

MIT 6.012 Microelectronic Devices and Circuits

Ian Poon

January 2025

Notes adapted from [1] which is the textbook reading for MIT 6.012 Microelectronic Devices and Circuits 2009 version taught by Professor Charles Sodini

Contents

| | | |
|-----|---|----|
| 1 | Semiconductor Physics and IC technology(2)..... | 2 |
| 1.1 | Pure Semi Conductors(2.1) | 2 |
| 1.2 | Generation, Recombination, Thermal Equilibrium(2.2) | 2 |
| 1.3 | Doping(2.3) | 4 |
| 1.4 | Carrier Transport(2.4) | 7 |
| 1.5 | IC resistors (2.6) | 11 |
| 2 | Pn Junction and MOS electostatics(3)..... | 11 |
| 2.1 | applied electrostatics (3.1) | 11 |
| 2.2 | Carrier Concentration and potential in thermal equilibrium(3.2) | 12 |
| 2.3 | the pn junction: a first pass..... | 13 |
| 2.4 | the pn junction in thermal equilibrium(3.4) | 15 |
| 3 | The electrostatics of the MOS Capacitor(3.8)..... | 20 |
| 3.1 | MOS Electrostatics in Thermal Equilibrium | 20 |
| 3.2 | MOS Electrostatics under applied bias..... | 23 |
| 4 | The MOS Field Effect Transistor(4) | 27 |
| 4.1 | Gradual Channel Approximation(4.4) | 28 |
| 4.2 | MOSFET Circuit Models..... | 34 |
| 5 | The pn Junction Diode(6) | 40 |
| 5.1 | The pn Diode Circuit Symbol and Terminal Characteristics(6.1) | 40 |
| 5.2 | The pn junction a first pass(6.3)..... | 43 |
| 6 | The Bipolar Junction Transistor(7) | 48 |
| 6.1 | BJT: a first pass(7.2) | 48 |
| 6.2 | Reverse active and saturation operating regions(7.3)..... | 53 |
| 7 | Single Stage Bipolar/MOS transistor amplifiers(8) | 57 |
| 7.1 | General Amplifier Concepts(8.1) | 57 |

1 Semiconductor Physics and IC technology(2)

1.1 Pure Semi Conductors(2.1)

Fact 1

Silicon has 4 tetrahedral hybrid orbitals half filled with electrons

$$(1s)^2(2s)^2(2p)^6(3sp)^4$$

Each of the four outer hybrid orbitals have one electron each and are called **valence electrons**

Definition 2

A **intrinsic semiconductor** is one in which there are no impurities i.e pure

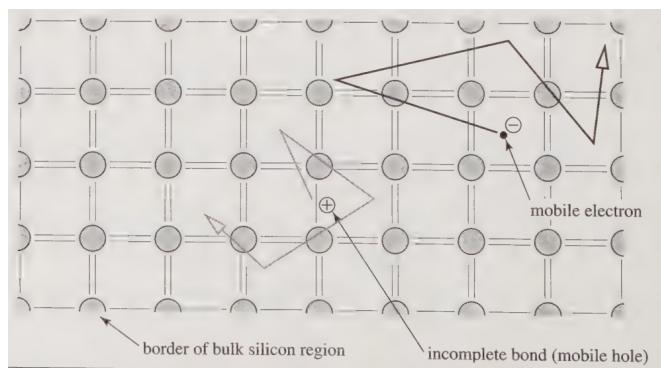


Figure 1: A hole is the incomplete bond that results when a bond is broken, freeing an electron of negative charge $-q$. The hole hence has positive charge $+q$. We take $q = 1.6 \times 10^{-19} C$ i.e the typical magnitude of charge of an electron

Fact 3

Since electrons have particle like behavior in a vacuum it seems plausible that the mobile electron is then capable of moving through the crystal lattice as verified by cathode ray tube experiments. Interestingly it was also found that the **hole** also demonstrated similar behavior due hyperconjugation(if you recall organic chemistry)

1.2 Generation, Recombination, Thermal Equilibrium(2.2)

We consider 2 ways in which bonds are broken which results in the creation and release of holes and electrons respectively. This is known as **generation**

1. **Thermal generation** - lattice vibrations that impart enough energy to valence electrons to break away from bonds
2. **optical generation** - bond absorbs photon with sufficient energy to break it

Proposition 4

The total generation rate is the sum of the thermal and optical rates

$$G = G_{\text{th}}(T) + G_{\text{op}}$$

Likewise the reverse process i.e the elimination of electrons and holes from the crystal by completing one of the broken bonds is known as **Recombination**. Types of recombination are classified according to what happens to the energy released when the bonds formed

1. **Thermal recombination** - energy goes to vibrations in the crystal lattice i.e the silicon crystal is heated
2. **optical recombination** - energy goes into photon emission

Proposition 5

For recombination to occur both an electron and a hole are necessary so it seems plausible that the recombination rate is given by

$$R = k(n \cdot p)$$

where n and p is the electron and hole concentration cm^{-3} respectively. k is the recombination rate constant $\text{cm}^{-3}\text{s}^{-1}$

Fact 6

Essentially **generation**(bond breaking) absorbs energy while **recombination**(bond forming) releases energy.

Definition 7

The situation in which recombination and generation rates are equal is known as the **steady state**. That is we have

$$R = G$$

Substituting 4 and 5, $R = G$ can be expressed as

$$n \cdot p = \frac{G_{\text{th}}(T) + G_{\text{op}}}{k} \quad (1)$$

Definition 8

An idealized special case of the steady state condition is called **thermal equilibrium** which is defined by a *prolonged absence of external energy sources*

Example 9

for example by placing the silicon sample in a "black box" at a constant temperature which is completely isolated from optical or electrical sources etc.

Theorem 10 (Mass Action Law)

In the case of **thermal equilibrium**, the RHS of 1 simplifies to a single function dependent only on temperature

$$n_o \cdot p_o = \frac{G_{\text{th}}^o(T) + G_{\text{op}}^o}{k_o} = n_i^2(T)$$

where the o subscripts or superscripts indicates thermal equilibrium.

The function of temperature defined $n_i^2(T)$ takes inspiration from equilibrium equations in chemistry. For example

$$[H^+] \cdot [OH^+] = K_{\text{eq}}(T)$$

where K_{eq} is the equilibrium constant. For intrinsic silicon in the absence of holes, $n = p = 0$. During generation we know electrons and hole concentrations are produced in pairs. But during thermal equilibrium we also know they are removed in pairs at the exact same rate. Therefore for *pure/intrinsic* silicon in thermal equilibrium essentially $n_0 = p_0$ the whole time. This is not true to impure silicon as we will see later. Therefore the **mass action law** may be rewritten as

$$n_0 \cdot p_0 = n_i^2 = p_o^2 = n_i^2(T)$$

where $n_i(T)$ is the **intrinsic concentration** in thermal equilibrium. Notice we now have $n_o = n_i = p_o$

1.3 Doping(2.3)

Donors

Example 11

Typical **donors** for silicon include *phosphorous(P), arsenic(As)* and *antimony(Sb)* all of which are in group V in the periodic table with 5 valence electrons. Consider the case of arsenic(As) below

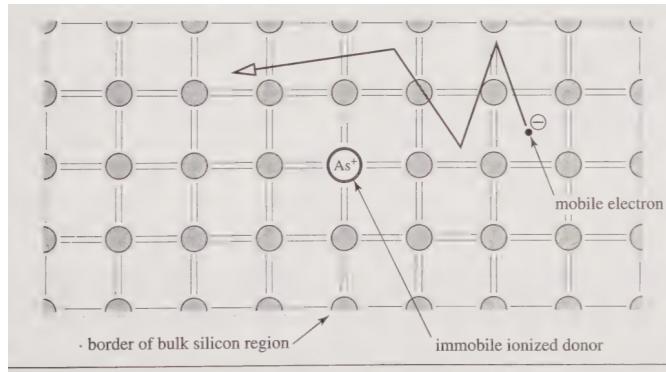


Figure 2: Arsenic substitutes one of the silicon atoms. Since it has 5 valence electrons, 4 are used to form bonds with the neighbouring silicon atoms while its last remaining valence electron is donated to the silicon lattice as mobile electron leaving behind an immobile positive charged arsenic atom. In this depiction notice the number of holes, p (incomplete bonds) remained the same at 0 but the number of mobile electrons n has increased by 1

This means unlike the case of pure silicon in the absence of holes, the number of mobile electrons is not zero. Therefore during thermal equilibrium, although likewise a pair of holes and electrons are created and removed at the same rate, the total number of holes and electrons are no longer the same.

Fact 12

As a fact we know that after doping the net charge density must still be zero. That is because as mentioned, although doping results in donation of mobile electrons it leaves behind an equivalent positive charge on the immobile dopant atoms in the lattice

Now assume thermal equilibrium and as usual denote

- n_0 concentration of free electrons in thermal equilibrium
- p_0 concentration of holes in thermal equilibrium
- N_d dopant concentration in thermal equilibrium

Then net charge density ρ being neutral condition requires that

$$\rho = -qn_0 + qp_0 + qN_d = 0$$

where qN_d clearly represents the positive charge contributions on the immobile dopants. Assuming the concentration of dopants is extremely low and their presence in the lattice does not significantly alter the *mass action law* we may invoke the mass action law by substituting $n_0 \cdot p_0 = n_i^2$ into the above equation to get

$$\frac{n_i^2}{n_0} + N_d - n_0 = 0$$

which we may solve using the quadratic formula to obtain

$$n_0 = \frac{N_d + \sqrt{N_d^2 + 4n_i^2}}{2} = \frac{N_d}{2} + \frac{N_d}{2} \sqrt{1 + \frac{4n_i^2}{N_d^2}} \quad (2)$$

However the concentration of dopants although low relative to the volume of the lattice(so mass action law still applies), is not necessarily low relative to intrinsic concentration.

Fact 13

For context, $n_i \simeq 10^{10} \text{ cm}^{-3}$ at 300K (room temperature). Given the number of bonds in the lattice is approximately $4 \times N_{\text{Si}} = 2 \times 10^{23} \text{ cm}^{-3}$, we can say that at room temperature an extremely small fraction of the bonds are broken

Definition 14

We call such a state where $N_d \gg n_i$ **extrinsic**

Notice when $N_d \gg n_i$ 2 simplifies to

$$n_0 \approx \frac{N_d}{2} + \frac{N_d}{2} = N_d$$

Because mass action law still holds, we have

$$p_0 = \frac{n_i^2}{n_0} = \frac{n_i^2}{N_d}$$

With this we finally see the effect of doping.

Theorem 15

Doping with donors depresses the hole concentration of silicon in thermal equilibrium conditions

Proof. Consider the case of intrinsic vs extrinsic silicon in thermal equilibrium at the same temperature. That is to say n_i which we recall to be a function only of T is the same for both cases. As we recall from for the pure case from mass action law

$$p_o = n_o = n_i$$

However for doped silicon case where $p_o \neq n_o$ we have from mass action law that $p_o = n_i^2/N_d$ where we know $N_d \gg n_i$. Therefore the hole concentration has been depressed!

Definition 16

We refer to silicon doped with donors as **n-type** in which electrons are **majority carriers** and holes are the **minority carriers**(as they are depressed relative to mobile electrons)

Acceptors

Example 17

A substitutional impurity from Group III of the periodic table where they have 3 valence electrons is called an **acceptor**

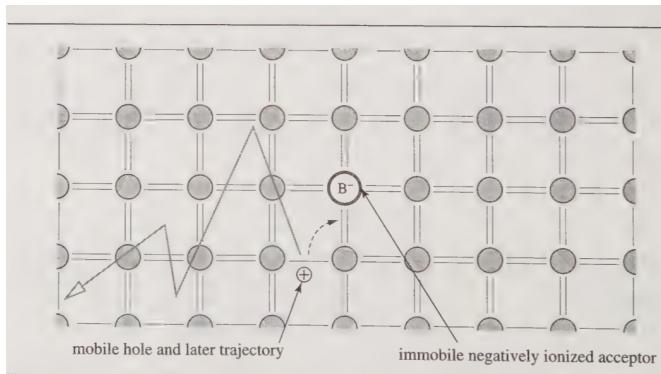


Figure 3: Bond model showing Boron on a substitutional lattice in a bulk region of the silicon crystal. A hole is created when the incomplete fourth bond is filled by acceptance of an electron from a neighboring bond. So this means in the absence of mobile electrons $n = 0$ the number of holes p is not identically 0. This is the reverse of the case for donors earlier

Using the same reasoning as for acceptors above we know that charge density must still be equal zero that is

$$\rho = -qn_o + qp_o - qN_a$$

where N_a refers now to the acceptor concentration. By the mass action law, subbing in $n_o = n_i^2/p_o$ then as usual using the quadratic formula we obtain

$$p_o = \frac{N_a}{2} + \frac{N_a}{2} \sqrt{1 + \frac{4n_i^2}{N_a^2}}$$

Analogously in the practice where $N_a \gg n_i$ (but still low relative to lattice volume so mass action law still applies) we have by the same reasoning as previously that

$$p_o \simeq N_a$$

therefore now by the mass action law we see that

$$n_o = \frac{n_i^2}{p_o} = \frac{n^2}{N_a}$$

so now clearly this time the concentration of mobile electrons is depressed instead.

Corollary 18

Doping with acceptors depresses the concentration of mobile electrons

Definition 19

We refer to silicon doped with acceptors as **p-type**

Donors and Acceptors: Compensation

Silicon is often doped with both donors and acceptors. By the same reasoning as before we still must have charge neutrality therefore

$$\rho = -qn_0 + qp_0 + qN_d - qN_a = 0$$

As usual subbing in $p_o = n_i^2/n_o$ from the mass action law we find

$$n_o^2 - (N_d - N_a)n_o - n_i^2 = 0$$

so by the quadratic formula we have

$$n_o = \frac{N_d - N_a}{2} + \frac{N_d - N_a}{2} \sqrt{1 + \frac{4n_i^2}{(N_d - N_a)^2}}$$

where we have assumed $N_d > N_a$. Similar reasoning for the other case...

1.4 Carrier Transport(2.4)

Drift Velocity

Definition 20

At room temperature the holes and electrons in silicon are in random motion and their average velocity is called the **thermal velocity** v_{th} which is about 10^7 cm/s or $1/3000$ that of light.

Definition 21

The average **collision time** τ_c at room temperature is about $1 \times 10^{-13} \text{ s} = 0.1 \text{ picoseconds}$.

So we see that between collisions the carries on average will travel a distance

$$\lambda = v_{th}\tau_c = 10 \text{ nm} = 0.01 \mu\text{m}$$

where λ is defined to be the **means free path**

Theorem 22

Experimentally drift velocity v_{dn} is found to be proportional to electric field

$$v_{dn} = \mu_n E$$

where μ_n is the **electron mobility** and $v_{dn} \ll v_{sat}$

Similarly for holes

Theorem 23

The drift velocity for holes is also proportional to the electric field

$$v_{dp} = \mu_p E$$

where μ_p is the **hole mobility** and $|v_{dp}| \ll v_{sat}$

Drift Current Density

Definition 24

Current density J is defined as the rate at which charge passes through a reference plane in the $+x$ direction per unit area

It has units $C/(s \text{ cm}^2)$. Then we have

$$J_p^{dr} = \frac{\Delta Q_p}{A\Delta t} = \frac{qp\Delta V}{A\Delta t} = qp v_{dp} = qp\mu_p E$$

where the last relation comes from substituting 23 in. Similarly we have

$$J_n^{dr} = -qnv_{dn} = qn\mu_n E$$

Note that the sign convention is chosen such that

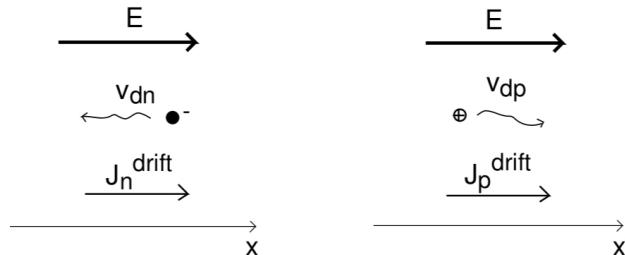
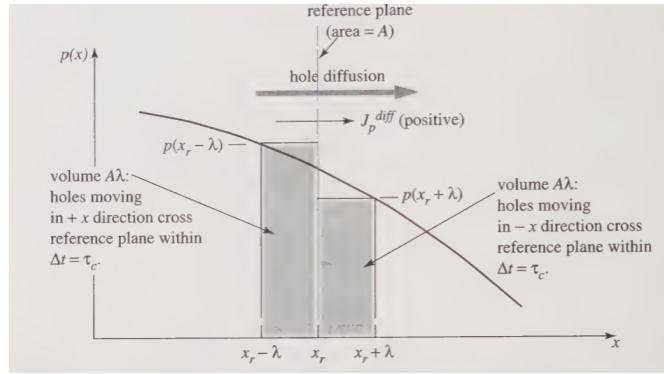


Figure 4: Notice that in these pairs of drift and electric field directions, J is positive

Diffusion Current Density



Probabilistically speaking only half of holes starting in either region across the reference plane will actually cross the reference plane while the other half will move away from the reference. Accounting the holes that do cross the reference plane and recalling current to be the charge passing through per unit area per unit time we have that

$$J_p^{\text{diff}}(x = x_r) = q \left[\frac{0.5p(x = x_r - \lambda)A\lambda - 0.5p(x = x_r + \lambda)A\lambda}{A\tau_c} \right]$$

where $p(x)$ denotes the concentration of holes at x . Now using the first order taylor expansions we then obtain

$$J_p^{\text{diff}}(x = x_r) = \frac{q\lambda}{\tau_c} \left(\frac{p(x_r) - \frac{dp}{dx}\Big|_{x_r} \lambda - \left(p(x_r) + \frac{dp}{dx}\Big|_{x_r} \lambda \right)}{2} \right)$$

canceling the hole concentration and collecting terms the above reduces to

$$J_p^{\text{diff}}(x = x_r) = -q \left(\frac{\lambda^2}{\tau_c} \frac{dp}{dx} \Big|_{x=x_r} \right) = -qD_p \frac{dp}{dx} \Big|_{x=x_r}$$

the term $\frac{\lambda^2}{\tau_c}$ is identified as the **hole diffusion constant** D_p with units cm^2/s .

