Semiconductors II

Dan Weatherill

Band Theory II

Outline

- Semiclassical Transport
  - Boltzmann equation
  - Drift-diffusion model
  - Energy Balance & Hydrodynamic Models
- Semiconductor Physics
  - Thermodynamics of Semiconductors, Doping, Fermi Level, QFLs
  - The non-linear Poisson Equation
  - Generation, Recombination, Mobility, Diffusivity, Resistivity
- Semiconductor Device Building Blocks
  - the pn junction
  - the MOS capacitor

General Point about accuracy

Most of the mathematics we do here is only loosely coupled to reality - even sophisticated and careful commercial device simulations typically get actual measured quantities wrong by factors of 2 or more. With the basic equations we show here you will **NOT** get quantitatively accurate calculations for real devices

However, most of the behaviours (e.g. with Temperature, Field etc) will be roughly right. So it is useful to know these things to start developing a "feel" for how a device structure you look at will behave.

This is only the beginning - it takes a long time of looking at device structures and thinking carefully about how the various implants and applied potentials are meant to work to get a real intuition for what a device will actually do.

Before we begin - Semiconductor modelling

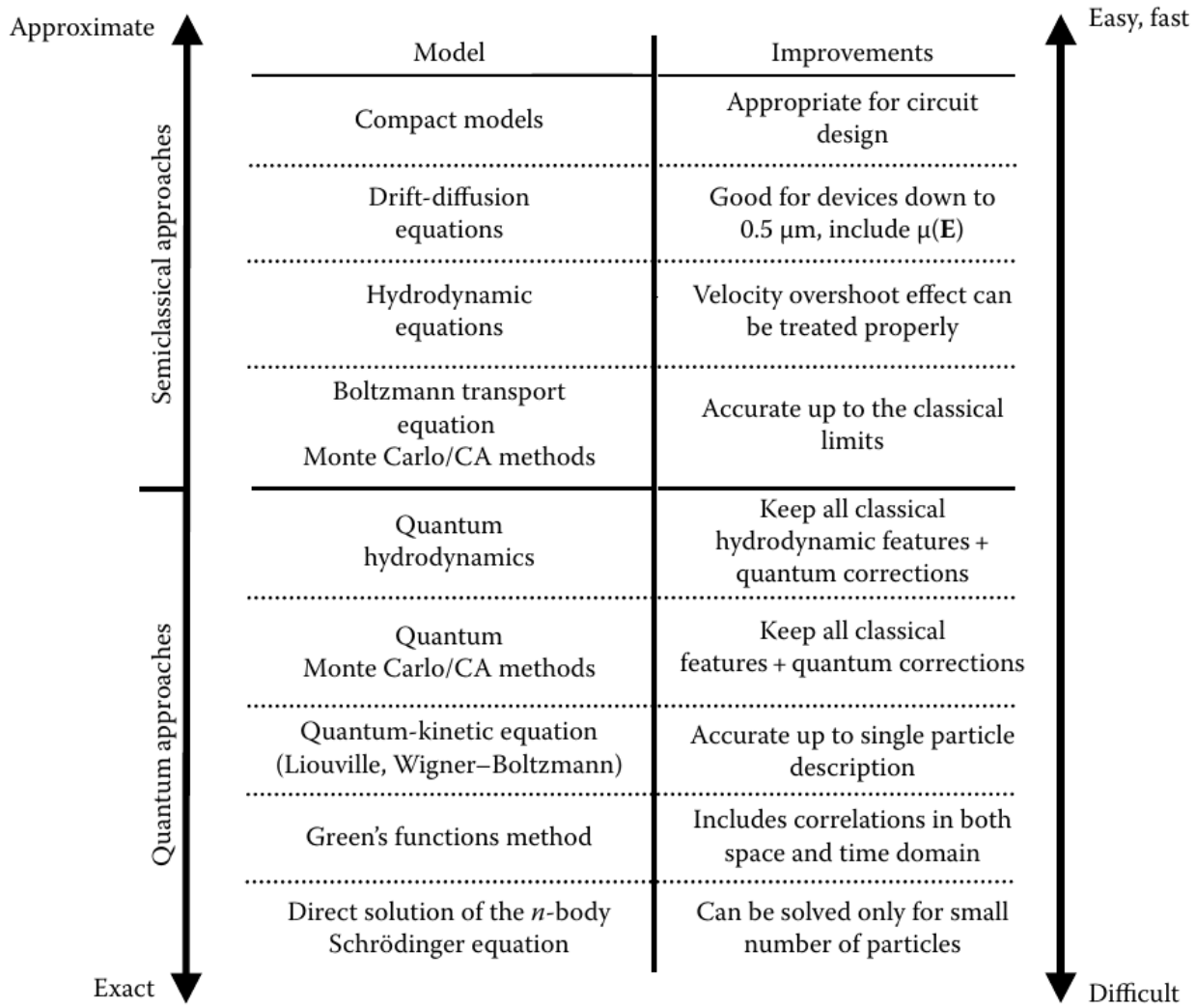
There are roughly 3 "levels" of semiconductor models:

- Quantum Mechanical: work out some actual solution (or approximate solution) to Schrodinger equation
- Kinetic: Treat electrons as scattering semi-classical particles, with various processes going on
- fluid-dynamical: Treat electrons and holes as (compressible, charged) fluids that move in response to applied fields

The **fluid-dynamical** models are accurate down to about ~500nm for the "Drift-Diffusion" model. Thus, for most semiconductor detector applications, they are adequate (at least in the pixel regions!). Trouble can happen at the times between when an energetic particle arrives and excites the lattice, and when the distribution settles down.

Hierarchy &
(image from Computational Electronics by Vasileska & Goodnick)

We will discuss (briefly!) the Boltzmann Transport Equation, and then concentrate on Drift-Diffusion (whilst mentioning in passing Hydrodynamic Equations)



Model	Improvements
Compact models	Appropriate for circuit design
Drift-diffusion equations	Good for devices down to 0.5 μm, include μ(**E**)
Hydrodynamic equations	Velocity overshoot effect can be treated properly
Boltzmann transport equation Monte Carlo/CA methods	Accurate up to the classical limits
Quantum hydrodynamics	Keep all classical hydrodynamic features + quantum corrections
Quantum Monte Carlo/CA methods	Keep all classical features + quantum corrections
Quantum-kinetic equation (Liouville, Wigner–Boltzmann)	Accurate up to single particle description
Green's functions method	Includes correlations in both space and time domain
Direct solution of the *n*-body Schrödinger equation	Can be solved only for small number of particles

The Boltzmann Equation

Recall, from last time, we have already given up on a fully quantum mechanical approach. The semi-classical approach starts from the point of assuming that we can validly talk about having single particle wavefunctions, which we can regard as having classical-like dynamics.

We start by introducing the **phase space distribution function** of electrons, $f(\mathbf{r}, \mathbf{p}, t)$, which represents the probability of finding an electron at position $\mathbf{r}$, with momentum $\mathbf{p}$ at time t.

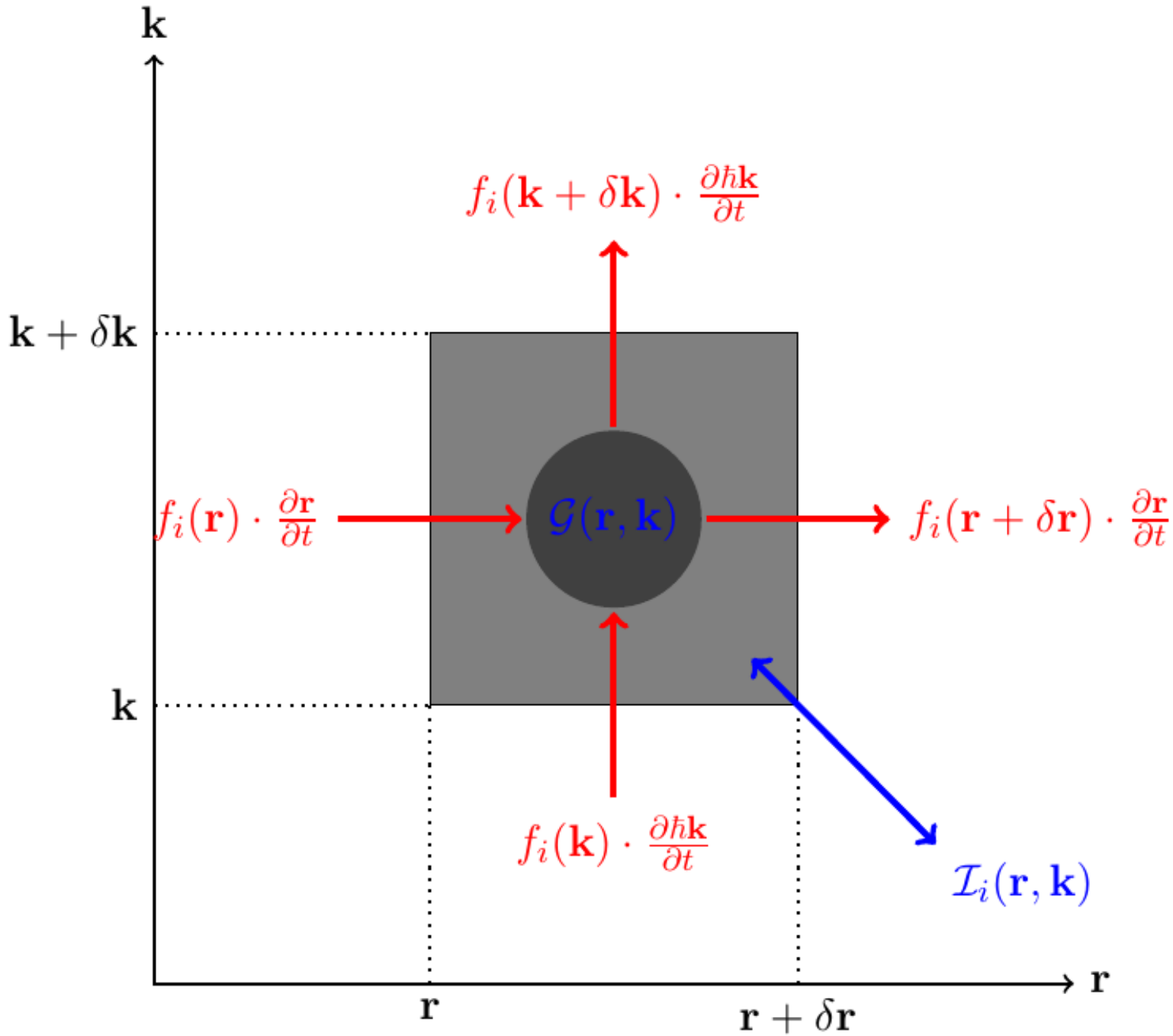
In the absence of applied fields or temperature gradients, f would be the Fermi-Dirac distribution.

We then consider processes which can change the phase space distribution. We treat the momentum and spatial axes of the phase space as separate (there goes Heisenberg's uncertainty principle and any hope of including spin!)

