8.2.2 Small Signal Response of p-n Junctions

Equivalent Circuit Description

- If we pretend for a moment that we are hard-core electrical systems engineers, we do not care at all about what a **pn-** junction diode consists off, how it works, or how it is made. We simply describe it as a **black box** and use the two equations from before, but now with a possible *frequency dependence* thrown in for the output current. For small signal behavior we have by definition

$$U(t) = U_0 + U_m \cdot \exp(i \cdot \omega t)$$

$$I(t) = I_0 + I_m(\omega) \cdot \exp(i \cdot [(\omega t) + \varphi(\omega)])$$

  - i.e. we consider a frequency dependent amplitude $I_m(\omega)$ of the current output signal and a frequency dependent phase shift $\varphi(\omega)$.
- The relation between U_0 and I_0 is simply given by the (**DC**) current voltage characteristics of our black box.
  - In principle, the output signal amplitude (i.e. the current amplitude) is also a function of U_0 or I_0; in other words of the **working point** chosen. All we have to do then is to tabulate (or represent graphically) the functions $I_m(\omega, I_0)$ and $\varphi(\omega, I_0)$ for our black box.
  - This is all. We always have an output signal looking exactly like the input signal, and with an amplitude proportional to the input amplitude because we have *by definition* a linear system.
- We know even more. We always will have

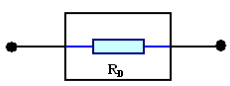
$$I_m(\omega) = \frac{dI}{dU} \cdot U(\omega)$$

$$\varphi = 0$$

  - as long as the output signal can be directly obtained from the **DC** characteristics, i.e. at *small* frequencies (leaving open at present what "small" means in numbers).
- Looking at our diode (or the small signal (i.e. linear) behavior of about anything else) in this way has a *large advantage*:
  - *All* black boxes can now be described by a suitable network of *basic linear electronic elements* - resistors, capacitors and inductors. In the most simple case all elements have constant values; more realistically their value depends on external parameters (e.g. the voltage).
  - Calculating the behavior of such a network is relatively simple - for a computer, that is.
- However, just looking at some network or **equivalent circuit diagram** as it is called, has *disadvantages* too:
  - Many different kinds of networks will give exactly the same response; their is no *unique* choice - but often *intelligent* choices.
  - If some network does behave like the actual device, it is not necessarily easy to extract the important physical parameters of the device from it. In other words, if some network behaves exactly like the diode you are interested in, it does tell you *how* you must change resistors and capacitors if you want to make it faster, but not which physical parameters of the device (doping, lifetime, dimensions, ...).
- The best way therefore is to construct an equivalent circuit - keeping it as simple as possible - by looking at the physical characteristics of the real device first. We are going to do this now for our junction diode.

Equivalent Circuit Construction for a Junction Diode

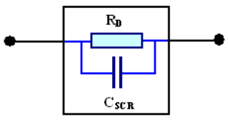
- If we start with the low frequency behavior of a junction diode, the equivalent circuit diagram is exceedingly simple:
  - It consists of a single resistor R_D with a value that depends on the voltage and is given by $R_D = dU_{diode}/dI_{diode} = R_D(U)$, i.e. by the (inverse) slope of the **DC** characteristic of the diode as illustrated before. This is shown below



- This "network" cannot possibly give *any* frequency dependence, so it can not be all there is to a diode. We now must remember that any **pn**-junction has a **capacitance** C_{SCR} associated with the space charge region. It arises from the **ionized dopants** that provide the net charge on both sides of the junction needed for any capacitor and its value (for a symmetric junction) is given by

$$C_{SCR} = \left(\frac{2e \cdot \epsilon\epsilon_0 \cdot N}{U}\right)^{1/2}$$

- C_{SCR} is obviously switched in parallel to R_S; our equivalent circuit diagram now transforms into this:



This network will give us a simple frequency response: High frequencies are essentially short-circuited by the capacitor; and the current amplitude will increase with frequency (accompanied by a phase shift of **90°** at the maximum).