Theorem 25

So we have shown that the relation for diffusion current density of holes is

$$J_p^{\text{diff}} = -qD_p \frac{dp}{dx}$$

this proportionality of the diffusion current desnity to the concentration gradient is known as **Fick's Law**

In a similar manner we derive the relation for electrons

$$J_n^{\text{diff}}(x = x_r) = q \left(\frac{\lambda^2}{\tau_c} \frac{dn}{dx} \Big|_{x=x_r} \right) = qD_n \frac{dn}{dx} \Big|_{x=x_r}$$

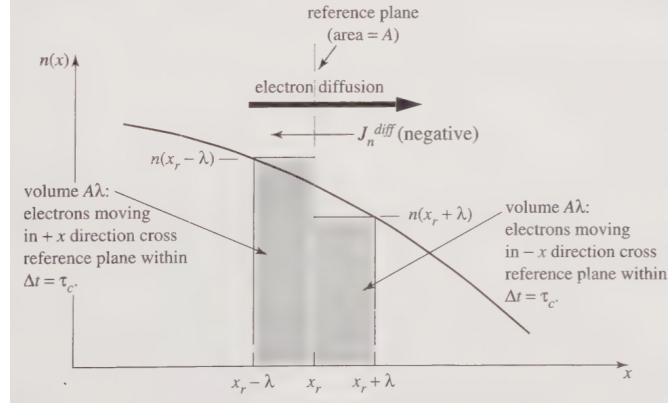


Figure 5: is just the same as the case for holes but the directions are reversed due to the charge

Theorem 26

Einstein Relation states that

$$\frac{D_p}{\mu_p} = \frac{KT}{q} \quad \text{and} \quad \frac{D_n}{\mu_n} = \frac{KT}{q}$$

Proof. The proof of this is out of scope for this course (check out solid state physics OCW!) but we will consider the implications of this below.

Definition 27

The **thermal voltage** is defined by

$$V_{th} = \frac{kT}{q}$$

total current density

Therefore the total current density is given by

$$J = J_n + J_p = J_n^{dr} + J_n^{\text{diff}} + J_p^{dr} + J_p^{\text{diff}}$$

which upon substitution with results from previous sections

$$J = qn\mu_n E + qD_n \frac{dn}{dx} + qp\mu_p E - qD_p \frac{dp}{dx}$$

Theorem 28

The total electron and hole densities J_n and J_p are hence given by

$$J_n = J_n^{dr} + J_n^{\text{diff}} = qn\mu_n E + qD_n \frac{dn}{dx}$$

and

$$J_p = J_p^{dr} + J_p^{\text{diff}} = qp\mu_p E - qD_p \frac{dp}{dx}$$

1.5 IC resistors (2.6)

2 Pn Junction and MOS electostatics(3)

2.1 applied electrostatics (3.1)

Gauss Law

Theorem 29

In differential form **Gauss Law** is given by

$$\frac{dE}{dx} = \frac{\rho}{\epsilon}$$

Proof. To see how this derived from the integral form where we have:

$$\oint_{\partial V} \mathbf{E} \cdot \mathbf{n} dA = \frac{Q_{\text{enc}}}{\epsilon_0},$$

where \mathbf{E} is the electric field, ∂V is a closed surface enclosing the volume V , \mathbf{n} is the unit normal to the surface, Q_{enc} is the total charge enclosed in V , and ϵ_0 is the permittivity of free space.

The total charge enclosed, Q_{enc} , can be expressed in terms of the charge density ρ :

$$Q_{\text{enc}} = \int_V \rho dV.$$

Using the divergence theorem, which relates the surface integral of a vector field to the volume integral of its divergence, we write:

$$\oint_{\partial V} \mathbf{E} \cdot \mathbf{n} dA = \int_V (\nabla \cdot \mathbf{E}) dV.$$

Substituting this result into Gauss's Law gives:

$$\int_V (\nabla \cdot \mathbf{E}) dV = \frac{1}{\epsilon_0} \int_V \rho dV.$$

Since this equality holds for any arbitrary volume V , the integrands must be equal. Thus, we obtain:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}.$$

For a one-dimensional case, where the electric field \mathbf{E} varies only along the x -axis, the divergence simplifies to a partial derivative:

$$\nabla \cdot \mathbf{E} = \frac{\partial E}{\partial x}.$$

Hence, Gauss's Law in differential form becomes:

$$\frac{\partial E}{\partial x} = \frac{\rho}{\epsilon_0}.$$

□

Then integrating along the interval $x_a \leq x' \leq x$ we find

$$\int_{x_a}^x d[\epsilon E(x')] = \epsilon E(x) - \epsilon_a E(x_a) = \int_{x_a}^x \rho(x') dx' = Q(x)$$

Potential and Poisson Equation

Definition 30

The electrostatic potential $\phi(x)$ is defined with respect to its value at a point x_0 as the integral of the negative of the electric field E from x_0 to x

$$\phi(x) - \phi(x_0) = \int_{x_0}^x -E(x)dx$$

so we have

$$E(x) = -\frac{d\phi(x)}{dx}$$

Then incorporation of **Gauss Law** gives

$$\frac{d^2\phi(x)}{dx^2} = -\frac{dE(x)}{dx} - \frac{\rho(x)}{\epsilon}$$

this result is known as **Poisson's equation**

2.2 Carrier Concentration and potential in thermal equilibrium(3.2)

Recall 8 that the thermal equilibrium is equivalent to the absence of any stimulus to the device - zero applied voltage, external light source etc. Then this implies the electron and hole current densities must both be equal to zero in thermal equilibrium

To see this consider that in the absence of an external electric field or energy sources the charge distribution should not change over time if not that would imply some net energy intake occurred to cause such a redistribution.

$$0 = qn_o\mu_n E_o + qD_n \frac{dn_o}{dx}$$

Essentially this means

Theorem 31

diffusion precisely balances drift in thermal equilibrium so there is no net change in charge distribution spatially over time.

More explicitly if you plot n_0, p_0 against x you should not expect the shape of the graph to change over time. Now rearranging to solve for the $\frac{dn_o}{dx}$ (as in middle term) and then substituting **Einstein's Relation 26** and the definition of electric field 30(as in last term) we obtain

$$\frac{dn_o}{dx} = \left(-\frac{\mu_n}{D_n}\right) n_o E_o = \left(-\frac{q}{kT}\right) n_o \left(-\frac{d\phi_o}{dx}\right)$$

rearranging(middle term) and substituting the definition of **thermal voltage**(last term)27 we obtain

$$d\phi_o = \left(\frac{kT}{q}\right) \frac{dn_o}{n_o} = V_{th} \frac{dn_o}{n_o}$$

Then upon integrating from x to x_0 (arbitrary reference point) gives

$$\phi_o(x) - \phi_o(x_0) = V_{th} \ln \left(\frac{n_o(x)}{n_o(x_0)} \right)$$

Applying the boundary condition which by convention we choose to be

$$\phi_o(x_o) = 0 \quad \text{when} \quad n_o(x_o) = n_i$$

where n_i is the **intrinsic concentration** if you recall.

Remark 32. well in order to find electrostatic potential you must always take a reference potential. In this case it is helpful for future calculations to zero it at some point x_0 where $n_0(x_0) = n_i$ which is basically when $n_0 = p_0 = n_i$ in intrinsic silicon bulk region if you recall from the 1st section.

we finally get

Theorem 33

the potential in terms of electron concentration in thermal equilibrium is

$$\phi_o(x) = V_{th} \ln \left(\frac{n_o(x)}{n_i} \right)$$

which can be expressed equivalently as

$$n_o(x) = n_i e^{\phi_o(x)/V_{th}}$$

Corollary 34

Similarly for holes we get

$$\phi_o(x) = -V_{th} \ln \left(\frac{p_o(x)}{n_i} \right)$$

which can be expressed equivalently as

$$p_o(x) = n_i e^{-\phi_o(x)/V_{th}}$$

These are also known as the **Boltzmann relations**

2.3 the pn junction: a first pass

Fact 35

A **pn junction** consists of p-type silicon region immediately adjacent to an n-type silicon region.

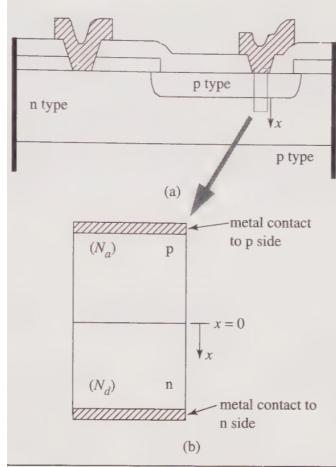
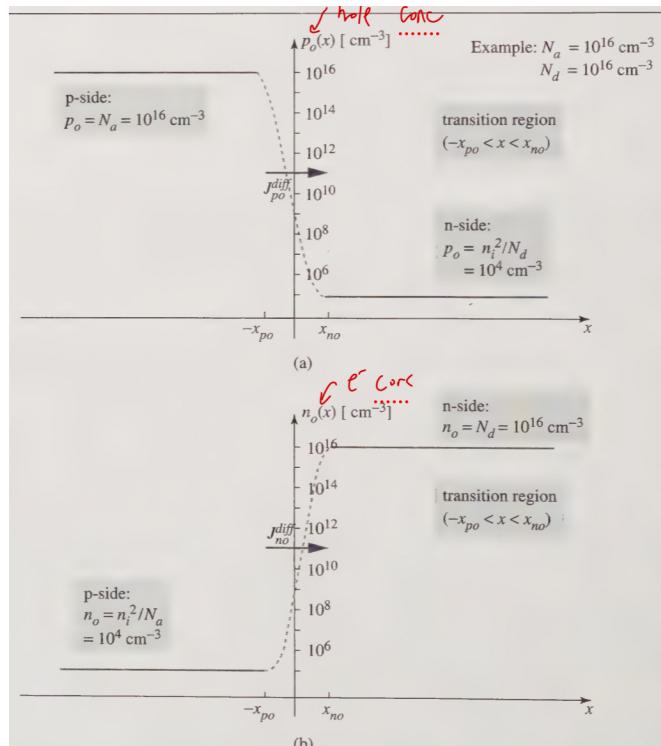


Figure 6: (a) cross section of a typical pn junction fabricated in an IC process. (b) one dimensional pn junction consisting of boxed region in (a) with metal contact to p side and n side added at the bottom

To analyze this first we consider the case of **thermal equilibrium**.



Recall from earlier section on doping that on the p-side, solving for net charge neutrality and applying the mass action law we have

$$p_0 = N_a \quad \text{and} \quad n_0 = \frac{n_i^2}{N_a} \quad (3)$$

and in similar fashion on the n side

$$n_0 = N_d \quad \text{and} \quad p_0 = \frac{n_i^2}{N_d} \quad (4)$$

However it turns out as seen from the image that there exists a "transition region" that extends x_{po} and x_{no} on the p and n side respectively instead of having the values n_0 and p_0 make sudden jumps at the boundary $x = 0$. How and why does this occur?

- Charge Redistribution at the Interface:** When two materials come into contact (e.g., a p-type and n-type semiconductor), there is initially a discontinuity in the carrier concentrations across the interface. For example:
 - In the **n-type region**, electrons are the majority carriers, and their concentration is high.
 - In the **p-type region**, holes are the majority carriers, and their concentration is high.

This discontinuity leads to:

- Diffusion of carriers:** Electrons from the n-side diffuse into the p-side, where their concentration is lower, and recombine with holes. Similarly, holes from the p-side diffuse into the n-side.
- As electrons and holes diffuse, they leave behind **fixed ionized donor atoms** (N_D^+) in the n-side and **fixed ionized acceptor atoms** (N_A^-) in the p-side, resulting in a net positive charge on the n-side and a net negative charge on the p-side.

- Formation of the Transition Region:** The accumulation of fixed charges near the interface creates an **electric field** pointing from the positively charged n-side to the negatively charged p-side. This electric field has two effects:

- Drift Motion Opposing Diffusion:** The electric field exerts a force on the remaining charge carriers, opposing further diffusion. Electrons are pulled back to the n-side, and holes are pulled back to the p-side. Eventually, a balance between the diffusion current and the drift current is reached, leading to **equilibrium**.
- Depletion of Mobile Carriers:** The electric field sweeps out mobile charge carriers near the interface, leaving behind a region dominated by fixed charges. This is the **depletion region**, which forms the core of the transition region.

2.4 the pn junction in thermal equilibrium(3.4)

Before we analyze this we first 28 where we know that the electron drift and diffusion current densities must cancel. Same for holes. In which case upon rearrangement of the equations in 28 we have

$$E_o = \frac{qD_p \frac{dp_0}{dx}}{qp_0\mu_p} = \frac{-qD_n \frac{dn_0}{dx}}{qn_0\mu_n}$$

Immediately we see that the electric field outside the transition region(i.e the bulk region) must be zero given that p_0 and n_0 are constant there. But what about the electric field in the transition region? There must be some points within the transition region where it is non zero for sure. Because electric field is a result of a net difference in charge over a distance it makes sense of analyze charge density next.

quantative analysis using the depletion approximation

The charge density of the transition area which is given by the below

$$\rho_o(x) = q(p_o - n_o + N_d - N_a) \equiv \begin{cases} q(p_o - N_a), & (-x_{po} \leq x \leq 0) \\ q(N_d - n_o), & (0 \leq x \leq x_{no}) \end{cases}$$

the cases are split like so because for the transition region on the p side we know $N_d = 0$ and n_0 is very small. Similarly for the n side we know $N_a = 0$ and p_0 is very small hence we may treat them as negible in our calculations. Now

substituting our results for p_0 and n_0 from 22 and 34 we have

$$\rho_o(x) \equiv \begin{cases} q(n_i e^{-\phi_o(x)/V_{th}} - N_a), & (-x_{po} \leq x \leq 0) \\ q(N_d - n_i e^{\phi_o(x)/V_{th}}), & (0 \leq x \leq x_{no}) \end{cases}$$

To further simplify this we invoke what is known as the **depletion approximation** which assumes that the transition region is completely depleted of holes and electrons i.e $n_0 = p_0 = 0$ (aka where the majority carriers are negligible compared to dopant concentrations. also notice the logarithmic scale on 2.3) So all in all we now have

$$\rho_o(x) \approx \begin{cases} -qN_a, & (-x_{po} \leq x \leq 0) \\ qN_d, & (0 \leq x \leq x_{no}) \end{cases} \quad \text{and} \quad \rho_o(x) = \begin{cases} 0, & (x < -x_{po}) \\ 0, & (x_{no} \leq x) \end{cases}$$

Remark 36. You should then immediately notice from 2.3 that this approximation may not be very accurate at $-x_{po}$ and x_{no} (the edges) where the majority carriers are approximately equal to the dopant concentration i.e $N_a \approx p_0$ and $N_d \approx n_0$ so we can't exactly treat them as relatively negligible to dopant concentrations there.

Note the second set of cases correspond to areas outside the transition region which are located in the bulk i.e the interior region of silicon which we know are areas away from edges, boundaries or defects. That is without dopants, the pure silicon lattice in that area all of complete octet forming four bonds each and so are charge neutral. As we know from earlier sections introducing dopants in that area will not affect net charge neutrality. Therefore $\rho_0 = 0$ outside the transition region. This is consistent with the n_0, p_0 values found there which are exactly the values found by solving for charge neutrality as we did in the earlier section on donors.