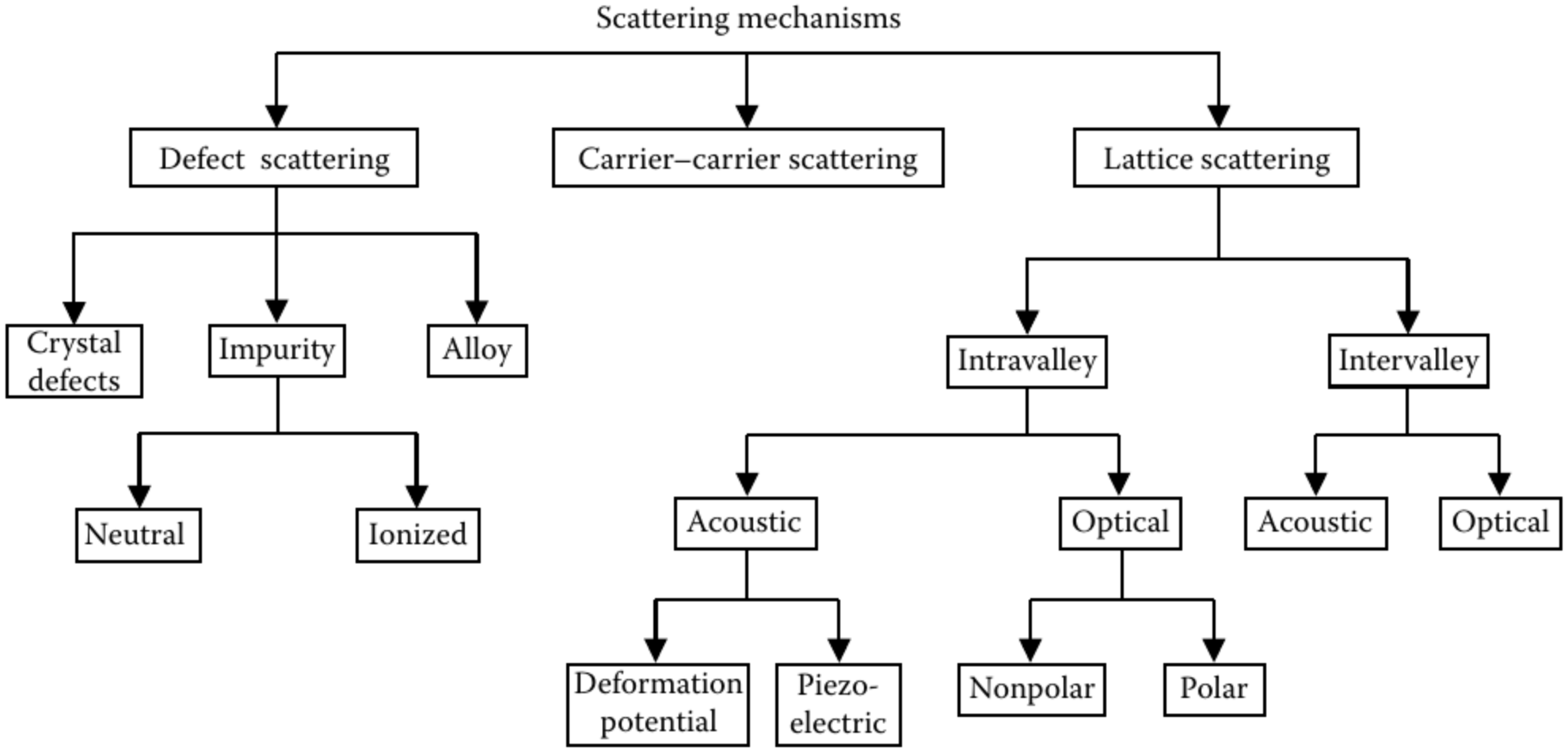
$\mathbf{v}$ - velocity; $\mathbf{F}$ - force

$$\frac{df}{dt} = \mathbf{v} \cdot \nabla_{\mathbf{r}} f + \mathbf{F} \cdot \nabla_{\mathbf{p}} f + \left.\frac{\partial f}{\partial t}\right|_{\text{coll}} + s(\mathbf{r}, \mathbf{p}, t)$$

- $\mathbf{v} \cdot \nabla_{\mathbf{r}} f$ - processes which depend on spatial variations in the distribution function (e.g. mechanical crystal strain, temperature gradients, thermal diffusion).
- $\mathbf{F} \cdot \nabla_{\mathbf{p}} f$ - responses to external force fields (i.e. F is the applied Lorentz force).
- $\left.\frac{\partial f}{\partial t}\right|_{\text{coll}}$ - the net rate of scattering processes contributing to the phase space density (interactions with the lattice, phonons).
- s - particle number change via generation or recombination (e.g. optical absorption, thermal generation).


$\mathbf{k}$
$\mathbf{k} + \delta\mathbf{k}$
$\mathbf{k}$
$f_i(\mathbf{k} + \delta\mathbf{k}) \cdot \frac{\partial \hbar\mathbf{k}}{\partial t}$
$f_i(\mathbf{r}) \cdot \frac{\partial \mathbf{r}}{\partial t}$
$\mathcal{G}(\mathbf{r}, \mathbf{k})$
$f_i(\mathbf{r} + \delta\mathbf{r}) \cdot \frac{\partial \mathbf{r}}{\partial t}$
$f_i(\mathbf{k}) \cdot \frac{\partial \hbar\mathbf{k}}{\partial t}$
$\mathcal{I}_i(\mathbf{r}, \mathbf{k})$
$\mathbf{r}$
$\mathbf{r}$
$\mathbf{r} + \delta\mathbf{r}$

- There are some approaches to device simulation etc that aim to directly solve some version of the Boltzmann equation, perhaps including quantum mechanical extensions. However, this is currently only possible for nano-scale devices (i.e. not particle detectors!)
- We can simplify the BTE substantially by using an approximation for the collision functional. In principle there are very many scattering processes that need to be included, but many are only separately relevant at small spatial, high field, or short time scales. Each process can be calculated using Fermi's Golden Rule, and included in a collision functional approximation.



(image above from Computational Electronics by Vasileska & Goodwin - an **excellent** book to read!)

Relaxation Time Approximation

Assume that for all the scattering processes, no net scattering occurs in equilibrium. Out of equilibrium, assume that the distribution function relaxes with a characteristic timescale τ back to the equilibrium distribution f_0 when scattering occurs

$$\left.\frac{\partial f}{\partial t}\right|_{\text{coll}} \approx -\frac{f - f_0}{\tau}$$

This is a 1st order differential equation in f which can be solved to give:

$$f(t) = f_0 + (f - f_0)e^{-\frac{t}{\tau}}$$

All the physics of the scattering processes have to be approximated in τ. Quite a tall order, but to be fair, we can allow it to be dependent on various things in indirect ways.

The Drift-Diffusion Model

First, recall some semi-classical transport basics from last time. We regard the velocity of the electron to be its group velocity, that is:

$$\mathbf{v} = \frac{1}{\hbar}\nabla_{\mathbf{k}} E(\mathbf{k})$$

The work δW done by a force $\mathbf{F}$ in time δt is $\delta W = \mathbf{F} \cdot \mathbf{v}\delta t$. Equating this work to a change in the energy of the electron then gives:

$$\delta W = \nabla_{\mathbf{k}} E(\mathbf{k}) \cdot \delta\mathbf{k} = \hbar\mathbf{v} \cdot \delta\mathbf{k} \quad \therefore \mathbf{F} = \hbar\frac{\partial \mathbf{k}}{\partial t}$$

One can stretch the analogy even further, take another derivative to get "acceleration" and recover the $\mathbf{F} = m\mathbf{a}$ form of Newton's 2nd law. Care must be taken, though, because the transport is diffusive (Field proportional to velocity not acceleration), so "acceleration" is a misleading term. **It is safest to never use the word "acceleration" when discussing semi-classical electron transport**

We take moments in $\mathbf{v}$ of the BTE and integrate out the momentum dependence. To fully understand each term in this, a more advanced source (e.g. Computational Electronics or "Transport Equations for Semiconductors" by A. Jungel).

Starting with the 0th moment:

$$\frac{\partial}{\partial t}\int f d\mathbf{p} = \int \mathbf{v}\cdot\nabla_{\mathbf{r}} f d\mathbf{p} + \int \mathbf{F}\cdot\nabla_{\mathbf{p}} f d\mathbf{p} - \int \frac{f-f_0}{\tau} d\mathbf{p} + \int s d\mathbf{p}$$

The integral in the LHS term is just the spatial part only distribution (we'll call it n for electrons). We can move $\mathbf{v}$ inside the divergence on the RHS. And using fundamental theorem of vector calculus, we can re-write the 2nd term on the RHS as well

$$\frac{\partial n}{\partial t} = \nabla_{\mathbf{r}}\int \mathbf{v} f d\mathbf{p} + \mathbf{F}\cdot\oint_{\partial\mathbf{p}} f d\mathbf{s} - \int \frac{f-f_0}{\tau} d\mathbf{p} + \int s d\mathbf{p}$$

Since there is a finite number of particles, and a finite total momentum, the surface integral on the RHS must vanish. All scattering processes preserve total particle number, and thus the $f - f_0$ term must also vanish. The integrand in the first term on the RHS is the velocity multiplied by the spatial distribution, which we can define to be the electron current,

$$\mathbf{J}_n \equiv q \int f \mathbf{v} \cdot d\mathbf{p}$$

and so the above complicated equation simplifies to:

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot \mathbf{J}_n + G$$

which is a continuity equation for electrons. A similar procedure on holes (which have opposite charge) yields:

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \nabla \cdot \mathbf{J}_p + G$$

Now (strap in) we take the 1st moment of the BTE. For electrons:

$$\int \mathbf{v}\frac{\partial f}{\partial t}d\mathbf{p} = \nabla_{\mathbf{r}}\int(\mathbf{v}\cdot\mathbf{v})fd\mathbf{p} + \mathbf{F}\int \mathbf{v}\cdot\nabla_{\mathbf{p}}fd\mathbf{p} - \int\frac{f-f_0}{\tau}\mathbf{v}\cdot d\mathbf{p} + \int s\mathbf{v}\cdot d\mathbf{p}$$

First, note that the equilibrium distribution function f_0 is even in k, and so multiplied by an odd function $\mathbf{v}$ and integrated, it must vanish. The same logic applies to the term on the LHS. If we **assume** that the generation-recombination processes do not depend on electron momentum, the same logic also applies to the last term. The first term on the RHS is equivalent to taking an (unnormalised) r.m.s average of the velocity. So, we now have:

$$0 = \left\langle v^2\right\rangle \nabla_{\mathbf{r}}\int fd\mathbf{p} + \mathbf{F}\int \mathbf{v}\cdot\nabla_{\mathbf{p}}fd\mathbf{p} - \frac{1}{\tau}\int f\mathbf{v}\cdot d\mathbf{p}$$

$$\Rightarrow 0 = \left\langle v^2\right\rangle \nabla n + \mathbf{F}\left(\oint_{\partial\mathbf{p}}\mathbf{v}fd\mathbf{s} - \int f\nabla_{\mathbf{p}}\mathbf{v}d\mathbf{p}\right) - \frac{1}{q\tau}\mathbf{J}_n$$

where the last step involved an integration by parts.

Finally we can return to our old friend $\mathbf{v} = \frac{1}{\hbar}\nabla_{\mathbf{k}} E(\mathbf{k})$ to do more work:

$$0 = q\tau \left\langle v^2 \right\rangle \nabla n + \frac{q\tau \mathbf{F}}{\hbar} \int f \nabla_{\mathbf{k}}^2 E(\mathbf{k}) d\mathbf{k} - \mathbf{J}_n$$

and invoking the effective mass approximation, we end up with:

$$\mathbf{J}_n = q\tau \left\langle v^2 \right\rangle \nabla n + \frac{q\tau \mathbf{F}}{m^*} \int f d\mathbf{p}$$

$$\Rightarrow \mathbf{J}_n = q\tau \left\langle v^2 \right\rangle \nabla n + \frac{q\tau \mathbf{F}}{m^*} n$$

The external force $\mathbf{F}$ just comes from the electric field due to the potential V (again we're pretending magnetism doesn't exist for a while). If we also pretend (for a second) that kinetic theory applied to these electrons, then we could invoke equipartition theorem to say that $\left\langle v^2 \right\rangle = \frac{k_b T}{m^*}$, again pretending that effective mass actually is a mass.

