Chapter 5

Excess Carriers and the Continuity Equations

Non-Equilibrium Conditions: Excess Carriers

- Whenever we have a departure from equilibrium, carrier concentrations are disturbed. We describe the deviation from equilibrium by introducing the concept of EXCESS CARRIERS. In the static case we can write:

$$n(x) = n_o(x) + \Delta n(x)$$
$$p(x) = p_o(x) + \Delta p(x)$$

...where $\Delta n(x)$ is the excess electron concentration and $\Delta p(x)$ is the excess hole concentration (either may be negative).

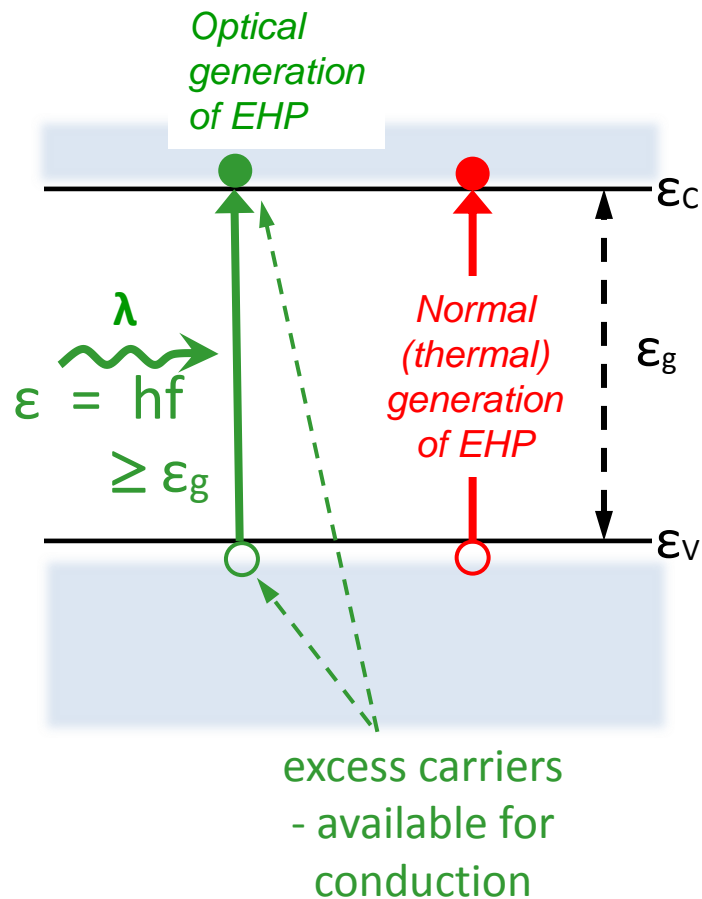- In the more general time-dependent case:

$$n(x,t) = n_o(x) + \Delta n(x,t)$$
$$p(x,t) = p_o(x) + \Delta p(x,t)$$

How Are Excess Carriers Produced?

- Any stimulus that perturbs thermal equilibrium will produce excess carriers, e.g.
    - Apply a voltage (this usually leads to "injection" of excess carriers from another region);
    - Irradiate with light (i.e. excess carriers are "generated");
    - Apply heat selectively;
    - Bombard with radioactivity…
- While $n_o(x) \neq p_o(x)$ in general, charge neutrality **means that we must always have equal concentrations of excess carriers** in a given material:

$$\Delta n(x, t) = \Delta p(x, t)$$

Example: Optical Generation of EHP's

Optical generation of EHP

Normal (thermal) generation of EHP

λ

$\varepsilon = hf$

$\geq \varepsilon_g$

ε_C

ε_g

ε_V

excess carriers - available for conduction

Photo-conductivity

- The increase in conductivity due to the excess carriers produced by the light is called *photoconductivity*

- This effect is used in light sensors, exposure meters, solar cells...

λ

z

semiconductor

More on this In SSE2
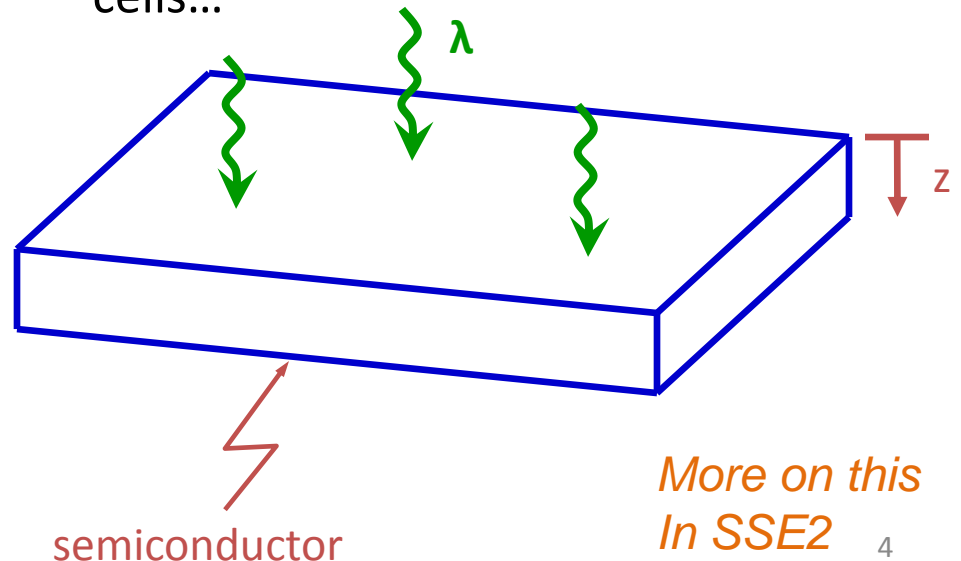
"Low-Level" of Excess Carriers

- As we have seen, in doped or extrinsic semiconductors typically $n_o \gg p_o$ (or $p_o \gg n_o$). Excess carriers are produced in equal numbers.

