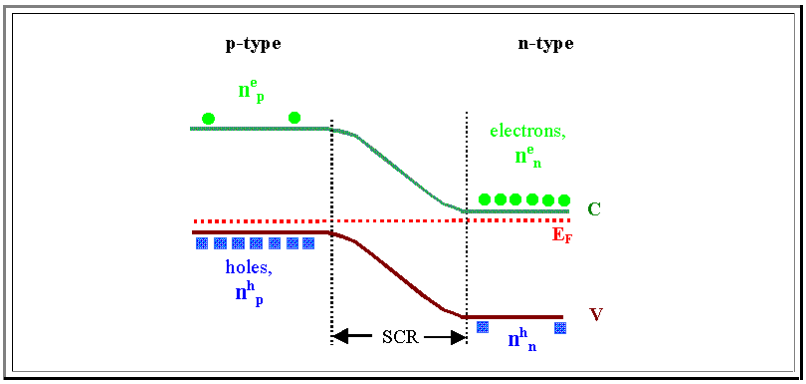
2.3.5 Junction Reconsidered

- In this section we will give the **p–n** junction a new look using a somewhat more advanced point of view.
  - A full treatment of **p–n** junctions in any kind of *three-dimensional* semiconductor, taking into account arbitrary doping *profiles*, *finite* size, and effects of the *interfaces* and *surfaces*, is one of the more difficult things to do in semiconductor theory; we will not attempt it here.
  - In all standard treatments of junctions, we always look at special (unrealistic) cases and use (lots of) approximations. This, admittedly, can become somewhat confusing.
  - It is thus very advisable to become really well acquainted with the simple treatments given in section 2.2.4; this will clear the mind for the essentials.

Junction Without Contributions from the Space Charge Region

- Here we look at an advanced treatment of a **p–n** junction, but we still will have to make some simplifications.
  - We consider a "*narrow*" "*abrupt*" one-dimensional junction in an infinitely long crystal, which is formed by a sudden change of doping, i.e. there are *no gradients* of the doping densities N_D and N_A to the left or right of the junction.
  - "Narrow" then means that the width of the space charge layer is much smaller than the diffusion length of minority carriers, but still much larger than the mean free path length needed for the thermalization of carriers.
  - In consequence, we do not consider recombination in the **SCR**, and we can assume thermal equilibrium in the bands, i.e. we can use the quasi-Fermi energies.
- First, we look at the junction in equilibrium, i.e. there is no net current and the Fermi energy is the same everywhere (we have the same situation as shown before; but for diversity's sake, the **p**- and **n**-side reversed).



- In what follows, there will be a lot of shuffling formulas around – and somehow, like by magic, the ***I–V*** characteristics of a **p–n** junction will emerge. So let's be clear about what we want to do in the major steps (highlighted by the cyan background).
- *The first basic goal* is to find an expression for the *carrier density at the edge of the space charge region*.
  - We know in a qualitative way from the consideration of pure diffusion currents that the minority carrier density around the edge of the space charge region is somewhat larger than in the bulk under equilibrium conditions – there is a Δn_{min} given by $\Delta n_{min} = n_{min}(x) - n_{min}(bulk)$, and Δn_{min} will increase for forward bias, i.e., for non-equilibrium conditions.
  - We also know that Δn_{min} induces a diffusion current and that we therefore need a "real" current to maintain a constant Δn_{min}. Finding Δn_{min} thus will *automatically give us the necessary currents* belonging to the non-equilibrium as defined by the voltage.
- We first look at the various energies involved:
  - The energy difference between the left side (= **p**-side, *raised index "**p**"*) and the right side (= **n**-side, *raised index "**n**"*) of the junction is given by

$$E_C^p - E_C^n = E_V^p - E_V^n = e \cdot \Delta V$$

  - Here, ΔV is the difference of the (yet-to-be-determined) electrostatic potential between the **n-** and **p**-side, taken far away from the junction; the details (especially about its sign) will be given below.
- The band energy levels $E_C^{n,p}(x)$ and the potential $V(x)$ are functions of x, which makes all densities functions of x, too. *We will, however, write all these quantities without the "(x)" from now on.*
  - As long as we discuss equilibrium, the Fermi energy is constant and the carrier densities are given by their usual expression. We consider them separately for the left- and right hand side of the junction, i.e., for the **n**- and **p**-part. (Note that here the usual minus sign in the exponent was used to change the order of the terms in the differences.)