The reason for doing so is that such an approximation yields nice solutions in solving the poisson equations. We now demonstrate it and solve for the electric field as desired. First consider gauss law

$$\frac{dE_o}{dx} = \frac{\rho_o(x)}{\epsilon_s}$$

Now perform integration on the p side then sub in the depletion approximation values for ρ_0 in the transition region. Note that we have approximated $E_o(x = -x_{po}) = 0$ once again because that is where the bulk region is

$$E_o(x) = \int_{-x_{po}}^x \frac{\rho_o(x')}{\epsilon_s} dx' + E_o(x = -x_{po}) = \int_{-x_{po}}^x \frac{-qN_a}{\epsilon_s} dx' + 0 = -\frac{qN_a}{\epsilon_s} (x - (-x_{po}))$$

Having now found an expression for the electric field we integrate it to find the potential recalling 30

$$\phi_o(x) = \int_{-x_{po}}^x -E_o(x') dx' + \phi_o(x = -x_{po}) = \frac{qN_a}{2\epsilon_s} (x + x_{po})^2 + \phi_p$$

where ϕ_p is the *potential of the p type bulk p region in thermal equilibrium* which we know from 34 is given by

$$\phi_p = -V_{th} \ln \left(\frac{N_a}{n_i} \right)$$

In the exact same way do the same for the n side

$$E_o(x = x_{no}) = E_o(x) + \int_x^{x_{no}} \frac{\rho_o(x')}{\epsilon_s} dx' = E_o(x) + \int_x^{x_{no}} \frac{qN_d}{\epsilon_s} dx' = E_o(x) + \frac{qN_d}{\epsilon_s} (x_{no} - x)$$

we again know $E_o(x = x_{no}) = 0$ so the above simplifies to

$$E_o(x) = -\frac{qN_d}{\epsilon_s} (x_{no} - x)$$

and now integrating to find the potential

$$\phi_o(x) = \phi_o(x_{no}) + \int_x^{x_{no}} -E_o(x') dx' = \phi_o(x_{no}) + \int_x^{x_{no}} \frac{qN_d}{\epsilon_s} (x_{no} - x') dx'$$

which is evaluated to

$$\phi_o(x) = \phi_n - \frac{qN_d}{2\epsilon_s} (x_{no} - x)^2, \quad (0 \leq x \leq x_{no})$$

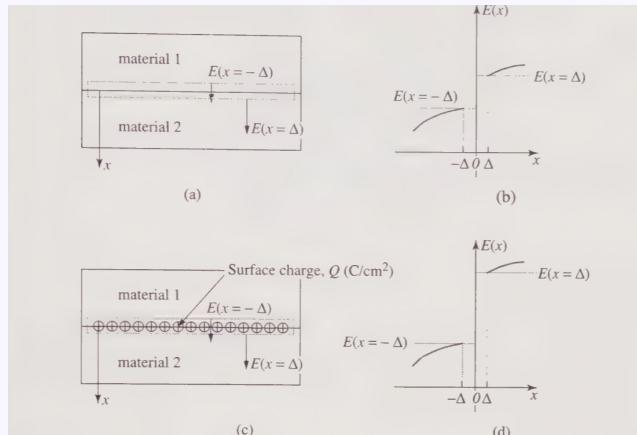
where ϕ_n is the *potential of the n type bulk n region in thermal equilibrium*

$$\phi_n = V_{th} \ln \left(\frac{N_d}{n_i} \right)$$

The two solutions for electric field must connect at $x = 0$ because there is neither a sheet of charge at $x = 0$ and on both sides of $x = 0$ they are the same material(silicon) and hence same permittivity(ϵ_s).

Lemma 37

Consider



► **Figure 3.4** (a) Boundary between materials 1 and 2 with permittivities $\epsilon_1 > \epsilon_2$ and (b) resulting jump in the electric field. (c) Boundary, with a surface charge Q located at the interface and (d) resulting jump in the electric field.

$$\text{Figure 7: where } E(0^+) = \left(\frac{\epsilon_1}{\epsilon_2} \right) E(0^-) + \frac{Q}{\epsilon_2}$$

Proof. We derive the relationship for the electric field at a boundary between two materials with permittivities ϵ_1 and ϵ_2 , where the boundary is located at $x = 0$. The electric field may jump discontinuously at the boundary, and we use Gauss's law to find this relationship.

The potential $\phi(x)$ must be continuous across the boundary:

$$\phi(x_b^+) = \phi(x_b^-),$$

to avoid infinite electric fields. However, the electric field \mathbf{E} , which is the gradient of the potential, may have a discontinuity due to differences in the material properties and any surface charge density present at the boundary.

To find the relationship for the electric field at $x = 0$, we consider the integral form of Gauss's law over a small region spanning the boundary:

$$\int_{-\Delta}^{+\Delta} \frac{d}{dx}(\epsilon(x)E(x)) dx = \int_{-\Delta}^{+\Delta} \rho(x) dx.$$

Expanding the derivative:

$$\int_{-\Delta}^{+\Delta} \epsilon(x) \frac{dE}{dx} dx + \int_{-\Delta}^{+\Delta} E(x) \frac{d\epsilon}{dx} dx = \int_{-\Delta}^{+\Delta} \rho(x) dx.$$

For a small pillbox centered on the boundary, we let $\Delta \rightarrow 0$, and the electric field is approximately constant in each region. The material properties ϵ_1 and ϵ_2 are piecewise constant, and $\rho(x)$ is a surface charge density Q at $x = 0$. Thus, the integrals reduce to:

$$\epsilon_2 E(0^+) - \epsilon_1 E(0^-) = Q.$$

Rearranging the above equation:

$$E(0^+) = \left(\frac{\epsilon_1}{\epsilon_2}\right) E(0^-) + \frac{Q}{\epsilon_2}.$$

□

So by gauss law the electric field must be continuous there. Equating the electric fields(see boxed equations above) to be the same at $x = 0$ we have

$$-\frac{qN_a}{\epsilon_s} (0 + x_{po}) = -\frac{qN_d}{\epsilon_s} (x_{no} - 0)$$

which simplifies to

$$N_a x_{po} = N_d x_{no}$$

Finally we would like to find an expression for x_{no} and x_{po} . Because the electric field is continuos at $x = 0$ so must be the potential given the electric field is the gradient of the potential. Recall $C^1 \rightarrow C^0$ from analysis. So equating the potentials(see boxed equations above) to be the same at $x = 0$ we get

$$\frac{qN_a}{2\epsilon_s} (0 + x_{po})^2 + \phi_p = \phi_n - \frac{qN_d}{2\epsilon_s} (x_{no} - 0)^2 = \phi_n - \frac{qN_d}{2\epsilon_s} \left(\frac{N_a}{N_d} x_{po}\right)^2$$

then using $N_a x_{po} = N_d x_{no}$ we can solve simulataneous equations to get

$$x_{po} = \sqrt{\frac{2\epsilon_s \phi_B}{qN_a} \left(\frac{N_d}{N_d + N_a}\right)}$$

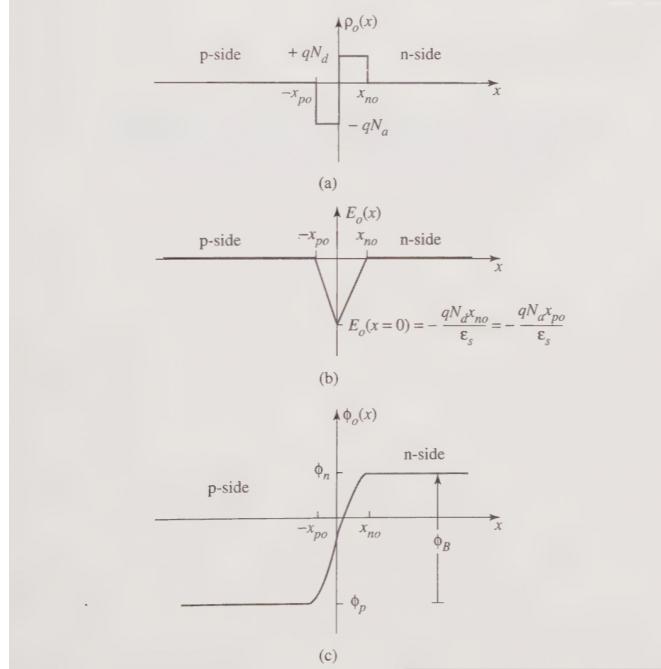
and

$$x_{no} = \sqrt{\frac{2\epsilon_s \phi_B}{qN_d} \left(\frac{N_a}{N_d + N_a}\right)}$$

Hence the width of the **depletion region** is found to be

$$X_{do} = x_{no} + x_{po} = \sqrt{\frac{2\epsilon_s \phi_B}{q} \left(\frac{1}{N_a} + \frac{1}{N_d}\right)}$$

All in graphically we have



► **Figure 3.11** (a) Charge density, (b) electric field, and (c) potential in thermal equilibrium based on the depletion approximation.

Definition 38

The **built-in potential** ϕ_B is defined as

$$\phi_B = \phi_{n^+} - \phi_p$$

where as defined above

- ϕ_{n^+} is the *potential of the n type bulk n region in thermal equilibrium*
- ϕ_p is the *potential of the p type bulk p region in thermal equilibrium*

Definition 39

The charge stored per unit area on the p-side is called the **depletion charge**. In thermal equilibrium the depletion charge Q_{jo} is given by

$$Q_{jo} = -qN_a x_{po}$$

Fact 40

Note that in summary depletion approximation is powerful tool to simplify the solving of poisson equations but it requires that carrier concentrations in **space charge region**(the transition region) is much smaller than the doping level.

space charge region, depletion region, transition zone are all synonymns of each other

3 The electrostatics of the MOS Capacitor(3.8)

3.1 MOS Electrostatics in Thermal Equilibrium

The analysis of MOS essentially also uses the thermal equilibrium and electrostatic relationships we found above. We also made use of boundary conditions like

1. $\phi_0(x) = \phi_{n^+}, x \leq t_{ox}$
2. $\phi_0(x) = \phi_p, x \geq X_{do}$
3. $\rho(x), E_0(x) = 0, x \geq X_{do}$

to elimate unknowns when we write out the relationships. Additionally we simplified our calculations as we also were able to invoke **depletion approximation** as X_{do} is designed to be very close to the gate oxide interface. This means the **space charge region** the density of the fixed charges occupying the depletion region is far larger relative to the mobile charges. More explanation of this follows below

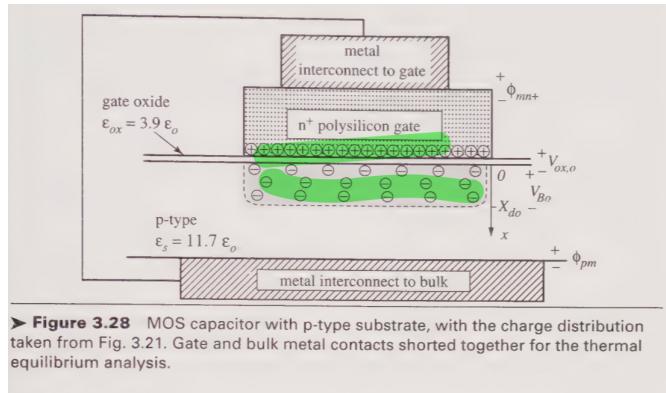


Figure 8: notice that V_{bo} is the PD across the depletion region while $V_{ox,o}$ is the PD across the oxide gate

Now recall under thermal equilibrium there can be no applied external voltage. In this case V_{GB} (the applied voltage) between the metal interconnect to gate and interconnect to bulk is zero. i.e they are shorted to one another. Now the n^+ means that the polysilicon gate is very heavily doped with donors such that its bulk region potential in thermal equilibrium takes the maximum possible

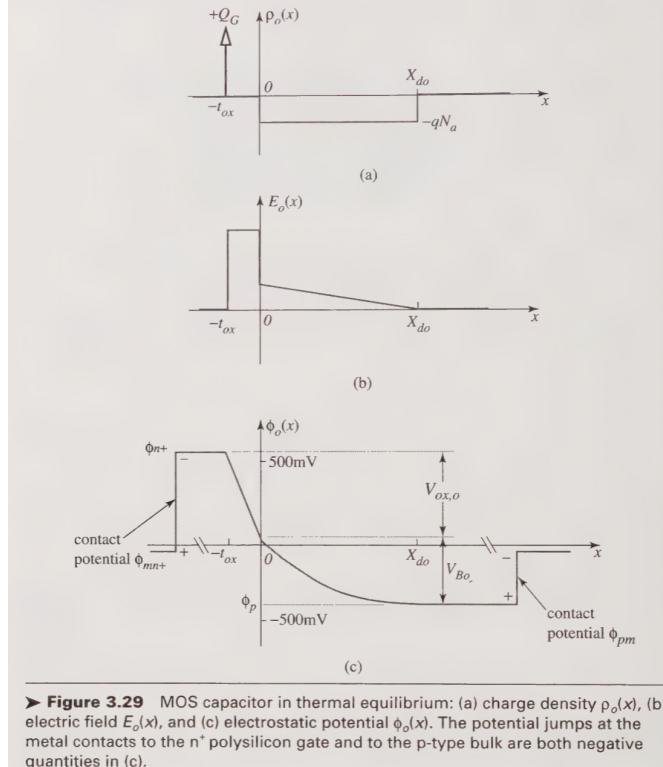
$$\phi_{n^+} = 550mV$$

As for the bulk potential in silicon it is given by the usual

$$\phi_p = -V_{th} \ln \left(\frac{N_a}{n_i} \right)$$

which is about $-420mV$ for a typical doping concentration of $10^{17} cm^{-3}$. This means the sign of the potential drop will point from gate to bulk as shown. Recall potential points from high to low. Since polysilicon being highly conductive, the **gate charge** Q_{Go} is depicted as a sheet charge located on its bottom surface. i.e all the mobile charges gather at the bottom. The gate oxide is considered as a charge free perfect insulator. Hence with no charges crossing, the positive sheet charge repels mobile holes in the silicon under the gate oxide, leaving behind immobile negatively charged acceptor ions to a depth X_{do} . This essentially creates the depletion region with **bulk charge** Q_{Bo} . By conservation of charge we know that

$$Q_{Go} = -Q_{Bo} = -qN_aX_{do} \quad (5)$$



► **Figure 3.29** MOS capacitor in thermal equilibrium: (a) charge density $\rho_o(x)$, (b) electric field $E_o(x)$, and (c) electrostatic potential $\phi_o(x)$. The potential jumps at the metal contacts to the n⁺ polysilicon gate and to the p-type bulk are both negative quantities in (c).

Figure 9: the values ϕ_p, ϕ_{n^+} should make sense from 2.4

Also notice $V_{ox,o} + V_{Bo} = \phi_B = \phi_{n^+} - \phi_p$

Definition 41

The dashed line on the $\phi_o(x)$ graph corresponds to the **surface potential** ϕ_{so} which is defined to be the potential at $x = 0$ (the SiO₂ interface) at thermal equilibrium.

First notice how this makes sense. Invoking **depletion approximation** (as justified at the start) we recall from the section on doping that in p side depletion region the charge density will be $-qN_a$ and then afterwards the charge density will be zero. Also recall that the gate oxide was assumed chargeless so $\rho_o(x) = 0$ from $-t_{ox} \leq x \leq 0$. The relationships for E_o, ϕ_o follow by direct integration. Notice by gauss law

$$E_o(0^+) = -\frac{Q_{Bo}}{\epsilon_s} = \frac{qN_a X_{do}}{\epsilon_s}$$

taking a pill box around the depletion region and noticing that the surface at $x = X_{do}$ does not contribute any flux because $E_o(X_{do}) = 0$ there.

Remark 42. its just like how you take gauss law across the surface of a conductor. No contribution by the interior surface as interior of conductor has zero electric field

Now applying the jump discontinuity relationship using the differential form of gauss law as found in 7 notice that

$$E_o(0^-) = \frac{\epsilon_s}{\epsilon_{ox}} \left(\frac{qN_a X_{do}}{\epsilon_s} \right) = \frac{qN_a X_{do}}{\epsilon_{ox}} = E_{ox}$$

where E_{ox} is the electric field inside the metal oxide gate. Across the thickness of the gate, t_{ox} we know that E_{ox} remains uniform as shown in the figure above because as mentioned E_{ox} is free of charge and has uniform permissibility

again by 7. Therefore assuming uniform E_{ox} we carry out integration over $-t_{ox} \leq x \leq 0$ to find the potential as follows

$$\phi_o(x) - \phi_{n^+} = \int_{-t_{ox}}^x -E_{ox} dx = -E_{ox}(x + t_{ox}) = -\frac{qN_a X_{do}}{\epsilon_{ox}}(x + t_{ox})$$

noticing that at the oxide/silicon interface at $x = 0$ the potential is

$$\phi_o(0) = \phi_{n^+} - \frac{qN_a X_{do} t_{ox}}{\epsilon_{ox}}$$

Therefore the potential drop across the oxide gate $V_{ox,o}$ 8 can be expressed as

$$V_{ox,o} = \phi_{p_n^+} - \phi_o(0) = \frac{qN_a X_{do}}{C_{ox}} = \frac{Q_{GO}}{C_{ox}}$$

since $C_{ox} = \frac{\epsilon_{ox}}{t_{ox}}$ by definition of capacitance and by kirchoff law we have

$$-\phi_{mn^+} - V_{ox,o} - V_{bo} - \phi_{pm} = 0$$

Now looking at $0 \leq x \leq X_{do}$

$$\epsilon_s E_o(x) - \epsilon_s E_o(0^+) = \int_{0^+}^x -qN_a dx = -qN_a x$$

Evaluating the above and subbing the relation we found for $E_0(0^+)$ above via gauss law we obtain

$$E_o(x) = \frac{qN_a(X_{do} - x)}{\epsilon_s}$$

Then as usual integrating our found expression for E_0 over this same range to find the potential

$$\phi_o(x) - \phi_o(0) = \int_0^x -E_o(x') dx' = -\frac{qN_a}{\epsilon_s} \left(X_{do}x - \frac{x^2}{2} \right)$$

which upon evaluation and subbing in the expression for $\phi_o(0)$ found previously(again we know potential cannot be discontinuous or infinite electric field will result) we get

$$\phi_o(x) = \phi_{p_n^+} - \frac{qN_a X_{do}}{C_{ox}} - \frac{qN_a}{\epsilon_s} \left(X_{do}x - \frac{x^2}{2} \right)$$