- The effect of excess carriers is usually much greater on the **minority** carrier concentration than on the majority carrier concentration

- E.g.: Consider a Si sample with N-type 10^{15} donors/cm^3 and with excess carrier concentrations of $\Delta n = \Delta p = 10^{13}$/cm^3 (say produced by shining light on the material). At 300K in thermal eq'm. we have $n_o = 10^{15}$/cm^3 and $p_o = [(1.5 \times 10^{10})^2/n_o] = 2.25 \times 10^5$/cm^3. With the excess carriers present the concentrations become: $n = 1.01 \times 10^{15}$/cm^3 and $p \approx 10^{13}$/cm^3 - i.e. the majority concentration is barely affected but the minority concentration is massively changed.

- **Whenever the excess carrier production is such that the effect on the <u>majority</u> carriers can be neglected, we speak of "low-level" injection/generation of excess carriers**

Recombination

- Recombination refers to the process whereby an electron transfers from the CB to the VB and annihilates a hole (EHP removal)

- Recombination is always present. Even when there is only thermal equilibrium, it then acts to eliminate the EHPs generated thermally

- It is also the mechanism by which carrier concentrations in excess of equilibrium (i.e. the excess carriers) are removed over time after the stimulus that produced them is removed
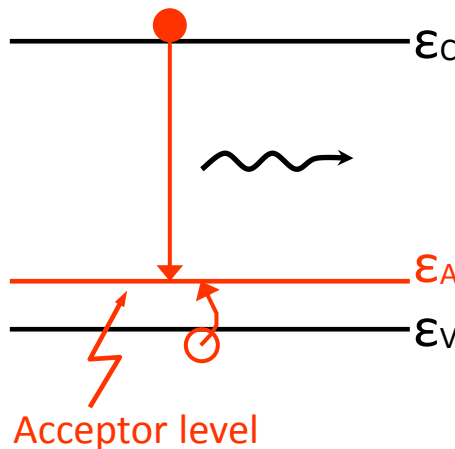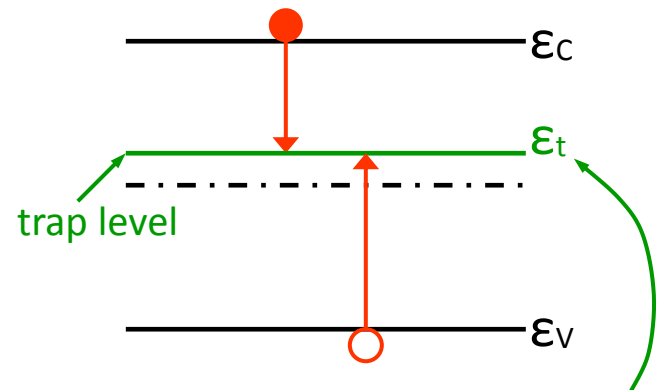
Methods of Recombination

Direct Recombination

$$\varepsilon_C$$

$$\varepsilon_g$$

Photon of light emitted

$$\varepsilon_V$$
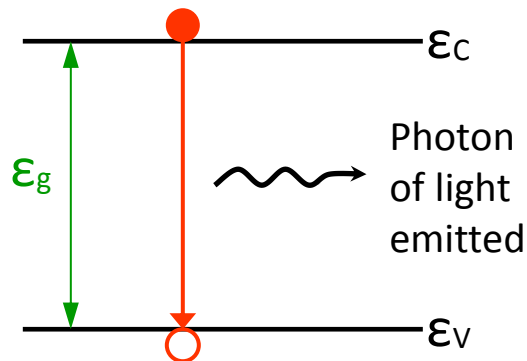
More common in direct band-gap materials e.g. GaAs.

Band-to-Impurity Recombination

$$\varepsilon_C$$

$$\varepsilon_A$$

$$\varepsilon_V$$

Acceptor level

Trap-Assisted Recombination

$$\varepsilon_C$$

$$\varepsilon_t$$

trap level

$$\varepsilon_V$$

A "trap level" is an energy level in the forbidden gap caused e.g. by Imperfections in the crystal

Common in Silicon

(Shockley-Read-Hall Recombination Analysis)

Generation and Recombination

- In a semiconductor under general conditions, electrons and holes (EHPs) are continually being generated and eliminated;

- We describe this mathematically using the rate term:

$$G - R \quad \text{or} \quad G(x,t) - R(x,t)$$

- Where G is the GENERATION RATE, defined as the average number of EHPs produced per unit volume per unit time (units of: /(cm^3.sec) or /(m^3.sec));

  …and R is the RECOMBINATION RATE, defined as the average number of EHPs removed per unit volume per unit time;

- In *thermal equilibrium*, these must exactly balance:

$$G_o = R_o$$

The Recombination Rate R

- Both an electron and a hole are required for a recombination event to take place, and so it seems reasonable to propose that R should be proportional to the product of "n" and "p":

