6.2 Si Oxides and LOCOS Process

6.2.1 Si Oxide

The Importance of Silicon Dioxide

- Silicon would not be the "miracle material" without its oxide, ***SiO_2***, also known as ordinary **quartz** in its bulk form, or as **rock crystal** if you find it in single crystal form (and relatively pure) somewhere out there in the mountains.
  - Not only the properties of **Si** - especially the achievable crystalline perfection in combination with the bandgap and easy doping - but also the properties of **SiO_2** are pretty much as one would have wished them to be for making integrated circuits and other devices.
  - From the beginning of integrated circuit technology - say **1970** - to the end of the millennium, **SiO_2** was a key material that was used for many different purposes. Only now (around **2004**), industry has started to replace it for some applications with other, highly "specialized" dielectrics.
- What is so special about **SiO_2**?
  - First of all, it comes in many forms: There are several *allotropes* (meaning different crystal types) of **SiO_2** crystals; the most common (stable at room temperature and ambient pressure) is "low quartz" or α-quartz, the quartz crystals found everywhere. **But SiO_2** is also rather stable and easy to make in amorphous form. It is amorphous **SiO_2** - homogeneous and isotropic - that is used in integrated circuit technology. The link provides the phase diagram of **SiO_2** and lists some of its allotropes.
  - **SiO_2** has excellent *dielectric properties*. Its dielectric constant ϵ_r is about **3.7 - 3.9** (depending somewhat on its structure). This *was* a value *large* enough to allow decent capacitances if **SiO_2** is used as capacitor dielectric, but *small* enough so that the time constant $R \cdot C$ (which describes the time delay in wires having a resistance R and being insulated by **SiO_2**, and thus endowed with a parasitic capacitance C that scales with ϵ_r) does not limit the maximum frequency of the devices. It is here that successors for **SiO_2** with larger or smaller ϵ_r values are needed to make the most advanced devices (which will hit the market around **2002**).
  - It is among the *best insulators* known and has one of the highest *break-through field strengths* of all materials investigated so far (it can be as high as **15 MV/cm** for very thin layers; compare that with the values given for normal "bulk" dielectrics).
  - The *electrical properties of the **Si** - **SiO_2** interface are excellent*. This means that the interface has a very low density of energy states (akin to surface states) in the bandgap and thus does neither provide recombination centers nor introduce fixed charges.
  - **SiO_2** is relatively *easy to make* with several quite different methods, thus allowing a large degree of process leeway.
  - It is also relatively *easy to structure*, i.e. unwanted **SiO_2** can be removed selectively to **Si** (and some other materials) without many problems.
  - It is very *stable and chemically inert*. It essentially protects the **Si** or the whole integrated circuit from rapid deterioration in a chemically hostile environment (provided simply by humid air which will attack most "unprotected" materials).
- What are the uses of **SiO_2**? Above, some of them were already mentioned, here we just list them a bit more systematically. If you do not quite understand some of the uses - do not worry, we will come back to it.
  - **Gate oxides**: As we have seen before, we need a thin dielectric material to insulate the gate from the channel area. We want the channel to open at low threshold voltages and this requires large dielectric constants and especially no charges in the dielectric or at the two interfaces. Of course, we always need high break through field strength, too. No dielectric so far can match the properties of **SiO_2** in total.
  - *Dielectrics in integrated capacitors*. Capacitors with high capacitance values at small dimensions are needed for so-called **dynamic random access memories** (***DRAM***), one of the most important integrated circuits (in terms of volume production). You want something like **30 fF** (femtofarad) on an area of **0.25 µm^2**. The same issues as above are crucial, except that a large dielectric constant is even more important. While **SiO_2** was the material of choice for many **DRAM** generations (from the **16 kbit DRAM** to the **1 Mbit DRAM**), starting with the **4 Mbit** generation in about **1990**, it was replaced by a triple layer

of **SiO_2 - Si_3N_4 - SiO_2**, universally known as "***ONO***" (short for oxide - nitride - oxide); a compromise that takes not only advantage of the relatively large dielectric constant of silicon nitride (around **7.5**) while still keeping the superior quality of the **Si - SiO_2** interface, but has a few added benefits - at added costs, of course.

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

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

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

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

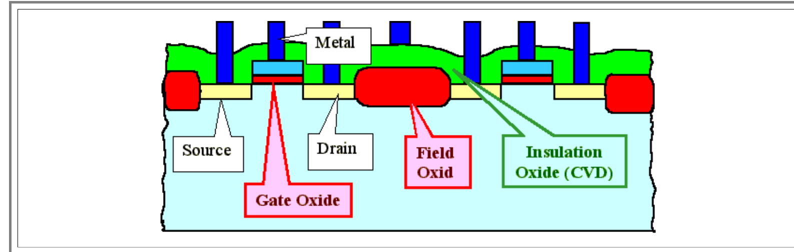
- **Thermal oxidation**. This means that a *solid state reaction* (**$Si + O_2 \Rightarrow SiO_2$**) is used: Just expose **Si** to **O_2** at sufficiently high temperatures and an oxide will grow to a thickness determined by the temperature and the oxidation time.
- **CVD oxide deposition**. In complete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

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

- Anodic oxidation uses a current impressed on a **Si** - electrolyte junction that leads to an oxidation reaction. While easy to understand in principle, it is not very well understood in practice and an object of growing basic research interest.

