3.3.2 Scaling Laws

General Consideration of Making Devices Smaller

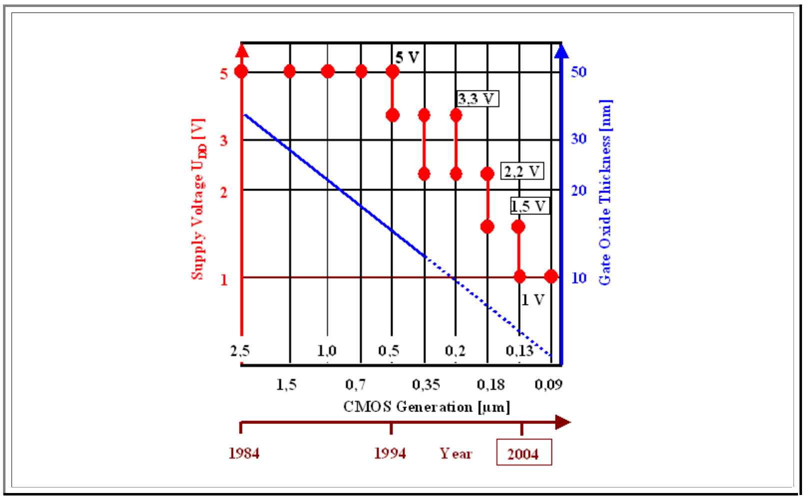
- All theories introduced so far (i.e. all of chapter 2), always assumed "infinite" or "semi-infinite" crystals. For example, the size of the crystals did not matter for the characteristics of a **pn**-junction; the dimensions of the **n**- and **p**-doped regions did not enter the equations.
  - However, if we reconsider for a moment the simple derivation of the **I-U**-characteristics of a **pn**-junction, we (hopefully) remember that the carriers responsible for the reverse current originated from a region defined by the diffusion length of the minority carriers.
  - What happens if the device is much smaller than the diffusion length ***L*** ? This will be almost always the case, considering that ***L*** is around **100 µm** and typical integrated transistor occupy hardly **1 µm^2**?
- While this question can still be answered relatively easily by conventional device physics, it nicely illustrates that we can not expect the properties of *any* device to be independent of its size as suggested by simple semiconductor physics.
  - The basic question in micro-technology thus is: What happens if you make an existing (and functioning) device smaller?
  - And making something smaller can be done in different ways: You may simply decrease the lateral extensions while leaving the depth dimension unchanged, or more realistically, you scale the lateral and depth dimensions by different factors. As an example, while you may reduce the lateral size of a source-drain region by a factor of **2**, the depth of the **pn**-junctions and the thickness of the gate oxide may scale with only a factor of **1,3**.
  - How do you find the optimum? Where are limits and how can they be overcome? In other words: What are the relevant scaling laws and when do we hit a brickwall **-** you can't make it smaller any more without insurmountable problems!
- There are some billion Dollar questions hidden in this scenario, and there are no easy answers for some of the details. There are also, however, some simple laws and rules which we will consider briefly in this subchapter.

Linear Scaling and Problems

- Lets look at some general **scaling laws**. We simply assume that we decrease all linear dimensions of an existing device by the **scaling factor *K***.
  - A first obvious conclusion is that the field strength in some insulating layer, e.g. a gate oxide, increases ***K*** -fold. We may accept that, or we might scale the voltage, too. In this case we would decrease U_{DD}, the external driving voltage, to U_{DD}/K.
  - Going through all important parameters (with some approximations if necessary), we obtain the following table

Property	Scaling	
All lateral and vertical dimensions	$1/K$	
Doping concentration	K	
	$U_{DD} \Rightarrow U_{DD}/K$	U_{DD} = constant
Packing density (No. transistor/cm^2)	K^2	K^2
Current densities	K	K^3
Field strenghts	1	K
Power loss density	1	K^3
Power loss per transistor	$1/K^2$	K
Time delay per transistor	$1/K$	$1/K^2$