- We write: $R(x,t) = \alpha.n(x,t).p(x,t)$ where α is a constant

- In thermal equilibrium, $R_o = \alpha.n_o.p_o = \alpha.n_i^2 = G_o$

- Consider an N-type material with low-level excess carriers. We will focus on the minority carriers (i.e. the holes). Substituting for n and p (and using the fact that $\Delta n = \Delta p$):

$$R = \alpha.(n_o + \Delta n).(p_o + \Delta p)$$

$$= \alpha.n_o.p_o + \alpha.(n_o+p_o).\Delta p + \alpha.\Delta n.\Delta p$$

(i.e. the last term is negligible under low-level excess carrier conditions, where Δn (and Δp) $<< n_o$)

G – R: Minority Holes in N-Type Material

- An important special case is where low-level excess carriers exist but <u>no generation</u> mechanism (other than thermal generation) is at work;

- Consider minority holes in an N-type material in this case with a donor concentration of N_D, so that $n_o = N_D$. Then:

$$(G - R) = (G_o - R)$$

$$= (G_o - \alpha \cdot n_o \cdot p_o - \alpha \cdot (n_o + p_o) \cdot \Delta p)$$

$$\approx - \alpha \cdot N_D \cdot \Delta p$$

(these terms cancel)

(since $n_o \gg p_o$ in N-type, and $n_o \approx N_D$)

- Earlier we saw: $G_o = \alpha \cdot n_i^2$, and so:

$$\alpha = \frac{G_o}{n_i^2}$$

G – R: Minority Holes in N-Type Material

- This means that the net or effective Generation-Recombination rate (G – R) is given by:

$$G - R = -\left(\frac{G_o \cdot N_D \cdot \Delta p}{n_i^2} \right) = -\frac{\Delta p}{\tau_p}$$

where τ_p is defined as the ***minority hole lifetime***

$$\tau_p = \frac{n_i^2}{G_o \cdot N_D}$$

(units of seconds)

- The *effective recombination rate* is: $\left(\Delta p / \tau_p \right)$

- Using the definition of Δp, another way of writing (G - R) is:

$$G - R = -\frac{\Delta p}{\tau_p} = -\left(\frac{p(x,t) - p_o(x)}{\tau_p} \right)$$

G – R: Minority Electrons in P-Type Material

- A similar analysis applies to excess minority electrons in a P-type material with an acceptor concentration of N_A so that $p_o = N_A$.

- We then have:

$$G - R = -\frac{\Delta n}{\tau_n} = -\left(\frac{n(x,t) - n_o(x)}{\tau_n}\right)$$

- Where τ_n is defined as the ***minority electron lifetime***

$$\tau_n = \frac{n_i^2}{G_o \cdot N_A}$$

- The *effective recombination rate* in this case $= \left(\frac{\Delta n}{\tau_n}\right)$

Example 5.1

Effect of Excess Carriers on Conductivity

- In a semiconductor in thermal equilibrium:
$$\sigma_o = q \cdot \left[\mu_n \cdot n_o + \mu_p \cdot p_o\right]$$

- With Excess Carriers present, the conductivity becomes:
$$\sigma = q \cdot \left[\mu_n \cdot (n_o + \Delta n) + \mu_p \cdot (p_o + \Delta p)\right]$$

- Consider the special case of an N-type doped semiconductor with N_D donors per unit volume (not lightly doped, T > ~100K):

$$\sigma_o = q \cdot \left[\mu_n \cdot n_o + \mu_p \cdot p_o\right] = q \cdot \left[\mu_n \cdot N_D + \mu_p \cdot \left(n_i^2/N_D\right)\right]$$

- Normally, we can neglect the minority (hole) concentration p_o compared to the majority electron conc.: $\sigma_o \cong q \cdot N_D \cdot \mu_n$

- With Excess Carriers present:
$$\sigma \cong q \cdot \left[\mu_n \cdot (N_D + \Delta n) + \mu_p \cdot (\Delta p)\right]$$

Example 5.2

Quasi-Fermi Levels

- In a non-degenerate semiconductor in thermal equilibrium, we have seen that:

$$n_o = n_i \cdot \exp\left[\frac{\varepsilon_f - \varepsilon_i}{kT}\right] \qquad p_o = n_i \cdot \exp\left[\frac{\varepsilon_i - \varepsilon_f}{kT}\right]$$

Note : $n_o p_o = n_i^2$

ε_f :   Fermi energy level in material : **note that thermodynamic arguments show that it is everywhere uniform in space in thermal equilibrium**.

ε_i :   Intrinsic level : i.e. the Fermi level in a pure or intrinsic material

n_i :   the concentration of EHPs in Intrinsic material (intrinsic concentration)

- We can extend these concepts to the _non-equilibrium_ case using the idea of _quasi-Fermi_ levels (for electrons and holes)

Quasi-Fermi Levels (or "imref"s)

- In the *non-equilibrium* case we have *excess carriers* and write:

$$n = n_o + \Delta n \qquad p = p_o + \Delta p$$

- Then the **electron quasi-Fermi** level ε_{fn} is defined as:

$$n = n_o + \Delta n = n_i \exp\left[\frac{\varepsilon_{fn} - \varepsilon_i}{kT}\right]$$

- In equilibrium, $\varepsilon_{fn} = \varepsilon_f$ but the greater the concentration of excess electrons Δn, the more they deviate.