Knowing that $\phi_o(X_{do}) = \phi_p$ we sub that into the equation above to get

$$\phi_{n^+} - \phi_p = \frac{qN_a X_{do}}{C_{ox}} + \frac{qN_a X_{do}^2}{2\epsilon_s}$$

Finally all unknowns are eliminated apart from X_{do} which we may rearrange to get

$$X_{do} = t_{ox} \left(\frac{\epsilon_s}{\epsilon_{ox}} \right) \left(\sqrt{1 + \frac{2C_{ox}^2(\phi_{n^+} - \phi_p)}{q\epsilon_s N_a}} - 1 \right)$$

To find $\phi(0) = \phi_{so}$ we simply sub $x = 0$ into $\phi_o(x)$ above to obtain

$$\phi_{so} = \phi_{n^+} - \frac{qN_a X_{do}}{C_{ox}}$$

However by definition(8) we also know that

$$V_B = \phi_{so} - \phi_p \quad \text{and} \quad V_{ox} = \phi_{n^+} - \phi_s$$

so therefore immediately on rearrangement of the equation for ϕ_{so} we obtain

$$V_{ox} = \frac{qN_a X_{do}}{C_{ox}}$$

and therefore from the equation for $\phi_{n^+} - \phi_p$ we see that

$$\phi_{n^+} - \phi_p = \frac{qN_a X_{do}}{C_{ox}} + \frac{qN_a X_{do}^2}{2\epsilon_s} = V_{ox} + \frac{qN_a X_{do}^2}{2\epsilon_s}$$

so we get

$$\frac{qN_a X_{do}^2}{2\epsilon_s} = \phi_{n^+} - \phi_p - V_{ox} = \phi_s - \phi_p = V_B \quad (6)$$

3.2 MOS Electrostatics under applied bias

We now consider the case where $V_{GB} \neq 0$

Depletion Regime

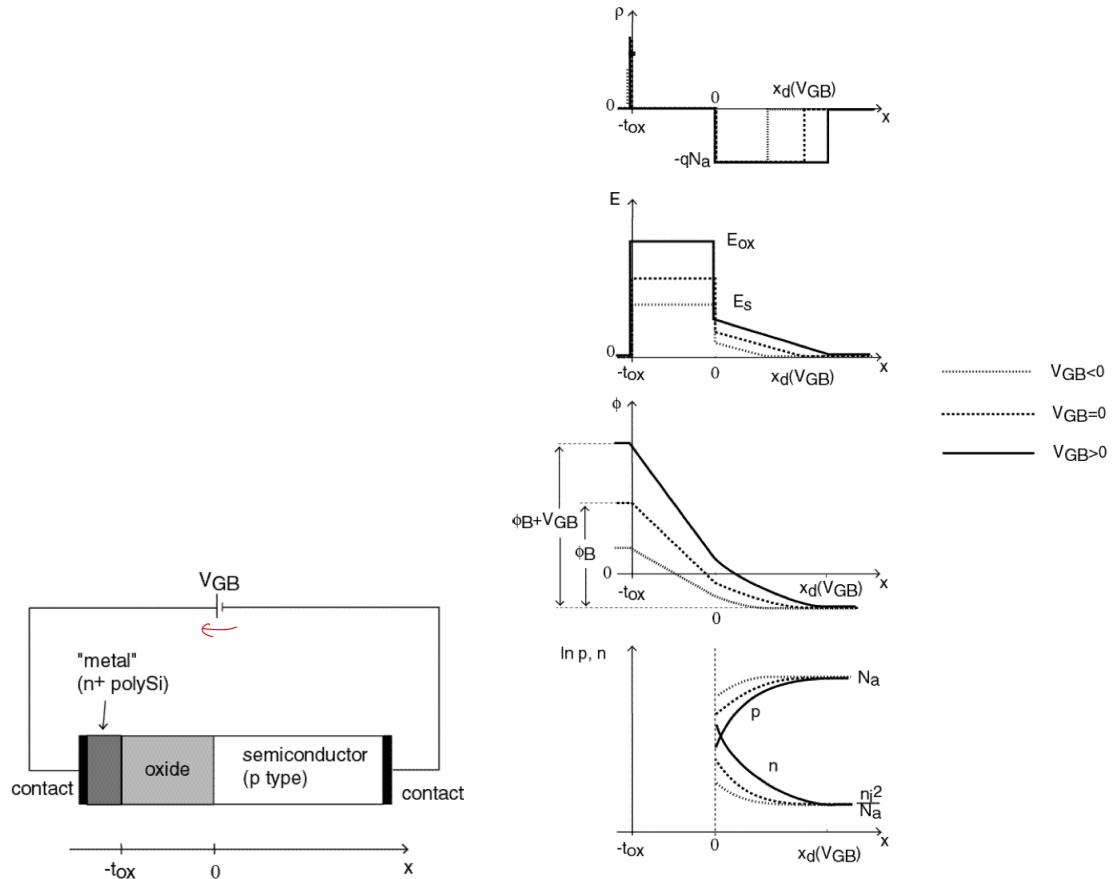


Figure 10: Depletion Regime

Essentially for a positive V_{GB} , the current direction as seen by the red arrows can be seen as a movement of positive charges to the metal. Reminder that the oxide is electrically insulating

- For $V_{GB} > 0$ the metal attracts electrons and repels holes so the *depletion region widens*
- For $V_{GB} < 0$ the metal repels electrons and repels holes so the *depletion region shrinks*

Also notice now that $V_{ox} + V_B = \phi_B + V_{GB}$. Look carefully at 8. So in depletion regime all results obtained for thermal equilibrium apply. Just substitute ϕ_{n^+} with $\phi_{n^+} + V_{GB}$ in the equations.

$$X_d(V_{GB}) = t_{ox} \left(\frac{\epsilon_s}{\epsilon_{ox}} \right) \left(\sqrt{1 + \frac{2C_{ox}^2(V_{GB} + \phi_{n^+} - \phi_p)}{q\epsilon_s N_a}} - 1 \right) = t_{ox} \left(\frac{\epsilon_s}{\epsilon_{ox}} \right) \left(\sqrt{1 + \frac{2C_{ox}^2(V_{GB} - V_{FB})}{q\epsilon_s N_a}} - 1 \right)$$

and

$$\phi_s(V_{GB}) = (V_{GB} + \phi_{n^+}) - \frac{qN_a X_d(V_{GB})}{C_{ox}}$$

Flatband

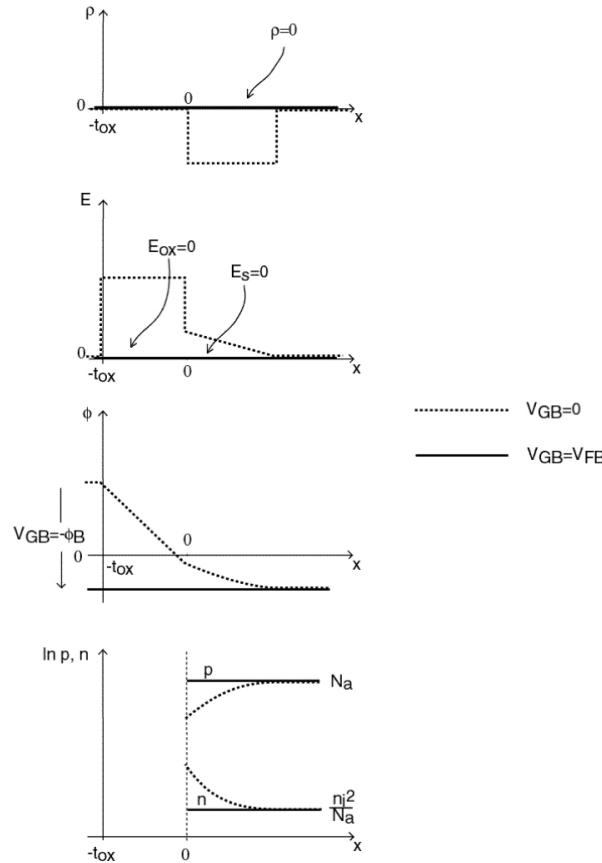
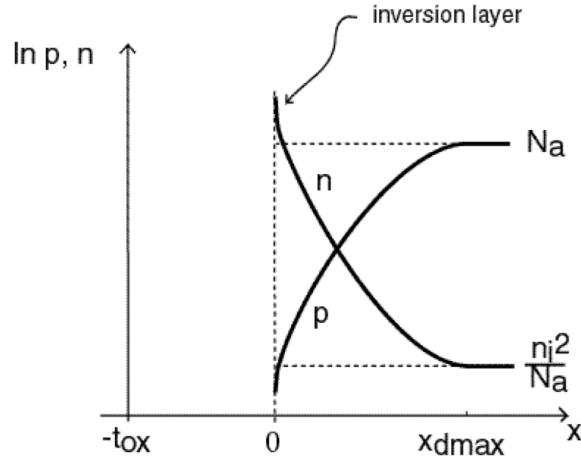


Figure 11: At a certain negative V_{GB} the depletion region is wiped out

Well intuitively by KVL we know that it must be

$$V_{GB} = V_{FB} = -\phi_B = -(\phi_{n^+} - \phi_p)$$

Threshold



At threshold V_T essentially $V_{GB} = V_T > 0$ such that $n_0(0) = N_a$. Essentially we can no longer assume n_0 is negligible on the p side relative to the dopant concentration. To apply this boundary condition and solve for threshold voltage. We first compute the potential drop in *semiconductor* at threshold starting from

$$n(0) = n_i e^{\frac{q\phi(0)}{kT}}$$

now solving for $\phi(0)$ at $V_{GB} = V_T$ we have

$$\phi(0) \Big|_{V_{GB}=V_T} = \frac{kT}{q} \cdot \ln \left(\frac{n(0)}{n_i} \right) \Big|_{V_{GB}=V_T} = \frac{kT}{q} \ln \left(\frac{N_a}{n_i} \right) = -\phi_p$$

that is at inversion we have

$$\phi_{so} = -\phi_p$$

we define **onset of inversion** when surface potential is equal and opposite to the potential of the bulk p-type substrate. Next we compute the potential drop in *oxide* at threshold instead

$$V_B(V_{GB} = V_T) = \frac{qN_a x_d^2(V_T)}{2\epsilon_s} = -2\phi_p \quad (7)$$

which is obtained if you recall 6 that $V_B = \phi_s - \phi_p$. Now solving for x_d at $V_{GB} = V_T$ we have

$$x_d(V_T) = x_{d_{max}} = \sqrt{\frac{2\epsilon_s(-2\phi_p)}{qN_a}}$$

then

$$V_{ox}(V_T) = E_{ox}(V_T)t_{ox} = \frac{qN_a x_d(V_T)t_{ox}}{\epsilon_{ox}} = \frac{1}{C_{ox}} \sqrt{2\epsilon_s qN_a (-2\phi_p)}$$

finally

$$V_T + \phi_B = V_B(V_T) + V_{ox}(V_T) = -2\phi_p + \frac{1}{C_{ox}} \sqrt{2\epsilon_s qN_a (-2\phi_p)}$$

Remark 43. Observe that $V_{ox}(V_T) = -\frac{Q_B}{C_{ox}}$. This is because

$$Q_{B,max} = -qN_a X_{d,max} = -\sqrt{2q\epsilon_s N_a(-2\phi_p)}$$

where in the extreme RHS we have subbed in $X_{d,max} = X_d(V_T)$. The significance of "max" will be explained in the next section on inversion (specifically we will assume surface potential is pinned at the onset of inversion value even as V_{GS} increases past V_T)

Remark 44. Essentially this means V_{ox} equals the magnitude of depletion charge (due to the immobile holes) at inversion divided by the oxide capacitance

then solving for V_T recalling that $V_{FB} = -\phi_B$ we have

$$V_{GB} = V_T = V_{FB} - 2\phi_p + \frac{1}{C_{ox}} \sqrt{2\epsilon_s q N_a(-2\phi_p)} \quad (8)$$

Fact 45

Now notice that

- if $N_a \uparrow \Rightarrow V_T \uparrow$. The higher the doping the more voltage required to produce $n(0) = N_a$
- if $C_{ox} \uparrow (t_{ox} \downarrow) \Rightarrow V_T \downarrow$. The thinner the oxide the less voltage dropped across the oxide

inversion

Now inversion layer is essentially when $V_{GB} > V_T$. As mentioned above since beyond threshold we can no longer assume n_0 is negligible with respect to N_a . Say we now want to find the total inversion charge Q_N (i.e the excess charge due to electrons that cannot be neglected). First we assume something called **delta-depletion approximation**, that is **surface potential** is pinned at $-\phi_p$ at inversion. Because of this we clearly can't use 33 to find the electron concentration at the surface in the inversion region because that takes the live actual ϕ_0 in consideration not a pinned value/constant. Therefore we try to find it through other means, in particular we invoke gauss law. Knowing that

$$V_{GB} - V_{FB} = V_{ox} - 2\phi_p$$

where we still have $V_B = -2\phi_p$ because surface potential is pinned at $-\phi_p$ (i.e the surface potential during onset of inversion). Consequently recalling that $X_{d,max}$ depends on V_B , the depletion region width still remains at $X_{d,max}$ found during the onset of inversion too. Obviously same goes for the depletion charge $Q_{B,max} = -qN_a X_d$. The depletion region contains now both the inversion and depletion charge. Taking gauss law around the depletion region and recalling from above that E_{ox} is uniform along t_{ox} we have

$$V_{ox} = E_{ox} t_{ox} = \left(\frac{t_{ox}}{\epsilon_{ox}} \right) (-Q_{B,max} - Q_N)$$

putting this all together we have

$$V_{GB} - V_{FB} = V_{ox} - 2\phi_p = \frac{-Q_{B,max} - Q_N}{C_{ox}} - 2\phi_p$$

upon rearrangement we obtain

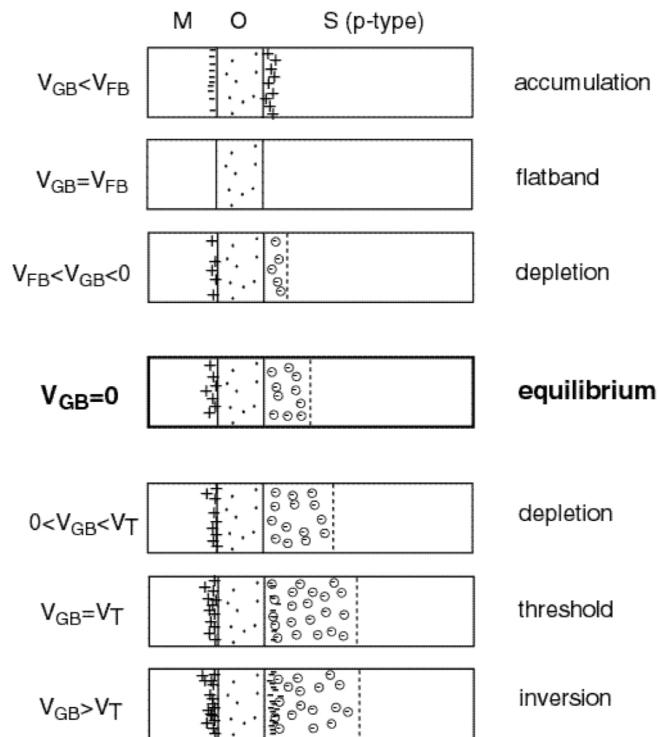
$$Q_N = -C_{ox} \left(V_{GB} - V_{FB} + 2\phi_p - \frac{-Q_{B,max}}{C_{ox}} \right)$$

but notice $-V_{FB} + 2\phi_p - \frac{-Q_{B,max}}{C_{ox}} = -V_T$ (which is the threshold voltage we found at onset of inversion. So we finally

$$Q_N = -C_{ox}(V_{GB} - V_T) \quad \text{for } V_{GB} > V_T$$

by definition of capacitance. Refer below for an intuitive visualization of this "capacitor" gate oxide. Intuitively we may think of the gate as a capacitor with capacitance C_{ox} . Then we know that extra potential beyond V_T is used to increase inversion charge only because we have pinned the surface potential at a fixed value which also pins the depletion charge constant as well.