- The current amplitude will have increased twofold if the **AC** resistance R_C of C_{SCR} equals R_D , this happens for

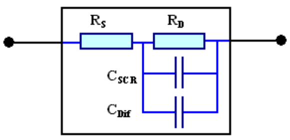
$$R_C = \frac{1}{\omega \cdot C_{SCR}}$$

With our formulas for $I = I(U,$ **doping, ...)** and for $C_{SCR} = C_{SCR}(U,$ **doping, ...)** , we could calculate the limit frequency as a function of the prime material parameters like doping, lifetime and so on - but we are not yet done in constructing the equivalent circuit diagram:

- There is a *second capacitance* hidden in the junction diode called "**diffusion capacitance** ". Lets see what it is.
- *Any* net charges to the left and right of the junction form a capacitance. Above, we considered the space charge region capacitance stemming from ionized dopants in the **SCR** which are not compensated by majority carriers as we have it in the bulk. For diodes biased in *reverse* direction, i.e. with only a small reverse current flowing, this is all there is.
- But if we look at the distribution of *all* carriers for a diode in *forward* direction, we notice that we have an *excess of minority carriers at the edge of the SCR*; a schematic picture was given before.
- These excess minorities - only present while current flows - are coming from the *injection of majority carriers* over the potential barrier which become minorities in the other part of the junction. Their concentration is increased because a concentration gradient is needed to remove these carriers by diffusion. They are *not* compensated by other charges and thus form a **diffusion capacitance** C_{Dif} that increases with current.
- This diffusion capacitance must also be switched *in parallel* to R_D and C_{SCR} .

Moreover, the ohmic resistance of the bulk **Si** is not zero, but has some finite value R_S which is constant and easily determined by the resistivity of the **Si** and all other ohmic resistors in the circuit.

- Clearly, R_S must be switched *in series* to everything else, and we obtain the final equivalent circuit diagram of a junction diode valid for all cases.



- Of course, you always could lump together the two capacitors into just one, but then you loose the connection to physical reality. Note that their numerical values will depend on physical parameters in very different ways - as we will see below.

If we would know the value of C_{Dif} , we could now calculate the small signal response of a junction diode. We have no ready formula, so we must derive one.

- We will do that right away; it will prove to be particular enlightening for the understanding of the correspondence of *primary material parameters* and their transformation to *equivalent circuits*.

The Diffusion Capacitance and its Relation to Time Constants of the Device

- The excess Q_{min} at to the left and right of the space charge region increases with increasing (forward) voltage ***U*** . If we have a linear relation between ***U*** and Q_{min}, the capacitance C_{Dif} would be given by

$$C_{Dif} = \frac{Q_{min}}{U}$$

  - Since this is not necessarily the case, C_{Dif} is a function of ***U*** and must be defined *differentially* as .

$$C_{Dif} = \frac{dQ_{min}}{dU}$$

  - Considering that the current ***I*** is a function of ***U*** (given by the basic diode equation), we may rewrite this expression as

$$C_{Dif} = \frac{dQ_{min}}{dU} = \frac{dQ_{min}}{dI} \cdot \frac{dI}{dU}$$

  - The second term is simply the derivative of the ***I-U*** characteristics, and thus equal to the inverse diode resistance.

- dQ_{min}/dI is the interesting term. This differential quotient must have the dimension of time, and thus can be seen as the *time constant* connected to C_{Dif} . In order to compute it, we need Q_{min} as a function of the current flowing through the junction, i.e.

$$Q_{min} = Q_{min}(I)$$

  - We have not encountered this functional relationship so far - but we came close.