- A similar definition can be used for the **hole quasi-Fermi** level ε_{fp}:

$$p = p_o + \Delta p = n_i \exp\left[\frac{\varepsilon_i - \varepsilon_{fp}}{kT}\right]$$

- Note that in non-equilibrium conditions: $n \cdot p \neq n_i^2$

Continuity Equations for Electrons & Holes

- The continuity equations are a kind of book-keeping or accounting device that keep track of carrier numbers within a given volume

- We consider a (1D) electron concentration *n(x,t)* and a hole concentration *p(x,t)* and suppose initially we have a bulk region of semiconductor with no current flows. Then for electrons:

$$\frac{\partial n(x,t)}{\partial t} = G(x,t) - R(x,t)$$

$$\begin{pmatrix} \text{Rate of change} \\ \text{of no. of electrons} \\ \text{per unit volume} \end{pmatrix} = \begin{pmatrix} \text{Generation} \\ \text{rate} \end{pmatrix} - \begin{pmatrix} \text{Recombination} \\ \text{rate} \end{pmatrix}$$

- While a similar equation also applies to holes:

$$\frac{\partial p(x,t)}{\partial t} = G(x,t) - R(x,t)$$

Continuity Equations (no Current Flows)

- An example of the situation just described would be the uniform illumination of a region of uniformly-doped N-type semiconductor with light (with $\mathcal{E} > \mathcal{E}_g$): excess carriers are produced but no currents will flow if we assume that no voltage is applied.

- Suppose the light is switched off at time t = 0. Let the initial concentration of excess electrons and holes be Δn_o and Δp_o, respectively. How do these excess concentrations vary with time?

- Earlier we showed that in this situation with t >= 0, for the minority holes:

$$G - R = -\frac{\Delta p(x,t)}{\tau_p}$$

19

Continuity Equation for Minority Carriers (no Current Flows)

- Note that the equilibrium concentrations can vary with distance in general, but <u>not</u> with time e.g. *p(x,t)=p$_o$(x)+Δp(x,t),* so that:

$$\frac{\partial p(x,t)}{\partial t} = G(x,t) - R(x,t) \implies \frac{\partial \Delta p(x,t)}{\partial t} = -\frac{\Delta p(x,t)}{\tau_p}$$

- If all concentrations are uniform in space, this becomes an ordinary differential equation:

$$\frac{d\Delta p(t)}{dt} = -\frac{\Delta p(t)}{\tau_p}$$

- For t >= 0, the solution is well-known and takes the form of an exponential decay:
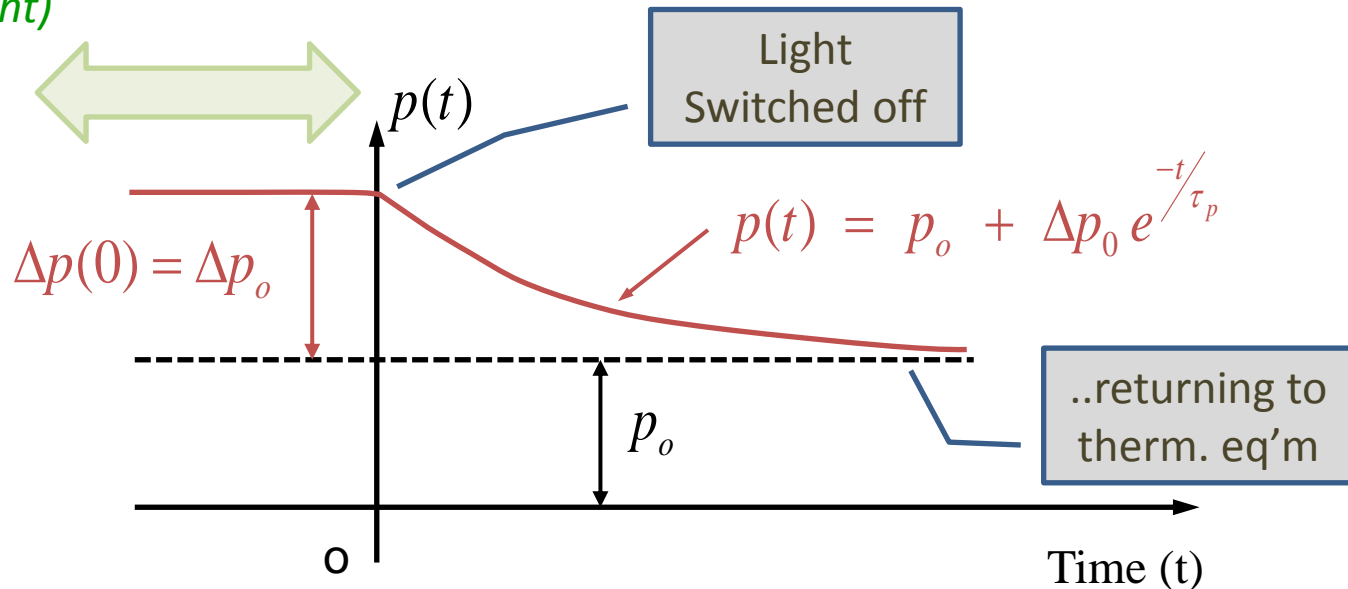
$$\Delta p(t) = \Delta p_o \cdot \exp\left(-\frac{t}{\tau_p}\right)$$

Continuity Equation for Minority Carriers (no Current Flows)