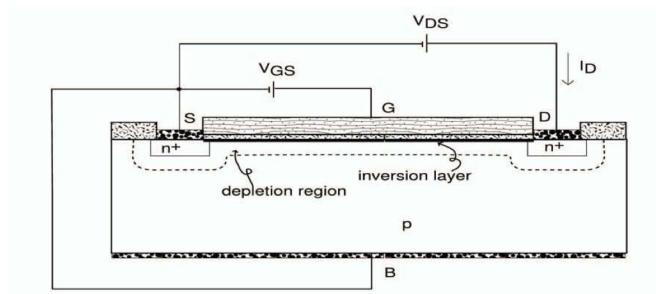
Summary



where in inversion

$$|Q_N| = C_{ox}(V_{GB} - V_T) \quad \text{for } V_{GB} > V_T$$

4 The MOS Field Effect Transistor(4)



The MOSFET as 3 regions of operation

1. the cut off region
2. linear or triode region
3. saturation region

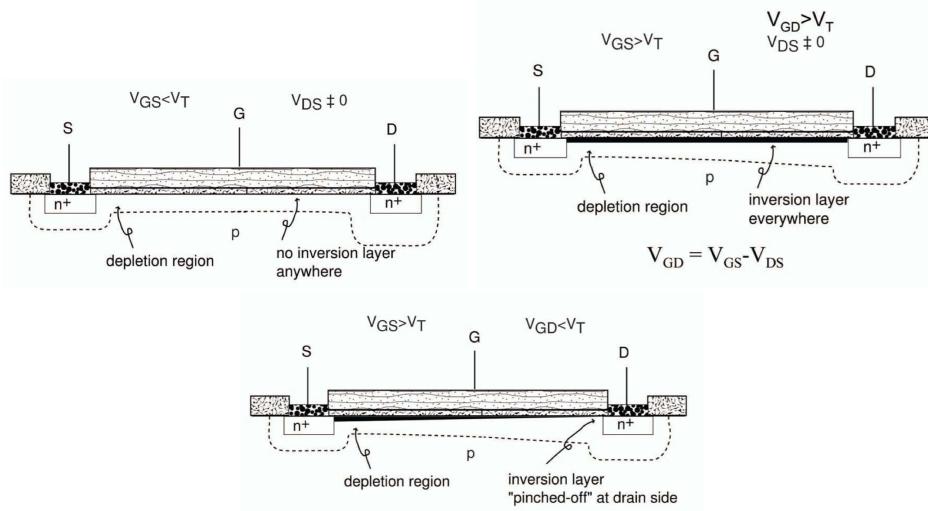


Figure 12: the 3 regions of operation in the order above(left to right top to bottom order respectively). Key observations:(1) for cutoff region there has no inversion layer.(2),(3): notice the depletion region is increasing for the triod and saturation region as indicated by the dotted line. However for the saturation region the inversion region as indicated by the solid gradient is decreasing along y before getting "pinched off"

4.1 Gradual Channel Approximation(4.4)

We now attempt to find an expression for the drain current for above. First we know that

$$I_D = -Wv_y(y)Q_N(y)$$

where W is width of the channel. For a small V_{DS} we may hence assume that the electric field along the channel between the source and drain is uniform in our calculations.

Proposition 46

Note that we assumed **steady state** conditions. That implies that the drain current I_D is constant along the channel length.

Proof. The drain current I_D remains constant along the MOSFET channel due to charge conservation, which is governed by the continuity equation.

The general form of the continuity equation is:

$$\frac{dQ}{dt} + \nabla \cdot J = 0, \quad (9)$$

where:

- Q is the charge density,

- J is the current density,
- $\nabla \cdot J$ represents the rate of current divergence.

The RHS is zero due to charge conservation (recall **Reynold Transport theorem**). Notice this equation is essentially the same form just that it is applied to charges Since we are considering a steady-state scenario, the charge density does not change with time, meaning:

$$\frac{dQ}{dt} = 0. \quad (10)$$

Thus, the continuity equation simplifies to:

$$\nabla \cdot J = 0. \quad (11)$$

This implies that the current density does not vary along the length of the channel. The total current is given by:

$$I_D = W \int_0^t J(y) dy, \quad (12)$$

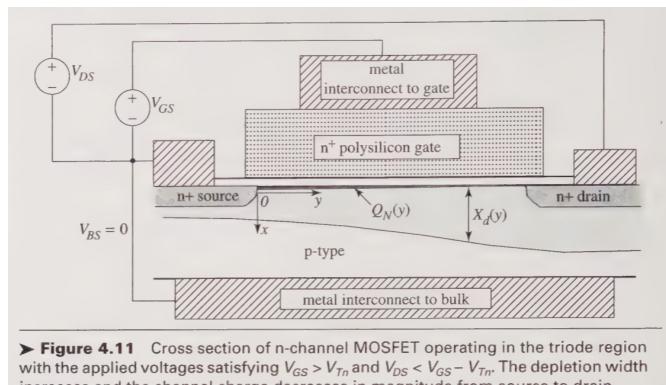
where:

- W is the channel width,
- $J(y)$ is the current density at position y along the channel.

Since $\nabla \cdot J = 0$, the current density $J(y)$ is uniform along the channel. Therefore, I_D remains constant for all y □

Remark 47. That is to say that although $v_y(y)$ and $Q_N(y)$, their product like in $-Wv_y(y)Q_N(y)$ is constant and equal to I_d along the channel length. Also recall that thermal equilibrium is a special case of steady state

Now let us first consider the triode region as an example. The cutoff region isn't too interesting. It is essentially when $V_{GS} < V_T$ with $V_{DS} \geq 0$ so there is no inversion charge so no drain current i.e $I_D = 0$



► **Figure 4.11** Cross section of n-channel MOSFET operating in the triode region with the applied voltages satisfying $V_{GS} > V_T$ and $V_{DS} < V_{GS} - V_T$. The depletion width increases and the channel charge decreases in magnitude from source to drain.

Figure 13: consider the triode region

Now because $V_{BS} = 0$ we know that $V_{GB} = V_{GS}$. Now taking KVL loops respectively we have

$$V_{GB} - V_{FB} = V_{GS} - V_{FB} = V_{ox}(y) + V_B(y) \quad (13)$$

Remark 48. KVL loops for capacitor circuits essentially if you recall basic circuit theory. The channel is seen as a capacitor. The current doesn't go through the channel in the x direction to complete a loop due to the insulating gate oxide.

Again by definition V_B is given by

$$V_B(y) = \phi(0, y) - \phi_p$$

and that $\phi(0, 0)$ is essentially the surface potential for the pn junction case where we derived the following relation if you recall

$$V_B(0) = \phi(0, 0) - \phi_p = -2\phi_p$$

$\phi(0, y)$ is different from the pn junction case as along y horizontally right there is an additional potential drop of V_{DS}

Remark 49. Yes we are still using **delta depletion approximation** as explained earlier which involves pinning the surface potential down to the value during onset of inversion even as V_{GS} increases past the threshold voltage which then results in accumulation of inversion charge solely while depletion charge remains constant. But you see that was for only one $y = 0$. We need to consider that surface potential at onset of inversion will be different along different values of y . As a matter of fact due to V_{DS} along the channel the surface potential increases. That is

- $\phi(0, y) \geq \phi(0, 0)$
- which then implies $V_B(y) \geq V_B(0)$
- consequently $X_{d,max}(y)$ increases as y increases (i.e. depletion width widens along the channel as shown in the above figure)

this is essentially the crux of gradual channel approximation.

Now we attempt to find an expression for $Q_N(y)$ the inversion charge along y . The depletion region contains now both the inversion and depletion charge. Taking gauss law around the depelition region like we have before we get

$$V_{ox}(y) = E_{ox}(y)t_{ox} = \left(\frac{t_{ox}}{\epsilon_{ox}} \right) (-Q_N(y) - Q_B(y)) = \frac{-Q_N(y) - Q_B(y)}{C_{ox}}$$

where the blue terms is an application of gauss law and E_{ox} is assumed uniform along t_{ox} since it is modelled as uniform field capacitor if you recall 42. Upon rearrangment of the above equation and then subbing in the relation for $V_{ox}(y)$ above 13 we get

$$Q_N(y) = -C_{ox} \left(V_{ox}(y) + \frac{Q_B(y)}{C_{ox}} \right) = -C_{ox} \left(V_{GS} - V_{FB} - V_B(y) + \frac{Q_B(y)}{C_{ox}} \right)$$

and then subbing in definition of $V_B = \phi(0, y) - \phi_p$ we have

$$Q_N(y) = -C_{ox} \left(V_{GS} - V_{FB} - \phi(0, y) + \phi_p + \frac{Q_B(y)}{C_{ox}} \right)$$

Now let us find the threshold voltage V_{th} . Recall for the pn junction we found it to be 8

$$V_{Tn}(y) = V_{FB} - V_B(y) - \frac{Q_B(y)}{C_{ox}} = V_{FB} + (\phi(0, y) - \phi_p) + \frac{1}{C_{ox}} \sqrt{2q\epsilon_s N_a(\phi(0, y) - \phi_p)}$$

where the blue terms corresponds to $V_B(y)$. In the pn case that will be $-2\phi_p$ but here it would be the terms in blue. The 2 cases will be the same for $y = 0$ which is the familiar

$$V_{Tn}(0) = V_{FB} - V_B(0) - \frac{Q_B(0)}{C_{ox}} = V_{FB} + (-2\phi_p) + \frac{1}{C_{ox}} \sqrt{2q\epsilon_s N_a(-2\phi_p)}$$

Now expressing Q_N in terms of V_{Tn} we have Note that the variation of $V_{Tn}(y)$ along y is called the **body effect**.

$$Q_N(y) = -C_{ox} \left(V_{GS} - V_{FB} - \phi(0, y) + \phi_p + \frac{Q_B(y)}{C_{ox}} \right) = -C_{ox} (V_{GS} - V_{Tn}(y))$$

We now define the **channel charge**

$$V_C(y) = \phi(0, y) + \phi_p$$

This is a convenient variable to have since

- $V_C(0) = \phi(0, 0) + \phi_p = 0$
- $V_C(L) = \phi(0, L) - (-\phi_p) = \phi(0, L) - \phi(0, 0) = V_{DS}$

so we can have that in our integration interval as later seen below. To further simplify our analysis we neglect body effect in the square root term as shown in the green terms(since square root is small) and denote $V_{tn}(0) = V_{tn}$ In that case $Q_n(y)$ becomes

$$Q_n(y) = -C_{ox} \left(V_{GS} - \left(V_{FB} + (\phi(0, y) - \phi_p) + \frac{1}{C_{ox}} \sqrt{2q\epsilon_s N_a(-2\phi_p)} \right) \right)$$

which in terms of $V_C(y)$ is

$$Q_n(y) \approx -C_{ox} (V_{GS} - V_{Tn} - V_C(y))$$

Now recalling $I_D = -Wv_y(y)Q_N(y)$ as mentioned at the start we may substitute our relation for Q_N and that $-\mu_n \frac{dV_C(y)}{dy} = \mu_n E_y = v_y$ we find that

$$I_D = WC_{ox} (V_{GS} - V_{Tn} - V_C(y)) \mu_n \frac{dV_C(y)}{dy}$$

and as mentioned in the start since E_y is uniform along the channel we have $\frac{dV_C(y)}{dy} = \frac{V_{DS}}{L}$ taking the integral

$$\int_0^L I_D dy = \int_0^{V_{DS}} W \mu_n C_{ox} (V_{GS} - V_{Tn} - V_C(y)) dV_C$$

which by separation of variables I_D and V_C yields a *general equation for I_d* (yes applies to other saturation and cutoff as well since no specific conditions were applied in the derivation)

$$I_D = \left(\frac{W}{L} \right) \mu_n C_{ox} (V_{GS} - V_{TN} - V_{DS}/2) V_{DS}$$

Note that we have treated I_D to be constant with respect to y during integration because we have assumed **steady state** for this section as mentioned at the start. We will talk more about other models later

Backgate effect(4.4.1)

Now consider the case where $V_{BS} \neq 0$ in 13 and so we have

$$V_B(y) = \phi(0, y) - (\phi_p + V_{BS}) = V_C(y) - 2\phi_p - V_{BS}$$

and

$$V_{Tn} = V_{FB} - 2\phi_p + \frac{1}{C_{ox}} \sqrt{(2qN_a\epsilon_s)(-2\phi_p - V_{BS})}$$

and

$$V_{Tn} = V_{TO_n} + \gamma_n \left(\sqrt{-2\phi_p - V_{BS}} - \sqrt{-2\phi_p} \right)$$

where $V_{TO_n} = V_{To}(V_{BS} = 0)$ and

$$\gamma_n = \left(\sqrt{2qN_a\epsilon_s} \right) / C_{ox}$$

which we define to be the **backgate effect parameter**

MOFSET in Saturation(4.4.2)

Recalling our separation of variables integral from above. This time say we want to integrate from 0 to some specified L and V_C

$$\int_0^L \frac{I_D}{\mu_n W C_{ox}} dy = \int_0^{V_C} (V_{GS} - V_{Tn} - V_C) dV_C = (V_{GS} - V_{Tn}) V_C - \frac{V_C^2}{2}$$

which is evaluated to be

$$V_C(y) = V_{GS} - V_{Tn} - \sqrt{(V_{GS} - V_{Tn})^2 - 2(V_{GS} - V_{Tn} - V_{DS}/2)V_{DS}(y/L)}$$

But here have a problem. With our expression for V_C , we can find the expression for the electric field along the channel to be

$$E_y(y) = -\frac{dV_C}{dy} = \frac{-(V_{GS} - V_{Tn} - V_{DS}/2)V_{DS}}{L \sqrt{(V_{GS} - V_{Tn})^2 - 2(V_{GS} - V_{Tn} - V_{DS}/2)V_{DS}(y/L)}}$$

Definition 50

The saturation voltage is defined by when $V_{DS} = V_{DS_{SAT}}$ where

$$V_{DS_{SAT}} = V_{GS} - V_{Tn}$$

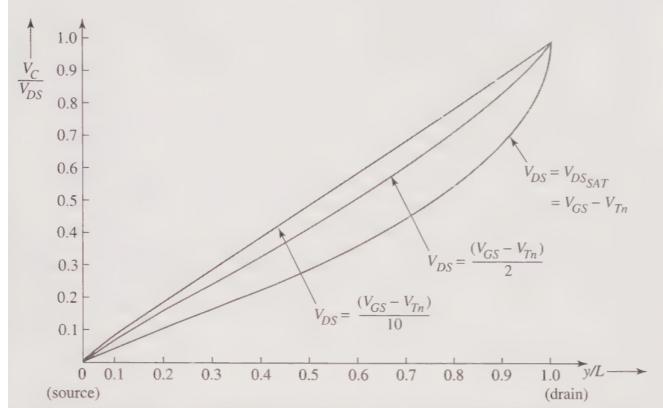
Fact 51 (Pinch Off explained)

recall that we have found a general expression for I_D

$$I_D = W C_{ox} (V_{GS} - V_{Tn} - V_C(y)) \mu_n \frac{dV_C(y)}{dy} = -W Q_N(y) v_y(y)$$

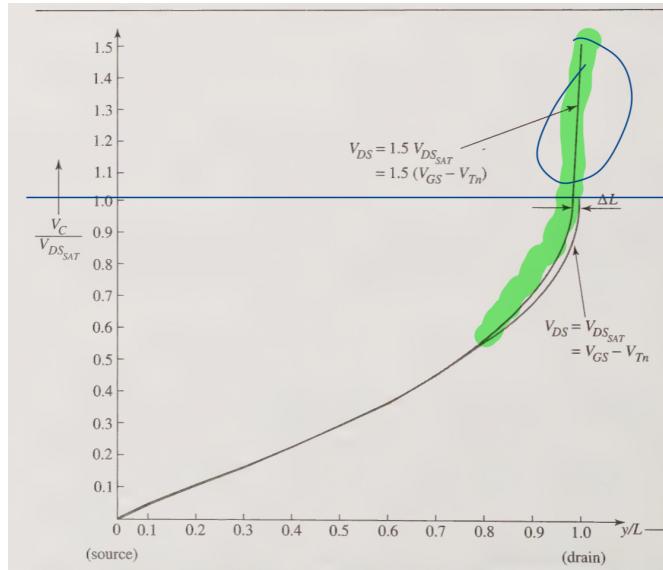
Immediately you see that when V_{DS} is at the saturation voltage the inversion charge $Q_N(y)$ becomes zero while $E_y = \frac{dV_C(y)}{dy}$ becomes high (especially for the very prominently non uniform E_y cases). So along y less inversion charges move faster so thus the inversion layer thins before being pinched off as $y \rightarrow L$

Now plotting V_C/V_{DS} against y we see that for V_{DS} values much smaller than $V_{DS_{SAT}} = V_{GS} - V_{Tn}$ we get a near uniform electric field along y as desired (recall our assumption at the start).



► **Figure 4.16** Normalized channel voltage $V_C(y)/V_{DS}$ along the channel for three values of drain-source voltage.

But as V_{DS} approaches $V_{DS_{SAT}}$ this relationship begins to breakdown especially at larger values along greater distances y down the channel. For example for $V_{DS_{SAT}} = V_{DS}$ not only is E_y not uniform, near the ends at $y = L$ we even have a near vertical slope suggesting an infinite electric field which is clearly not physical. To resolve this we do something called **channel length modulation**. Essentially



► **Figure 4.17** Channel voltage, normalized to the saturation drain-source voltage $V_{DS_{SAT}}$, for $V_{DS} = V_{DS_{SAT}}$ and for $V_{DS} = 1.5 V_{DS_{SAT}}$.