$$n_e^p = N_{eff}^e \cdot \exp\frac{E_F - E_C^p}{kT}$$ **density of electrons on the p-side**

$$n_h^p = N_{eff}^h \cdot \exp\frac{E_V^p - E_F}{kT}$$ **density of holes on the p-side**

$$n_e^n = N_{eff}^e \cdot \exp\frac{E_F - E_C^n}{kT}$$ **density of electrons on the n-side**

$$n_h^n = N_{eff}^h \cdot \exp\frac{E_V^n - E_F}{kT}$$ **density of holes on the n-side**

- We also have the mass action law, here applicable everywhere since we are in full equilibrium:

$$n_e^p \cdot n_h^p = n_h^n \cdot n_e^n = n_{min} \cdot n_{maj} = n_i^2 = N_{eff}^e \cdot N_{eff}^h \cdot \exp -\frac{E_g}{kT}$$

What we *need to know* to get on is the ***x*-dependence of the energies** or of the potential $V(x)$ – this simply means we need the quantitative band diagram that so far we always just drew "by feeling". (This is one of the **essential** points why we reconsider the p–n junction here; the other one will be the usage of the quasi-Fermi energies.)

- For this we need to solve the **Poisson equation** and this demands to specify the total charge $\rho(x)$ so that we can write down the charge as a function of x. *This is easy in principle*:

The total (space) charge $\rho(x)$ at any point along the junction is the sum of all charges: Electrons ($n_e(x)$), holes ($n_h(x)$), ionized donors ($N_D^+(x)$), and ionized aceptors ($N_A^-(x)$). We have as before

$$\rho(x) = e \cdot \left(n_h(x) - n_e(x) + N_D^+(x) - N_A^-(x) \right)$$

- Inserting $\rho(x)$ into the Poisson equation gives

$$-\frac{\epsilon\,\epsilon_0}{e} \cdot \frac{d^2V(x)}{dx^2} = n_h(x) - n_e(x) + N_D^+(x) - N_A^-(x)$$

- Solving this equation with the proper boundary conditions will yield $V(x)$ and everything else – *but not so easily* because the situation is complicated: Since $n_h(x)$ and $n_e(x)$ depend on $V(x)$ via the Fermi distribution, this is also an *implicit* equation for $V(x)$.
- It is, however, not too difficult to find good approximations for "normal", i.e. highly idealized junctions; this is shown in an advanced module accessible through the link.

For our final goal, which is to describe the **current–voltage characteristic** of a **p–n** junction, we use the same *approximations* and conventions, namely:

1. The *zero point of the electrostatic potential* is identical to the valence band edge in the **p**-side of the junction, i.e. $eV^p = E_V^p = 0$ as shown in the complete illustration to the situation shown in the picture above. This is a simple *convention* without any physical meaning.

2. *All dopants are ionized, their density is constant up to the junction, and there is only one kind on each side of the junction*, i.e.

$$n_h^p(\text{bulk}) = N_A = N_A^-$$

$$n_e^n(\text{bulk}) = N_D = N_D^+$$

- This is a *crucial assumption*. Note that while $n_{h,e}^{p,n}(\text{bulk})$ are constant, this is not required for $n_{h,e}^{p,n}(x)$ around the junction.

3. We also assume that away from the junction, the **Si** extends into infinity (or at least to a distance much larger than several diffusion lengths) to both sides of the junction – in total we use the "abrupt" "large" junction approach

- This gives us for the carrier densities in equilibrium anywhere in the junction:

$$n_h(x) = N_A \cdot \exp - \frac{e \cdot V(x)}{kT}$$

$$n_e(x) = N_D \cdot \exp - \frac{e \cdot [V^n - V(x)]}{kT}$$

- Here, V^n is the constant value of the potential deep in the **n**-type region. Note that, having chosen the zero point for $V(x)$ at the **p**-side of the junction where there is the negative pole of the electric field, it holds inside the **SCR** that $0 \le V(x) \le V^n$.
- These equations mean that the carrier density is whatever you have in the undisturbed **p**- or **n**-part (i.e., the dopant density) times the Boltzmann factor of the energy shifts relative to this situation.

V^n is the difference of the built-in potential for equilibrium conditions, it is thus determined by the difference in the Fermi energies of the n- and the p-side before contact (relative to the band edges) – our simple view of a junction is totally correct on this point.