- The time-variation of the total hole concentration can then be easily found:
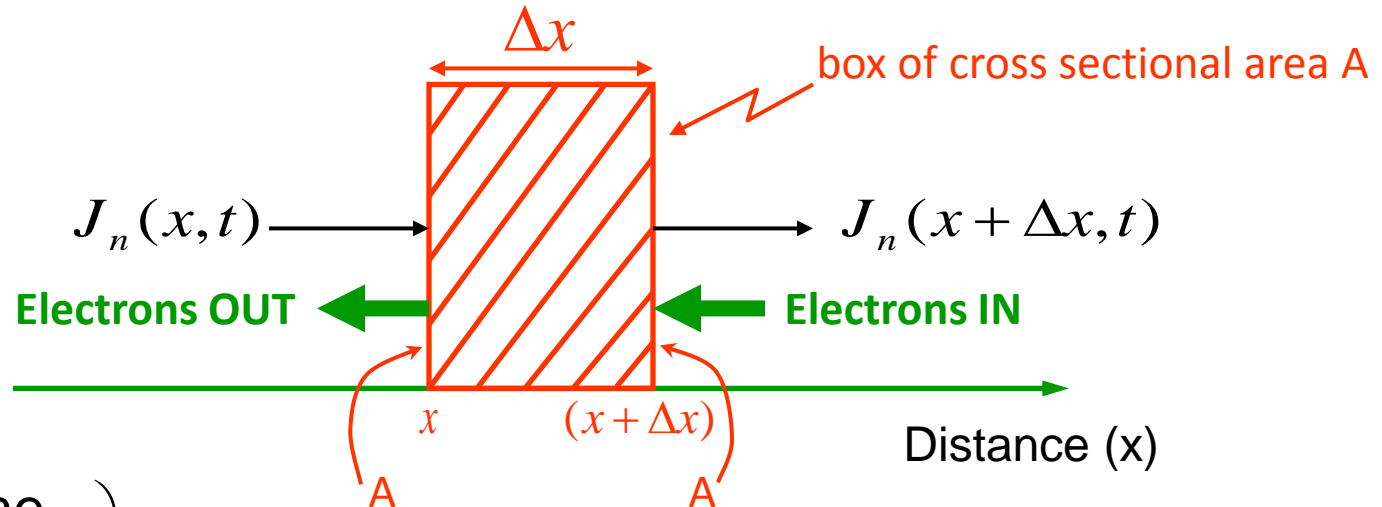
$$p(t) = p_o + \Delta p(t) = p_o + \Delta p_o \cdot \exp\left(-\frac{t}{\tau_p} \right)$$

Excess Carriers Produced

(e.g. by light)

$p(t)$

Light Switched off

$\Delta p(0) = \Delta p_o$

$p(t) = p_o + \Delta p_0\, e^{-t/\tau_p}$

p_o

..returning to therm. eq'm

o

Time (t)

Example 5.3

The Continuity Equations : General Case of Electrons in 1D

$$\Delta x$$

box of cross sectional area A

$$J_n(x, t)$$ $$J_n(x + \Delta x, t)$$

Electrons OUT **Electrons IN**

$$x \qquad (x + \Delta x)$$

Distance (x)

A A

$$\begin{pmatrix} \text{Overall time} \\ \text{rate of change} \\ \text{in the number} \\ \text{of electrons} \\ \text{in the box} \end{pmatrix} = \begin{pmatrix} \text{no. of electrons} \\ \text{entering the box} \\ \text{at } x + \Delta x \\ \text{per unit time} \end{pmatrix} - \begin{pmatrix} \text{no. of electrons} \\ \text{leaving the box} \\ \text{at } x \text{ per unit time} \end{pmatrix}$$

$$+ \begin{pmatrix} \text{no. generated} \\ \text{in box} \\ \text{per unit time} \end{pmatrix} - \begin{pmatrix} \text{no. recombined} \\ \text{in box} \\ \text{per unit time} \end{pmatrix}$$

Connecting Current Density and Particle Numbers

J$_n$

A
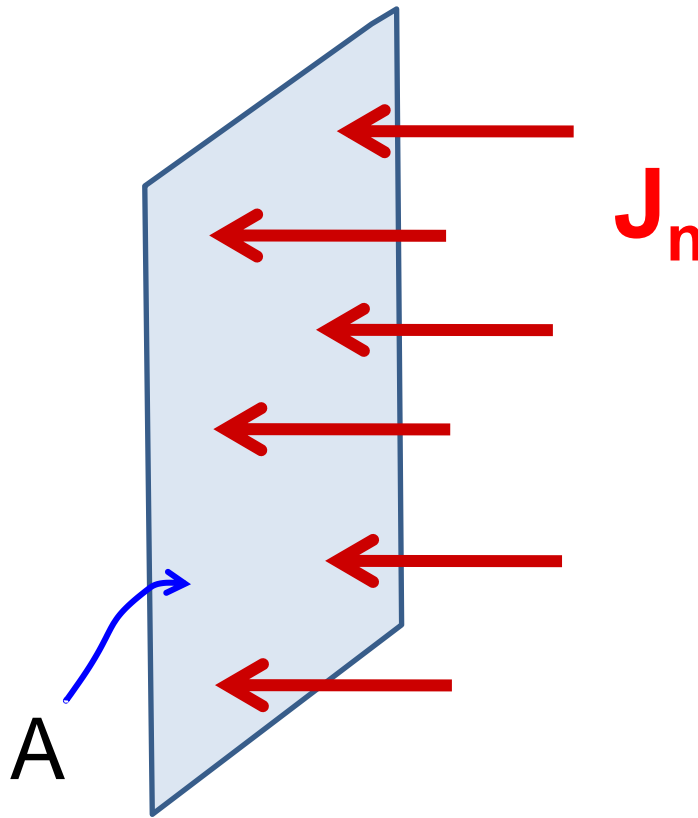
Units of Current Density J$_n$ are Amp/m^2 = Coul/(m^2.sec)