Figure 14: we are integrating from $V_C : 0 \rightarrow V_{DS} = V_{DS_{SAT}}$ that is all the way to the blue horizontal line. So when we shorten $L - \Delta L$ green highlighted section is cut out from the integration region

If we "shorten" the channel length, replacing L with $L - \Delta L$, we could essentially cut out the high/infinite electric field from our integration range i.e this unphysical infinite field region is not included in our calculations so may continue to use the assumptions on uniform electric field above. Now plugging this shortened channel length into our integral we have

$$\int_0^{(L-\Delta L)} I_D dy = I_D(L - \Delta L) = \mu_n C_{ox} W \int_0^{V_{DS_{SAT}}} (V_{GS} - V_{Tn} - V_C) dV_C$$

and recalling by definition that $V_{DS_{SAT}} = V_{DS}$ we obtain

$$I_{DSAT} = \frac{1}{2} \left(\frac{W}{L - \Delta L} \right) \mu_n C_{ox} (V_{GS} - V_{Tn})^2, \quad \text{for } V_{DS} \geq V_{DSAT} = V_{GS} - V_{Tn}$$

now making the 1st order approximation

$$\frac{1}{L - \Delta L} \approx \frac{1}{L} \left(1 + \frac{\Delta L}{L} \right)$$

and

$$\frac{\Delta L}{L} = \lambda_n V_{DS}$$

where we have defined the **channel-length modulation parameter** to be

$$\lambda_n = \frac{0.1 \mu mV}{L}$$

we arrive at a useful approximate expression for drain current in saturation given by

$$I_D = \frac{W}{2L} \mu_n C_{ox} (V_{GS} - V_{Tn})^2 (1 + \lambda_n V_{DS})$$

4.2 MOSFET Circuit Models

so far we have considered steady state models where we have derived the equations for drain current for cutoff, triode and saturation. As a recap they are

Theorem 5

Large signal model regimes of operation:

- Cut off region

$$I_D = 0$$

- linear/triode

$$I_D = \frac{W}{L} \mu_n C_{ox} \left[V_{GS} - \frac{V_{DS}}{2} - V_T \right] \cdot V_{DS}$$

- Saturation

$$I_D = I_{Dsat} = \frac{W}{2L} \mu_n C_{ox} [V_{GS} - V_T]^2 \cdot [1 + \lambda V_{DS}]$$

effect of back bias

$$V_T(V_{BS}) = V_{T0} + \gamma \left[\sqrt{-2\phi_p - V_{BS}} - \sqrt{-2\phi_p} \right]$$

These are actually classified under what we call the **large signal model** of NMOS transistors.

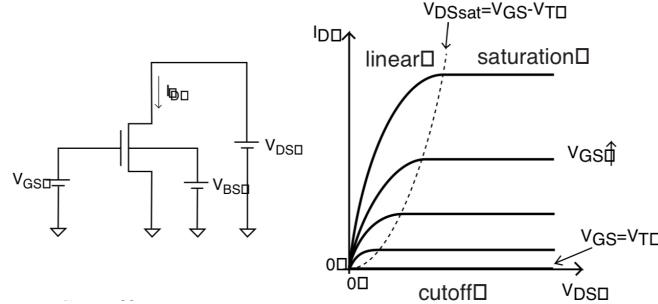


Figure 15: Large signal model

Now we extend our knowledge of models to what is called the **small signal model** where we are only interested in the response of the device to small signal applied on top of a bias

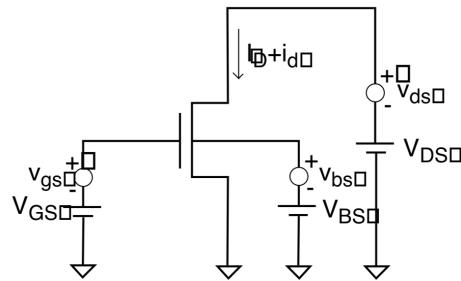


Figure 16: Small signal model

Now because signals are small, we can do a first order linear approximation of the response like so

$$i_D = I_D + i_d = I_d + \frac{\partial i_D}{\partial v_{GS}} |_Q v_{gs} + \frac{\partial i_D}{\partial v_{BS}} |_Q v_{bs} + \frac{\partial i_D}{\partial v_{DS}} |_Q v_{ds}$$

where v_{gs} , v_{ds} , v_{bs} are the small incremental voltages. i_d is the small incremental current. I_D is a variable while i_D is a particular value of the variable. Also i_D is the variable and i_d is the increment of that variable. Similarly v_{GS} is the variable and v_{gs} is the small increment of that variable. Same applies for v_{BS} and v_{DS} .

Example 53

And analogy for all these will be

- Δx - small increment of the variable
- x - the variable
- when $x = X_0$ - a particular value of that variable

You will know what I mean when you look at the graphs below.

Definition 54

We define

- the **transconductance** $g_m = \frac{\partial i_D}{\partial V_{GS}}|_Q$
- the **output/drain conductance** $g_o = \frac{\partial i_D}{\partial V_{DS}}|_Q$
- the **backgate transconductance** $g_{mb} = \frac{\partial i_D}{\partial V_{BS}}|_Q$

Fact 55

First we recap that

- A current source is independent of voltage
- A resistor follows Ohm law(i.e voltage dependent)

We now calculate the respective conductances and transconductances. For **transconductance** from our equation for saturation current derived above we have

$$i_D = \frac{W}{2L} \mu_n C_{ox} [v_{GS} - V_T]^2 \cdot [1 + \lambda V_{DS}] \approx \frac{W}{2L} \mu_n C_{ox} [v_{GS} - V_T]^2$$

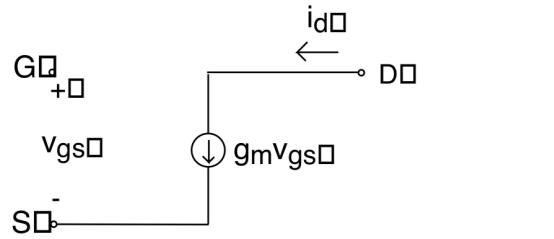
Remark 56. Notice the blue term used for i_D is the variable v_{GS} not the large signal V_{GS} as used in I_D which are treated as constants as they are particular values.

where we have neglected the contribution from the channel length modulation because it is small if $\lambda_n V_{DS} \ll 1$. so we get

$$g_m = \left. \frac{\partial i_D}{\partial v_{GS}} \right|_Q \approx \frac{W}{L} \mu_n C_{ox} (V_{GS} - V_T)$$

Expressing in terms of I_D we hence have

$$g_m = \sqrt{2 \frac{W}{L} \mu_n C_{ox} I_D}$$



B

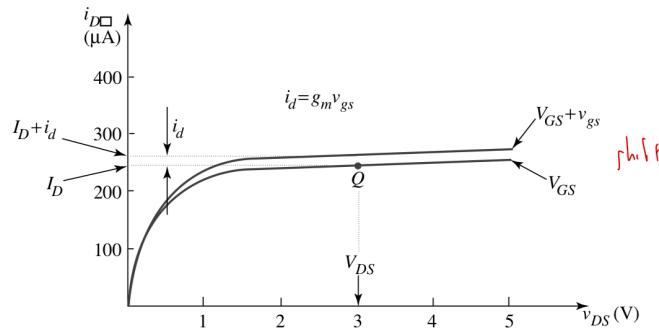


Figure 17: model as current source as clearly i_D is independent of v_{DS} . Again linear relation as signal small

Remark 57. For this and the other graphs notice the axes are variables i_D and v_{DS} . The large signal equivalents I_D and V_{DS} are treated as constants as they particular values of the variables as mentioned

For **output/drain conductance** we have again starting from

$$i_D = \frac{W}{2L} \mu_n C_{ox} [V_{GS} - V_T]^2 \cdot [1 + \lambda v_{DS}]$$

where we then obtain in terms of I_D

$$g_o = \left. \frac{\partial i_D}{\partial v_{DS}} \right|_Q = \frac{W}{2L} \mu_n C_{ox} (V_{GS} - V_T)^2 \cdot \lambda \approx \lambda I_D$$

and

$$r_o = \frac{1}{g_o} = \frac{1}{\lambda I_D}$$

and

$$\lambda \propto \frac{1}{L}$$

and

$$r_o \propto L$$

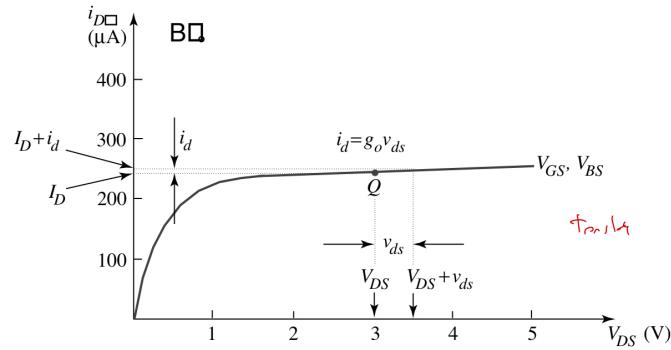
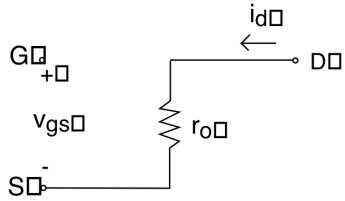


Figure 18: model as ohmic resistor as clearly variable i_D is dependent on variable v_{DS} . Notice that now g_o is the gradient. Again linear relation as signal small

Finally for **backgate transconductance**

$$i_D \approx \frac{W}{2L} \mu_n C_{ox} [v_{GS} - V_T]^2$$

where again we have neglected the contribution from the channel lenght modulation so then obtain

$$g_{mb} = \left. \frac{\partial i_D}{\partial v_{BS}} \right|_Q = -\frac{W}{L} \mu_n C_{ox} (V_{GS} - V_T) \cdot \left(\left. \frac{\partial V_T}{\partial v_{BS}} \right|_Q \right)$$

and recalling our equation for V_T that accounts for backgate effect derived earlier we find that

$$V_T(v_{BS}) = V_{T0} + \gamma \left[\sqrt{-2\phi_p - v_{BS}} - \sqrt{-2\phi_p} \right]$$

there althogether we have

$$\left. \frac{\partial V_T}{\partial v_{BS}} \right|_Q = \frac{-\gamma}{2\sqrt{-2\phi_p - V_{BS}}}$$

and

$$g_{mb} = \frac{\gamma g_m}{2\sqrt{-2\phi_p - V_{BS}}}$$

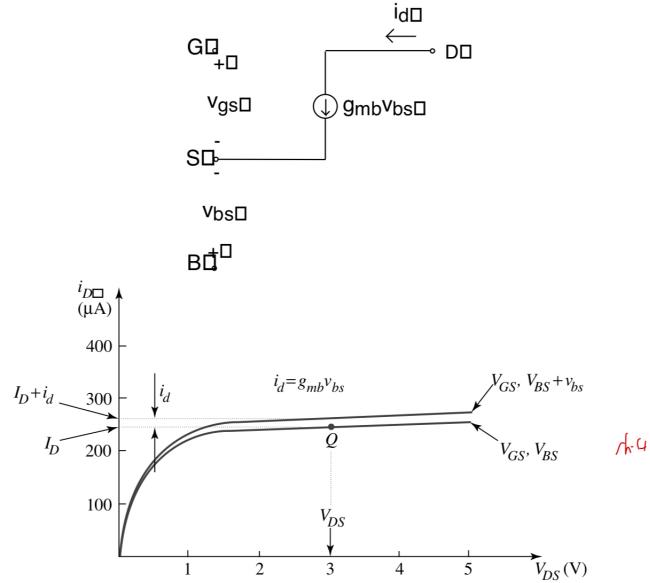
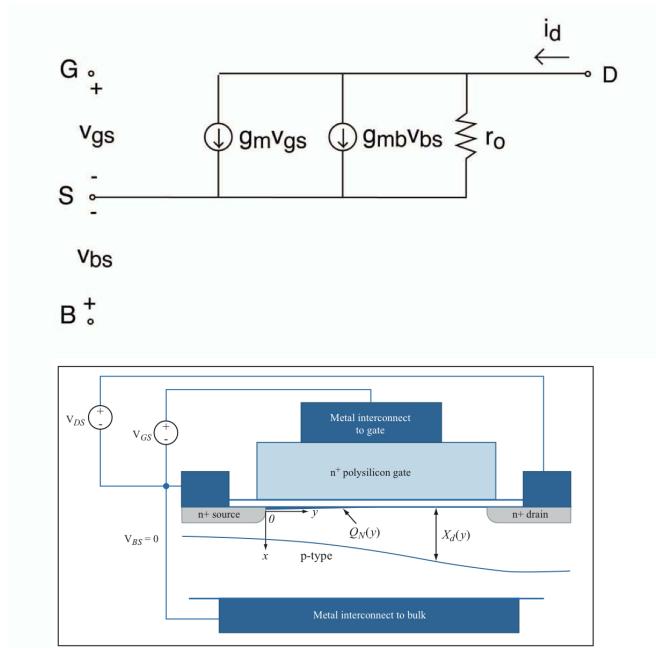


Figure 19: model as current source as clearly i_D is independent of V_{DS} . Again linear relation as signal small

All in all because the small increments of variables are related like so

$$i_d = g_m v_{gs} + g_{mb} v_{bs} + g_0 v_{ds}$$

kirchoff law suggests that these 3 components must have been connected in parallel that is

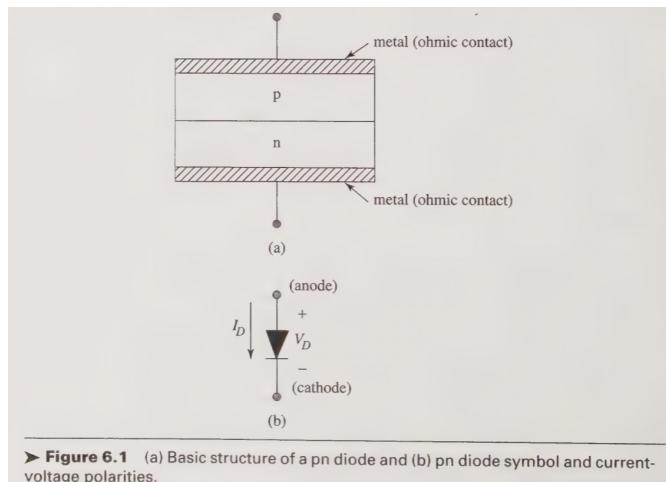


5 The pn Junction Diode(6)

5.1 The pn Diode Circuit Symbol and Terminal Characteristics(6.1)

There are 3 regions of operations of the pn diode

1. **forward bias** - diod current $I_D > 0$ with the diode voltage $V_D \approx 0.7V$
2. **Reverse bias** - diode current $I_D \approx 0$
3. **breakdown** - diode current $I_D < 0$ and the diode voltage $V_D < V_{BD}$ where V_{BD} is the diode's breakdown voltage



► **Figure 6.1** (a) Basic structure of a pn diode and (b) pn diode symbol and current-voltage polarities.

Definition 58

the transport of holes from the p side across the depletion region into the n side under a forward bias is called **hole injection**

Definition 59

the transport of electrons from the n side to p side of the junction under a forward bias is called **electron injection**

Let us see what is happening in the different regions

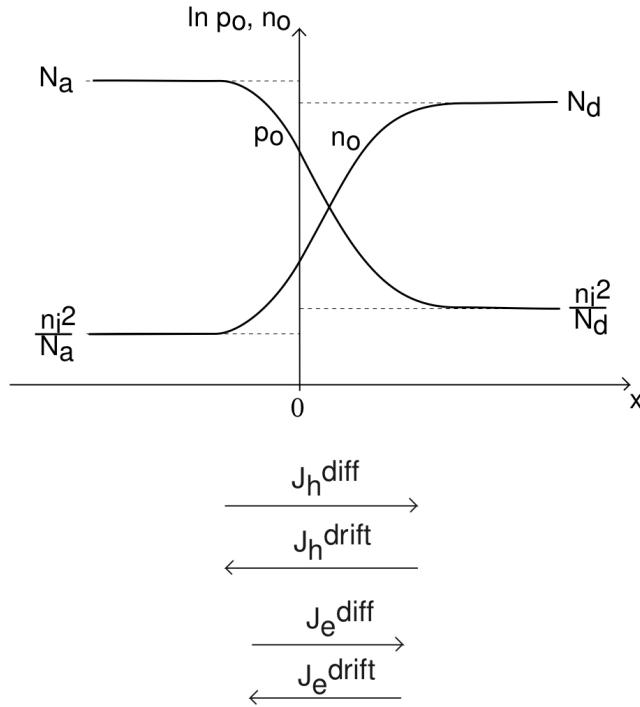


Figure 20: thermal equilibrium. recap that the right side is the n side while the left is the p side. S shape increasing graph will be the electrons while the S shape decreasing graph will be the holes. No net hole or electron current density. All these covered in the 1st section if you recall

We have

$$|J_{\text{drift}}| = |J_{\text{diff}}|$$

For $V > 0$, $\phi_B - V \downarrow \Rightarrow |E_{\text{SCR}}| \downarrow \Rightarrow |J_{\text{drift}}| \downarrow$

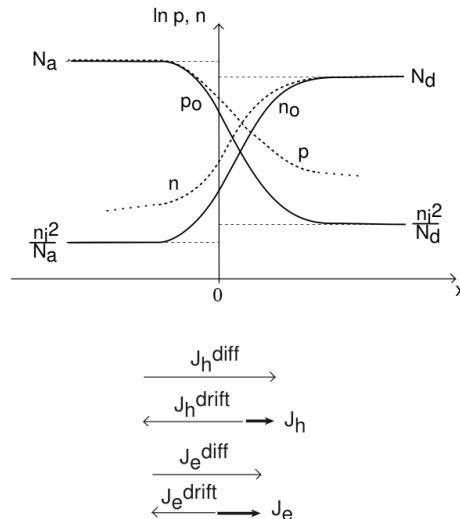


Figure 21: forward bias: i.e. net E field pointing right. expect net movement holes right and electrons left

Note that the forward bias raises the minority carrier concentrations at the edge of the depletion region relative to

that of thermal equilibrium values

- $n_p(-x_p) \gg n_{po}(-x_{po})$
- $p_n(x_n) \gg p_{no}(x_{no})$

We have

$$|J_{\text{drift}}| < |J_{\text{diff}}|$$

This means we have a net diffusion current in the space charge region or minority carrier **injection** into the QNRs