We now have:

$$\mathbf{J}_n = \frac{q\tau k_b T}{m^*}\nabla n - \frac{q^2\tau}{m^*}\nabla V \cdot n$$

you may recall, that in the free electron model from last time, we defined a quantity called mobility$\mu = \frac{q\tau}{m}$. If we squint a bit and pretend that's effective mass on the bottom we would then have:

$$\mathbf{J}_n = \mu_n k_b T \nabla n - q\mu_n n \nabla V$$

If we made the Boltzmann assumption earlier, we can now invoke the familiar Einstein-Smoluchowski relations (i.e. that $\mu k_b T = qD$). If we didn't, well, actually it turns out this still works but the reason that it does needs a few dozen pages and a more careful derivation (which you can find in Computational Electronics chapter 6). Anyway, we now have:

$$\mathbf{J}_n = qD_n \nabla n - q\mu_n n \nabla V$$

All the equations, written out for electrons and holes, are:

$$\mathbf{J}_n = -q\mu_n n\nabla V + qD_n\nabla n$$

$$\mathbf{J}_p = -q\mu_p p\nabla V - qD_p\nabla p$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla\cdot\mathbf{J}_n + G$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla\cdot\mathbf{J}_p + G$$

which, together, constitute what is known as the **Drift Diffusion Model**. When combined with a bit more semiconductor statistics (we're getting to it now), this full equation system is known as the **van Roosbroeck system**.

Drift Diffusion Model - Comments

Why bother going through all that hideous derivation from the BTE? Well, mainly because whilst drift and diffusion are physically reasonable processes intuitively, it is nice to see that they arise directly from a kinetic approach as the first approximation.

Do not underestimate the Drift Diffusion model's complexity. Though we have done major approximations to obtain it, it is still equivalent to two sets of fully compressible (though luckily non-turbulent) fluid dynamics, coupled by a generation / recombination term.

Drift Diffusion model is an excellent intuition tool - at any scale where quantum effects are not dominant, it gets things **basically right**, though quantitatively it will be wrong in many ways, and miss out many subtle effects in real semiconductors.

This system (with some extensions, and not for nanoscale devices) is what the commercial TCAD tools (Sentaurus, Silvaco) you will learn about in other lectures are solving when doing device simulation.

Beyond Drift Diffusion - Hydrodynamic Model

One way to go beyond the Drift Diffusion model is to simply take more moments of the BTE, and solve more coupled equations. This approach can never truly overcome the limit of semi-classical dynamics (because the BTE itself is a fundamentally classical kinetic formulation), though it can capture various dynamics and scattering processes more accurately.

In particular, the DD model relies on there being only one single temperature throughout the system. Higher order (so-called "hydrodynamic" or "energy balance") models can remove this limitation, and are much better therefore at modelling situations where the electron gas' temperature is different to the lattice's, (e.g. hot-carrier injection in MOSFET channels), or when there is a temperature gradient across the device.

Hydrodynamic Model derivation

Just kidding, we won't be doing this. See Computational Electronics, chapter 5.

The idea is basically this:
define a quantity $\phi(\mathbf{p})$ that is proportional to various powers of $\mathbf{p}$ (proportional with arbitrary constants at each power, because we want ϕ to represent e.g. carrier density, current density, etc etc). The 0th and 1st ϕ give us something that reduces to DD model, but if we left out some of the assumptions on collision functional etc, we would have:

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot \mathbf{J}_n + G$$

$$\frac{\partial \mathbf{J}_n}{\partial t} = \frac{2q}{m^*}\sum_i \frac{\partial W_{ij}}{\partial x_i} + \frac{nq^2}{m^*}\mathbf{E} - \left\langle \frac{1}{\tau_m} \right\rangle \mathbf{J}_n$$

where $\mathbf{W}$ is the so-called "energy density tensor".

The 2nd moment would give us:

$$\frac{\partial \mathbf{W}}{\partial t} = -\nabla \cdot \mathbf{F}_W + \mathbf{E} \cdot \mathbf{J} - \left\langle \frac{1}{\tau_E} \right\rangle (\mathbf{W} - \mathbf{W_0})$$

which has introduced another quantity called the "energy flux" $\nabla \cdot \mathbf{F}_W$. Each balance equation introduces another, ever more esoteric term. At some point we choose an approximation for the collision functional and thus the τ_i terms, and this will determine to which level of energy balance equation we will solve. Solving an infinite number of balance equations, would be equivalent to solving the BTE directly. In practice, it seems, most commercial simulators don't go beyond Energy Flux (3rd moment). At this point, the simulation is accurate enough until you hit the quantum level where totally different approaches are needed.

Semiconductor Physics - Thermodynamics

In the simplest case, we can use the parabolic band approximation for density of states (from last time). So, we have the density of states in the conduction band (with conduction band edge E_c):

$$g_c(E) = \frac{\sqrt{2}(m^*_{e,dos} k_B T)^{\frac{3}{2}}}{\pi^2 \hbar^3} \sqrt{E - E_c}$$

and effective density of states in the valence band (with valence band edge E_v):

$$g_v(E) = \frac{\sqrt{2}(m^*_{h,dos} k_B T)^{\frac{3}{2}}}{\pi^2 \hbar^3} \sqrt{E_v - E}$$

The Fermi-Dirac occupation function is given by (where E_f is the Fermi energy):

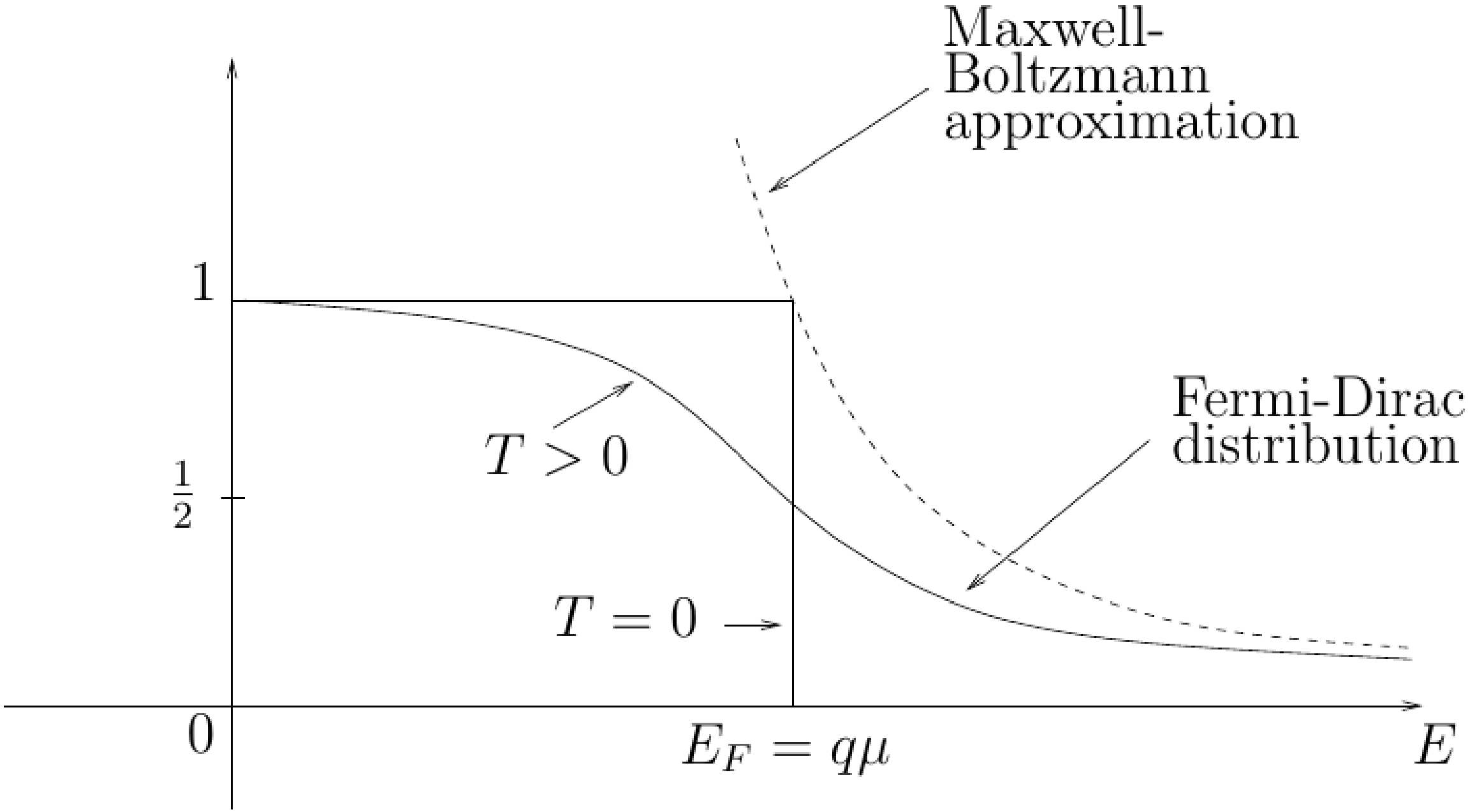
$$f(E) = \frac{1}{1 + e^{\frac{E - E_f}{k_b T}}}$$

Boltzmann Approximation

Yep, more stuff confusingly called "Boltzmann". Anyway, hopefully you have seen before that, at high temperatures (typically we say temperatures such that $E - E_f > 3k_bT$:

$$f(E) = \frac{1}{1 + e^{\frac{E-E_f}{k_bT}}} \approx e^{-\frac{E-E_f}{k_bT}}$$

Note the following derivations don't actually properly **rely** on this approximation. Unlike other approximations we have taken, you can form the full Drift Diffusion model without doing this (and it has some improvement in accuracy for so-called "hot carrier" and high applied bias situations). The solutions end up being not closed-form but in terms of the Fermi-Dirac integral, though, and the formulae are a bit more complicated. For full details, see e.g. Sze pp. 17-21


Maxwell-
Boltzmann
approximation
1
$\frac{1}{2}$
$T > 0$
Fermi-Dirac
distribution
$T = 0$
0
$E_F = q\mu$
E

Carrier Density

First thing to work out is the carrier densities n for electrons and p for holes. We use the convention of a bar above a quantity meaning that quantity divided by $k_b T$, i.e. $\bar{E}_f \equiv \frac{E_f}{k_b T}$. The carrier densities are given simply by:

$$n = \int_{E_c}^{\infty} g_c(E) f(E) dE \approx K_n \int_{E_c}^{\infty} \sqrt{E - E_c} e^{\bar{E} - \bar{E}_f} = K_n e^{\bar{E}_c - \bar{E}_f} \int_{E_c}^{\infty} \sqrt{E - E_c} e^{\bar{E} - \bar{E}_c} dE$$

and

$$p = \int_{0}^{E_v} g_v(E)(1 - f(E)) dE \approx K_p \int_{0}^{E_v} \sqrt{E_v - E} e^{\bar{E}_f - \bar{E}} dE = K_p e^{\bar{E}_v - \bar{E}} \int_{0}^{E_v} \sqrt{E_v - E} e^{\bar{E}_v - \bar{E}} dE$$

The integrals turn out to equate to constants - they are the form of the Gamma function (under the Boltzmann approximation, anyway), and we end up with the important equations for carrier densities:

$$n = N_c(T)e^{-(\bar{E}_c - \bar{E}_f)}$$

$$p = N_v(T)e^{-(\bar{E}_f - \bar{E}_v)}$$

> where N_c and N_v are called the "effective density of states" and depend on Temperature. They may be semi-empirical or use the full parabolic approximation

The above expressions are important because most of the following results are derived from simple manipulation of them.

Note that we **had to** assume a simplistic (parabolic) form of the density of states to get this to work. In part, this is why we hold on to the parabolic band & effective mass approximations so much!