...but each electron has a charge of "q". Hence the total number of electrons passing through "A" each second is:

$$\left(\frac{1}{q} \cdot A \cdot J_n \right)$$

Continuity Equations : General Case

Note that the total no. of electrons in the box at time 't' $= n(x,t) \cdot A \cdot \Delta x$

$$\left(\frac{\partial \left(n(x,t) \, A \, \Delta x \right)}{\partial t} \right) = \left(\frac{1}{q} A \, J_n(x+\Delta x, t) \right) - \left(\frac{1}{q} A \, J_n(x,t) \right)$$

$$+ \left(G \, A \, \Delta x \right) - \left(R \, A \, \Delta x \right)$$

Cancel A, divide across by Δx and take the limit $\Delta x \to 0$

$$\frac{\partial n(x,t)}{\partial t} = \frac{1}{q} \lim_{\Delta x \to 0} \left[\frac{J_n(x+\Delta x, t) - J_n(x,t)}{\Delta x} \right] + G - R$$

This leads directly to the general form of the
ELECTRON CONTINUITY EQUATION:

$$\boxed{\frac{\partial n(x,t)}{\partial t} = \frac{1}{q} \frac{\partial J_n(x,t)}{\partial x} + G(x,t) - R(x,t)}$$

Continuity Equations for Electrons and Holes

- A similar analysis can be carried out for holes. The result is the final form of the continuity equations for electrons and holes in semiconductors. These provide the final 2 equations we need:

$$\frac{\partial n(x,t)}{\partial t} = \frac{1}{q} \cdot \frac{\partial J_n(x,t)}{\partial x} + G(x,t) - R(x,t) \qquad [4]$$

$$\frac{\partial p(x,t)}{\partial t} = -\frac{1}{q} \cdot \frac{\partial J_p(x,t)}{\partial x} + G(x,t) - R(x,t) \qquad [5]$$

N.B.

"Drift-Diffusion" Approximation

- In principle, we can identify 5 functions (in a 1-D case) which, if we knew as a function of space (x) and time (t), would completely describe the behaviour of a semiconductor device. These are:

$$n(x,t), \ p(x,t), \ J_n(x,t), \ J_p(x,t) \text{ and } E(x,t).$$

- In practice, we would need to specify geometrical dimensions ("*boundary conditions*"), material properties, excitation signals (if any) and the starting or "*initial conditions*".

- To find the 5 functions, we would then need 5 independent equations - these we have already found: two current density equations, one Poisson equation and the two continuity equations...collectively often called the drift-diffusion equations for a semiconductor.

The "Drift-Diffusion" Framework for the Physical Analysis of Semiconductors

$$J_n(x,t) = q \cdot n(x,t) \cdot \mu_n \cdot E(x,t) + q \cdot D_n \cdot \frac{\partial n(x,t)}{\partial x} \qquad [1]$$

$$J_p(x,t) = q \cdot p(x,t) \cdot \mu_p \cdot E(x,t) - q \cdot D_p \cdot \frac{\partial p(x,t)}{\partial x} \qquad [2]$$

$$\frac{\partial E(x,t)}{\partial x} = \frac{\rho(x,t)}{\varepsilon} = \frac{q}{\varepsilon} \cdot \left[p(x,t) - n(x,t) + N_D^+ - N_A^- \right] \qquad [3]$$

$$\frac{\partial n(x,t)}{\partial t} = \frac{1}{q} \cdot \frac{\partial J_n(x,t)}{\partial x} + G(x,t) - R(x,t) \qquad [4]$$

$$\frac{\partial p(x,t)}{\partial t} = -\frac{1}{q} \cdot \frac{\partial J_p(x,t)}{\partial x} + G(x,t) - R(x,t) \qquad [5]$$

TCAD – Technology Computer Aided Design

- In principle, we could describe any device using these equations (generalised to 3D), provided we knew the full geometry of the problem, the doping profiles, certain basic material data, the initial conditions in time etc

- Certain commercial CAD products provide such a "TCAD" or *technology-CAD* infrastructure (e.g. ATLAS from Silvaco Corp.)

- This is useful for device developers but in spite of much effort, the simulation complexity and time makes this approach unsuitable for most general-purpose engineering design, where we rely more usually on *equivalent circuit models*

- Note that despite their apparent complexity, these equations are too simple for modern nano-scale semiconductor devices - generalisations include the *Boltzmann Transport Equations*

# Semiconductor Measurement and Characterisation

- It is important to be able to determine by measurement certain basic properties of a grown semiconductor sample – doping level, minority lifetime, conductivity etc

- This is vital for quality control in manufacturing and also as an input to physical device modelling

- Very sophisticated techniques have been developed for this purpose – we just mention a few simple techniques here.