- net diffusion hole current J_h to the right
- net diffusion electron current J_e to the left

both of which in the direction out of the SCR towards the ohmic contacts

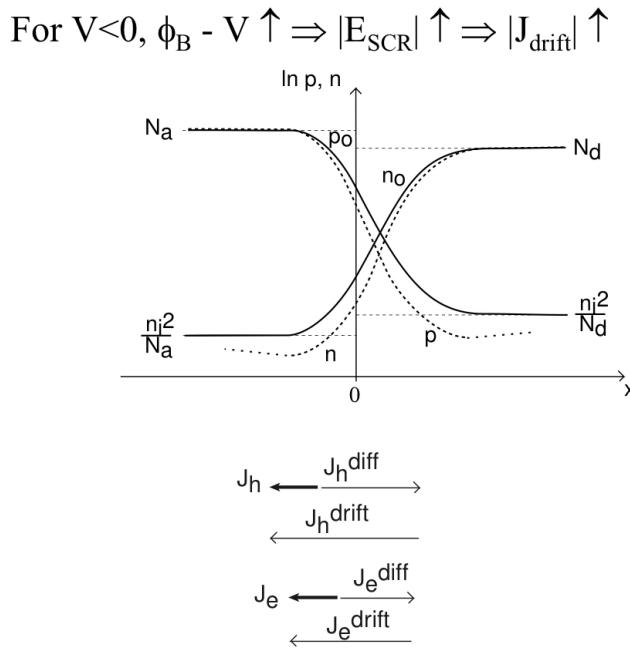


Figure 22: backward bias: i.e net E field pointing left. expect net movement holes left and electrons right

Note that the forward bias lowers the minority carrier concentrations at the edge of the depletion region relative to that of thermal equilibrium values

- $n_p(-x_p) \ll n_{po}(-x_{po})$
- $p_n(x_n) \ll p_{no}(x_{no})$

We have

$$|J_{\text{drift}}| > |J_{\text{diff}}|$$

This means we have a net drift current in the space charge region or minority carrier **extraction** from the QNRs

- net drift hole current J_h to the left
- net drift electron current J_e to the right

both of which in the direction of away from ohmic contacts into the SCR

Fact 60

To sum up

- in forward bias: injected minority carriers diffuse through QNR and recombine at semiconductor surface
- in reverse bias, minority carries generated at the semiconductor surface diffuse through the QNR and exracted by SCR

5.2 The pn junction a first pass(6.3)

Law of the junction

First recall the built in potential ϕ_B which by definition we will have

$$\phi_B = \phi_n - \phi_p = V_{th} \ln \left(\frac{N_d}{n_i} \right) - \left(V_{th} \ln \left(\frac{N_a}{n_i} \right) \right) = V_{th} \ln \left(\frac{N_d N_a}{n_i^2} \right)$$

note that we will adopt the following notation.

Example 61

The minority hole and majority electron concentrations on the n side of the junction are:

$$p_{no} = \frac{n_i^2}{N_d}, \quad n_{no} = N_d$$

that is, letters represent the charge carrier type while the first subscript indicates the region. The second subscript letter o indicates at thermal equilibrium as usual

similarly

Example 62

The minority electron and majority hole concentrations on the p side of the junction are:

$$n_{po} = \frac{n_i^2}{N_a}, \quad p_{po} = N_a$$

Now substituting these into our relation for built in potential above we find that

$$\phi_B = V_{th} \ln \left(\frac{p_{po}}{p_{no}} \right), \quad \phi_B = V_{th} \ln \left(\frac{n_{no}}{n_{po}} \right)$$

This gives us a new perspective upon rearrangment:

Proposition 63

The minority carrier concentrations are proportional to the majority carrier concentrations on the other side of the junction multiplied by an exponential factor.

Specifically we have

$$p_{no} = p_{po} e^{-\phi_B/V_{th}}, \quad n_{po} = n_{no} e^{-\phi_B/V_{th}}$$

Now converting this into base-10 logarithms using the 60mv rule we have

$$\phi_B = (60 \text{ mV}) \log \left(\frac{p_{po}}{p_{no}} \right) = (60 \text{ mV}) \log \left(\frac{n_{no}}{n_{po}} \right)$$

Theorem 64

Boltzmann Principle states that

$$\frac{N_2}{N_1} = e^{-(\phi_2 - \phi_1)/kT} = e^{-\Delta\phi/V_{th}}$$

Now consider when voltage V_D is applied to the diode. Now observe the following figure

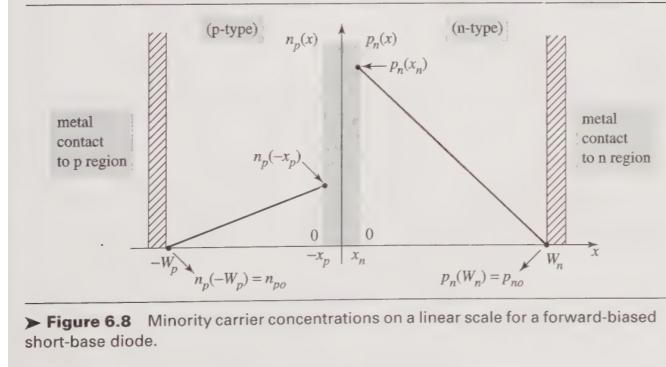


Figure 23: the linear profile will be explained in the next section

Recall from above

- $n_p(-x_p) \gg n_{po}(-x_{po}) = n_{po}$
- $p_n(x_n) \gg p_{no}(x_{no}) = p_{no}$

This is essentially the concentration difference that drives the diffusion. See below for more

Fact 65 (Boundary Conditions)

Note that at the ohmic contacts in this figure we ensure that

- $p_n(W_n) = p_{no}$
- $n_p(-W_p) = n_{po}$

that is the minority carrier concentrations are at their thermal equilibrium concentrations. Essentially we are saying that all excess carriers injected starting from the edge of SCR must all recombine at the ohmic contact after diffusing across the negative concentration gradient

assuming the balance between the very large diffusion and drift current densities in the depletion region in thermal equilibrium is not perturbed significantly by the forward bias (i.e. the carriers are still in near equilibrium with each other across the junction), we may then apply 63 noting that $n_p(-x_p)$ is the minority electron concentration on the p side edge of depletion region while $p_n(x_n)$ is the minority hole concentration on the n side edge of depletion region

$$p_n(x_n) = p_p(-x_p) e^{-\phi_j/V_{th}} = p_p(-x_p) e^{-(\phi_B - V_D)/V_{th}}$$

and

$$n_p(-x_p) = n_n(x_n)e^{-\phi_j/V_{th}} = n_n(x_n)e^{-(\phi_B - V_D)/V_{th}}$$

where $\phi_j = \phi_B - V_D$

Under conditions of **low level injection** which requires

1. $n_p(-x_p) \ll p_{po} \approx N_a, \quad p_n(x_n) \ll n_{no} \approx N_d$
2. $p_p(-x_p) \approx p_{po} = N_a, \quad n_n(x_n) \approx n_{no} = N_d$

Remark 66. Essentially we can neglect the slight increase in majority carrier concentration due to minority carrier injection

Our 2 equations relating the minority and majority carriers under bias the edges of the depletion region can now be rewritten as

$$p_n(x_n) = N_d e^{-\phi_B/V_{th}} e^{V_D/V_{th}}$$

and

$$n_p(-x_p) = N_a e^{-\phi_B/V_{th}} e^{V_D/V_{th}}$$

therefore we have just derived what is known as

Theorem 67 (Law of the Junction)

We have

$$p_n(x_n) = p_{no} e^{V_D/V_{th}}, \quad n_p(-x_p) = n_{po} e^{V_D/V_{th}}$$

pn Junction Currents under forward bias

During *steady state diffusion* where forward biased junction injects minority carriers continuously which diffuse across the bulk region to the ohmic contacts where they continuously recombine at the same rate.

The bulk regions are short enough that we can neglect the loss of minority carriers due to recombination outside the depletion region(this situation is known as the **short-base diode** case)

Proposition 68

Then concentration of minority charge carriers linearly decreases from the edge of the space charge region to the ohmic contacts on both the n and p sides for forward-biased short-base diode

Proof. By Fick's first law of diffusion, the diffusive flux J is given by:

$$J = -D \frac{dC}{dx}, \quad (14)$$

where D is the diffusion coefficient (assumed constant), and $C(x)$ is the concentration as a function of spatial coordinate x .

Under steady-state conditions, there is no accumulation or depletion of mass at any point in the medium, implying that the concentration does not change with time:

$$\frac{\partial C}{\partial t} = 0. \quad (15)$$

Using Fick's second law,

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} = 0. \quad (16)$$

Since D is constant and nonzero, this simplifies to the ordinary differential equation:

$$\frac{d^2 C}{dx^2} = 0. \quad (17)$$

Integrating once,

$$\frac{dC}{dx} = A, \quad (18)$$

where A is an integration constant. Integrating again,

$$C(x) = Ax + B, \quad (19)$$

where B is another integration constant.

The constants A and B are determined by boundary conditions. Since this equation is linear in x , the concentration profile is spatially linear under steady-state diffusion. \square

Now with this understanding of why we have a linear profile we attempt to find an equation given the graph as in 23 which we get using basic knowledge of linear graphs to be

$$p_n(x) = p_n(x_n) - \left(\frac{p_n(x_n) - p_{no}}{W_n - x_n} \right) (x - x_n) \quad \text{for } x_n \leq x \leq W_n$$

and

$$n_p(x) = n_p(-x_p) + \left(\frac{n_p(-x_p) - n_{po}}{W_p - x_p} \right) (x + x_p) \quad \text{for } -W_p \leq x \leq -x_p$$

and

$$p_p(x) = N_a + n_p(x) \quad \text{for } -W_p \leq x \leq -x_p$$

and

$$n_n(x) = N_d + p_n(x) \quad \text{for } x_n \leq x \leq W_n$$

Next we express for the total current density

$$J = J_p^{diff}(x_n) + J_n^{diff}(-x_p) = -qD_p \frac{dp_n}{dx} \Big|_{x=x_n} + qD_n \frac{dn_p}{dx} \Big|_{x=-x_p}$$

Note that we seemed to have ignored the contributions by drift currents or majority carriers. This is because in forward bias we recall there is a net diffusion current which is due to the injection of minority charge carriers. Therefore we only need to consider the diffusive current due to these minority charge carriers inserted which can be found by subtracting the edge minority carrier concentration with that of the thermal equilibrium as below. Finally because we know the relation of the diffusion profile will be linear we can express the gradients in our expression for J like so

$$J = -qD_p \left[\frac{p_n(W_n) - p_n(x_n)}{W_n - x_n} \right] + qD_n \left[\frac{n_p(-x_p) - n_p(-W_p)}{W_p - x_p} \right]$$

now using the fact that $p_n(W_n) = p_{no}$ and $n_p(-W_p) = n_{po}$ (recall 65) as well as our expressions for $p_n(x_n)$ and $n_p(-x_p)$ above we substitute them in to get

$$J = \left(\frac{-qD_p}{W_n - x_n} \right) (p_{no} - p_n e^{V_D/V_{th}}) + \left(\frac{qD_n}{W_p - x_p} \right) (n_{po} e^{V_D/V_{th}} - n_{po})$$

we may further simplify this equation if we have $x_n \ll W_n$ and $x_p \ll W_p$ so we may treat $x_n, x_p = 0$ (i.e. as negligible) in the denominator above .

$$J = qn_i^2 \left(\frac{D_p}{N_d W_n} + \frac{D_n}{N_a W_p} \right) (e^{V_D/V_{th}} - 1)$$

where we also used 3 and 4 for pn thermal equilibrium where we obtained

- $n_{po} = \frac{n_i^2}{N_a}$
- $p_{no} = \frac{n_i^2}{N_d}$

Now we multiply the total current density of the diode by the cross sectional area of the diode A to find the diode current I_D like so

$$I_D = qn_i^2 A \left(\frac{D_p}{N_d W_n} + \frac{D_n}{N_a W_p} \right) (e^{V_D/V_{th}} - 1) = I_o (e^{V_D/V_{th}} - 1)$$

where we have defined

Definition 69

We define the **short base diode saturation current** to be

$$I_o = qn_i^2 A \left(\frac{D_p}{N_d W_n} + \frac{D_n}{N_a W_p} \right)$$

pn Junction Currents under reverse bias

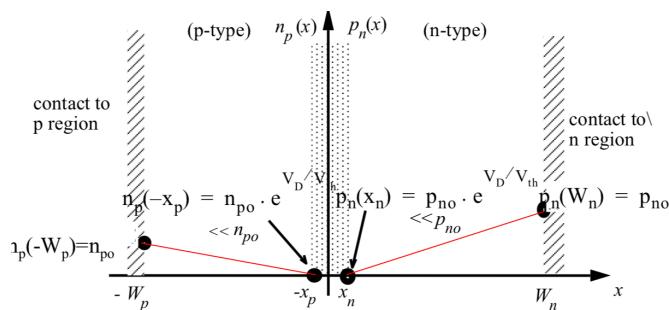


Figure 24: pn junction: reverse bias

In similar fashion to forward bias just that now

- $n_p(-x_p) \ll n_{po}(-x_{po})$
- $p_n(x_n) \ll p_{no}(x_{no})$

as mentioned earlier. Again linear profile is assumed under steady state short base diode conditions. In this case minority carriers are *continuously extracted* across the junction to maintain the value of the SCR edge set by the applied bias. The *same number continuously are generated* at the ohmic contact.

6 The Bipolar Junction Transistor(7)

Essentially BJT is just two neighboring pn junctions put back to back

6.1 BJT: a first pass(7.2)

Fact 70

In an IC npn transistor the emitter donor concentration N_{dE} is much greater than the base acceptor concentration N_{aB} which is much greater than the collector donor concentration N_{dC}

$$N_{dE} \gg N_{aB} \gg N_{dC}$$

therefore the thermal equilibrium minority carrier concentrations in the emitter,base and collector are

$$p_{nEo} = \frac{n_i^2}{N_{dE}} \ll n_{pBo} = \frac{n_i^2}{N_{aB}} \ll p_{nCo} = \frac{n_i^2}{N_{dC}}$$

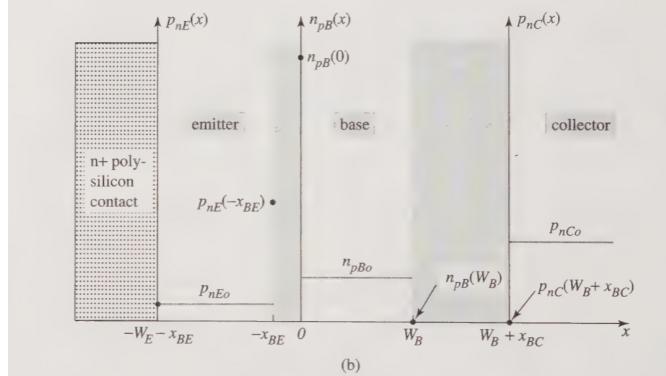


Figure 25: Forward-active regime: edge minority carrier concentration profile vs that of thermal equilibrium. Assume steady state short base diode conditions

You should realize the heights of the thermal equilibrium minority carrier concentrations drawn in diagram above is a direct consequence of 6.1. This will again will be represented in the diagram 6.1 where we account for flux paths of minority carriers.