- With and without an external voltage U_{ext} we have

$$V^n(U_{ext}=0) = \frac{1}{e} \cdot (E_F^n - E_F^p)$$

$$V^n(U_{ext}) = \frac{1}{e} \cdot (E_F^n - E_F^p + e \cdot U_{ext})$$

- Here, the sign of U_{ext} is such that a positive external voltage *increases* the built-in potential difference. Note that this is just an interim choice; later on we will replace it by the usual standard.
- In the general case, the maximum potential at the **n**-side, $V^n(\text{bulk})$, becomes

$$V^n(\text{bulk}) = V^n(U_{ext}=0) + U_{ext} = \frac{1}{e} \cdot \Delta E_F + U_{ext} = V^n + U_{ext} = \Delta V$$

- Looking at the proper solution of the Poisson equation for our case, we realize that the space charge region was defined as the part of the **Si** where the potential was not yet constant. This means that $V^n(\text{bulk}) = V^n|_{\text{SCR edge}}$ on the **n**-side, and $V^p|_{\text{SCR edge}} = 0$. This is an essential point, even so it is matter-of-course.

We now can move towards our primary goal and find an expression for the carrier density at the edge of the **SCR** by considering the ratio of a carrier species on both sides of the junction. From the equations above, we obtain *for the edge of the **SCR***:

$$\frac{n_e^p}{n_e^n}\Big|_{\text{SCR edge}} = \frac{n_h^n}{n_h^p}\Big|_{\text{SCR edge}} = \exp - \frac{e \cdot \Delta V}{kT} = \exp - \frac{e \cdot (V^n + U_{ext})}{kT}$$

- The *minority* carrier densities (*always at the edge of the **SCR** without indicating it anymore*) can now be written as

$$n_e^p(U_{ext}) = n_e^n(U_{ext}) \cdot \exp - \frac{e \cdot (V^n + U_{ext})}{kT} \quad \text{electrons on the p-side}$$

$$n_h^n(U_{ext}) = n_h^p(U_{ext}) \cdot \exp - \frac{e \cdot (V^n + U_{ext})}{kT} \quad \text{holes on the n-side}$$

- These equations are nothing but the Boltzmann distribution giving the number of particles (n_{min}) that make it to the energy $e(V^n + U_{ext})$ out of a total number n_{maj} – in thermal equilibrium. We used essentially the same equation before, but now we know the kind of approximations that were necessary and that means we also know what we would have to do for "better" solutions of the problem.

Since this is important, let's review the approximations we made:

- Besides the "abrupt" "large" junction, we used the approximations from the simple solution to the Poisson equation which implies that the potential stays constant right up to the edge of the **SCR** and then changes monotonously.
- This means that for equilibrium we must obtain the same equations by computing the minority carrier density from the mass action law, i.e.

$$n_e^p(U_{ext}=0) = \frac{n_i^2}{n_h^p(U_{ext}=0)}$$

- We will see if this is true in a little exercise:

Exercise 2.3.5-1

Show the equivalence of the two equations
for the minority carrier density!

Now comes a crucial point: We are looking at *stationary non-equilibrium*. We first review the starting point again:

- At equilibrium ($U_{ext} = 0$), the *majority* carrier densities $n_e^n|_{SCR\ edge}$ and $n_h^p|_{SCR\ edge}$ are given by

$$n_h^p = N_{eff}^p \cdot \exp - \frac{E_F}{kT}$$

$$n_e^n = N_{eff}^e \cdot \exp + \frac{E_F - E_C^n}{kT}$$

- Do you remember them? These are two of our first equations from above, but given here for the choice of $E_V^p = 0$.

The essential point for the majority carrier density at the edge of the space charge region for *non-equilibrium* is that it *remains practically unchanged* (approximately at its bulk value) if we now apply a voltage U_{ext}, i.e.

$$n_{e,h}^{n,p}(U_{ext})\Big|_{edge}^{SCR} = n_{e,h}^{n,p}(equ)\Big|_{edge}^{SCR} = n_{e,h}^{n,p}(bulk)$$

- The trick here is that we consider the majority carrier density *at the **SCR** edge* – and the position of the latter may vary with the applied voltage!
- Nevertheless, beyond that point we have the bulk behaviour of the majorities – because that's how we have defined the **SCR** edge: The bulk potential stays constant right up to the edge, and this is only possible for a constant density of majority carriers.

The minority carrier densities $n_e^p|_{SCR\ edge}$ and $n_h^n|_{SCR\ edge}$, however, depend *very much on the applied voltage* as expressed in the formulae above.