The Haynes-Schockley Experiment

- This is a classical semiconductor experiment. For an applet demonstration see:

 http://jas.eng.buffalo.edu/education/semicon/diffusion/diffusion.html

- The experiment works best when the minority carrier lifetimes τ_p or τ_n are fairly long (e.g. In Germanium (Ge))

- We inject a very short pulse of excess carriers at a plane of semiconductor (e.g. using an intense flash of light)

- Then monitor the flow of excess minority carriers. E.g:
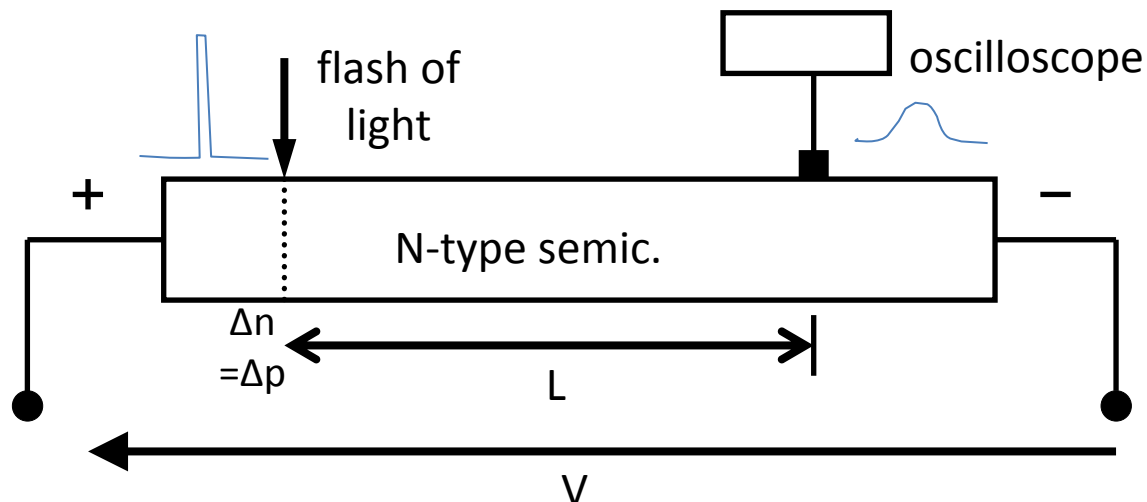
The Haynes-Shockley Experiment

It is found that we can solve the continuity equation analytically for this case.
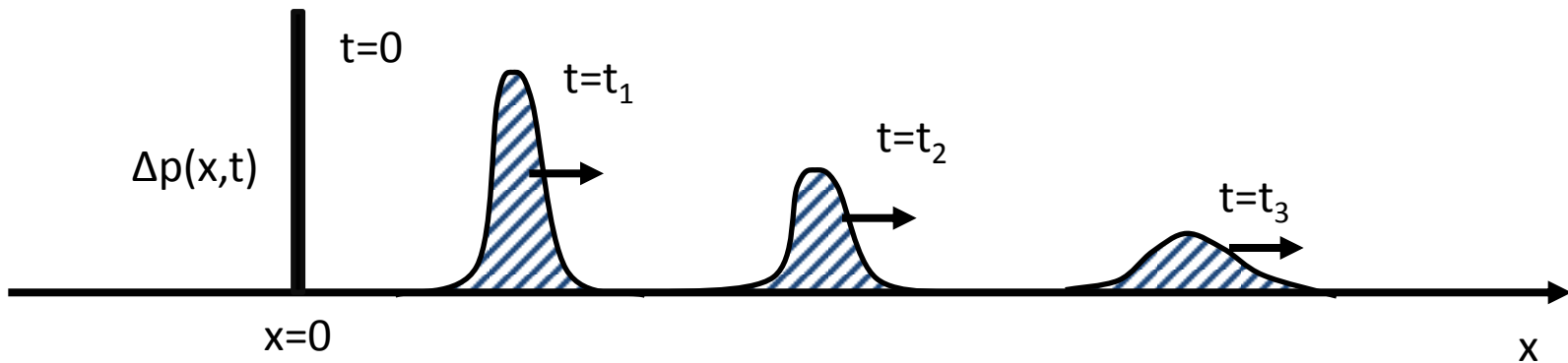
3 processes are at work:

1) drift - measure time taken as T, then since the distance L is known
   we can estimate the drift velocity as: $v_p \approx L/T$

2) diffusion - by measuring the shape of the pulse at a known T, it is
   possible to estimate D_p

3) recombination - the area under the pulse decreases overtime →
   can use this to estimate τ_p

Four-Point Probe Method for Measuring Conductivity

- Suppose we have a "semi-infinite" semiconductor half-plane. We set up 4 equi-spaced needle probes in a line pressed against the surface.   A current is forced through the two outer probes and the voltage dropped is sensed between the inner two probes

