.e. looking through it you hardly notice it is there; for example optical glass), light will come out in all directions and we call it **translucent**, for example etched glass or milky glass. If a lot of the light is refelcts and absorbed internally, nothing much will come out at all and we call this material **opaque**, for example china ware or "milky" ice.

Of course we can put some numbers on the properties *transparency*, *translucency*, and *opacity*, but we will not do that here.

- Suffice it to say that good transparency is a rather rare property -even for materials with intrinsically small κ and thus little absorption. Most ionic crystals and oxides, e.g. $\mathbf{Al_2O_3}$ are perfectly transparent as ideal crystals; this is also true for many polymers. Your chinaware and almost everything else made from those materials, however, tends to be rather opaque and at best somewhat translucent if it is thin.
- The reason are imperfections of surface and in the bulk of the material. Any defects not far smaller than the wavelength of the light will make their presence felt by scattering some of the light in unwanted directions.

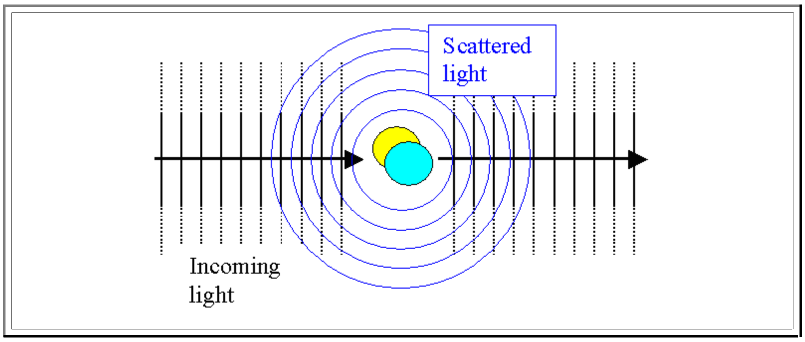
Metals are also rather opaque but for a different reason. Since their free electrons can absorb arbitrary energies and momenta, incoming photons quickly find a suitable electron that will "take" their energy and momentum, "absorbing" the photon within a few nanometers below the surface.

- Scattering or no scattering - nothing will come out on the back surface in either case.

- The picture above contains a small puzzle - did you notice?
  - The main beam is drawn nicely refracted at the air - material boundary. For doing this you need an index of refraction for the material. What is the index of refraction for an **optical compound material**? The question is reminescent to, e.g., Young's modulus for compound materials.
  - The answer, not unexpectedly I hope, is: As long as the constituents of the compond are small enough, it is possible to define a compound index of refraction that is a weighted average of the individual indexes.
  - Eaxtly how you do this might be tricky but there is nothing special here.

Rayleigh Scattering

- Why is the sky blue? Because the air **molecules** scatter blue light stonger than red light.
  - So if you don't look at the sun directly, only scattered light scattered somehow in the atmosphere can reach your eye. The more you look in a direction away from the sun, the bluer the light will be. Actually, it should be violet but because the eye is not very sensitive to the shortest still visible wave lengths you are biased towards blue.
- Atoms and molecules are far smaller than the wave length of light so we are talking Rayleigh scattering here. In contrast to Mie scattering occuring for particles with sizes comparable to the wave length, Rayleigh scattering can be treated analytically. Here we only look at the major points of this treatment.
  - The figure shows the basic situation. An incoming light beam in the form of a plane wave with some energy or intensity (per **cm^2**) is transmitted to some extent and scattered to some other extent. We assume that scattering happens evenly in all directions, symbolized by the spherical wave emanating from the atom / molecule.
  - In the particle pictures, a stream of photons with some flux density or power per **cm^2** and second hits the atom / molecule. A given photon is either not affected at all and just continues going straigtht, or deflected (= scattered) into some other directions with equal probabilities for all directions.



- If we relate the total power (= energy per second) P_{sc} contained in the scattered part to the intensity I = power per **cm^2** contained in the incoming beam, i.e. from $P_{sc}/I = \sigma$, the quantity σ has the dimension **cm^2** and is therefore called the **scattering cross section**.
  - It is easy to understand the meaning of the σ. The energy or the power contained in the incoming beam that flows through an an area of σ **cm^2** is what will be scattered out of it.
  - We might expect that σ scales roughly with the (two-dimensional) size of the scattering particles, i.e. we expect it very roughly to be found in the **10^{-19} cm^2** region.
- Why does an atom / molecule scatter light?
  - Read up chapter 3.3.3 and you know. The electrical field of the incoming wave jiggles the electrons of the particle ("electronic polarization"). Accelerated electrons (that's what jiggled electrons are) emit electromagnetic radiation (= light) with the frequency they are jiggled with, which is the frequency of the incoming light. The total effect is the scattering of the light.
  - Looking a bit deeper into the characteristics of radiation emittance of jiggled electrons (the word "antenna" comes up in this context if we look at it wiht electrical engineers eyes), one finds that the power radiated into space scales with ω^4 or λ^{-4}. Going through the math for a particle with volume V one obtains

$$\boxed{\sigma = \frac{P}{I} \approx \frac{\omega^4 V^2}{6\pi c^4}}$$

- If we have density of n particles (= atoms / molecules / whatevers) per **cm^3**, we can estimate the penetration depth or absorption length l_{sc}; i.e. the length after most of the incoming radiation has been scattered off and nothing comes out anymore, to

$$I_{sc} = \frac{1}{n\sigma}$$

- This calls for an exercise

Exercise 5.2.5

Rayleigh Scattering

A Bit More to Scattering

We have now dealt with most of the questions from above directly or indirectly:

- A piece of paper has a fairly flat surface but still you can see it from all directions. Why? Because it is far from being *optically* flat and we have *diffuse* reflection and not just *specular* reflection. On top of that we might have some Mie and Rayleigh scattering at the small and very small inhomogeneities.
- Clouds are formed by condensed water in the atmosphere and you can clearly see them radiantly white against a blue sky. Why? - considering that you can't see the uncondensed water vapor even so the water concentration is the same.. We have Rayleigh scattering at the small water droplets and since the scattering cross section scales with V^2, small water particles containing n water molecules scatter the light far more strongly than n water molecules far apart.
- Why, exactly, is the sky blue? Because Rayleigh scattering increases sharply (fourth power) with decreasing wave length. It is blue and not violet because the sun emits more blue than violet photons and our eyes are more sensitive to blue than to violet.
- Why is milk (= water with a few percent of fat) white but water colorless and transparent? Even if it contains far more than a few percent of e.g. alcohol? See above.

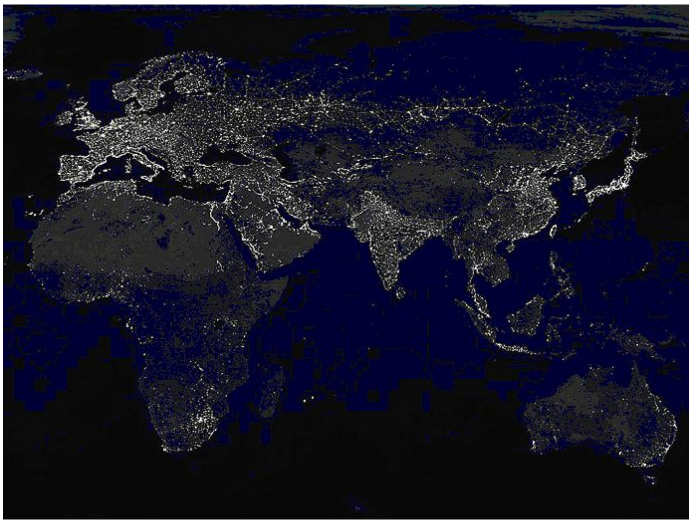
One last point needs to be made. If we consider *large* water droplets or simply a pool of water (huge droplet), it doesn't seem to scatter light very much. If the surface would be perfect (consider a perfect ice crystal for the sake of the argument) there would be some perfectly specular reflection but nothing else. *Why?*

- The reason is that for Rayleigh scattering from water droplets far smaller than the wavelengths, the electrical fields of the waves coming from each of the n water molecules of a small droplet have all pretty much the same phase. The phases of the scattered waves from randomly distributed molecules then are random, and you know what that means.), so it is small wonder that scattering increase with volume or number of molecules squared.
- If the droplets get larger we first enter the (difficult) regime of Mie scattering and we don't know off hand what we are going to see. For really large volumes, e.g. a visible ice crystal, we know what we will see. however. What has changed now is that for any wave send out by some atom at the surface via scattering on some direction other than the direction of specular reflection, some other atom at the surface produces an identical wave but with *reversed* phase. So all intensities cancel - except in the specular reflection direction.
- The transmitted light cannot be scattered at all at the atoms of a perfect crystal (or "liquid"). Only imperfections like small precipitates or voids in crystal or fat globules in milk will scatter.

5.2.6 Principles of Generating Light

General Considerations

- So far we have looked at how light interacts with matter; eventually disappearing by some kind of absorption. Here we look - very briefly - at the principles of **generating light**.
  - If we were to look at all the electromagnetic radiation there is - from long wave radio to γ rays - we would now have to start a long lecture course by itself. Even if we restrict ourselves to *visible* light plus a bit of infrared and ultraviolet, we still have a large task ahead of us.
  - In the context of this lecture course we can do no more than to enumerate major light generating principles together with a few key properties.
- Any light source will be characterized by the kind of light it produces. For that we look first at the **properties of the light produced** :
  - **Monochromatic** or **polychromatic**. In the latter case we want to know the **spectrum**. The link gives an example, it shows the spectrum of our most important light source.
  - **Spectral details**. Even for monochromatic light of wavelength λ we need to know details like the spread $\pm\Delta\lambda$ and the stability in time, i.e. (t).
  - **Emission characteristic**. Is the light emitted in just one direction (as in a Laser beam), in all directions evenly, or with some angular characteristics?
  - **Polarization**. Is the light polarized linearly, circular, elliptical or not at all (meaning that all polarization directions occur with the same probability).
  - **Intensity** or energy density. Possibly as a function of λ, emittance angle, polarization and so on.
- Being technically minded we are just as interested in **technical properties**:
  - **Power efficiency** η_{light}, telling us how many percent of the energy flowing into a light source comes out as energy of light we want.
  - **Luminous efficacy**, telling us how the eye perceives efficiency. In other words, if a green and violet light source have the same efficiency and produce the same number of photons per second that are entering your eye, the green one will have a far higher *efficacy*, appearing brighter, because the eye is far more sensitive to green than to violet light.
  - **Device lifetime**. After how many hours of operation do you have to replace your light source by a new one?
  - **Maintenance**. Is some regular service needed including, e.g., replacements of parts or re-calibration?
  - **Costs**. How much do you have to pay for the device up front? How large are the operating costs?
  - The technical properties are the more interesting ones for everyday life. For very basic research it might be the other way around.
- Roughly **20 %** of the electrical energy produced on the planet goes into light production. This is more electricity than all nuclear and water power plants produce.
  - The CO_2 produced just for illumination is about **70 %** of that from cars and three times more than that from air traffic. The picture gives an idea of what that implies.