- In the "Useful Relations" subchapter we have an equation giving the excess minority carrier density $\Delta n(x)$ as a function of the distance ***x*** from the edge of the depletion layer:

$$\Delta n(x) = \Delta n_0 \cdot \exp - \frac{x}{L}$$

  - With Δn_0 = excess density at the edge of the depletion zone, ***L*** = diffusion length.
  - In the "Junction Reconsidered" subchapter we derived an equation relating the current density ***j*** (for one of the two current components) through the junction to the excess carrier density Δn_0 at the edge of the depletion layer:

$$j = \frac{e \cdot D}{L} \cdot \Delta n_0$$

- So all that remains to do is to express the charge Δn_0 at the edge of the **SCR** as a function of the total excess charge Q_{min}, substitute it in the current equation, and do the differentiation dQ_{min}/dI. This is easy:
  - Q_{min} is simply the integral over $e \cdot \Delta n(x)$ taken from $x = 0$ to $x = \infty$; i.e.

$$e \cdot Q_{min} = \int_{x=0}^{x=\infty} e \cdot \Delta n(x) \cdot dx = \int_{x=0}^{x=\infty} e \cdot \Delta n_0 \cdot \exp - \frac{x}{L} \cdot dx = e \cdot \Delta n_0 \cdot L$$

- Inserting $\Delta n_0 = Q_{min} / L \cdot e$ in the current equation from above gives

$$j = \frac{e \cdot D}{L} \cdot \frac{Q_{min}}{e \cdot L} = \frac{D \cdot Q_{min}}{L^2}$$

- The differentiation (turning to current densities j instead of currents I and interpreting Q_{min} as charge density) finally yields

$$\frac{dQ_{min}}{dj} = \frac{L^2}{D} = \tau$$

And this, of course, is nothing but the minority carrier life time τ!

- The time constant associated with the diffusion capacitance thus has a very clear physical meaning - it is the minority carrier life time. It is, if you like, the fundamental property that *causes* the diffusion capacitance.
- Interpreted in other words: While separated charges Q always cause a (static) capacitance C given by $C = Q/U = \epsilon\epsilon_0 Q \cdot A/d$, you always will find a (dynamic) capacitance, too, if it *takes time* to change charge concentrations. The capacitance associated with some *given* time constant τ than is more appropriately expressed as

$$C_{dyn} = \frac{\tau}{R}$$

- With R being some ohmic resistor which limits current flow.

Now we have all terms; the diffusion capacitance can be finally expressed as

$$C_{Dif} = \frac{dQ_{min}}{dI} \cdot \frac{dI}{dU} = \frac{\tau}{R_D}$$

Finally, we look at the diffusion capacitance from yet another angle: It results from the necessity to "store" some charge on both sides of the junction for forward current flow.

- The concept of **stored charge** and the time constant associated with its removal will come up again later.

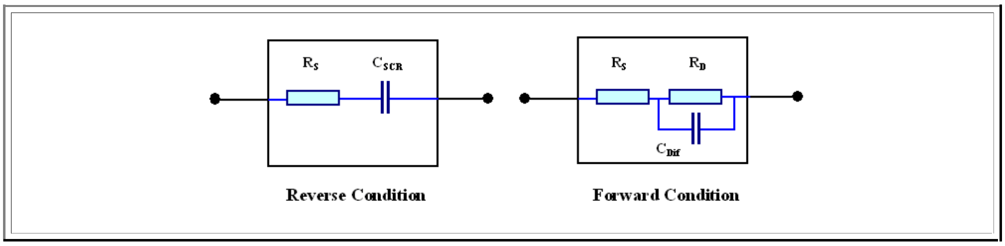
Consequences

We now have all ingredients to analyze the small signal response of a **pn**-junction - as long as it is ideal, infinitely extended, and symmetric.

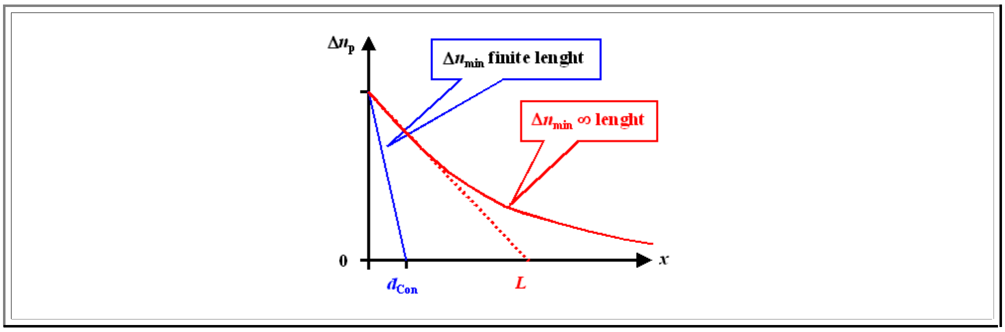
The first question coming to mind is the relation between the two capacitances. Electrically, they add up - but which one dominates?