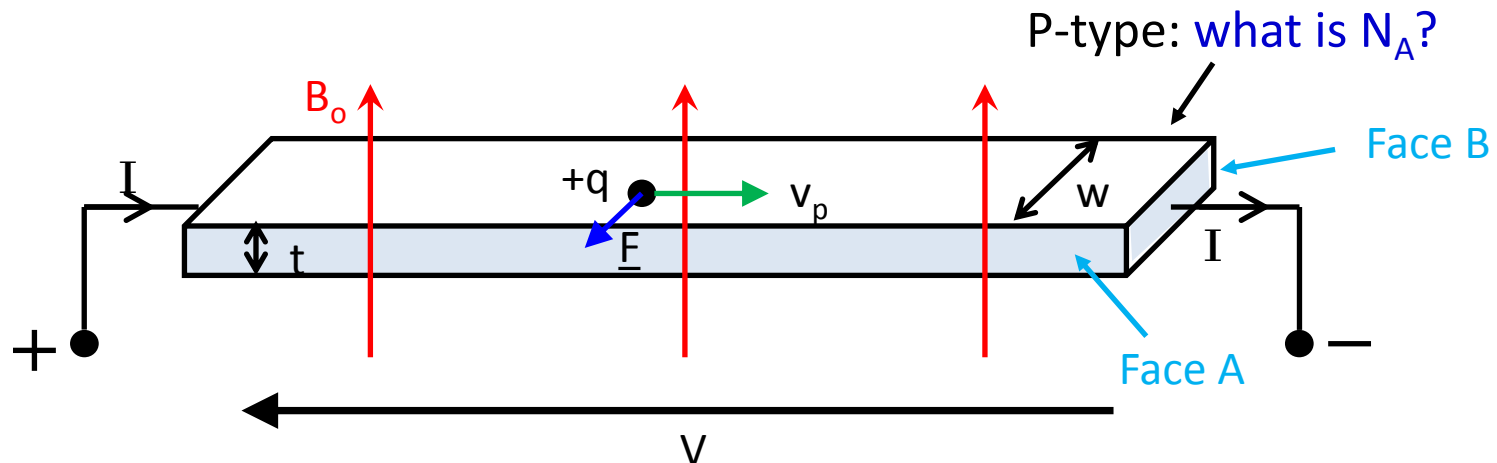"force needle"

"force needle"

I

~0

V

~0

I

d   d   d

semiconductor

"sense"   measure V (open-circuit)
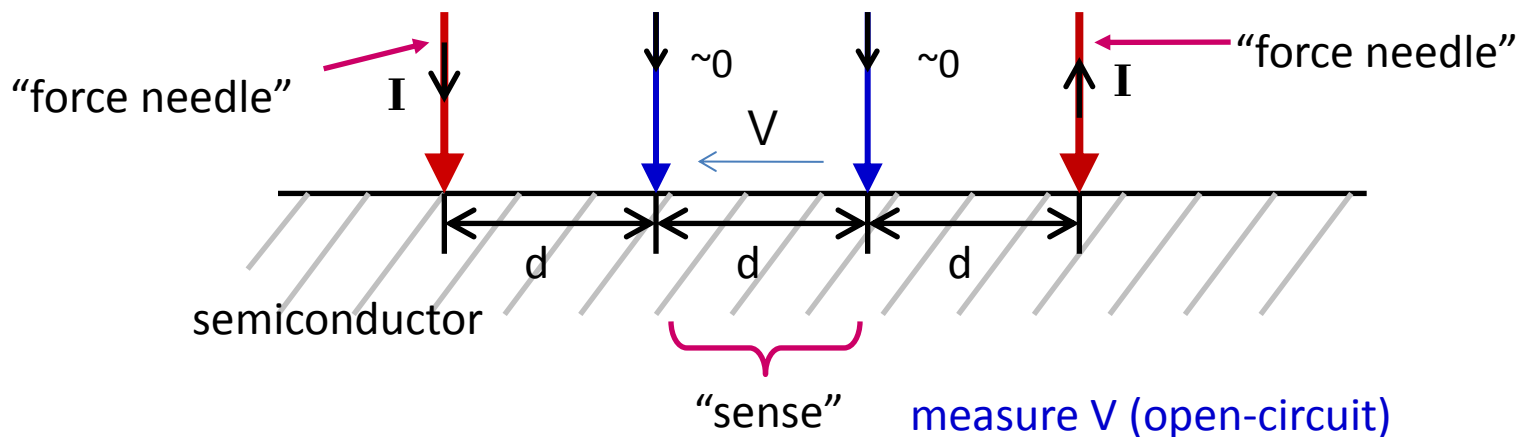
- It can be shown (assuming a semi-infinite semiconductor):

$$\sigma = \frac{1}{2\pi d}\left(\frac{I}{V}\right)$$

... so by measuring I and V, (and "d") we can determine the conductivity "σ"

The Hall Effect

The Hall Effect is used as a measurement technique to estimate doping concentrations in the semiconductor industry

With the magnetic flux density B_o present, a force acts on each moving (drifting) hole:

$$\underline{F} = q \cdot (\underline{v}_p \times \underline{B}_0)$$

This leads to an accumulation of holes at the sidewall of the block (Face A) and a deficit at Face B → we can measure an open-circuit voltage due to this phenomenon (the "Hall Voltage" V_H).

Semiconductor Measurements

V_H is the result of a lateral electric field E_y which opposes the Lorentz Force:

$$qv_p B_o = qE_y$$

Let J = current density through block, then we define:

$$R_H = \frac{E_y}{J \cdot B_o} = \frac{v_p B_o}{J \cdot B_o} = \frac{v_p}{qv_p p} = \frac{1}{qp}$$

neglecting diffusion

The Hall Coefficient

Knowing the dimensions of the block and measuring V_H, I, B_o we find:

$$\left(E_y = \frac{V_H}{w} \right) \quad \left(J = \frac{I}{wt} \right)$$

=> can determine R_H and therefore measure the majority carrier concentration p which is ≈ N_A

Example 5.4