Intrinsic Density

The above definitions imply that the product np is given by:

$$np = N_c N_v e^{\bar{E}_v - \bar{E}_c} \equiv N_c N_v e^{-\bar{E}_g} = n_i^2; \quad \therefore n_i = \sqrt{N_c N_v} e^{-\frac{\bar{E}_g}{2}}$$

with E_g the band gap. We call the quantity n_i the **intrinsic concentration** or **intrinsic density**, and it allows us to write new equations for n and p that are more often used (because this quantity n_i is relatively easy to actually measure).

$$n = n_i e^{\bar{E}_f - \bar{E}_i}; \quad p = n_i e^{\bar{E}_i - \bar{E}_f}$$

where E_i is called the **intrinsic Fermi Level**, i.e. the Fermi level of an intrinsic semiconductor (which just means without any doping - see later)

We can solve the above equation for the location of the Fermi Level:

$$E_f = E_i = \frac{E_c + E_v}{2} + \frac{k_b T}{2} \ln \frac{N_v}{N_c}$$

The Fermi level therefore moves (pretty slowly) with temperature. Slowly enough that you can't really make an insulator into a semiconductor without it melting. But fast enough that this contributes a large part of the variation in carrier concentrations with temperature. The band gap itself also varies (again, slowly) with temperature (a reasonably good model is the **Varshni Equation**).

Law of Mass Action

The kinetics of electrons & holes are simple. Whenever we create an electron, we also create a hole! Thus, for an intrinsic piece of material, we must have

$$n = p = n_i$$

An alternative expression for the Fermi level is:

$$E_F = E_i + k_b T \ln \frac{n}{n_i}$$

Quasi-Fermi Levels ("Imrefs")

Everything we have talked about up to now in terms of thermodynamics assumed equilibrium (meaning, no applied external field). Imagine if there is such a field, that can impart energy to electrons and holes.

Thinking in terms of our drift-diffusion model, electrons and holes can thermalise very rapidly (meaning, the energy distribution function within a band can return to something that looks like equilibrium within a few scattering times (i.e. picoseconds). The bands can only "talk to each other" via generation and recombination processes, and these take on the order of the carrier lifetime, which is between nanoseconds & milliseconds in most devices.

Even when the system is quite far from equilibrium, because of this mismatch in timescales, there can exist a quasi thermal equilibrium state for electrons and holes **separately**. So, even though we cannot define a single Fermi level for the system out of equilibrium, we can define **Quasi Fermi Levels** (also known as "Imrefs").

We can adapt our previous relations for n and p to include QFLs very simply. Suppose we're applying a voltage V which therefore imparts energy qV to all the carriers. By definition, we no longer have a Fermi level (which is an equilibrium concept), but if the energy change is not too large, we can assume a similar distribution for electrons and holes **separately**.

$$n = n_i e^{q\bar{V} - q\bar{\phi}_n - \bar{E}_i}; \quad p = n_i e^{\bar{E}_i + q\bar{\phi}_p - q\bar{V}}$$

ϕ_n and ϕ_p are the QFLs for electrons and holes, respectively.

The new "law of mass action" is:

$$np = n_i^2 e^{q\bar{\phi}_p - q\bar{\phi}_n}$$

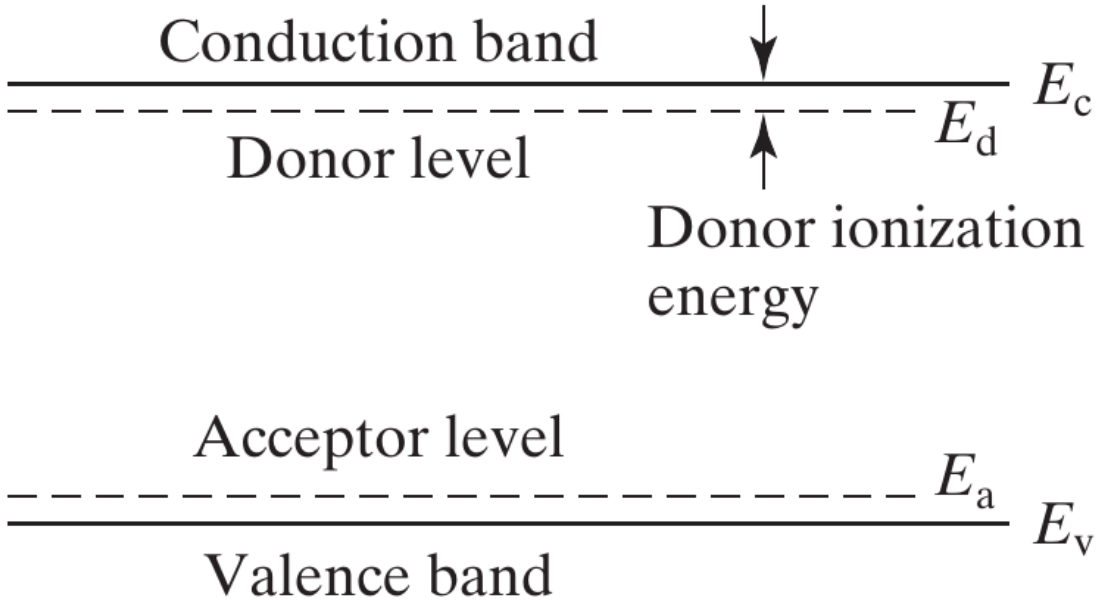
so we can see that if $\phi_n = \phi_p = 0$ we recover the original situation. It is often convenient to choose the zero reference energy level to be E_i, and then the above become even simpler.

Doping

When we add dopant to the semiconductor, we are (conceptually) adding impurity levels in the band gap either near the conduction (for n-type) or valence (for p-type) bands.

The semiconductor is now called "extrinsic". In equilibrium, we can solve for the concentrations by adding the condition of charge neutrality, that is:

$$p + N_D^+ - n - N_A^- = 0$$



going back to our (equilibrium) equations for n and p, we find for the extrinsic case (and assuming all donors and acceptors are ionised for now):

$$n = \frac{N_D - N_A}{2} + \sqrt{\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2}$$

$$p = \frac{N_A - N_D}{2} + \sqrt{\left(\frac{N_A - N_D}{2}\right)^2 + n_i^2}$$

The above expressions are rarely used, because in practice we usually have much more of one type of dopant than the other present. For n-type we usually have $N_D - N_A \gg n_i$, which leads to:

$$n = N_D - N_A; \quad p = \frac{n_i^2}{n}$$

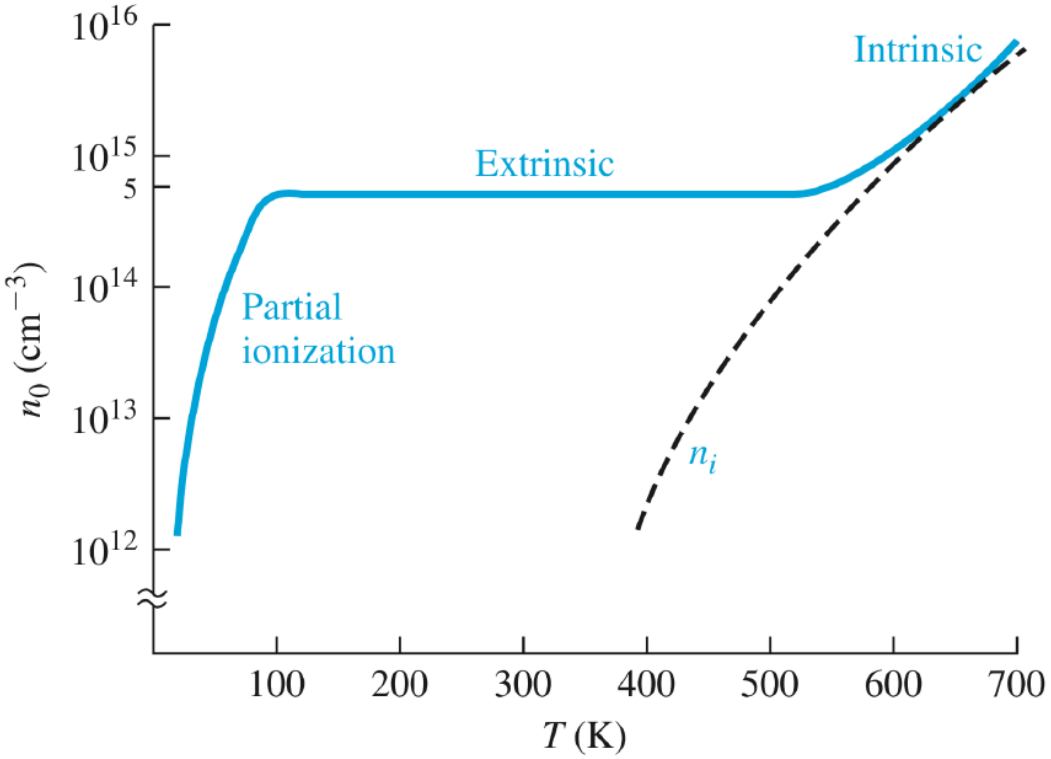
if we additionally have that $N_D \gg N_A$ then the expressions are even simpler:

$$n = N_D; \quad p = \frac{n_i^2}{N_D}$$

vice versa for $N_A \gg N_D$. Note that in an extrinsic situation the carrier with the high concentration is called (unsurprisingly) the **majority carrier** and the other the **minority carrier**. Importantly, (including in particle physics detectors) it is very often the minority carriers we care about.

Carrier Freeze out

At low temperatures, there can be insufficient energy to ionize all the dopants (referred to as "carrier freeze out"). At high temperatures, the intrinsic density n_i increases, and eventually will become higher than the extrinsic (dopant induced) concentrations. Below plot from "Modern Semiconductor Devices for Integrated Circuits" by Chenming Hu.



The Non-Linear Poisson Equation

Recall that in the drift diffusion model, we had the potential V and the carrier densities n and p included. But, also, above we showed that the influence of an external potential was to modify the carrier distribution functions. Poisson's equation must (as always) be satisfied:

$$\epsilon_0 \nabla \cdot (\epsilon \nabla V) = -q(p - n + N_D^+ - N_A^-)$$

where N_D^+ and N_A^- are donor and acceptor densities, respectively.

substituting in the thermodynamic definitions of the densities (and referencing all energies relative to E_i), we find:

$$\epsilon_0 \nabla \cdot (\epsilon \nabla V) = -q n_i (e^{-q\bar{V} + \bar{\phi}_p} - e^{q\bar{V} - \bar{\phi}_n}) + q(N_D^+ - N_A^-)$$

(note the $\nabla^2 V$ on the LHS and the $e^{f(V}$ terms on the RHS, this is a powerfully nasty non-linear equation now!)