- Thus, we have to adjust the minority carrier density independent of the majority density, which means we have to use the **quasi-Fermi energies**.
- In other words: While the *quasi-Fermi energy* E_F^{maj} for majority carriers remains at the equilibrium value E_F near the **SCR**, the *quasi-Fermi energy* for the minority carriers, E_F^{min}, branches off early; the details will be shown below.

We now ask about the *difference of the minority carrier density relative to equilibrium*, i.e. we look at

$$\Delta n_{e,h}^{p,n}\Big|_{\text{edge}}^{\text{SCR}} = n_{e,h}^{p,n}(U_{ext}) - n_{e,h}^{p,n}(U_{ext}=0)$$

- It comes out as

$$\Delta n_{e,h}^{p,n} = n_{e,h}^{n,p} \cdot \left(\exp - \frac{e \cdot (V^n + U_{ext})}{kT} - \exp - \frac{eV^n}{kT} \right)$$

$$= n_{e,h}^{n,p} \cdot \exp - \frac{eV^n}{kT} \cdot \left(\exp - \frac{eU_{ext}}{kT} - 1 \right)$$

Inserting the general expressions for the minority carrier density from above for the case $U_{ext} = 0$ yields the final formula for our first goal:

$$\Delta n_{e,h}^{p,n}\Big|_{\text{edge}}^{\text{SCR}} = n_{e,h}^{p,n}(equ) \cdot \left(\exp - \frac{eU_{ext}}{kT} - 1 \right)$$

- In other words: The density of minority carriers at the edge of the **SCR** will be changed by an external voltage.

In steady state conditions (which does *not* imply equilibrium, just that nothing changes) this density must remain constant as a function of time.

- Since deep in the material the minority carrier density is unchanged and has its equilibrium value, we now must have a current, driven by the density gradient alone, and *this current must be maintained by the voltage/current source* if we want steady state.
- Physically speaking, the excess density of minority carriers will diffuse around and disappear after some diffusion lengths – deep in the material they are not noticeable any more.

This is exactly the situation treated under "useful relations" for pure diffusion currents.

- We can take the formula derived there with $\Delta n_{p,n}^{e,h}(x=0)$ given by the equation from above and obtain immediately for the **current–voltage relationship** of a **p–n** junction (just considering the absolute magnitudes):

$$|j_e(U_{ext})| = \frac{e \cdot D_e}{L_e} \cdot \Delta n_e\Big|_{\text{edge}}^{\text{SCR}} \quad \text{or}$$

$$|j_e(U_{ext})| = \frac{e \cdot D_e}{L_e} \cdot n_e^p(equ) \cdot \left(\exp - \frac{eU_{ext}}{kT} - 1 \right)$$

$$|j_h(U_{ext})| = \frac{e \cdot D_h}{L_h} \cdot n_h^n(equ) \cdot \left(\exp - \frac{eU_{ext}}{kT} - 1 \right)$$

We now see that the external voltage, as we have introduced it, raises the potential barrier and therefore decreases the minority carrier density – and, thus, also the current flow.

- This means that, in order to *enhance* the current flow over the p–n junction, we have to apply the external voltage in a way that it *lowers* the barrier.
- Therefore, the forward voltage is $U_D := -U_{ext}$, and since it is the forward voltage, it is also the one which is taken as positive; the subscript "D" refers to the p–n junction functioning as a **diode**.

For the *final result* we add the electron and hole currents, drop suffixes and functional arguments now unnecessary, and obtain the **diode equation** (giving the total current density, including the reverse current, counted in the standard way):

$$j_D(U_D) = \left(\frac{e \cdot n_e^p \cdot D_e}{L_e} + \frac{e \cdot n_h^n \cdot D_h}{L_h} \right) \cdot \left(\exp \frac{eU_D}{kT} - 1 \right)$$

This is the same equation as before if we take into account *that the pre-exponential factor can be written in many ways*. To see that, we use the following identies:

- For the diffusion length we have

$$L_{e,h} = \left(D_{e,h} \cdot \tau_{e,h} \right)^{1/2}$$

$$D_{e,h} = \frac{L_{e,h}^2}{\tau_{e,h}}$$

$$\tau_{e,h} = \frac{L_{e,h}^2}{D_{e,h}}$$

- From the mass action law, which is still valid for the bulk, and the general approximation for the majority carrier density (that is already contained in our equations) we get

$$n_{e,h}^{p,n} = \frac{n_i^2}{n_{h,e}^{p,n}}$$

$$n_e^p = \frac{n_i^2}{N_A}$$

$$n_h^n = \frac{n_i^2}{N_D}$$

Shuffling everything around with these identities gives us – among many other equivalent formulations – . . .