- The problems you are running into are obvious. Lets look at the transition from a **1 µm** process to a **0,25 µm** process, i.e. ***K* = 4**
  - Without a fourfold reduction in the supply voltage, we would have a **64** fold increase in current densities and power loss, and a **4** fold increase in field strength. This is *not* going to work
  - Decreasing U_{DD} fourfold (from **5 V** to **1,25 V**) still increases the current density fourfold, but keeps all other parameters manageable. However, our device only speeds up **4** fold, compared to **16** fold for constant. U_{DD}.
- So simply lets scale down U_{DD} some more? Well, yes **-** but: You can't just decrease all dimensions and U_{DD} just so!
  - The tjickness of gate and capacitor dielectrics might be close to absolute limits (e.g. imposed by tunneling) and simply cannot be scaled down much more.
  - Internal voltages might only be fractions of the supply voltage and reducing U_{DD} may decrease signal to noise levels to unacceptable values.
  - Voltage swings for switching transistor must at least be in the order of the band gap, i.e. **1 V** for Si. Voltages thus cannot be reduced to arbitrarily small levels.
- If we look at the actual scaling of devices, much ingenuity and many additional process steps were used to avoid the simple rules of scaling. Here are a few examples:
  - Trench capacitor instead of planar capacitor. The thickness of the dielectric thus could stay relatively constant, which allowed higher supply voltages than required by scaling.
  - Electromigration resistant metallization (addition of **Cu** or other atoms to the **Al** lines, multi layers etc.) allowed larger current densities
  - "Lightly doped drain", i.e. complicated dopant profiles of the source/drain region below the gate allowed higher field strengths in the **MOS** transistors.
- The list is easily extended by **10** or more points; but you get the drift.
  - But all tricks notwithstanding: the supply voltage had to come down. The historical development is shown in the figure below.



- There are several remarkable features:
  - For almost **10** years the supply voltage was kept constant at U_{DD} **= 5 V** despite a scaling of ***K* = 5**. This was possible by a dramatic increase of process complexity and materials engineering.
  - The end in scaling U_{DD} is near. Now (**2001**), supply voltages are as low as **1, 5 V**, and we simply cannot go much below **1 V** with **Si**.
  - New principles are needed, because we still can make functioning transistors far below **0,1 µm**. One possible solution is to make vertical transistors. Demonstrators (Bell Labs) work with gate length of **30 nm** or less.
- What happens: You will not only live to see it, but possibly help to establish it.

Fundamental Limits

- Whatever clever tricks are used to make just one more step towards smaller devices, there are some ultimate limits. One simple example shall be given; it concerns doping:
  - Lets say we need doped areas with a typical dopant concentrations of ρ **= 10^{16} cm^{-3}** or **10^{18} cm^{-3}** and maximum deviations of **± 10%.** This implies that the doped area must contain at least **100** dopant atoms because the statistical fluctuations are then **$(100)^{1/2}$ = 10** giving the **10%** allowed.
  - **100** atoms at the prescribed density need a volume of (**100/10^{16}) cm^3** or **(100/10^{18}) cm^3** , respectively, which equals **10^7 nm^3** or **10^5 nm^3**, respectively.
  - These volumes correspond to cubes with a linear dimension of **215 nm** or **46,4 nm**, respectively. What does this mean?
- Look at it in a different way: A typical source/drain region in a modern integrated circuit may be **0,5 µm x 0,5µm x 0,2 µm = 5 · 10^{-2} µm^3 = 5 · 10^7 nm^3**.
  - At a doping level of **10^{16} cm^3** - corresponding to the perfectly reasonable resistivity of **1.4 Ωcm** (**p**-type) or **0.5 Ωcm** (**n**-type) - we have about **500** doping atoms in there! Doping to a precision better than **$(500)^{1/2}$ = 22,4 or 4,47 %** is principially not possible assuming a statistical distribution of the doping atoms.
  - Decreasing the size to e.g. **0,1 µm x 0,1µm x 0,02 µm = 2 · 10^5 nm^3** leaves us **2** doping atoms in there - obviously absurd. Again we have to turn to novel structure - vertical transistors may do the trick once more.
- There are more "fundamental limits" - just how fundamental they are, is a matter of present day research (and a seminar topic).