- This is most easily seen if we consider the forward and reverse current direction separately.
- The *reverse* current mode is easy: The diffusion capacitance is small because dI/dU is practically zero and can be neglected. We only will find the junction capacitance C_{SCR}. Moreover, R_D is very large, certainly much larger than R_S , and may be taken to be infinite.
- In *forward* direction, C_{SCR} increases with $U^{1/2}$ while C_{Dif} increases *exponentially* with U (look at the formulas!). This means that in forward direction C_{Dif} will always win and dominate the total capacitance.

We thus can simplify the equivalent circuit diagrams for the two extremes:



- This means that the small signal behavior of a **pn**-junction is quite different in reverse or forward direction, and that it also depends very much on the working point (the **DC** current) in forward direction. It also means that we only can transmit **AC** signals in reverse direction.
- If we now try to deduce some numbers for limiting frequencies of real diodes, we run into problems:
  - The diffusion capacitance will short-circuit all signal with frequencies considerably larger than **1/τ**. Since the minority carrier life time in good **Si** can be as large as **1 ms**, we must expect problems in the **kHz** region - and even for life times of **10 µs** we would not get far beyond the **MHz** region. This is obviously *far below* limiting frequencies in actual devices.
- *What went wrong*? Not so easy to unravel, lets just consider the most important point:
  - Dimensions in real (**Si**) devices, especially integrated circuits, are *much, much smaller* than the diffusion length ***L***. For the excess minority carrier distribution as shown for a large diode, this has profound consequences.
  - The excess concentration must go down to zero at the contact, i.e. after a distance d_{Con} that gives the length of the (almost) neutral piece of **Si** (the distance from the edge of the space charge layer to the contact). In a *first approximation* we have to replace the diffusion length ***L*** (which is in the order of magnitude of **100 µm**) by d_{Con}, which can be a fraction of a **µm**.
  - The minority carrier excess density now decays rapidly and reaches zero at the contact, i.e. after a distance d_{Con}. For all practical purposes it is sufficient to assume a linear relationsship as shown below. In practice, the difference between ***L*** and d_{Con} is much larger.



  - Accordingly, we have to replace the bulk lifetime $\tau = L^2/D$ by

$$\tau_{tran} = \frac{(d_{Con})^2}{D} = \text{base transit time}$$

  - This is something we have encountered before by looking at real diodes.
- A small device thus could be faster by several orders of magnitude.
  - And this is where, quite generally, the *dimensions* come in. We always need to move carriers from here to there, and this takes some time determined by the distance and the velocity (or mobility), respectively, as pointed out before.
  - For our case of *small diodes*, we may use the equation from before in the form

$$\tau_{tran} = \frac{(d_{Con})^2}{\mu \cdot U}$$

  - This tells us that the small signal frequency response of a small junction diode is ultimately controlled by the dimensions of the device and the mobility of the carriers that carry the current.
- Of course, *real* small devices are even more complicated. But all kinds of parameters neglected in this simple consideration will make the frequency response worse, not better. The absolute limits are always determined by dimensions and mobility.

- A few last words of caution:
  - While we replaced the (bulk) lifetime by a transit time in this simple consideration of the real device world, it would be premature to conclude that the life time is of no importance in small devices.
  - It also would be premature to conclude that we should reduce τ as much as possible, because this carries heavy penalties in other areas - see the links 1, 2,
  - However, that does not mean that "*lifetime killing*" is not used on occasion. For some devices - particularly (large) power devices - some **Au** might be diffused into a junction to reduce the life time as much as possible.