$$j_D(U_D) = \left(\frac{e \cdot L_e \cdot n_i^2}{\tau_e \cdot N_A} + \frac{e \cdot L_h \cdot n_i^2}{\tau_h \cdot N_D} \right) \cdot \left(\exp \frac{eU_D}{kT} - 1 \right)$$

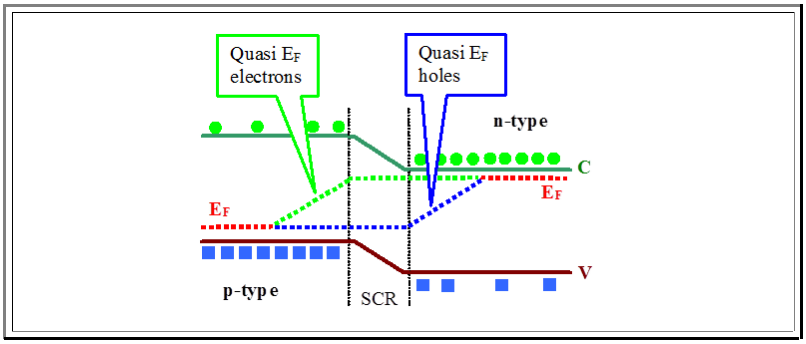
- . . . and that is exactly the equation we got before! However, we did not have to "cut corners" this time and we did not have to assume that some proportionality constant equals **1**!

More important, however: The interpretation of what happens may now be different. *Different* in the sense of looking at one and the same situation from a *different point of view*, not different in the sense that it is something else. The two points of view are complementary and not mutually exclusive; neither one is wrong!

- In the simple picture we looked at the minority carriers that had to be *generated* to account for the *loss of carriers accounting for the reverse current* and running down the energy slope.
- Here we looked at the *surplus of minorities accounting for the forward current* and which has to be moved away from the junction.
- Think about why this is the same thing! (Hint: Start from $U_D = 0$.)

What is left is just to consider the *quasi-Fermi energies* relevant for the forward direction; not only was the relevant drawing promised already above, it will also show explicitly what is meant by "surplus of minorities, having to be moved away from the junction" – because it will show us where those minorities end up.

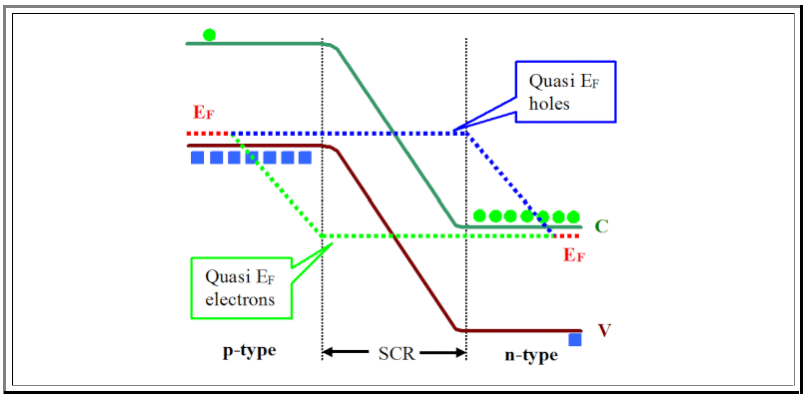
- To cut a long story short, here it is:



  - That the quasi-Fermi energies of the majorities remain constant throughout the **SCR** corresponds to the expressions giving the ratio of each carrier type on both sides of the junction.
- Note that in the dawing, deliberately there are more minority carriers close to the **SCR** edges than deeper in the bulk. Yes, that's where the surplus minorities go. But that's not the end of the story:
  - That the quasi-Fermi energies of the minorities outside the **SCR** linearly merge towards the majorities' ones corresponds to the exponential decay of the surplus minority density away from the **SCR**, with the decay constant given by the diffusion length – as already discussed for the case of pure diffusion currents.
  - Think for yourself about why all this is the case! And think about the possible consequences of the surplus minorities' presence in the case of a direct semiconductor.