The above is called the Non-Linear Poisson Equation (of course "Linear Poisson equation with non-linear source term" would be more accurate). It is the penultimate piece of the van-Roosbroeck system puzzle. The final piece comes from checking what happens when we substitute the thermodynamic definition of n into the DD current equation. First note that:

$$\nabla n = \frac{q}{k_b T}(\nabla V - \nabla \phi_n) n$$

and then we get:

$$\mathbf{J}_n = -q\mu_n n \nabla V + qD_n \frac{q}{k_b T}(\nabla V - \nabla \phi_n) n$$

invoking the Einstein relation again $\mu k_b T = Dq$, we have:

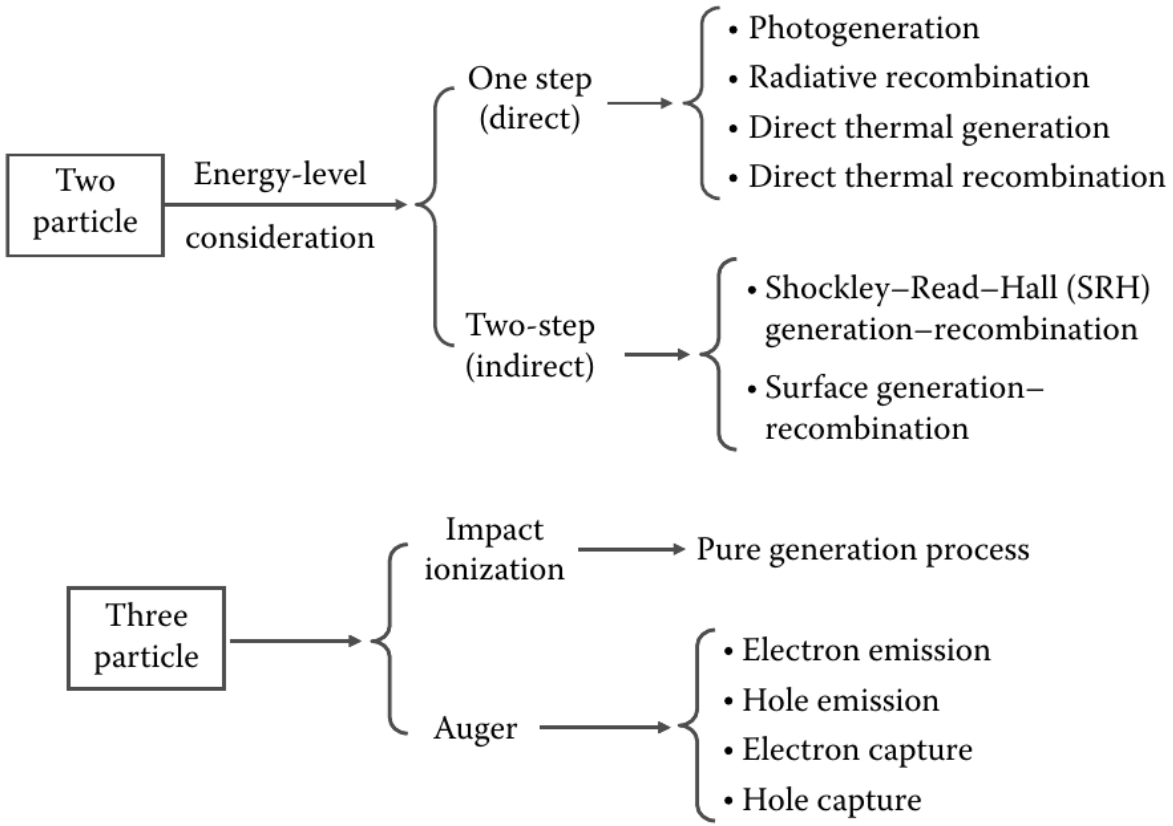
$$\mathbf{J}_n = -q\mu_n n \nabla V + q\mu_n (\nabla V - \nabla \phi_n) n = -q\mu_n n \nabla \phi_n$$

a similar relation applies to holes. In other words, the currents are given by the QFLs in the DD model.

So, finally, how to solve a semiconductor problem (one way):

1. solve the NLPE in equilibrium (where $\phi_n = \phi_p = 0$) to get a self-consistent potential V. In equilibrium this also tells you n and p. No need to solve continuity as no net current flows in equilibrium
2. Apply some small bias, solve NLPE again to get a potential V. This potential will no longer be consistent with the continuity equations. So, solve the continuity equations to get ϕ_n and ϕ_p (or equivalently, to get n and p). As a by-product, you will learn $\mathbf{J}_n$ and $\mathbf{J}_p$. Keep solving until these two sets of equations are consistent
3. Apply some more bias, repeat step 2, and keep applying more bias until you reach the desired conditions

The above is called **Gummel's Method**, and it is still a common flow in modern device simulation (though others are available).



Generation & Recombination

Having thrown away most of quantum mechanics and complicated scattering processes to get the DD model, we need to re-inject some physical reality to get good matches to real behaviour. One place we do this is in the generation & recombination terms. There are many possible effects to include (see left from Computational Electronics).

Most important (for detector physics!!) are: photogeneration, thermal generation, SRH G/R, impact ionization. (Note that the processes responsible for generating primaries from high energy particles e.g. Bragg scattering, compton scattering, pair production are not included here!).

Most unintuitive is that direct band to band recombination is *very* slow in silicon, because it requires phonon interactions to account for the crystal momentum conservation.

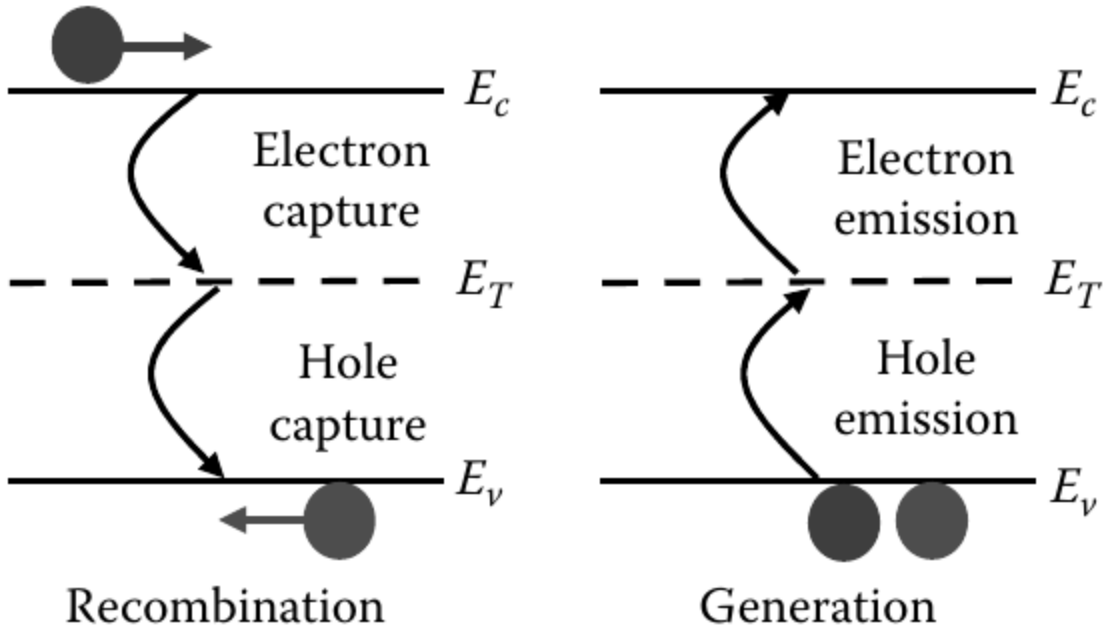
Intuition about Recombination - electrons & holes must be in the same place (spatially) to recombine. Therefore expect most recombination mechanisms rate to depend on product np. In thermal equilibrium, there is no *net* recombination, and in thermal equilibrium, $np = n_i^2$, therefore the recombination rate:

$$u \propto np - n_i^2$$

This is true for SRH, thermal generation, Auger, etc etc.

SRH generation / recombination

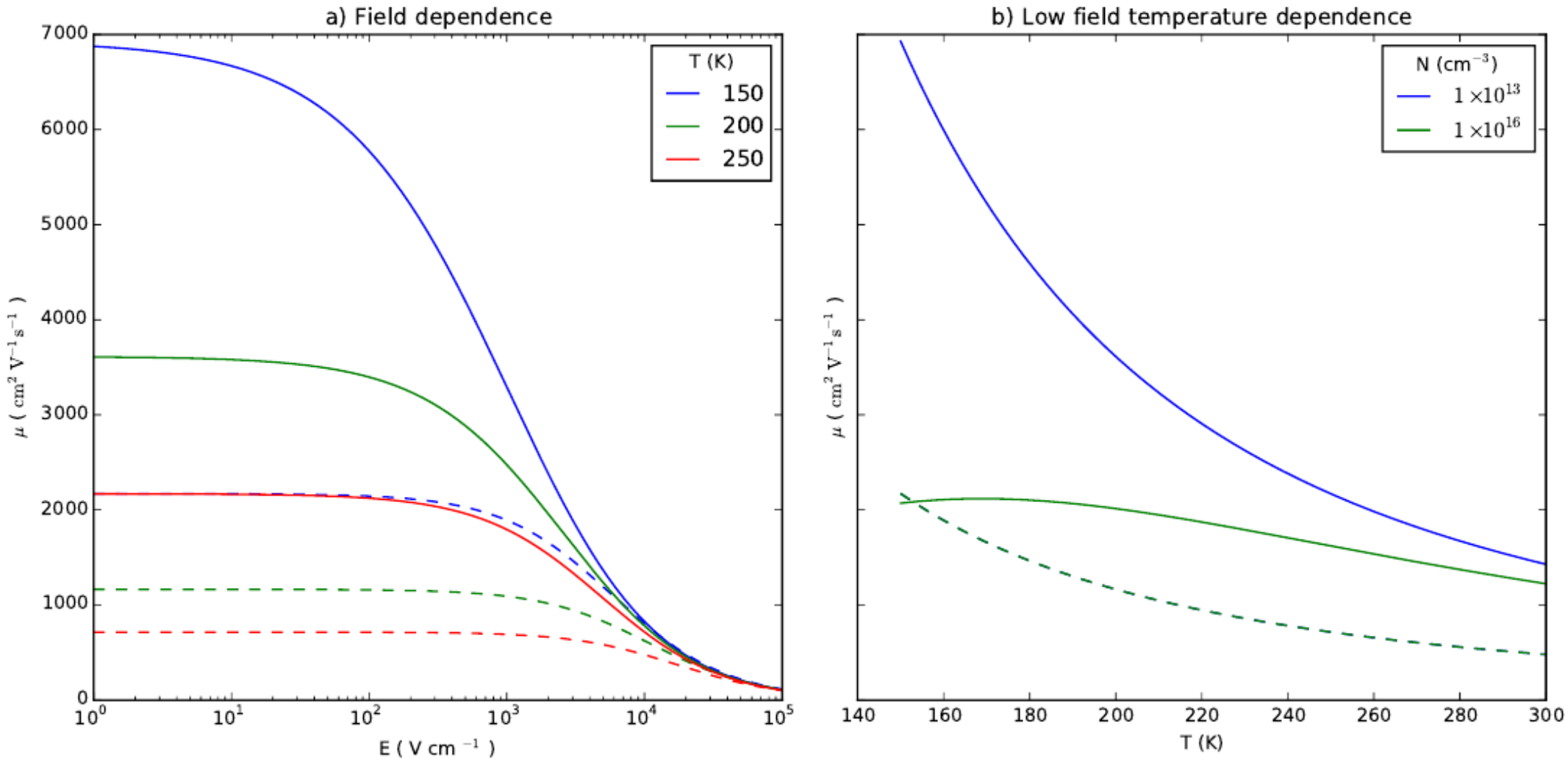
Most important recombination mechanism in Silicon. "Trap" levels (arising due to impurities or radiation damage) capture/emit electrons or holes, and then later capture/emit the other species. The phonon accounting is much easier than direct recombination / generation because the trap level can "hold on" to the electron for quite a while. These trap levels are typically "deep" (i.e. near the middle of the band gap, not like impurity levels for doping)



Diffusivity & Mobility

Perhaps the most important extension to DD model to get some realistic results is allowing the mobility (and diffusivity) to depend on other quantities, in particular the applied Field. What is observed is that electrons exhibit "velocity saturation" in semiconductors. After they reach a certain velocity, it becomes harder and harder to get faster. This is due to a combination of band structure effects and energy loss mechanisms like phonons and (in the extreme case) impact ionisation. There are literally hundreds of different empirical and theoretical mobility models around, but the simplest reasonably good one is probably the Caughey-Thomas formula using the Canaali measurements:

$$\mu_i = \frac{\mu_{i,0}}{\left(1 + \left(\frac{\mu_{i,0}|\mathbf{E}|}{v_{sat}}\right)^{\beta_i}\right)^{\frac{1}{\beta_i}}}$$



Mobility in Silicon (solid lines are electrons, dashed lines are holes)