The planet at night showing artificial illumination.

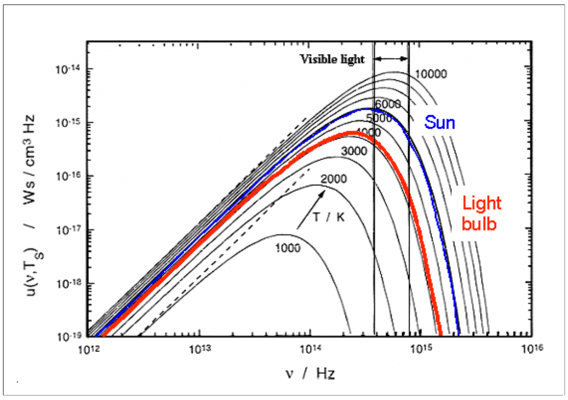
- Given the energy and climate crisis, the need for new light sources with high efficiency / efficacy is obvious.
  - Helping to save the planet in this way is one of the major jobs for Material Scientists and Engineers. Right now, and for many years to come.

Hot Bodies as Light Sources

Anything very hot emits light and if the "anything" is a "black-body" we know the spectrum emitted as a function of temperature because this is is given **Planck's** famous **equation**

$$E_\nu \cdot d\nu \quad = \quad \frac{8 \cdot \pi \cdot n^3 \cdot (h\nu)^2}{h^3 \cdot c^3 \cdot kT} \cdot d(h\nu)$$

For an elegant derivation see this link



- Light generated by hot bodies we call "**black body radiation**" or **incandescend light**, resulting from **incandescence**
- The temperature of the sun surface is about **5800 K**. Tungsten (**W**), the typical filament material of a light bulb, melts at **3683 K**. The temperature of the light emitting part of a light bulb is thus around **3000 K**. If you look at the spectrum above, one conclusion is inevitable:

Black body radiation light sources will always have a lousy efficiency

- Most of the radiation emitted is not in the frequency range of interest, and there is little you can do. Moreover, quite a bit of the energy input is wasted by simply heating the device. Ideal black body radiators at **4000 K** or **7000 K** have efficiencies around **7 %** or **14 %**, respectively. Our ubiquitous light bulb converts the electrical power ***UI*** flowing through with an efficiency of about **5 %** to light energy.
- Since a lot of the electrical energy produced goes into light, and given the current and future energy crisis, this needs to be changed presto. There is thus no choice but to employ another principle for producing light

Light from "Cold" Bodies

Hot bodies emit light because thermal energy is sufficient to move electrons up to all kinds of high energy levels E_{high} from which they transit to all kinds of lower lying levels E_{low}. Since all kinds of $E_{\text{high}} - E_{\text{low}}$ occur, the spectrum emitted covers a large wavelength region.

- If we want to make light more efficiently we have to make sure that $\Delta E = h\nu$'s available for electrons are in the visible range. Do we have examples for that?
- Words like "**Luminescence**" or "**Phosphorescence**" come to mind. Certain materials (called "**phosphors**" on occasion) have convenient energy levels for visible light production and as soon as you feed them "somehow" with energy so their electrons can populate the upper energy levels, they start to *luminesce* or *phosphoresce*. The difference between those modes is simple:

Luminescence:	**General name for "cold" light production**
Fluorescence:	**Light production shortly after energy input** **Short life time of excited level (< µs)**
Phosphorescence:	**Light production long after energy input** **Long life time of excited level (> ms)**

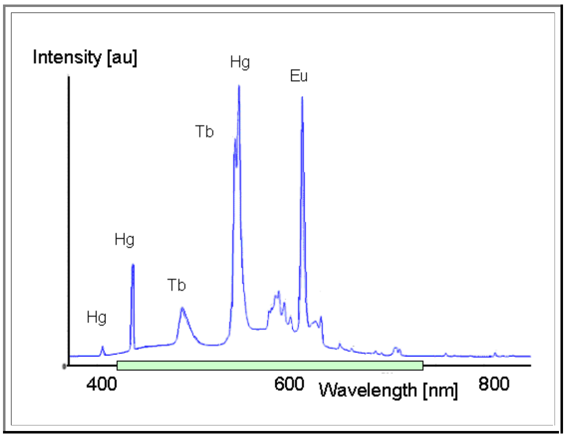
So what we want is *fluorescence*. The big question now is how we "excite" the luminescent material. In other words, how do we supply the energy flow necessary for kicking those electrons up to the proper energy levels all the time. Let's enumerate the *major* possibilities. A more complete list can be found via the link

1. **Cathodoluminescence** refers to the use of "cathode rays" or simple electrons with sufficient energy. The light generated by good old TV tubes (before the advent of flat panel displays) is generated by cathodoluminescence and so is the light from "fluorescent tubes".
2. **Electroluminescence** refers to excitation by simply running a current through the system that neither heats nor generates a plasma but moves carriers to the high energy level called *conduction band* in this case.
3. **Photoluminescence** refers to excitation by light of a somewhat higher energy than what we want to generate. Seems to defy the purpose but is nonetheless an important mechanism as we shall see.

- This link lists about **10** more types of luminescence; some with quite interesting properties.

Cathodoluminescence is the principle behind what we generally know as **fluorescent lamp** or **fluorescent tube**. We have a gas-discharge lamp and the electrons in the plasma have enough energy to excite the mercury vapor in the plasma produced. The excited mercury atoms then produce ultraviolet light that then causes a phosphor to fluoresce, producing visible light. So we do have indirectly *photoluminescence* in there as well.

- The major advantage is a good efficiency around **20 %**. A typical spectrum is shown below, the green bar marks the visible region.



- The picture makes clear why luminescence can have a high efficiency: A lot of the energy going in comes out as light with sharp frequencies in the visible range. There is no need to always produce a lot of infrared and ultraviolet light in the process.
- The disadvantages are clear, too. Mercury (**Hg**) is needed, causing environmental hazards, and the rare earths Terbium (**Tb**) and Europium (**Eu**) are called "rare" for a reason. Right now (**2011**) China controls around **90 %** of the rare earth market and what that means is that prices are bound to go up in years to come. You also have to condier that teh ligh tmay appear white because it haas trhe right moc of wavelnegth but that its spectrum is rather different form that of the sun or any other black body radiator.
- Using *direct* photoluminescence as light source is of course (?) ridiculous so we won't discuss it any more.

This leaves us with ***electroluminescence*** or, to use another word for essentially the same thing, *radiant electron - hole recombination* in semiconductors. In yet other words: I'm now turning to **LED**s as light sources.

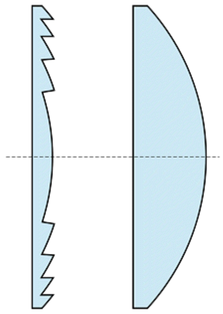
- This can be done with very large efficiencies. We are talking the future of lighting here.
- How it's done we will see briefly in chapter 5.3. Otherwise use these links
  - **LED's simple**
  - **LED's involved**

5.2.7 Specialities

Before we move on to more technical optics, we need to consider a few special effects that are of some importance to optical apparatus and not yet directly covered like (Rayleigh) scattering or indirectly like diffraction gratings. The list could be rather large, we restrict ourselves to a few effects here, however

Fresnel Lens

- Recalling that a lens "bends" the light rays impinging on it only at the air - lens interface while nothing much happens inside the glass, **Augustin-Jean Fresnel**, whom we encountered before, had the great idea to turn this principle into practical use. Here is what a Fresnel lens looks like; the principle is obvious:

	 Pictures from Wikipedia
Principle of Fresnel lens	Large Fresnel lens in lighthouse

- It wasn't so easy way back to shape glass like that. Nowadays it is easy, especially if you use plastic material and just press it into form. Fresnel lenses are no all over the place - as long as the ultimate in resolution doesn't matter. The "tips" at the singularities scatter light to some extent. and if you make the structure too dense, you run into diffraction effects at the periodic structure.

Optical Activity

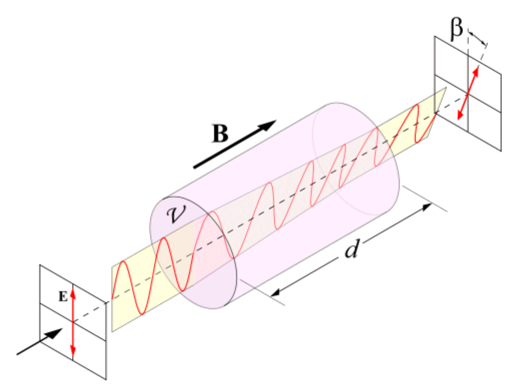
- Optical Activity generally is a type of birefringence. It describes materials that rotate a linearly polarized wave as it travels through the material. All materials with some chiral symmetry (i.e. where you find some kind of "spiral" inside show this behavior (provided, of course, they are sufficiently transparent)
- The typical material is quartz but its internal spiral is not so obvious., So take DNA, where the spiral is obvious. Run light through such a material and its polarization vector either turns clockwise ("right") or counterclockwise ("left"), depending on how the internal spirals are "wound"
- Do not confuse optical activity with circular polarization, In both cases the electrical field vector rotates but for circular polarized light a full rotation occurs in the space of a wavelength, where it might take many mm of active material for that.