Contributions from the Space Charge Region

- We now should include the **generation currents from the space charge region**, as we did (in a somewhat fishy way) in our simple consideration of a junction.
  - This, however, is not so easy to do in a correct (albeit still very approximate) fashion.
- For the *reverse part of the generation current* from the **SCR**, we can obtain an equation directly from the Shockley-Read-Hall theory. All we have to do is to consider the **quasi-Fermi energies** of a *junction in reverse bias*. This is schematically shown in the following picture:



  - The quasi-Fermi energies must behave in the way shown (the details do not matter), because otherwise the density of charge carriers (especially minority carriers!) in the junction would be too high.
  - Note that in the dawing there are no minority carriers close to the **SCR** edges (deliberately!); only in those regions away from the **SCR**, where there is a single Fermi energy (shown in red), minority carriers are depicted. There, the standard full-equilibrium mass action law holds.
- The decisive point is that we may consider any given thin slice of the **SCR** to be in *local equilibrium*, and that *the quasi-Fermi energy of the electrons is lower than that of the holes* throughout the **SCR**.
  - The latter is a direct consequence of the applied reverse bias, increasing and steepening the potential barrier in the **SCR**, in combination with the diffusion length of the minorities being larger than the width of the **SCR** (remember the *narrow junction approximation* from above).
  - This ordering of the quasi-Fermi energies is the exact opposite of the situation that we have considered so far in the recombination business, where we looked at an *increased density of carriers*, e.g. produced by irradiation with light. Then recombination outweighs generation and $\boldsymbol{U_{DL}}$, the difference between recombination and generation, was positive.
- Hence, in the case we are considering here, $\boldsymbol{U_{DL}}$ *is negative*, i.e. there is more generation than recombination. And this means that the space charge region is busily producing carriers, always in pairs because of neutrality, which will run down the energy barrier producing *an additional reverse current*.

- Pair production means that a deep level first emits a hole to the valence band, and then an electron to the conduction band.
- Let's look at this using the formula for U_{DL}:

$$U_{DL} = \frac{v \cdot \sigma \cdot N_{DL} \cdot (n_e \cdot n_h - n_i^2)}{n_e + n_h + 2n_i \cdot \cosh \frac{E_{DL} - E_{MB}}{kT}}$$

- For making estimates easier, we assume a mid-band level (i.e., $\cosh[(E_{DL} - E_{MB})/(kT)] = 1$) and $n_e, n_h << n_i$. This leaves us with

$$U_{DL} = -\frac{v \cdot \sigma \cdot N_{DL} \cdot n_i}{2}$$

- For these assumptions we have seen that, treating holes and electrons on equal footing, $1/(v \cdot \sigma \cdot N_{DL}) = \tau$.
- However, because we now have *more generation* than recombination, τ is now called the **generation life time** τ_G for this case. (More to that topic in the link.)
- This leaves us with a *net generation of one kind of carrier* of

$$|U_{DL}| = G = \frac{n_i}{2\tau_G}$$

The current density from the net generation of carriers in the **SCR** is then given by the product of the net generation rate with the width ***d*** of the **SCR**; adding up the holes and the electrons yields

$$j_R(SCR) = \frac{e \cdot n_i \cdot d}{\tau_G}$$

- This is exactly the same formula (give or take a factor of **2**) as in our "quick and dirty" estimate from before. The physical reasoning wasn't so different either, if you think about it.

How about the contribution of the **SCR** to the *forward current*?

- The proper treatment is much more complicated and physically different from our simple explanation. The physical reasoning is as follows:
- We have seen that we need to sustain a certain density of surplus minority carriers, $\Delta n_{e,h}^{p,n}$, at the edges of the **SCR** to maintain local equilibrium. The surplus carriers needed were injected from the other side of the junction and crossed the junction *without losses* – at least in our present approximation.
- In reality, however, *some injected holes from the **p**-side will recombine with the injected electrons from the **n**-side*. Recombination in the **SCR** thus reduces the current needed to maintain $\Delta n_{e,h}^{p,n}$ and an additional current has to be produced which exactly compensates the losses.

The necessary calculations are shown in an advanced module, suffice it to state here that the final result for the forward current from the **SCR** is (in a rather crude approximation)

$$j_F(SCR) = \frac{e \cdot n_i \cdot d}{2\tau_G} \cdot \exp \frac{eU_D}{2kT}$$

- Again, besides the factor **2** (and the new kind of life time), the same formula as before. But this time it was a kind of lucky coincidence, not really very well justified.
- *Or was it*? Think about it!