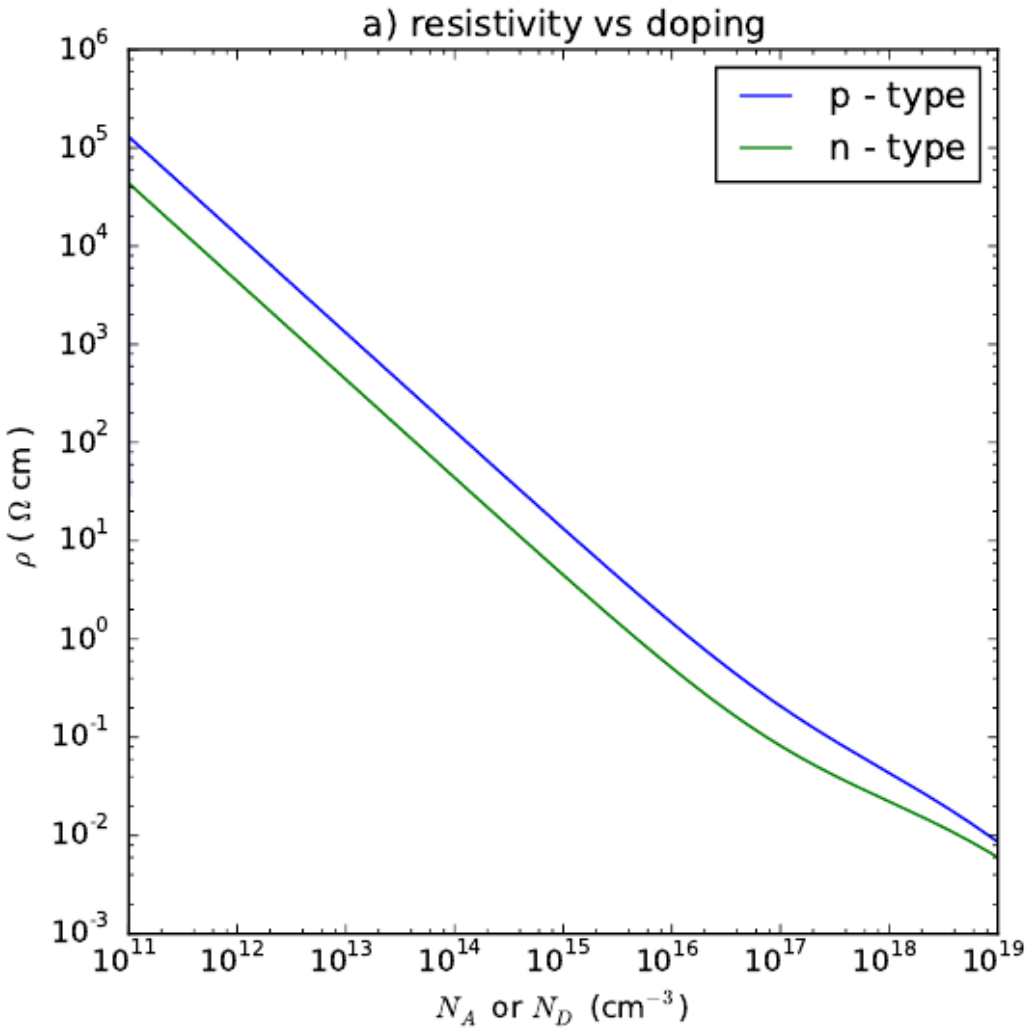
Resistivity

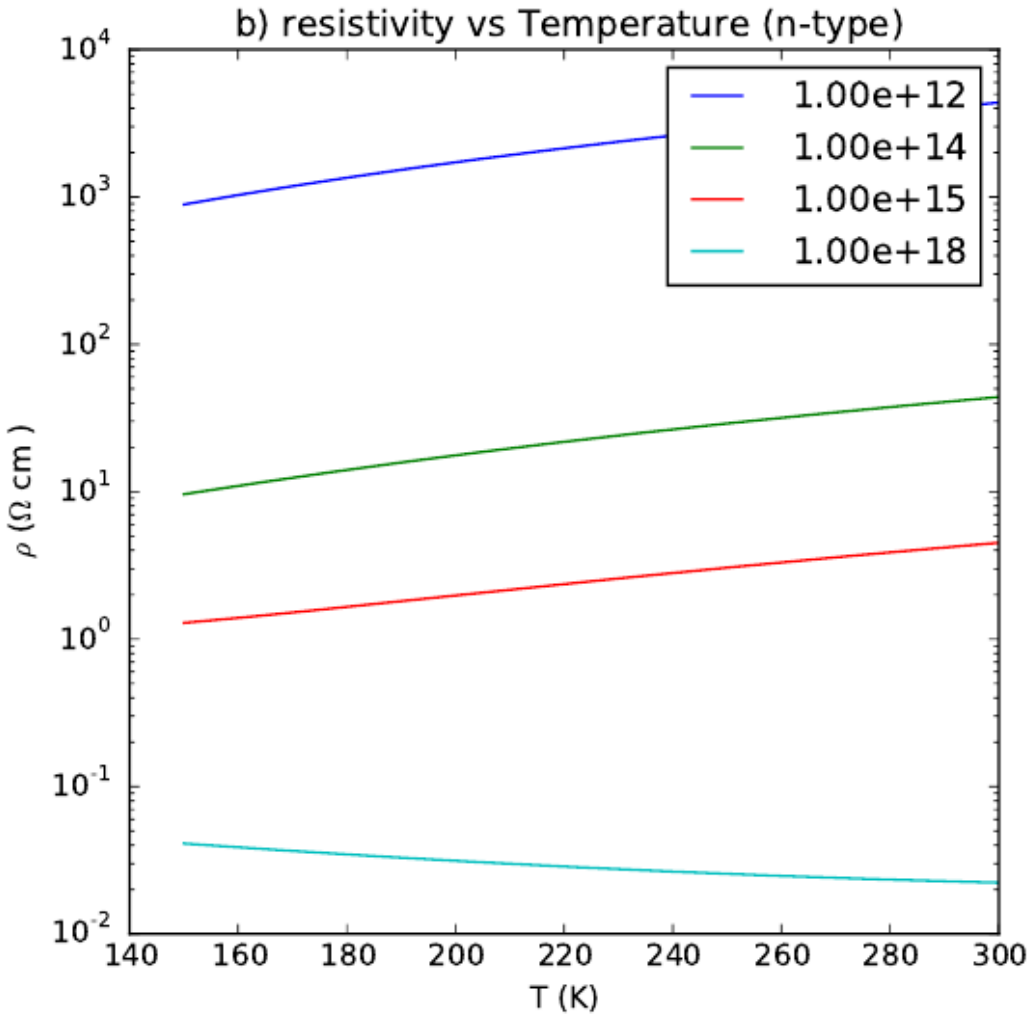
Recall from last time that we could relate conductivity (via Drude model) to carrier density:

$$\sigma = nq\mu_n + pq\mu_p$$

Thus, conductivity (more commonly, resistivity) is a useful quantity to measure in semiconductors. It is not *quite* the same conceptually as the resistivity of a metal (where you can't alter n much by applying voltages!).

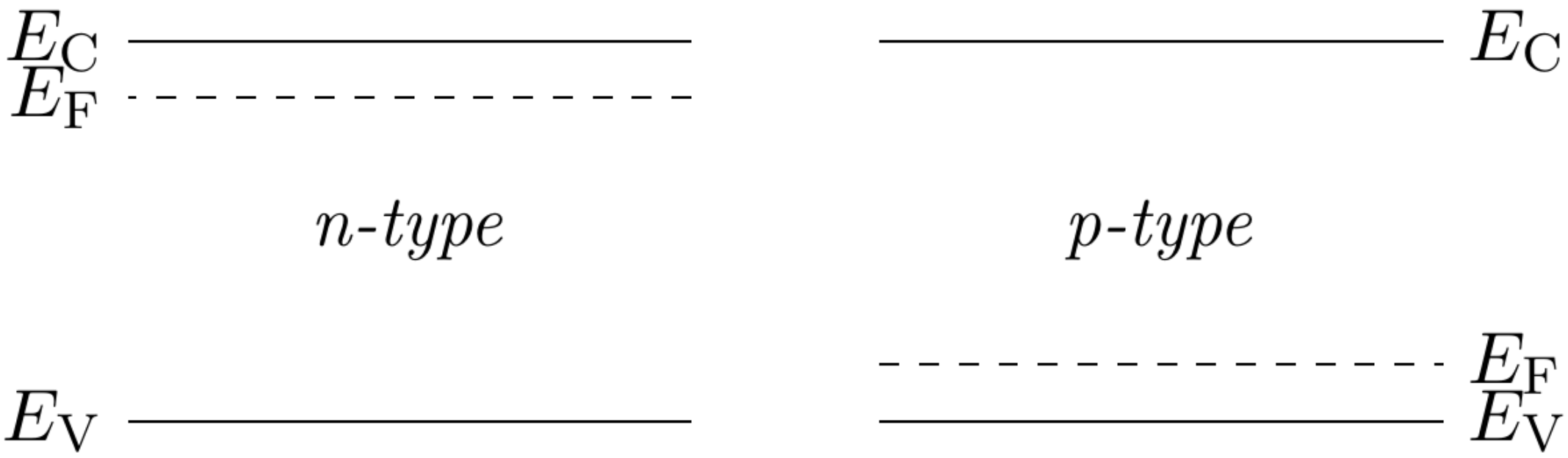
As you'll hear in other lectures, high resistivity is crucial to many modern particle (& astro) detector devices, because it allows us to obtain maximum depletion for minimum applied bias & thus leakage current. The resistivity of the wafer depends critically on manufacturing processes (see other lectures).


a) resistivity vs doping
p - type
n - type
ρ (Ω cm)
N_A or N_D (cm^{-3})

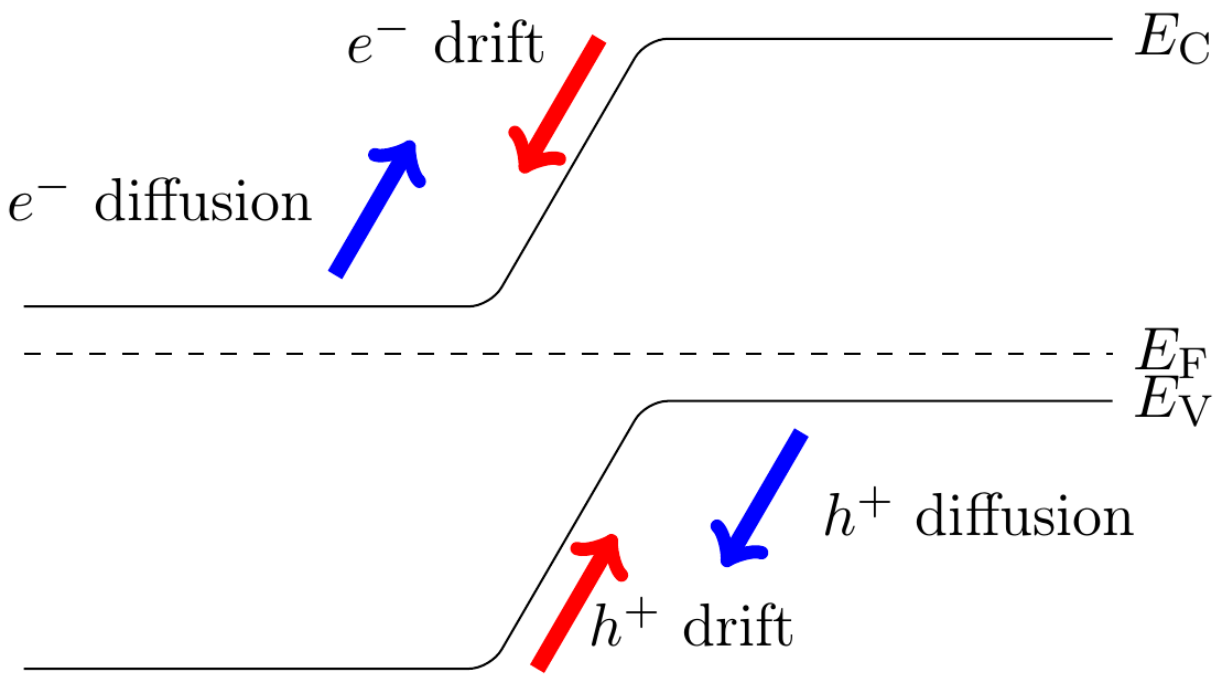

b) resistivity vs Temperature (n-type)
1.00e+12
1.00e+14
1.00e+15
1.00e+18
ρ (Ω cm)
T (K)