There are plenty of applications:

- Most prominent, perhaps, is what "**Polarimeters**" do. The measure how much a given substance at controlled conditions rotate light. This is very valuable in biology or medicine because most "biological" molecules are optical active - but invariably only "leftish". This means that all biomolecular spirals are always wound in just one way. Amazing! Think about that. So be careful when that sexy alien arrives. If she or he is wound the other way, all kinds of trouble will ensue on close contact. If you measure the amount of rotation in a controlled experiment, you can derive the concentration of the active molecules.
- Next we note that in the presence of magnetic fields, *all* molecules have optical activity. This is called the *Faraday effect* and is treated right below.

Faraday effect

- The picture shows it all. An incoming linearly polarized beam turns its polarization direction by an angle given by $\beta = bd \cdot V$. The **Verdet constant** V contains the material property as some number in minutes of angle per Oersted/cm (nobody really working in magnetics uses **SI** units). We have for example: V(diamond) **= 0,012**; V(glass) ≈ **0,015 - 0,025**; but crystals like terbium gallium garnets might have far larger values.



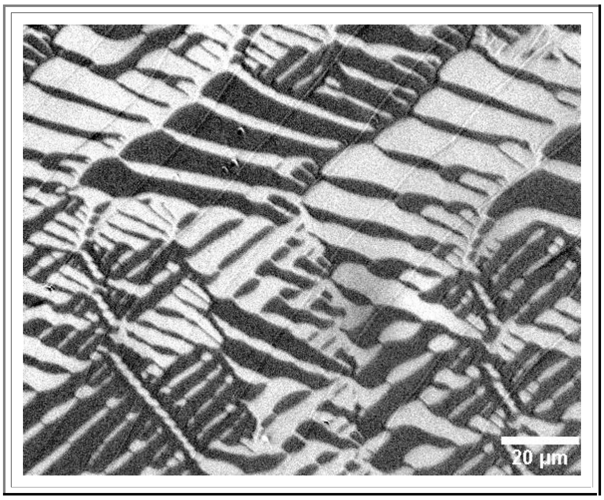
Picture from Wikipedia

The obvious thing to do now is to derive the equations from above and to calculate the Verdet constant. Well, *forget it*. Just to give you an idea of what it would involve:

- What we would need to do is to go over the driven damped harmonic oscillator once more (we looked at that in some detail in the context of dielectric polarization mechanisms). But now we must add the force that the magnetic field exerts on the charge. Instead of just $-e\mathbf{E}(\omega)$ as (oscillating) driving force we have now $-e\mathbf{E}(\omega) - e \cdot (\mathrm{d}x/\mathrm{d}t \times \underline{\mathbf{B}})$. This will change the dielectric susceptibility χ resulting from the solution - it will turn it into a tensor! It's not difficult to see why: The magnetic field produces a force that is at right angles to the principal axis of motion. The problem is thus no longer one-dimensional and the amplitude and thus the polarization will have components in other than the ***x***-direction.
- In other words: An electric field in ***x***-direction applied to a cubic crystal will only produce polarization on the crystal surface perpendicular to the ***x***-direction, let's say the **(100)** sides, and it doesn't matter if the field oscillates or not for that. The dielectric susceptibility is proportional to the oscillation amplitude with the known frequency dependence. Now add a magnetic field - static or not - and the other surfaces of the cube will now become charged or polarized, too. A simple consequence is already that in a static magnetic field *all* materials are anisotropic because the polarization vector is no longer parallel to the electrical field vector.
- What is the Faraday effect good for? Well, you can measure magnetic fields with it. In particular if those fields are far away like in the next sun system or galaxy. Astrophysics is very big on using the Faraday effect but in terrestrial terms it is not used very much.

Kerr Effect

- There are *two* Kerr effects:
    - The **magneto-optic Kerr effect** (***MOKE***). It describes how an applied *magnetic field* changes the *reflection* from some material.
    - The **electro-optic Kerr effect**, also called the *quadratic electro-optic effect* (***QEO*** effect), describes the change in the *refractive index* of a material in response to an applied *electric field*.
- If a magnetic field is turned on (or is present because the material is ferromagnetic), the polarization and the intensity of reflected light might be different from what on would calculate with the Fresnel equations. How exactly depends on the material and the components of the magnetic field on the surface
- The magneto-optic Kerr effect is identical to the *Faraday effect* from above but deals with the reflected and not the transmitted light. It is very useful because it allows to image magnetic domains. The link shows examples and here is one from the Uppsala University / Materials Physics:



The electro-optical Kerr effect describes an electrical field induced index change that is proportional to the *square* of the electric field; i.e. the *intensity*. If the index changes linearly with the field, the effect is called *Pockels effect*, see right below.

- All materials show this *quadratic electro-optic effect* but certain liquids display it more strongly than others. The electro-optical Kerr effect can and has been used to modulate light because it still works at very high frequencies.
- There is an fascinating emerging special use. If one does not apply the electrical field by conventional means (i.e. by applying a voltage across two electrodes) but by (intense) light, all kinds of strange things can happen. One needs strong electrical field not usually found in normal light (look at exercise 5.1.-2). So we use an intense Laser beam. The resulting (always nonlinear) optical effects go by names like self-focusing, self-phase modulation, modulational instability, and Kerr-lens modelocking. Whatever that might be (google it), we can expect to find applications based on that in the not too distant future

Pockels Effect

- The Pockels effect produces birefringence in an optical medium by applying a constant or varying electric field. It is distinguished from the Kerr effect above by the fact that the birefringence is proportional to the electric field, whereas in the Kerr effect it is quadratic in the field.
- The Pockels effect occurs only in crystals that lack inversion symmetry, such as lithium niobate or gallium arsenide or special polymers or glasses, while the Kerr effect occurs in all materials.
- Concerning applications one only needs to consider that optical communication via glass fibres depends on **Pockels cells**. Modulating light intensities with GHz or even THz simply cannot be done otherwise.

Concluding Remarks:

1. Above several effects are listed where electric or magnetic fields superimposed on some optic material does "something" to light. There are even more of those "named" effects not shown here, and they all have one thing in common:

- All that light can do inside material is to change what the electrons are doing. Since any change of what an electron is doing involves a change of it state, only electrons at the Fermi edge can do something, as we have asserted many times before.
- What a simple oscillating electrical field will do (i.e. a light wave) we have already treated with a simple mechanical model, when we discussed the frequency dependence of polarization mechanisms. Straight away we got the complex dielectric function and thus also the complex index of refraction.

All we need to do to cover *all* the effects of electric and magnetic fields on optics is to include the forces on the electron they add *in addition* to the force produced by the electric field of the light wave. The equation above for the Faraday effect hints at that. The new equations, however, are not so easy to solve any more. What we generally find is

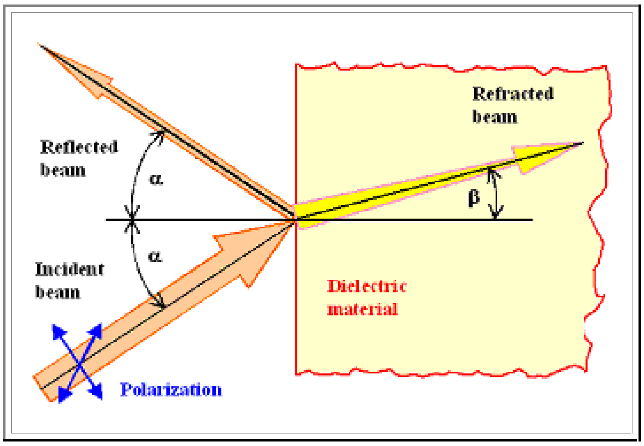
  - In the case of a isotropic material it now becomes anisotropic. Dielectric constant ϵ_r, susceptibility χ and index of refraction n (all simply related as you know) become tensors.
  - Those tensors may have imaginary components.
  - Any light wave can always be decomposed into two light waves with orthogonal polarization. They will travel through the material now with different wave vectors . We have two beams now.
  - In the case of an anisotropic material things get really complicated
  - In the case of high field strength of the extra field, things get also non-linear.

- So we have been lucky. Even if you hate the mathematical complexity, you must appreciate that a lot of useful effects result that are not even remotely exhausted at present. That's where some of you might find useful (and well paid) work to do!

- If you look at the applications listed above, one very major application is conspicuously missing. Close your eyes and think. For which product that *you* use daily is it absolutely essentialy that you can influence the transmittance of light by an electrical field?
  - Right - liquid crystal displays (***LCD***) or practically all computer monitors, many TV's and plenty of cell phone displays. An electrical field between the transparent electrodes) of a pixel change the polarization properties of the liquid crystals inside that capacitor and those how much light is transmitted. Are we using the electro-optical Kerr effect here? Or the Pockel effect? Personally, I don't know for sure. But I know that it doesn't matter. It's time to leave the past behind and look at the general picture and not at enumerated experimental descriptions of effects observed some 100 years ago.