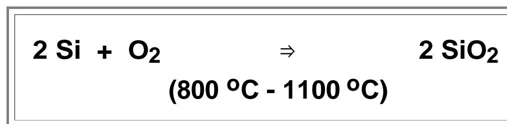
Thermal Oxidation

- In this paragraph we will restrict ourselves to **thermal oxidation**. It was (and to a large extent still is) one of the key processes for making integrated circuits. While it may be used for "secondary" purposes like protecting the bare **Si** surface during some critical process (remember the "screen oxide" from above?), its major use is in three areas:
  - **Gate oxide** (often known as "***GOX***")
  - **Capacitor dielectric**, either as "simple" oxide or as the "bread" in an "**ONO**" (**=** oxide-nitride-oxide sandwich)
  - **Field oxide** (***FOX***) - the lateral insulation between transistors.
- We can use a picture from chapter 4.1.4 to illustrate **GOX** and **FOX**; the capacitor dielectric can also be found in this chapter.

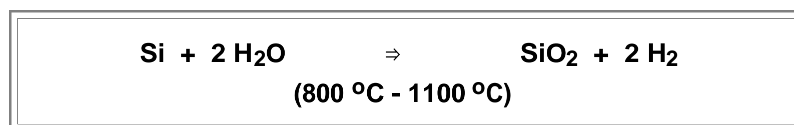


  - We must realize, however, that those drawing are *never to scale*. The gate oxide is only around **10 nm** thick (actually, it "just" (**2007**) petered out at **1.2 nm** accoding to Intel and is now replaced by a thicked $\mathbf{HfO_2}$), whereas the field oxide (and the insulating oxide) is in the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

.



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



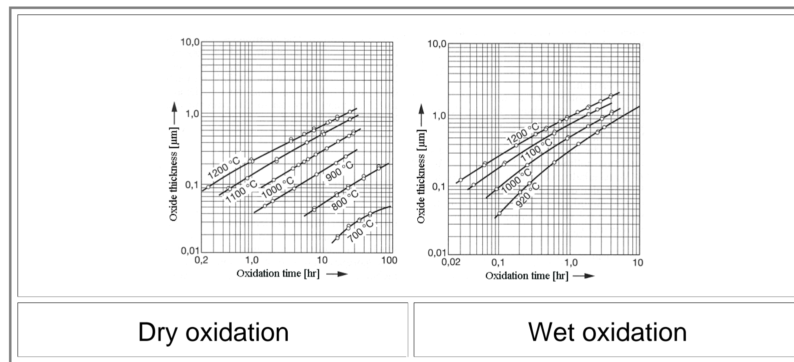
  - The growth kinetics are about **10x** faster than for dry oxidations; this is the process used for the thick *field oxides*.
  - Growing **700 nm** oxide at **1000 °C** now takes about **1.5 hr** - still pretty long but tolerable. Thin oxides are never made this way.
- In both cases the oxygen (in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

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

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

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

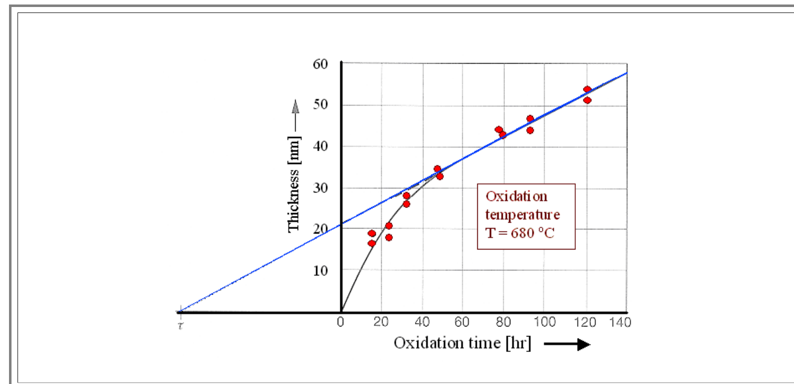
  - In this case the limiting factor is the rate at which **O** can be incorporated into the $\mathbf{Si - SiO_2}$ interface.
- This kind of behavior - linear growth switching to square root growth - can be modelled quite nicely by a not too complicated theory known as the **Deal-Grove model**. Some results of experiments and modeling are shown below.



Dry oxidation | Wet oxidation

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

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

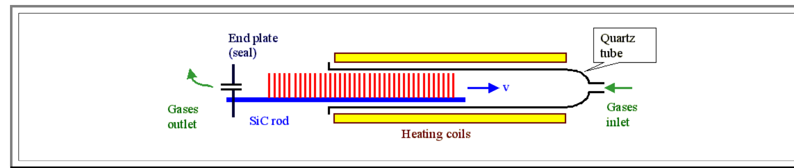


- The red points are data points from an experiment (at an unusually low temperature); the blue curve is the best Deal-Grove fit under the (not justified) assumption that at $t = 0$ an oxide with a thickness of about **20 nm** was already present.
- The Deal-Grove model is clearly inadequate for the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

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

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

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



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

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

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

Questionaire
Multiple Choice questions to 6.2.1