Physics of the PN Junction

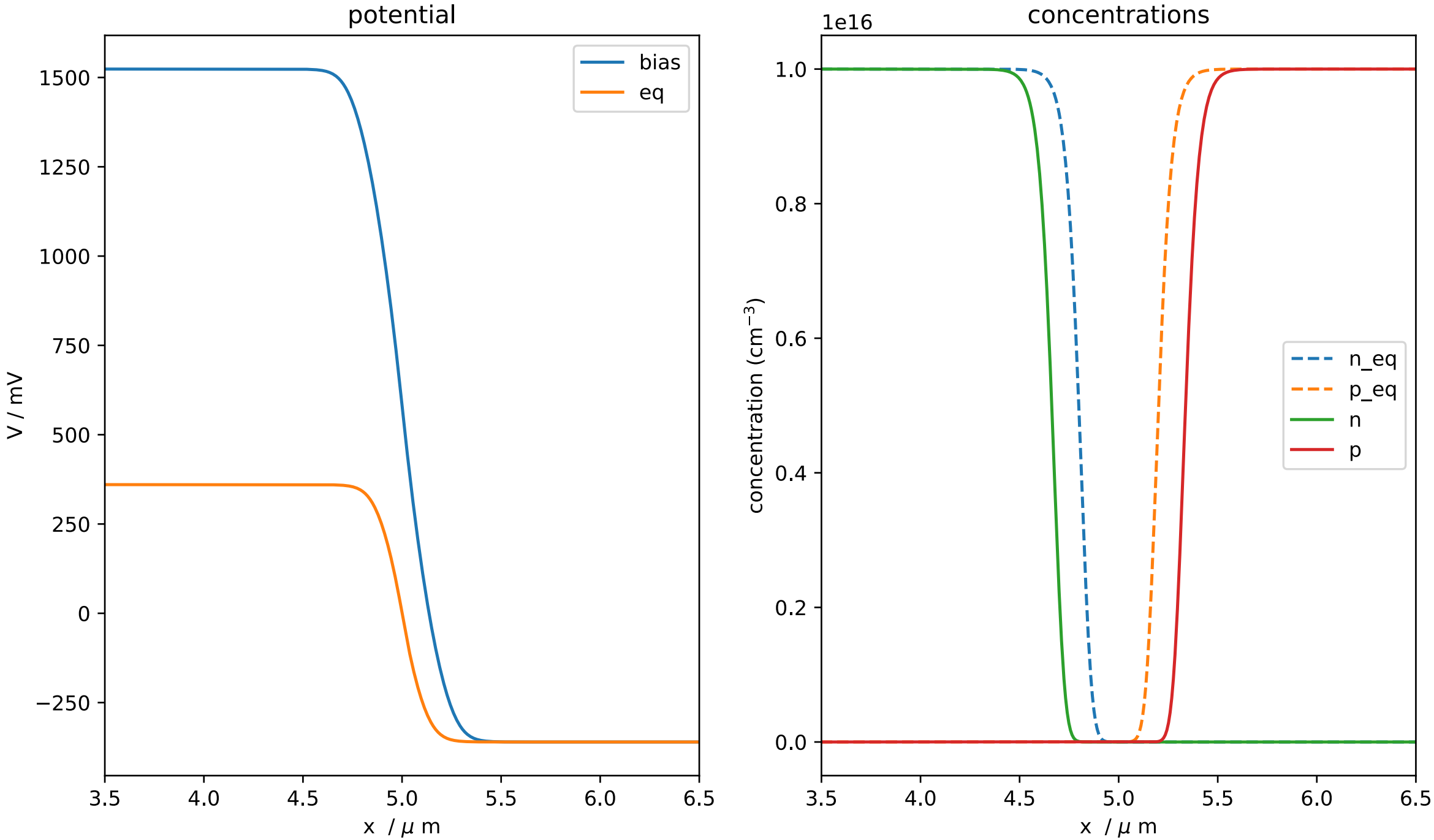
Finally! We can discuss what happens when we put together two differently doped semiconductor regions:

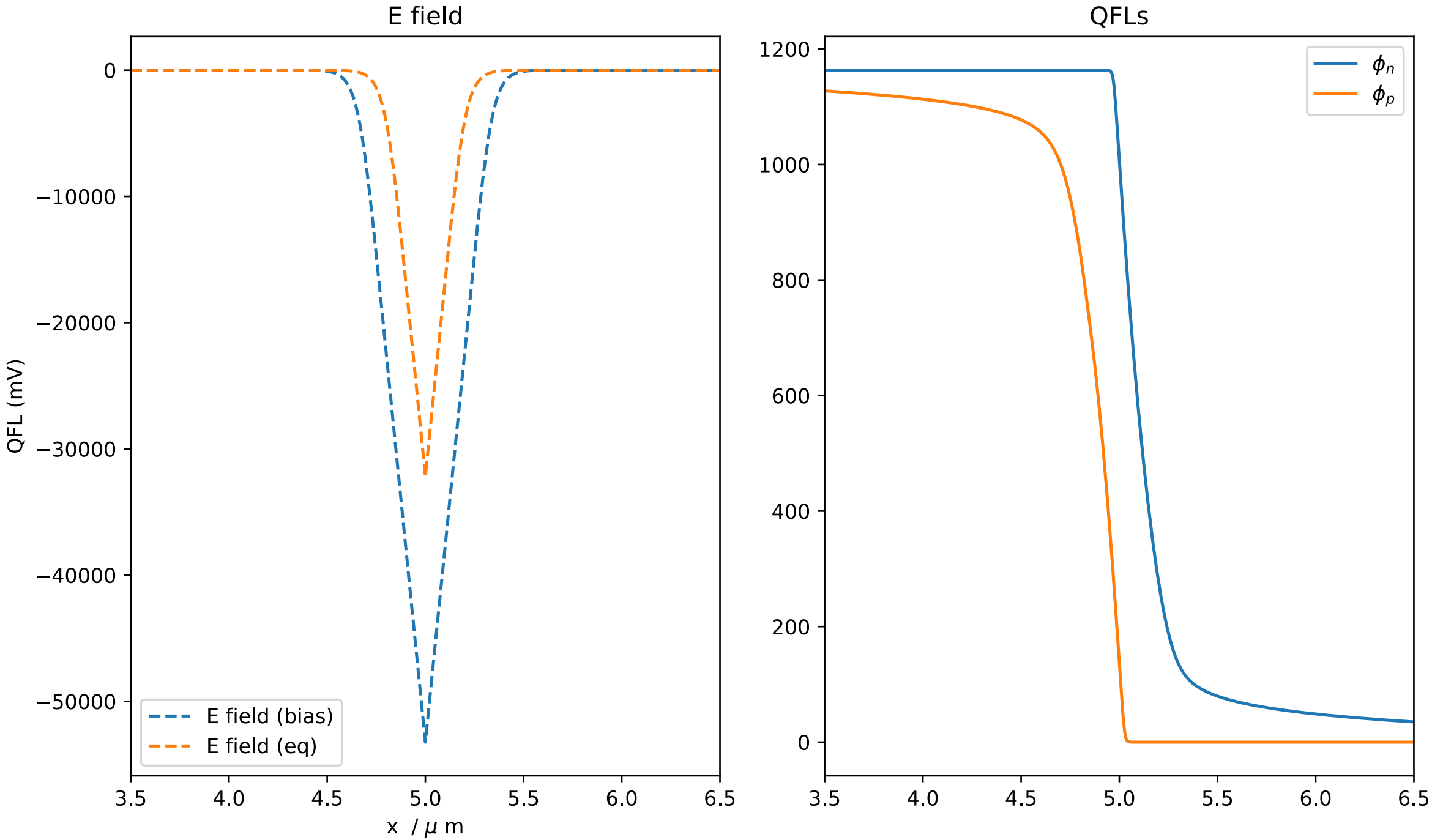


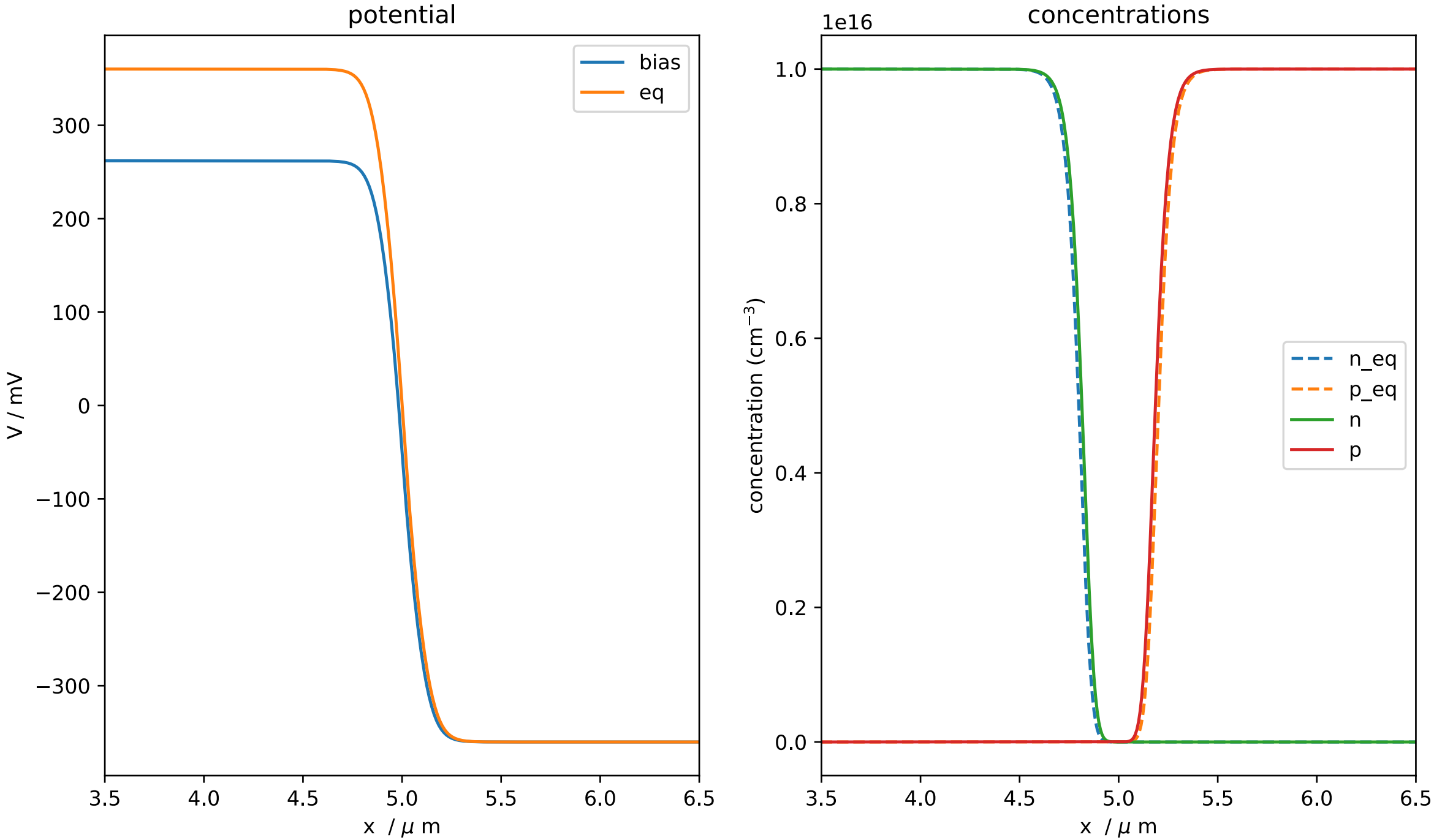
The important thing to realise is that the Fermi level will equilibrate accross the device...