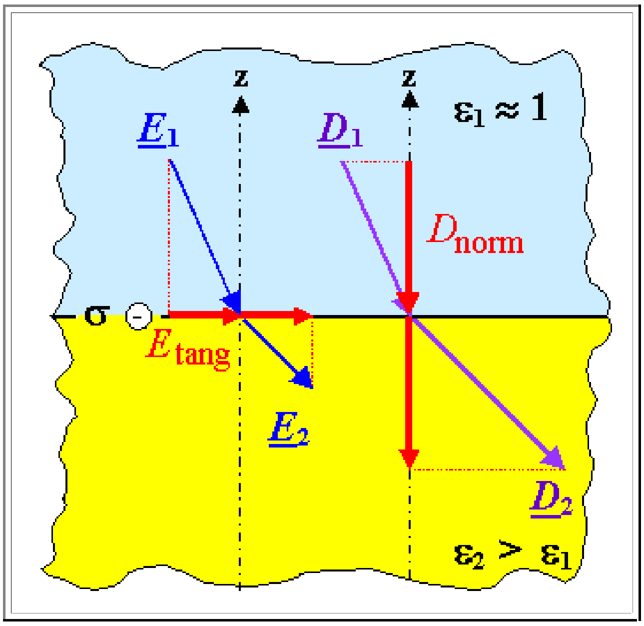
5.2.8 Summary to: Optics and Materials

- **The task:**
  Calculate and understand intensities, angles, phases, polarization and attenuation (damping) of the various light beams shown from the materials properties
  - Still assuming a perfectly flat surface



- First step: Decompose impinging light into two waves with polarization in he interface plane (**TE** case) or at right angles (**TM** case)
  - Energy conservation yields for the intensities:

$$\boxed{I_{tr}(z=0) \;=\; I_{in} - I_{re}}$$



- Boundary conditions as shown in the figure involve the "dielectric constant ϵ and thus the so far only relevant material property.
  - Considering energy (proportional to E^2) and momentum (proportional to k^2") conservation for the TE and TM case separately yields the **Fresnel equations** that provide the answers to the questions above
  - A wealth of insights and relations follow, e.g. or field strength ***E*** or intensities *I*:

$$\frac{E_{ref}}{E_{in}} = -\frac{n-1}{n+1}$$

$$\frac{I_{ref}}{I_{in}} = \left(\frac{n-1}{n+1}\right)^2$$

  - one consequence as example for the power of these equations: $n = 2$ means that almost **10 %** of the intensity will be reflected, implying that for optical instruments you *must* provide some "anti-reflection" coating.

- Using the complex (and frequency dependent "dielectric constant **ϵ(ο) = ϵ' + iϵ"** yields the **complex index of refraction**

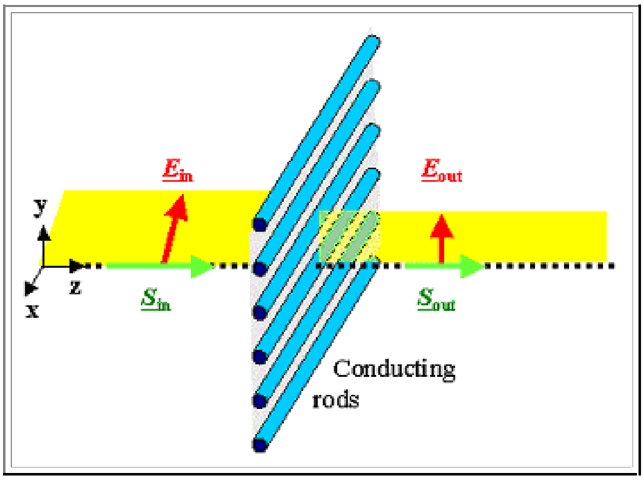
$$\boxed{n^*(\omega) \;=\; n + i\,\kappa}$$

$$n^2 = \frac{1}{2}\left(\left(\epsilon'^2 + \epsilon''^2\right)^{1/2} + \epsilon'\right)$$

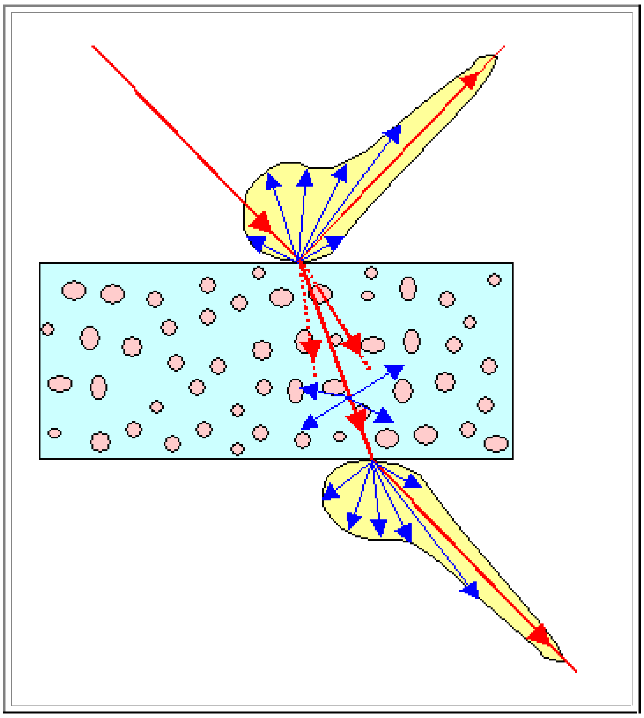
$$\kappa^2 = \frac{1}{2}\left(\left(\epsilon'^2 + \epsilon''^2\right)^{1/2} - \epsilon'\right)$$

  - The imaginary part κ describes the attenuation (damping) of the transmitted wave in the material.

- **Polarization and Material**2. How to polarize a light beam
  - **1. Geometry**. Use Fresnel equations to produce a polarzed beam under specific angles ("Brewsater angle")
  - **1. Polarization foils** = alined conducting rods (of possibly molecular size) "short-circuiting" the electrical field in on direction.
  - **3. "Tensor" materials** with optical anisotropy
- Theory can an get rather involved; products can be extremely simple and cheap (e.g. circular polarizer in 3-D movie glasses)



- **Not so perfect materials** and properties like specular and diffuse Reflection, transparency, Translucency, Opacity.
  - Light is scattered at small things in all directions and the scattering of light is the major topic encountered if we look at not-so-perfect materials
  - The picture illustrates:
    Specular and diffuse reflection at the surface.
    Scattering of the transmitted light (running in different directions) at defects or imperfections contained in the material (fat droplets in milk, air bubbles in glass, ...).
    Specular and diffuse reflection at the internal surface the light is coming out off. This is described by a (different) polar diagram characterizing this surface.
- Scatter mechanism depend on the size l_{sca} of the scatterer" relative to the wavelength:
  $l_{\text{sca}} << \lambda$: The extreme case would be scattering at single atoms or molecules. Proper **nanoparticles** also belong into this group. This kind of scattering is called **Rayleigh scattering**
  $l_{\text{sca}} >> \lambda$: No problem, we covered that already. Just look at any part of the sample by itself.
  $l_{\text{mat}} \approx \lambda$: Now we have a problem. What will happen in this case is difficult to deal with and no general rules apply. This kind of scattering is called **Mie scattering**



- **Generating Light**
  Two basic cases:
  - Light from **hot bodies**. Planck radiation law applies. Efficiency tends to be low
- Light from **"cold" bodies** or luminescence
  - There are many types of cold light production. Of utmost importance is electroluminescence or, to use another word for essentially the same thing, radiant electron - hole recombination in semiconductors.
    In yet other words: It's the LED, the **light emitting diode**..

Luminescence:	**General name for "cold" light production**
Fluorescence:	**Light production shortly after energy input. Short life time of excited level (< µs)**
Phosphorescence:	**Light production long after energy input. Long life time of excited level (> ms)**

Specialities

- Forget it. The list names some, there are many more.

That's where serious "optics and material" starts. This would need another full lecture course

- **Fresnel Lens**
- **Optical Activity**
- **Faraday effect**
- **Kerr Effect**
- **Pockels Effect**

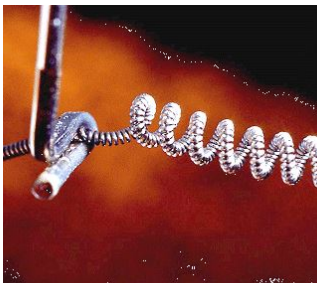
5.3 Optical Components

5.3.1 Light Sources

Conventional Light Sources

- There is not much to say about conventional light sources like simple light bulbs, "halogen" light bulbs, gas-discharge sources and so on. You all are quite familiar with them. What follows gives the bare essentials.
    - **Thomas Edison** is usually credited as inventor of the light bulb in **1880** but there were many others working on "light bulbs" as early as **1840**. Edison's breakthrough probably was due to a combination of three factors: an effective incandescent material, a higher vacuum compared to others, and a high resistance that made power distribution at high voltages from a centralized source economically viable.
    - Consider a **100 W** bulb operated at **230 V**. It draws **0,44 A** and thus has a resistance of **227** Ω. This is not easily achieved with the metal wires then available. Edison of course, used carbon. It took until about **1905** before tungsten (**W**) filaments were used and until about **1913** before an inert gas like $\mathbf{N_2}$ was inside the bulb instead of vacuum.
    - In fact, present day light bulbs are high-tech objects despite their lowly image. If you have doubts about this consider: How would you make a "coiled coil" filament as shown below for a standard **1 €** light bulb a from an extremely hard to shape material like **W** in such a way that it is extremely cheap?

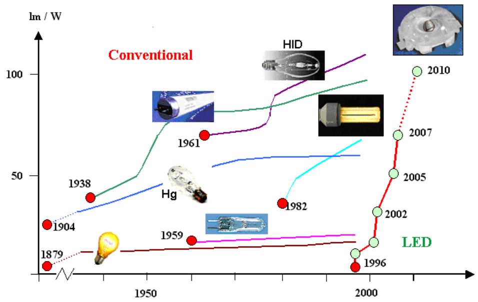
Edisons light bulb

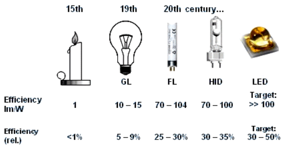


Modern double coiled **W** filament

Source: Wikipedia

- It is hard for us to imagine the impact of "easy" light on humankind. Nevertheless, the **120+** years of illumination by **incandescent light** has to come to an end *right now* for reasons already given
- Fluorescent and gas discharge light sources have better efficiencies (and efficacies) than "black body radiators" but are not without problems of their own.
    - The pictures tells it all, just look at the LED branch. No more needs to be said about "conventional light sources".





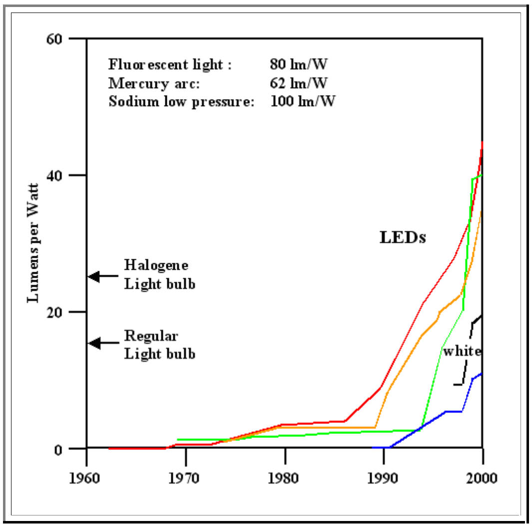
	15th	GL	FL	HID	LED
Efficiency lm/W	1	10 – 15	70 – 104	70 – 100	Target: >> 100
Efficiency (rel.)	<1%	5 – 9%	25 – 30%	30 – 35%	Target: 30 – 50%

HID = "high intensity discharge" light bulb; the "Xenon" light in your more expensive car.
FL = "fluorescent light"
Hg = "mercury vapor lamp"
GL = "Glühlampe " (Glowing light); light bulb
LED = light emitting diode

Data from Osram

Light Emitting Diode

- **Light emitting diodes** or ***LED***'s nowadays come in two variants: "Standard" **LED's** made from *inorganic* crystalline semiconductors based on, e.g., **GAAlAs**, **GaP** or **GaN** and **"organic" LED's** or ***OLED***'s.
  - **OLED** devices are coming into their own right now (2011). They are not yet mass products for general lightning applications but we will find out how far they will go in the near future (based on the work of possibly you and other materials scientists and engineers; who else?).
  - Standard **LED's** have been around for more than 40 years by now. However, they used to be only red in the beginning, see the picture below, and their efficiencies were lousy. The breakthrough came around **1990** when **Shuji Nakamura** of Nichia Corporation almost single-handled introduced the **GaN** based blue **LED.** This started the ongoing revolution of world wide lighting that will contribute in a major way to saving the planet from the climate crisis. Of course, if you google "Nakamura" you will find a soccer player first.
- The picture below gives an idea of what was happening. Nobody seem to have updated this picture but the trends continued. The **LED** market is growing rapidly



  - In analogy to "Moore's law", "**Haitz's Law**" has been proposed: In every decade, the cost per lumen (unit of useful light emitted) falls by a factor of **10**, the amount of light generated per LED package increases by a factor of **20**, for a given wavelength (color) of light. Haitz also predicted that the efficiency of **LED**-based lighting could reach **200 lm/W** (lumen per Watt) in **2020** crossing **100 lm/W in 2010**.
  - **This is important:**

More than **50%** of the electricity consumption
for lighting or
20% of the totally consumed electrical energy
would be saved reaching **200 lm/W**

  - So get going, young Material Scientist!
- What does on need to do to make better (and cheaper) **LED**'s?
  - As a first step you must learn a minimum about semiconductor physics or Halbleiterphysik and semiconductor technology.
  - The links provide starting points because *we are not going to do that here*.

Laser

All the light sources discussed so far share certain broad characteristics:

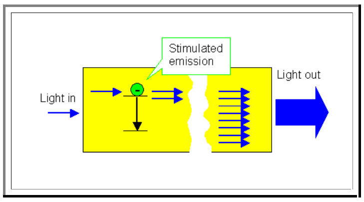
- They emit either a whole *spectrum*, i.e. light with many colors, several spectral lines, or in the case of **LED's** only one line but with a rather large half-width.
- Their light may come from a small area ("point source"; e.g. standard LED), from a longish area ("fluorescent tubes") or even from a large area (OLED's) and cannot really be processed into that *parallel beam* always used for illustrating optical stuff
- The light is emitted in *many directions* with various characteristics but *never* in only one direction.
- The light is never fully coherent and mostly *rather incoherent*.
- The light is mostly not *polarized*

Negate everything in that list (except, maybe, polarization) and you have a *Laser*, a device that operates on the principle of **L**ight **A**mplification by **S**timulated **E**mission of **R**adiation.

Lasers are rather recent light sources; the first one was built by **Maiman** in **1960**; for a short history use the link

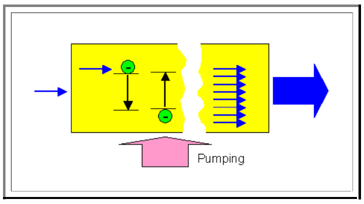
- We cannot go much into the principles of Lasers here. We only look at a few basic concepts and keywords..

The name "LASER" says it all. To understand the very basic principles of Lasers, we look at a sequence of a few simple pictures



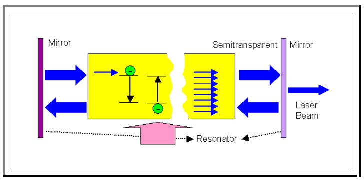
First we need **L**ight **A**mplification. For that we need a material with two suitable energy levels, $\Delta E = h\nu$ apart. Light results whenever the electron jumps from the higher level to the lower (ground) level one with a basic frequency of ν Hz. Note that this is not true for just any levels; the electron may get rid of its energy in other ways, too, e.g. in indirect semiconductors.

- Second we need **stimulated emission**, a phenomenon that was calculated and predicted by **Albert Einstein** in **1916**. In simple terms, stimulated emission means that a photon with the energy ΔE, when encountering an electron sitting on the upper energy level, stimulates it to "fall down" and to emit a photon that is identical in wave vector, and phase to the one that stimulates the process (and does not get absorbed!)
- Instead of one photon we have now two identical one. We have achieved **light amplification**. The two photons now stimulate other electrons along their way to produce more photons, all being *fully coherent.*. A lot of light now merges from the output.

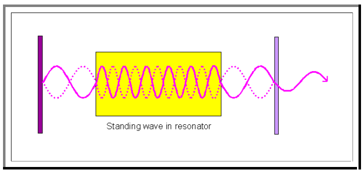


The process from above, however, only works once - until all electrons that happens to populate the upper energy level are "down".

- For a material with a dimension of $1 = 1$ **cm** this takes about $t = c_{mat} // \approx [1/(2 \cdot 10^9)]\,\mathrm{s} = 0{,}5\,\mathrm{ns}$, so we would have a rather short light flash.
- or a "***cw***" or **continuous wave Laser** we obviously need to kick the electrons up to the higher energy level - just as fast as they come down - by "**pumping**" the Laser. In fact, we need to have more elctrons sitting at the high energy level all the times than at the lower level. This is a very unusual state for electrons called **inversion**.
- *Pumping* requires that we put plenty of energy into the system all the time. This can be done by intense illumination (obviously with light of somewhat higher energy than ΔE). Some Lasers of the US military were supposed to be pumped by **X**-rays produced by a nuclear explosion (no joke). They would not live long but still be able to produce a short-lived ultra-high intensity Laser beam suitable for shooting down missiles.
- Our cheap, simple and long-lasting semiconductor lasers, in contrast, are "simply" pumped by running a *very large* current density (**> 1000 A/cm^2**) through a suitable **pn**-junction in some *direct* semiconductors. This link gives an idea of what that means.
- Note that the incoming photon could just as well kick a lower electron up, than it would be **absorbed**. The photon generated at random some time later when the electron moves back down again is *not* adding to the desired output, it just adds noise.

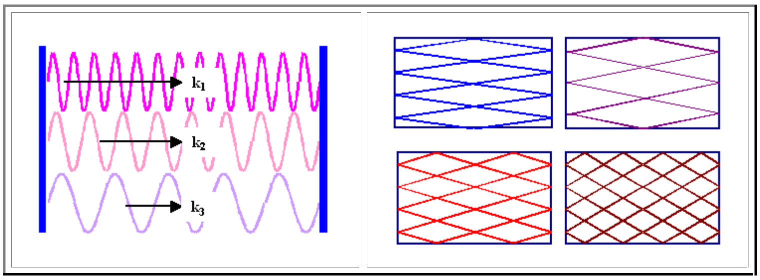


- We are not done yet. The picture above are greatly simplified because in reality we would produce light beams running in all kinds of directions. That's not what we want.
  - Just as important, the energy of the light produced would not be exactly ΔE but, roughly, $\Delta E \pm kT$ since our excited electrons would also have some thermal energy. For a good monochromatic light, an energy or frequency spread of about **1/40 eV** at room temperature is ridiculously large, so we must do something.
  - What we do is putting the pumped material inside a "Fabry Perot" **resonator**. This is nothing else but two mirrors (one with a reflectivity less than **100 %**, i.e. "semi" transparent) that are exactly parallel (within fractions of a **µm**) and at a distance ***L*** from each other.
  - The light generated then is reflected back and forth. For reasons clear to us now, only waves with $\lambda = 2L/m$; $m = 1,2,3,...$ will "fit".
  - A certain part of the light impinging on the "semi" transparent mirror leaks out, forming our .now fully monochromatic and coherent Laser beam.It propagates in one directiononly (here perpendicular to the mirrors).



- The way to visualize that is shown here.
  - We have one standing wave right between the two mirrors. Note that the wave length in the material is different from that in air; you must take that into account when going through numbers.
  - Note that the picture for an organ pipe with an acoustic wave inside would, in principlelook exactly the same. The pipe would leak some of the wave and you hear a tone with a well defined frequency.
  - This looks pretty involved, so how come that we have ultra-cheap Lasers in DVD drives? Because you don't need extra mirrors, you just use the internal surface of your semiconductor single crystal that reflect parts of the beam according to the Fresnel equations. If you obtain those surfaces by cleaving down a low-index plane, they are automatically exactly plane parallel. That makes Lasers more *simple*.
  - However, typically Lasers are far more complicated than shown here

- An organ pipe or any longish musical instrument will not only produce a tone with one frequency ν_0 but also the harmonics or overtones $\mathbf{m} \cdot \nu_0$. Same for our Laser, of course, as shown below left.