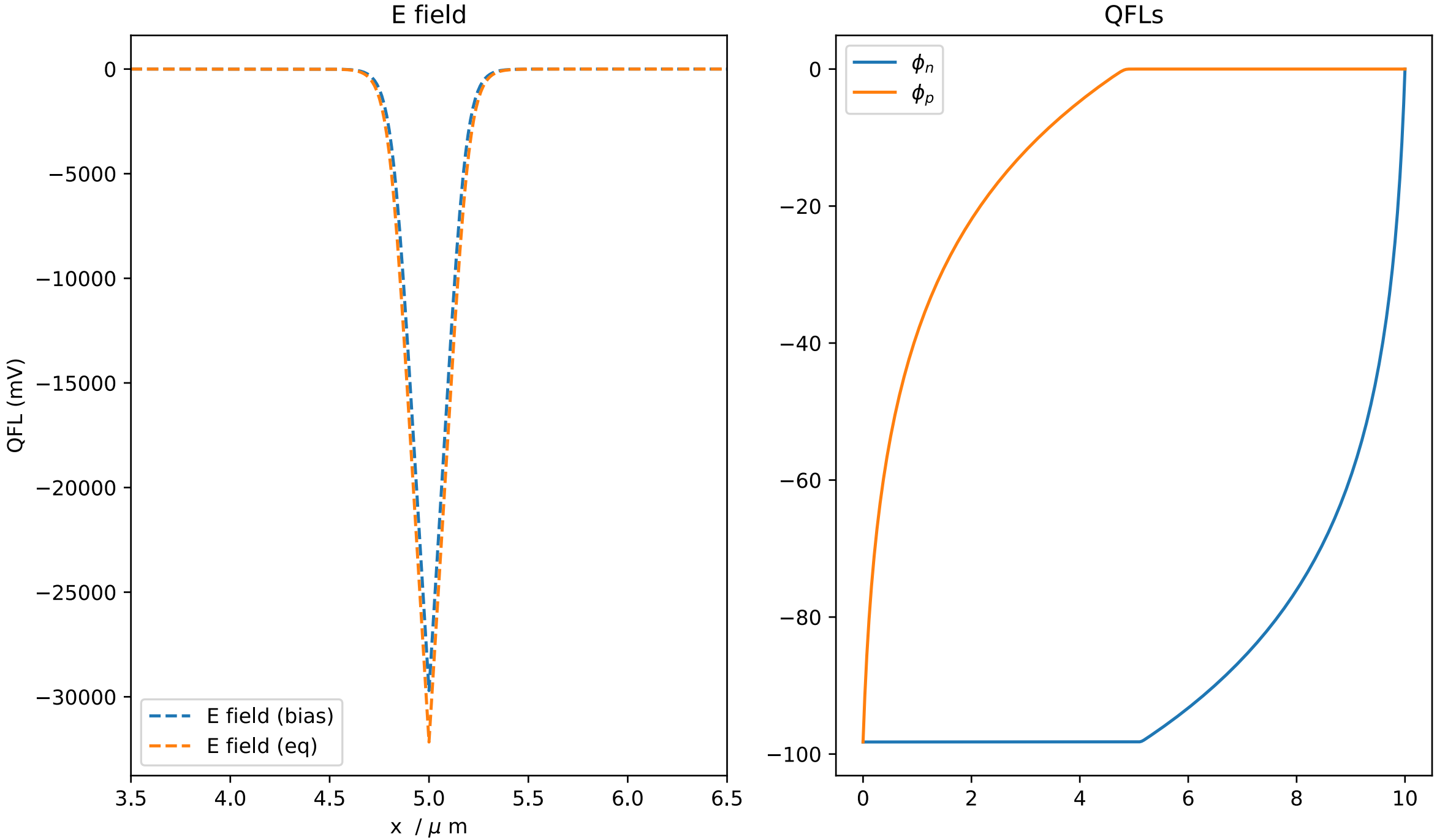


The carriers flow as shown above, and meet at the centre of the junction, where recombination happens. This recombination serves to create a region of unscreened dopants (called a "depletion region"). When the junction is put in reverse bias, the size of this region increases. The following slides show a simple 1D simulation of solving the DD model that I put together for a reverse biased pn junction with $N_A = N_D = 10^{16}\mathrm{cm}^-3$. They include basic SRH recombination and Klaassen mobility model.


potential
bias
eq
1500
1250
1000
750
500
250
0
−250
V / mV
3.5
4.0
4.5
5.0
5.5
6.0
6.5
x / μ m
concentrations
1e16
1.0
0.8
0.6
0.4
0.2
0.0
concentration (cm^{-3})
n_eq
p_eq
n
p
3.5
4.0
4.5
5.0
5.5
6.0
6.5
x / μ m


E field
QFL (mV)
0
−10000
−20000
−30000
−40000
−50000
E field (bias)
E field (eq)
3.5
4.0
4.5
5.0
5.5
6.0
6.5
x / μ m
QFLs
1200
1000
800
600
400
200
0
ϕ_n
ϕ_p


potential
bias
eq
300
200
100
0
−100
−200
−300
V / mV
3.5
4.0
4.5
5.0
5.5
6.0
6.5
x / μ m
concentrations
1e16
1.0
0.8
0.6
0.4
0.2
0.0
concentration (cm^{-3})
n_eq
p_eq
n
p
3.5
4.0
4.5
5.0
5.5
6.0
6.5
x / μ m


E field
QFL (mV)
0
−5000
−10000
−15000
−20000
−25000
−30000
3.5
4.0
4.5
5.0
5.5
6.0
6.5
x / μ m
E field (bias)
E field (eq)
QFLs
ϕ_n
ϕ_p
0
−20
−40
−60
−80
−100
0
2
4
6
8
10

Depletion Layer Approximation

It is possible using a little extra simplification to calculate an approximate analytical solution to Poisson's equation which gives the width of the depletion region and the built in potential (which is useful for a lot of things in particular IV measurements). We don't have time to go into full details but basically the potentials are quadratic and the depletion width W is given by:

$$W = \sqrt{\frac{2\epsilon}{q} V_0 \left(\frac{1}{N_a} + \frac{1}{N_D} \right)}$$

with

$$V_0 = \frac{k_b T}{q} \ln \frac{N_A N_D}{n_i^2}$$

the built in potential

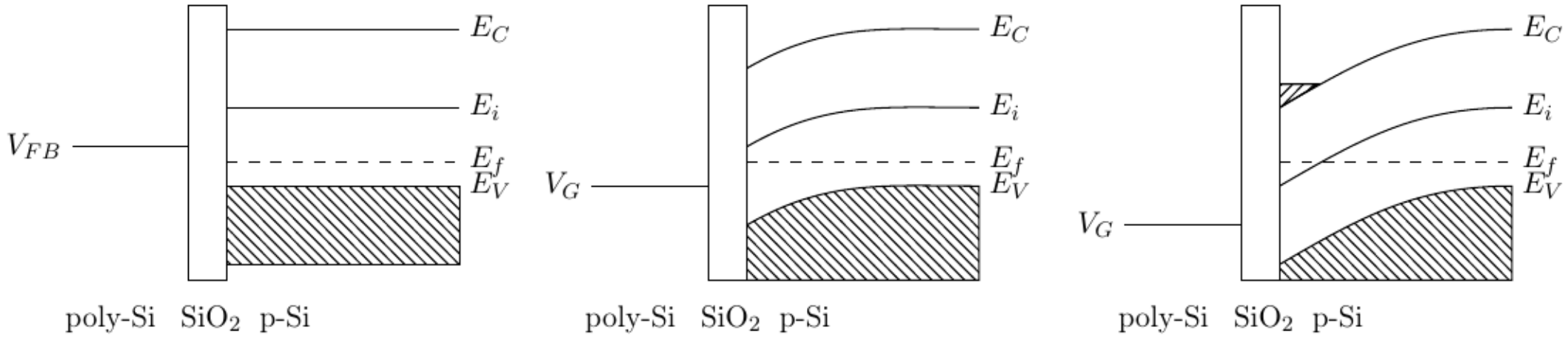
Breakdown - Zener Effect & Avalanche

A thing you will become intimately familiar with if you test or operate any silicon detectors is **breakdown**. That is, when you reverse bias a pn junction too far and it eventually starts conducting. There are two mechanisms:

1. Zener breakdown - in heavily doped & narrow depletion regions, what is happening is the field inside junction has become high enough to enable some electrons to quantum mechanically tunnel through the potential barrier. This is called the Zener effect.
2. Avalanche breakdown - in wider, lower doped regions, a random thermal fluctuation or optical generation may release a minority carrier into the depletion region, which will accelerate due to the high field. If the region is wide enough and it reaches a high enough velocity, it will have enough energy to cause impact ionisation, releasing further carriers.

Physics of the MOS capacitor

The MOS capacitor is an incredibly important device building block for MOSFET transistors and some detectors (e.g. CCDs).



The models we talked about earlier didn't really include insulating layer boundary conditions, so I won't go into detail on simulation etc. It is important to understand the four "operating regions" of the device, though. This diagram shows an n-channel MOS capacitor.

1. accumulation - pretty useless for detectors and not shown in diagram above. Applying a voltage less than the flatband voltage leads to the gate attracting holes. Only a small difference below V_{FB} is needed such that essentially all the potential is dropped across the insulator, the majority holes "screening out" the potential entirely from the semiconductor
2. flatband - setting the potential on the gate *exactly* right to have no free charges in the device at all.
3. depletion - apply more voltage, majority carriers are pushed away from the gate, forming a depletion region (similar to pn junction though mathematically slightly different). This can be used to collect signal charges (e.g. like in a CCD or modern CMOS camera).
4. inversion - the magic happens. The band gets bend so much at the insulator interface that generation of minority carriers starts happening at the interface. This is both how the channel is formed in a MOSFET, and how a CCD can keep electrons buried in a packet below the surface so they don't recombine.

Important Things we Missed Out

actual detailed physics of the generation & recombination processes. This is especialy crucial e.g. when thinking about radiation damage. Hopefully another lecture in this series covers it in more detail.

any consideration of frequency dependence particularly crucial in the MOS capacitor system

The Klaassen model of mobility the most widely used and best "basic" mobility model

discretization and solution process of the DD model really just for enthusiasts but is very interesting and helps to understand what the TCAD is actually doing (would need several more hours though!)

Recommended Reading

Solid State Physics - Ashcroft & Mermin - absolutely comprehensive presentation of solid state physics including band structure

Computational Electronics: Semiclassical and Quantum Device Modeling and Simulation - Vasileska & Goodnick - the absolute best text to follow on how to model & treat semiconductors

Physics of Semiconductor Devices - Simon Sze - reference text on operations of various types of semiconductor devices

Understanding Semiconductor Devices - Sima Dimitrijev - alternative and comprehensive presentation of semiconductor physics from a devices perspective

Transport Equations for Semiconductors - Ansgar Jungel - deep dive discussion into Boltzmann transport and higher order semi-classical models

Thanks

Sorry these are rough, this is the first year we've run this and they are brand new prepared lectures!

Any comments & questions greatly received:

daniel.weatherill@physics.ox.ac.uk