  - A musical instrument that isn't long and slender like an organ pipe or a flute (i.e. an essentially 1-dim. system) but a rectangular box (or a complex-shaped body shape like a violin, can contain standing waves in all directions with many possible wavelengths. Same for our Laser; cf.. the situation in the figure on the upper right.
- So depending on the exact shape of the laser, the way it's pumped, and so on and so forth, there can be more than just one **Laser mode**
  - We needed to get to that word. so let's repeat: There can be more than just one standing wave inside a Laser resonator, or a real laser might emit more than just *one* mode.
- We will not discuss what kinds of Lasers we find for all kinds of applications here. There is a bewildering variety and more and more different kinds are introduced. We just note one important item:
  - Increasing the frequency / energy of Lasers becomes "exponentially" difficult because with increasing photon energy the number of ways it can be absorbed increases rapidly (there are lot of empty states far above some densely populated ground level onto which electrons could be "kicked") but only one state is useful for lasing!
  - That's why there aren't so many **UV** Laser around and no **X**-ray Lasers yet.

5.3.2 Processing Light

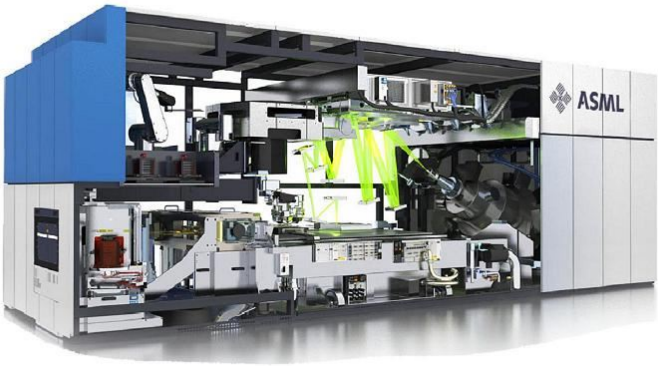
Conventional Lenses, Mirrors and Prisms

Little needs to be said about the Lenses, Mirrors, and Prisms. The basics have been covered before, here we just look at a few specifics to illustrate a few additional points

- Below are two pictures that demonstrate what one can do with lenses and mirrors. They are, after all, still the most important components of most optical systems

Here is a "lens" for a modern production "stepper" i.e. the machine that projects the desired structure onto a light sensitive layer (the resist) on a Si wafer. Steppers are crucial for making **Si microchips**. Here is the link for details.
The producer of this lens is Carl Zeiss SMT AG , Germany.

Although designed for manufacturing on a nanoscopic scale, a lithography stepper lens is not small. The Starlith 1900i weighs more than a metric ton, stands several feet tall and is as big around as a tree trunk. A catadioptric lens consisting of reflecting mirrors and refractive optics enables volume semiconductor production at **40 nm** resolution.
The stepper lens has a numerical aperture of NA = **1.35** (*huge!*) and is intended for use in immersion lithography, a technique that replaces the air gap between wafer and stepper with water or another fluid. Zeiss notes that the device is, in some sense, the *end of the art*.
The lens is designed for an **UV** light source with a wave length of **193 nm**, which needs an **ArF** (yes, "Argon Fluoride") Laser. Remember that the visible spectrum ends around **400 nm**. **UV** radiation from the sun, for comparison, spans the range (**315** - **400**) **nm** ("Ultraviolet A") and (**280** - **315**) **nm** ("Ultraviolet B").

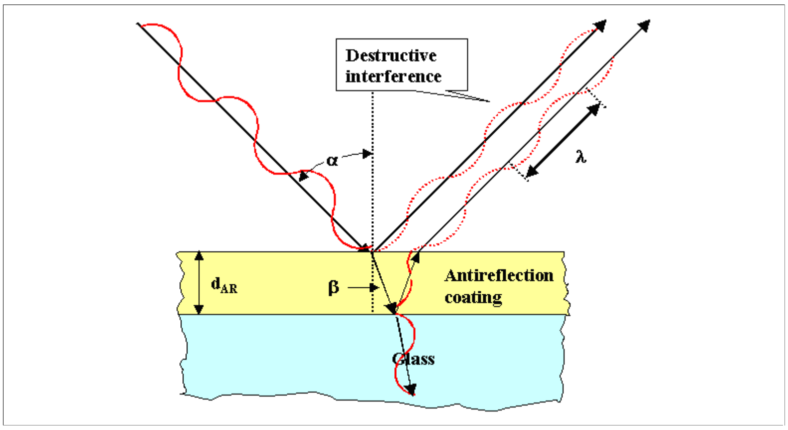


That is the kind of "projector" that will move into chip production after lenses have reached the *end of the art*. It needs to operate in vacuum because air starts to absorb light below around **185 nm**.

The " NXE:3100" lithography machine from ASM corporation, employs extreme-ultra-violet (EUV) light with a wave length of **13,5 nm** to provide an imaging capability close to 20 nm. EUV will enabke 27 nm resolution down to to below 10 nm eventually. How one makes an intensive 12.5 nm deep-UV light source is a rather interesting topic in its own right.
The "optics" is based entirel on mirrors - with essentially atomically flat surfaces .
The machine shown is a kind of pre-production prototype that is presently (2011) tested.
It is essentially still a slide projector but a bit more expensive (around (10 - 20) Mio € would be my guess).

Note added July 2021:
As it turns out, I was a bit conservative in my coist estimation. You can buy the "most complicated machine" (New York Times) now for about 150.Mio $

- From a Materials Science and Engineering point of view, making those machines is a big challenge but nothing more shall be said about them here.
- It goes without saying, however, that for any non-trivial system employing lenses you need **anti-reflection coatings**; check your Fresnel equations!
  - You also need anti-reflection coatings for solar cells (reflected light cannot produce electricity) and in numerous or better almost all "optical" devices. .
    Here is the (known) working principle of a simple antireflection coating (for one wavelength and angle of incidence). Since the two beams reflected from the surface and the interface cancel each other exactly because of the **180°** phase difference, the incoming beam must go into the material in its entirety. If you noticed the little paradox contained in this statement, activate this link.



Schematic view (don't look too closely at phases) of the working of an antireflection coating.

  - Of course, if you want to minimize reflection for a whole *range* of wavelengths and angles of incidence, you have a problem. The answer to the problem, as ever so often in Materials Science is: compromise! Achieving perfect antireflection for those conditions is next to impossible or at least expensive.

Polarizers, Diffraction Gratings and Filters

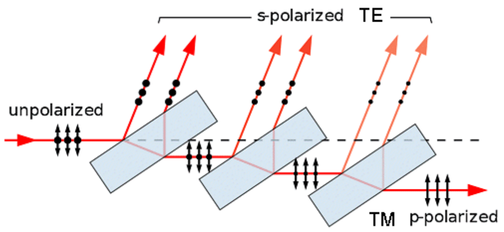
- Now that we have lenses and mirrors covered, we need polarizers, diffraction gratings, and filters next.
- We have already covered a lot of ground with respect to **polarization** and encountered some ways to produce a polarized beam. There are two basic ways to achieve **linear polarization**:

Absorbing polarizers: the unwanted polarization is absorbed
Beam-splitting polarizers: the unpolarized beam is split into two beams with opposite polarization states.

  - Absorbing linear polarizers are essentially of the "array of conducting rods" type as outlined before. Not much more needs to be said here.
- **Foil polarizers** of this type are used most of the time- whenever utmost quality is not the concern. They are essentially based on Lang's old invention.
  - Instead of (birefringent) herapathite crystals embedded in a stretched plastic foil, we now use aligned (again by stretching) polyvinyl alcohol (***PVA***) foils and dope the molecules with Iodine. In other words, we produce a more or less conducting polymer in one direction. Polarizing foils of this type are most common type of polarizers in use, for example for sunglasses, photographic filters, and liquid crystal displays. They are also much cheaper than other types of polarizer.
  - A modern type of absorptive polarizer is made of elongated silver nanoparticles embedded in thin (≈ **0.5 mm**) glass plates. These polarizers are more durable, and can polarize light much better than plastic Polaroid film, achieving polarization ratios as high as 100,000:1 and absorption of correctly-polarized light as low as 1.5%. Such glass polarizers perform best for short-wavelength infrared light, and are widely used in optical fiber communications.
- **Beam-splitting polarizers** come in many varieties and two basic types:

1. Use simple materials and employ the Brewster angle.

- The reflected light then will be linearly polarized (**TE** case) because the **TM** components of the incident light are not reflected at all (figure it out yourself!)

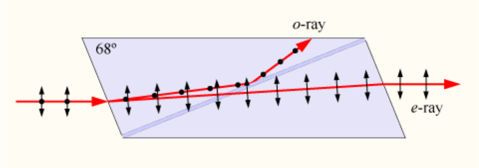


The quality improves with the number of reflections but the intensity goes down.

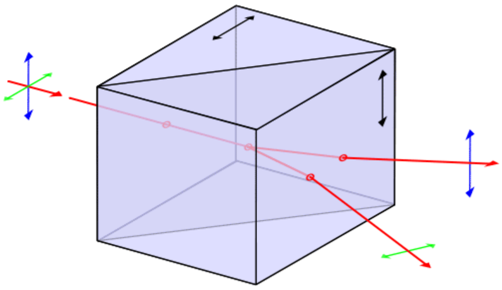
- Doable but not very elegant. Think about using this method for sunglasses or **3-D** glasses. On the other hand, if you need to polarize in the deep **UV** or **IR**, it might be your only choice.

2. Use **tensor materials** or in other words effects like birefringence.

- Use birefringence, e.g. in the form of a **Nicole prism**, **Wollaston prism**, or a number of other "Prisms".



Nicol prism



Wollaston prism

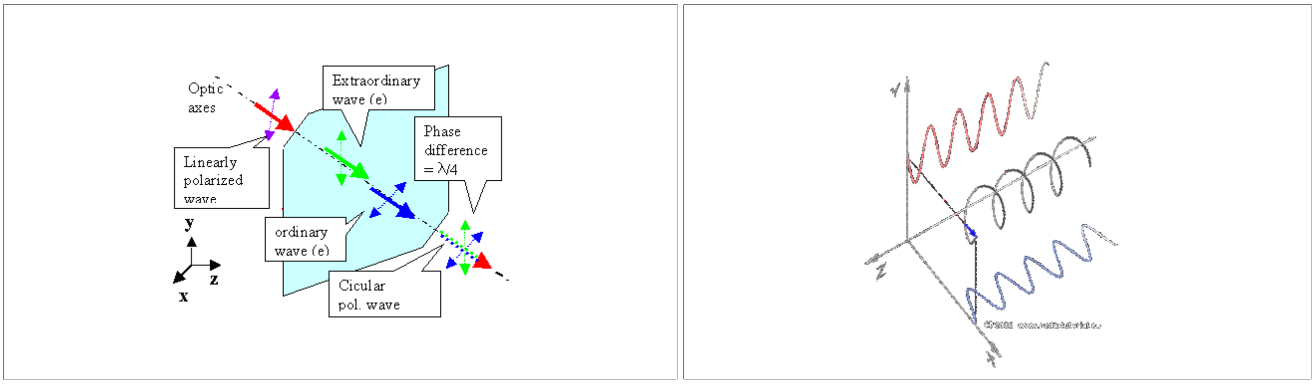
- All those "prisms" use birefringent (or tensor) materials, typically the easy to get or make calcite **$CaCO_3$**. The incoming beam splits into an ordinary and extraordinary beam that can be fully polarized. In the Nicol prism the geometry is chosen in such a way that the extraordinary beam undergoes total reflection at the interface where the two parts of the crystal are joined. The ordinary beam is not only full polarized but continues in the same direction as the incoming beam. The Nicole prism is therefore relatively easy top use in optical equipment.

Achieving **circular polarization** is also "easy" in principle. All you need is a linear polarizer and a "**quarter wave plate**".

- A "quarter wave plate" is a (typically thin) piece of material, where a polarized beam goes in, and two beams come out with the following properties:

  1. They two beams are linearly polarized with polarization directions perpendicular to each other
  2. The two beams have equal intensities
  3. One beam is phase shifted by exactly a quarter wave length (**λ/4**) with respect to the other.

- The two waves thus produced superimpose to a circular polarized wave as shown below:



The incoming beam is polarized at **45°** relative to the principal polarization directions in the anisotropic medium. The two beams exciting the medium produce a circular polarization as shown in the animation (© Wikipedia)

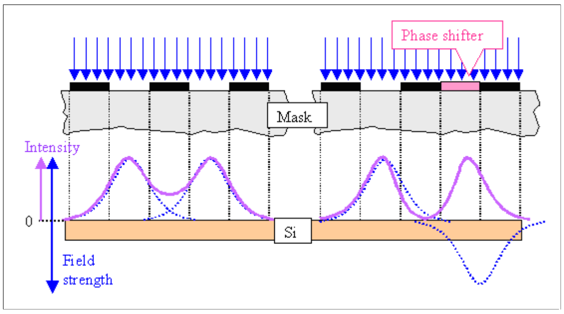
- How is it done? Let's look at the "old" and slightly modified picture above to understand how it is done in principal.
  - We need an anisotropic material oriented with respect to the optical axis in such a way that the ordinary and extraordinary beam are parallel. The two beam than will automatically have defined linear polarizations at right angle to each other.; condition 1 is met.
  - We polarize the incoming beam linearly in such a way that its polarization direction is at **45°** to the polarization direction in the anisotropic crystal. It then will split in two beams that have equal intensities. Condition 2 is met.
  - The ordinary and extraordinary beam travel with different velocities inside the material. The ordinary beam - that's why is is called "ordinary" - travels with $\boldsymbol{c_O = c_o/n_o}$ but the extraordinary beam does not; it travels with a speed $\boldsymbol{c_{EO} = c_o/n_e}$. Whatever the "extraordinary" index of refraction $\boldsymbol{n_e}$ will be, after traveling some distance $\boldsymbol{d}$ the phase shift between the two waves will be **λ/4** Obviously we have

$$d = \frac{\lambda}{4(n_e - n_o)}$$

  - So all we have to do is to cut our anisotropic material to the thickness ***d*** and condition 3 is met.
- Looks complicated? Well, that's because it is complicated. In principle and whenever you make your "**lambda quarter plate**" from a single crystal (like mica; the most prominent crystal for doing this)
  - So next time you watch a **3-D** movie, gives those (obviously cheap) glasses you're being handed a closer look. They contain two circular polarizers: one eye with a left-handed polarization, the other one with a right-handed one. And obviously they are really cheap. So how is it done?

Phase Shifters and Holograms

- We want to shift the phase of some light beam for various reasons; here we look at just two:
  - We wan to do sub-micron lithography for making microelectronic chips with structure sizes $\boldsymbol{d_{min}}$ considerably smaller than the wave length λ
  - We wan to make a **hologram**
- Let's look at **phase shift masks** for the ultimate in lithography first.
  - Right above we have data for the ultimate lens for lithography: Numerical aperture **NA = 1,35**, **λ = 193 nm**, so $\boldsymbol{d_{min} = \lambda / 2NA = 71{,}5\ nm}$; larger than what we want to get. So how are we going to beat the limits to resolution dictated by diffraction optics? By using a phase shifting mask (***PSM***). The principle is shown in the figure below:



- Remembering that waves "bend" around corners, it becomes clear that if two corners are very close together as in the schematic outline of a mask (or reticle) used for making the smallest possible structures on a chip, the "around the corner" waves overlap and from an electrical field strength and intensity (field strength squared) profile as schematically shown. The two structures are no longer fully resolved; there is an appreciable intensity below the middle light blocking layer on the mask.
- Now we introduce a "phase shifter", something the shited the phase of the light going through the right part of the struntre by **180°**. This changes the sign of the electricl field strength as shown.

There is far more but let's forget it for this leture course. Go to the next one.

5.3.3 Detecting Light

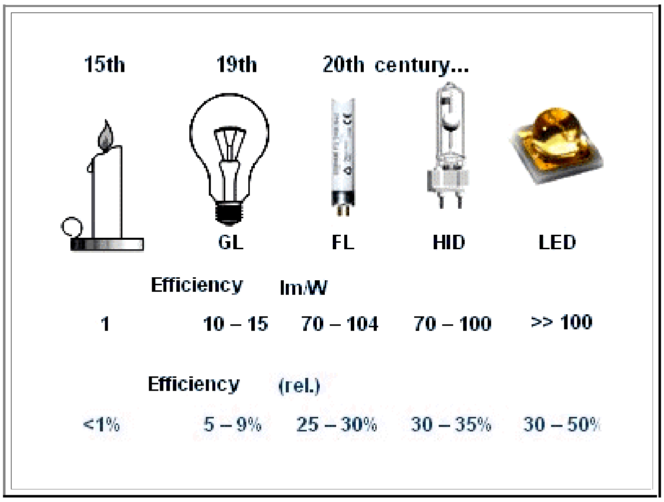
The topic would be quite interesting. It is one of the corner stones of modern communication based on light running down fibre optical cables. It encompasses topics like:

- How the eye works.
- The Eye and Quantum Efficiency.
- Photographic film
- Detecting with photomultipliers
- Detecting with semiconductors.
- Bolometers.
- Single photon detection
- Detecting high frequency modulations.
- Specialities

However! The powers that are in their undoubtedly (almost) infinite wisdom decided to shorten the time for lecture courses by 2 - 3 weeks. There is simply not enough tome to look at this topic.
Sorry

5.3.4 Summary to: Optical Components

- **Light Sources**
  Hot bodies (tungsten filaments) in light bulbs and plasma discharge in fluorescent tubes
  - Inefficient light bulbs still dominates when this lecture course started (2010)
  - LEDs have taken over when this hyperscript was finalized (2019)
- Mot included above is the **Laser**.
  - You must learn about the Laser somewhere else



- **Processing light**
  with, for example, conventional lenses, mirrors and prism, anti-reflection coatings
  - Even simple light processors like lenses (and the rest from above) might be extremely complex materials engineering products today. Just look at the picture of a (by now (2019) outdated lens for microelectronic lithography-
- Polarizers, diffraction gratings and filters add another layer of complexity.
  - The list goes on, with, e.g. phase shifters and whatever is needed for doing holgraphy or...
  - Laser "beam forming", modulation, fultr-high speed detection, ...

5.4.1 Summary to: 5.1 Optics

Know your numbers and relations for visible light!

- Wavelengths: $\lambda \approx$ **400 nm - 800 nm.**

  $\lambda_{mat} = \lambda_0 / n$.
- Frequency: $\nu \approx$ **10^{15} Hz**.
- Index of refraction: $n = \epsilon_r^{1/2} \approx$ **1,5 - 2,5**
- Energy $E \approx$ **1,8 eV - 3,2 eV**.
- Dispersion relation: $c_0 = \nu\, \lambda_0 = 3 \cdot 10^8$ **m/s**

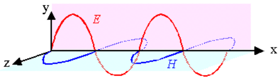
  $c_{Mat} = \nu\, \lambda_0 / n(\lambda)$

For the ***propagation*** **of light:**
use the ***wave model***
For the ***generation*** **and disappearance (=** ***absorption*****) of light:**
use the ***photon model***

Snellius law:
$n = \sin\alpha / \sin\beta$ with α, β the angle of incidence or propagation, resp.

Know yout basic equations and terminology

$$\begin{matrix} \underline{E}(\underline{r},t) \\ \underline{H}(\underline{r},t) \end{matrix} \;=\; \begin{matrix} \underline{E}_0 \\ \underline{H}_0 \end{matrix} \;\cdot \exp\{i(\underline{kr} - \omega t)\}$$

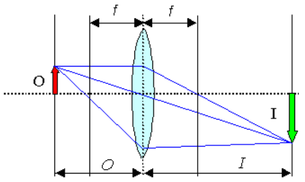


Coherent monochromatic plane wave
E and ***H*** perpendicular and in phase

- **Reflection** always with "angle in" = "angle out".
- **Refraction** is the sudden "***bending***" or "flexing" of light beams at the interface
- **Diffraction** is the continous "***bending***" of light beams around corners; interference effects.

Geometric optics
Key paramters

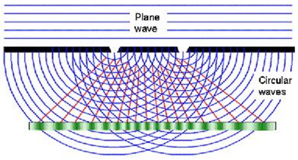
- Focal length ***f*** and
  numerical aperture **NA** of lenses, mirrors.
- Image formation by simple geometric constration
- Various aberrations (spherical. chromatic, astigmatism, coma, ...) limit performance.



Wave optics
Huygens principle: and interference

- Ultimate limit to resolution

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



Know your basic types of waves:

- (Running, coherent, monochromatic) *plane wave*.
- *Standing waves* = superposition of plane waves.
- Incoherent, multichromatic *real* waves

- Relation s between electrical field ***E***, magnetic field ***H*** and **Poynting vector** (energy flow vector) $\underline{\boldsymbol{S}} = \underline{\boldsymbol{E}} \times \underline{\boldsymbol{H}}$

$$\langle S \rangle = \frac{E_0 H_0}{2} = \frac{E_0^2}{Z_W}$$

  This equation links *energy flow* (easy in photon picture) to *field strength* in wave picture.

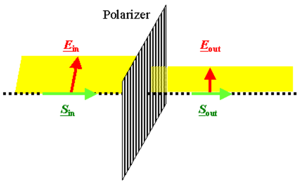
  - Z_W = **wave impedance** of the medium.
    Z_W(vacuum) = 376,7 Ω

$$W_{elect} = \frac{\epsilon_0 \cdot E^2}{2} \qquad W_{mag} = \frac{\mu_0 \cdot H^2}{2}$$

$$[W_{elect;\ magn}] = [\mathrm{Ws\ m^{-3}}]$$

$$E_0 = \left(\frac{\mu_r \mu_0}{\epsilon_r \epsilon_0}\right)^{1/2} \cdot H_0 = Z_W \cdot H_0$$

- **Polarization** = key to "advanced" optics.
  Simple case: **linear polarization**.
  - Plane of polarization contains ***E***-vector and ***S*** (***k***) vector.
  - Any (coherent) wave is polarized but *net polarization* of many waves with random polarization is zero!
  - Light *intensity* ($\propto \underline{E}^2$) between polarizers at angle α scales with **(cosα)^2**.

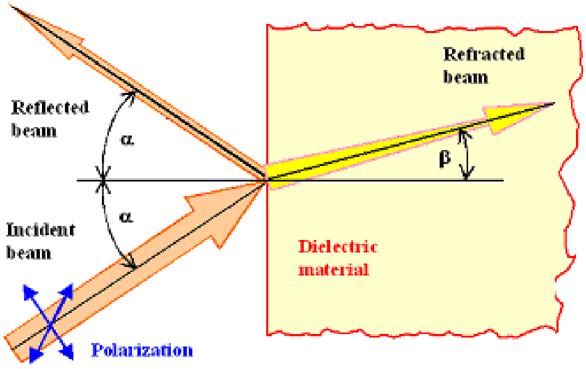


- General case: *elliptical* polarization; important are the two extremes: *linear* and **circular** polarization.
  - For circular polarizaiton the ***E***-vector rotates on a circle while moving "forward". This results from a superposition of two plane waves with ***E***-vectors ar right angles and a *phase difference* of **π/2**.
  - Technically important (3-dim Cinema; Lab optics)

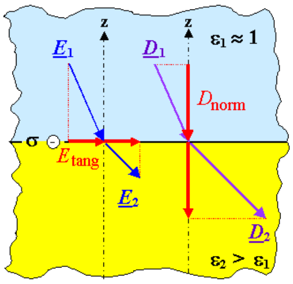
- **The task:**
  Calculate and understand intensities, angles, phases, polarization and attenuation (damping) of the various light beams shown from the materials properties
  - Still assuming a perfectly flat surface

- First step: Decompose impinging light into two waves with polarization in he interface plane (**TE** case) or at right angles (**TM** case)
  - Energy conservation yields for the intensities:

$$I_{tr}(z = 0) = I_{in} - I_{re}$$



- Boundary conditions as shown in the figure involve the "dielectric constant ϵ and thus the so far only relevant material property.
  - Considering energy (proportional to $\boldsymbol{E}^2$) and momentum (proportional to $\underline{\boldsymbol{k}}^2$") conservation for the TE and TM case separately yields the **Fresnel equations** that provide the answers to the questions above
  - A wealth of insights and relations follow, e.g. or field strength ***E*** or intensities ***I***:



$$\frac{E_{ref}}{E_{in}} = -\frac{n-1}{n+1}$$

$$\frac{I_{ref}}{I_{in}} = \left(\frac{n-1}{n+1}\right)^2$$

  - one consequence as example for the power of these equations: ***n* = 2** means that almost **10 %** of the intensity will be reflected, implying that for optical instruments you *must* provide some "anti-reflection" coating.

- Using the complex (and frequency dependent "dielectric constant $\epsilon(\omega) = \epsilon' + i\epsilon''$" yields the **complex index of refraction**

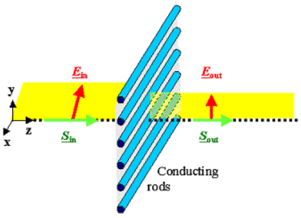
$$n^*(\omega) = n + i\,\kappa$$

  - The imaginary part κ describes the attenuation (damping) of the transmitted wave in the material.

$$n^2 = \frac{1}{2}\left(\left(\epsilon'^{\,2} + \epsilon''^{\,2}\right)^{½} + \epsilon'\right)$$

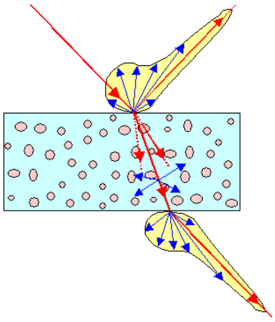
$$\kappa^2 = \frac{1}{2}\left(\left(\epsilon'^{\,2} + \epsilon''^{\,2}\right)^{½} - \epsilon'\right)$$

- **Polarization and Material**2. How to polarize a light beam
  - **1. Geometry**. Use Fresnel equations to produce a polarzed beam under specific angles ("Brewsater angle")
  - **1. Polarization foils** = alined conducting rods (of possibly molecular size) "short-circuiting" the electrical field in on direction.
  - **3. "Tensor" materials** with optical anisotropy



- Theory can an get rather involved; products can be extremely simple and cheap (e.g. circular polarizer in 3-D movie glasses)

- **Not so perfect materials** and properties like specular and diffuse Reflection, transparency, Translucency, Opacity.
  - Light is scattered at small things in all directions and the scattering of light is the major topic encountered if we look at not-so-perfect materials
  - The picture illustrates:
    Specular and diffuse reflection at the surface.
    Scattering of the transmitted light (running in different directions) at defects or imperfections contained in the material (fat droplets in milk, air bubbles in glass, ...).
    Specular and diffuse reflection at the internal surface the light is coming out off. This is described by a (different) polar diagram characterizing this surface.



- Scatter mechanism depend on the size ***l***

case is difficult to deal with and no general rules apply. This kind of scattering is called Mie **scattering**

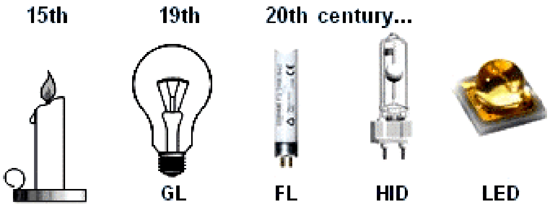
- **Generating Light**
  Two basic cases:
  - Light from **hot bodies.** Planck radiation law applies. Efficiency tends to be low
- Light from **"cold" bodies** or luminescence
  - There are many types of cold light production. Of utmost importance is electroluminescence or, to use another word for essentially the same thing, radiant electron - hole recombination in semiconductors.
    In yet other words: It's the LED, the **light emitting diode**..

Luminescence:	**General name for "cold" light production**
Fluorescence:	**Light production shortly after energy input. Short life time of excited level (< µs)**
Phosphorescence:	**Light production long after energy input. Long life time of excited level (> ms)**

- **Specialities**
  - Forget it. The list names some, there are many more.
- That's where serious "optics and material" starts. This would need another full lecture course

- **Fresnel Lens**
- **Optical Activity**
- **Faraday effect**
- **Kerr Effect**
- **Pockels Effect**

- **Light Sources**
  Hot bodies (tungsten filaments) in light bulbs and plasma discharge in fluorescent tubes
  - Inefficient light bulbs still dominates when this lecture course started (2010)
  - LEDs have taken over when this hyperscript was finalized (2019)
- Mot included above is the **Laser**.
  - You must learn about the Laser somewhere else



Efficiency lm/W

1	10 – 15	70 – 104	70 – 100	>> 100

Efficiency (rel.)

<1%	5 – 9%	25 – 30%	30 – 35%	30 – 50%

Processing light
with, for example, conventional lenses, mirrors and prism, anti-reflection coatings

- Even simple light processors like lenses (and the rest from above) might be extremely complex materials engineering products today. Just look at the picture of a (by now (2019) outdated lens for microelectronic lithography-

Polarizers, diffraction gratings and filters add another layer of complexity.

- The list goes on, with, e.g. phase shifters and whatever is needed for doing holgraphy or...

- Laser "beam forming", modulation, fultr-high speed detection, ...