Fact 71 (Law of junctions: BJT)

Like we have for pn junctions assuming low level injection and hence invoking **law of junctions** we have

$$p_{nE}(-x_{BE}) = p_{nEo} e^{V_{BE}/V_{th}} \quad n_{pB}(0) = n_{pBo} e^{V_{BE}/V_{th}}$$

and

$$n_{pB}(W_B) = n_{pBo} e^{V_{BC}/V_{th}} \quad p_{nC}(W_B + x_{BC}) = p_{nCo} e^{V_{BC}/V_{th}}$$

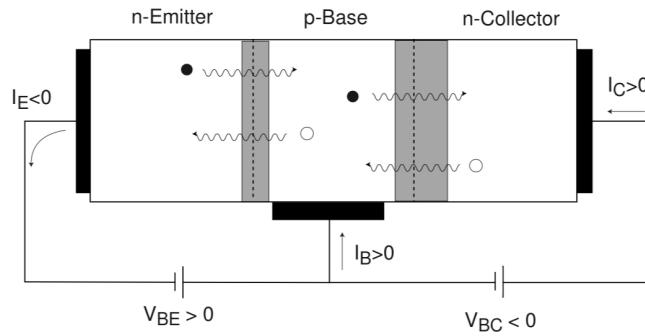


Figure 26: Circuit Representation: Forward active regime: flux of *minority carriers* injected/extracted due to bias

Remark 72. It is important to note as mentioned that these minority carrier fluxes are due to injections/extractions as a result of the bias. These fluxes then result in the currents I_C, I_E, I_B . So if $V_{BE}, V_{BC} = 0$ then we will have no such fluxes. Therefore $I_C, I_E, I_B = 0$ as well

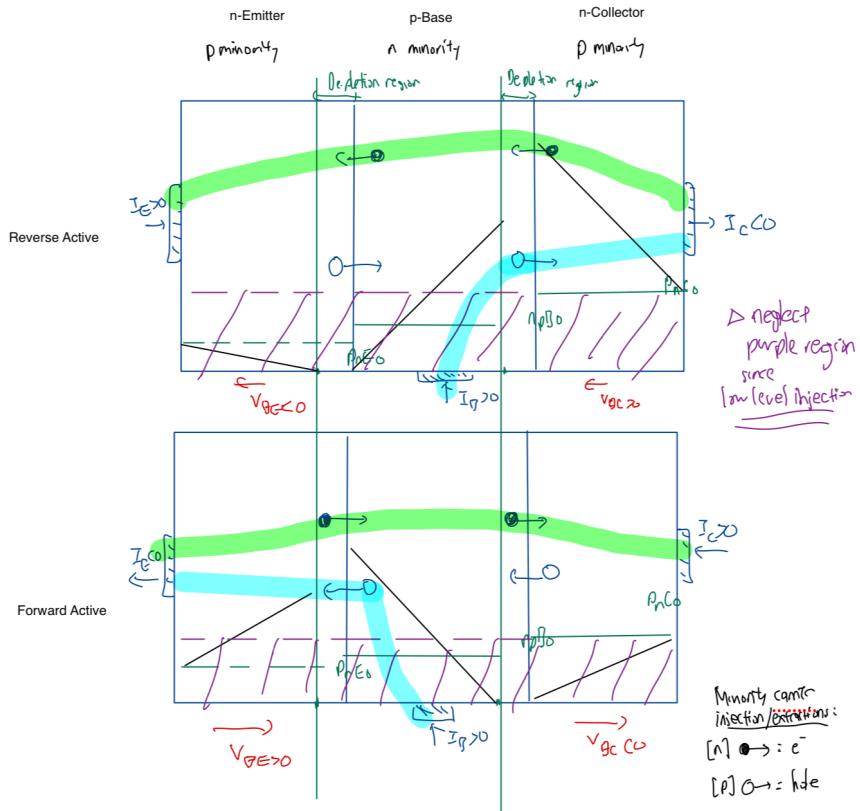
where we have for the forward active regime

- $V_{BE} > 0 \Rightarrow$
 - injection of electrons from the **emitter** to the **base**
 - injection of holes from the **base** to the **emitter**
 - essentially the emitter base junction is forward biased
- $V_{BC} < 0 \Rightarrow$
 - extraction of electrons from the **base** to the **collector**
 - extraction of holes from the **collector** to the **base**
 - essentially the base collector junction is reversed biased

Fact 73

For reversed active it is basically just $V_{BE} < 0, V_{BC} > 0$ (i.e reverse the polarity of the bias) so that the EB junction is reverse active while the BC junction is forward biased instead(which is the exact opposite of the forward active case)

Before we proceed let us analyze the currents contributions by the minority carriers.



Minority Carriers contribution to current for Reverse Active

Firstly the linear profiles and gradients should be expected knowing that the EB junction is reverse biased while the BC junction is forward biased in the reverse active mode recalling our discussion for short base steady state pn junctions. Next we omit the purple region entirely by treating it as zero. That is the thermal equilibrium minority carrier concentration levels are to be treated as indistinguishable from zero. So now only considering the region above the purple region notice that

- the minority carrier concentration profile in the n-collector is primarily due to injection of holes from the p-base into collector
- the minority carrier concentration profile in the p-base is primarily due to injection of electrons from the n-collector to base and extraction of electrons to emitter from base. This also explains why the extreme right side of the concentration profile in p base is above thermal equilibrium level(due to injection) while the left side is below(due to extraction)
- the concentration profile in the n-emitter region is similarly due to extraction of holes into the p-Base. In this case we don't have to consider it because it is in the purple region.

Now let us attempt to account for the currents. We can connect the flux stream of electrons(highlighted green) from collector to emitter contacts only. This is because recall we have assumed short base diode. If we connect it to be the base contact that means that recombination occurred outside the endpoints of diffusion(refers to the collector and emitter contacts). Specifically it happened in the QNR at the base contact which is not allowed. Next let us connect the hole flux. Considering the direction of the hole diffusion and that the hole diffusion on the left is not considered as mentioned due to low level injection, the above flux path highlighted in blue is the only possible path. We will now proceed to show how the currents are now calculated having rationalized the flux paths using the case for forward

active regime below as an example

Minority Carriers contribution to current for Forward Active

Again the same assumptions as before apply. You should be able to rationalize the flux path indicated in 6.1 for the forward active case. As usual omitting the purple region will leave us 6.1 below which we will now analyze in detail with calculations for the currents as promised.

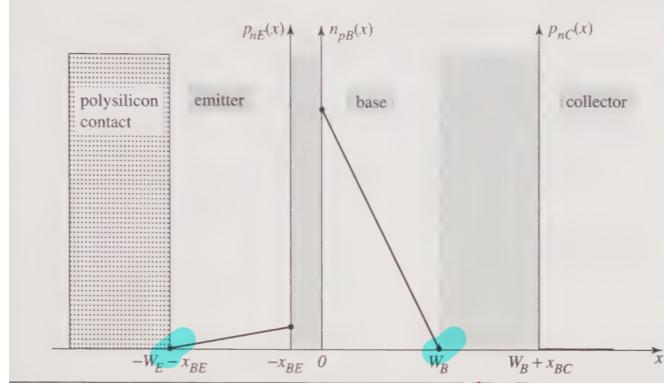


Figure 27: Forward active region BJT: EB junction is forward biased while BC junction is reversed biased so this linear minority carrier profile should be expected recall steady state short base pn diodes. Also notice that low level injection assumption was applied so the thermal equilibrium minority concentration levels are treated as zero as discussed just previously

Now again using **short base diode** steady state assumptions the current concentration profile will be linear like we did for the pn diode previously. In similar fashion starting with Ficks 1st law we have

$$J_{nB}^{diff} = qD_n \frac{dn_{pB}}{dx}$$

which is by substituting our relations from law of junctions for BJT 71 we have

$$J_{nB}^{diff} = qD_n \left(\frac{n_{pB}(W_B) - n_{pB}(0)}{W_B} \right) = qD_n \frac{n_{pB0} (e^{V_{BC}/V_{th}} - e^{V_{BE}/V_{th}})}{W_B}$$

because the base collector junction is reversed biased $n_{pB}(W_B)$ is negligible relative to the thermal equilibrium value if you recall so we may ignore the contribution by it(treat as zero as in 6.1) to obtain

$$J_{nB}^{diff} = - \left(\frac{qD_n n_{pB0}}{W_B} \right) e^{V_{BE}/V_{th}}$$

For the emitter region

$$J_{pE}^{diff} = -qD_p \frac{dp_{nE}}{dx}$$

we similarly get

$$J_{pE}^{diff} = -qD_p \left(\frac{p_{nE}(-x_{BE}) - p_{nE}(-W_E - x_{BE})}{W_E} \right)$$

and again using 71 we obtain

$$J_{pE}^{diff} = - \left(\frac{qD_p p_{nE0}}{W_E} \right) (e^{V_{BE}/V_{th}} - 1)$$

Definition 74

We define **Electron Flux** to be

$$F_n = \frac{J_h}{-q}$$

Before you proceed please recap 6.1 and make sure 77 makes sense to you. Then the following should be relatively straightforward as to find the respective currents from the above current densities we have found, simply multiply by the area in which the current density passes through once you have determined which diffusion flux to use.

For Base Current:

$$I_B = -J_{pE}^{diff} A_E = \left(\frac{qD_p p_{nE0} A_E}{W_E} \right) \left(e^{V_{BE}/V_{th}} - 1 \right)$$

Collector Current:

$$I_C = -J_{nB}^{diff} A_E = \left(\frac{qD_n n_{pB0} A_E}{W_B} \right) e^{V_{BE}/V_{th}}$$

Note that for the above two currents we included the negative sign because the direction of A is oriented such that it is positive *into* the base and the collector terminal for the base and collector currents respectively. However observing the flux directions in 6.1(as indicated the minority carrier movement arrows drawn) they precisely in the direction outwards from A. Note that because minority carriers are assumed to have no recombination until the ends of the diffusion at the emitter and collector contacts, the flux remains constant from the start where it originates from. That is when multiplying the flux density with the area, the area used should be the source/start of the flux. This is why we used $A = A_E$ (emitter contact area) above.

Finally the emitter current is just $I_E = -(I_C + I_B)$ by kirchoff current law see 26

$$I_E = (J_{nB}^{diff} + J_{pE}^{diff}) A_E = - \left[\left(\frac{qD_p p_{nE0} A_E}{W_E} \right) + \left(\frac{qD_n n_{pB0} A_E}{W_B} \right) e^{V_{BE}/V_{th}} \right]$$

Definition 75

Define **Common-Base Current Gain**:

$$\alpha_F = \frac{I_C}{-I_E} = \frac{\frac{qD_n n_{pB0} A_E}{W_B}}{\frac{qD_p p_{nE0} A_E}{W_E} + \frac{qD_n n_{pB0} A_E}{W_B}}$$

Simplifying our expression for α_F we obtain

$$\alpha_F = \frac{1}{1 + \frac{D_p N_{nB} W_B}{D_n N_{nE} W_E}}$$

Base Current:

$$I_B = -I_E - I_C = \frac{I_C}{\alpha_F} - I_C = I_C \left(\frac{1 - \alpha_F}{\alpha_F} \right)$$

Definition 76

Define **Current Gain**:

$$\frac{I_C}{I_B} = \left(\frac{\alpha_F}{1 - \alpha_F} \right) = \beta_F$$

6.2 Reverse active and saturation operating regions(7.3)

Fact 77

As a sanity check you should be aware from our previous discussions that

- Forward Active:
 - I_C : found by analyzing concentration profile in p-base
 - I_B : found by analyzing concentration profile in n-emitter
 - I_E : kirchoff current law
- Reverse Active:
 - I_C kirchoff current law
 - I_B found by analyzing concentration profile in n-collector
 - I_E found by analyzing concentration profile in p-base

As a recap we look again at the forward active region

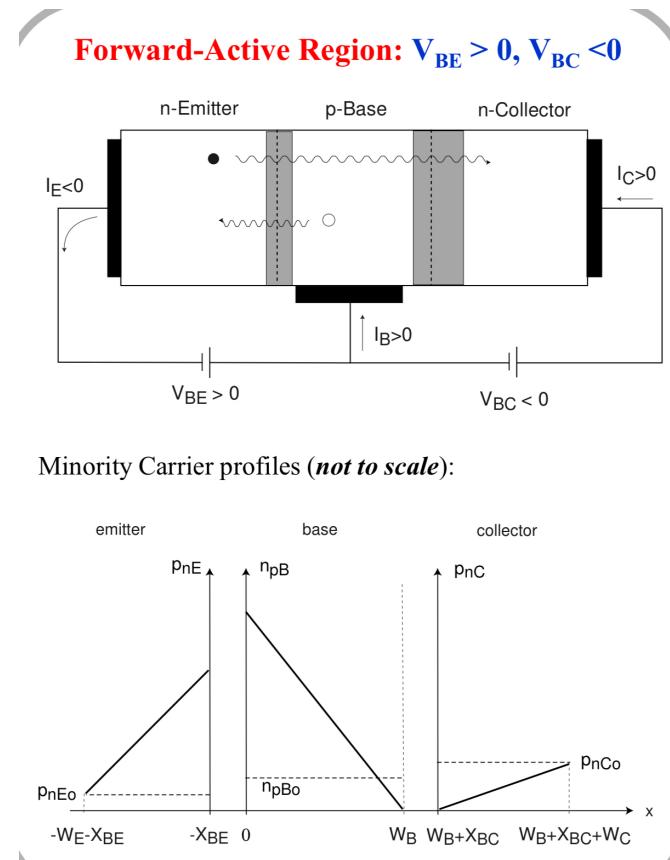


Figure 28: Forward Active Region

$$I_C = I_S e^{\frac{V_{BE}}{V_{th}}} \quad (20)$$

$$I_S = \frac{qA_E n_{pBo} D_n}{W_B} \quad (21)$$

$$I_B = \frac{I_S}{\beta_F} \left[e^{\frac{V_{BE}}{V_{th}}} - 1 \right] \quad (22)$$

$$\frac{I_S}{\beta_F} = \frac{qA_E p_{nEo} D_p}{W_E} \quad (23)$$

$$I_E = -I_C - I_B = -I_S e^{\frac{V_{BE}}{V_{th}}} - \frac{I_S}{\beta_F} \left(e^{\frac{V_{BE}}{V_{th}}} - 1 \right) \quad (24)$$

$$\beta_F = \frac{I_C}{I_B} = \frac{n_{pBo}}{p_{nEo}} \cdot \frac{D_n}{W_B} \cdot \frac{W_E}{D_p} = \frac{N_{dE} D_n W_E}{N_{aB} D_p W_B} \quad (25)$$

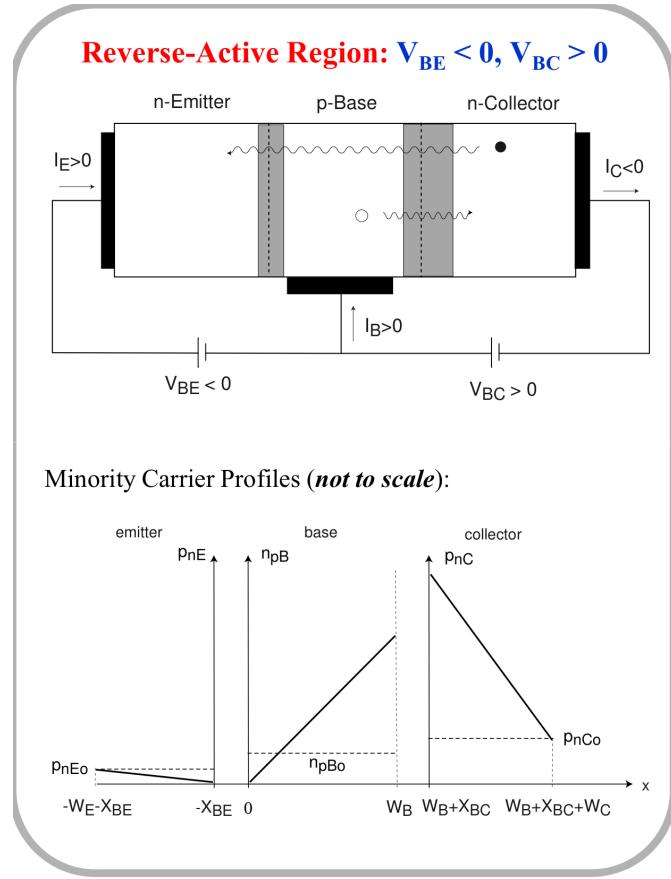


Figure 29: Reverse Active Region

Although not explicitly derived as we did in the forward active case you should be able to see the parallel. For example I_E swaps with I_C compared with the forward active case(that is now I_C is found from kirchoff current law after finding I_E and I_B first instead). Also the area A (start of flux) used in the equations is now A_C

$$I_E = I_S e^{\frac{V_{BC}}{V_{th}}} \quad (26)$$

$$I_S = \frac{qA_C n_{pBo} D_n}{W_B} \quad (27)$$

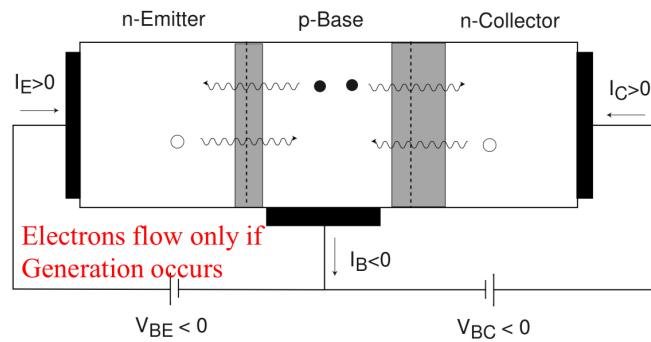
$$I_B = \frac{I_S}{\beta_R} \left[e^{\frac{V_{BC}}{V_{th}}} - 1 \right] \quad (28)$$

$$\frac{I_S}{\beta_R} = \frac{qA_C p_{nCo} D_p}{W_C} \quad (29)$$

$$I_C = -I_E - I_B = -I_S e^{\frac{V_{BC}}{V_{th}}} - \frac{I_S}{\beta_R} \left(e^{\frac{V_{BC}}{V_{th}}} - 1 \right) \quad (30)$$

$$\beta_R = \frac{I_E}{I_B} = \frac{n_{pBo}}{p_{nCo}} \cdot \frac{D_n}{W_B} \cdot \frac{W_C}{D_p} = \frac{N_{dC} D_n W_C}{N_{aB} D_p W_B} \quad (31)$$

Cut-Off Region: $V_{BE} < 0, V_{BC} < 0$



Minority Carrier Profiles (*not to scale*):

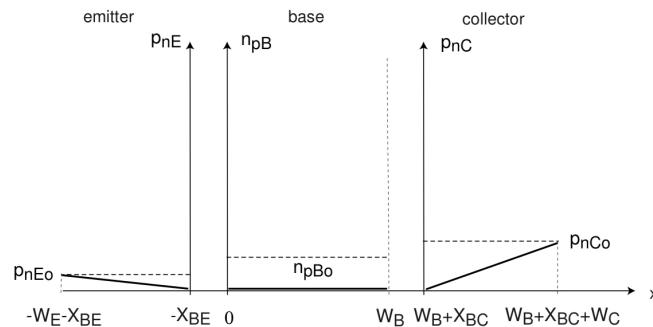
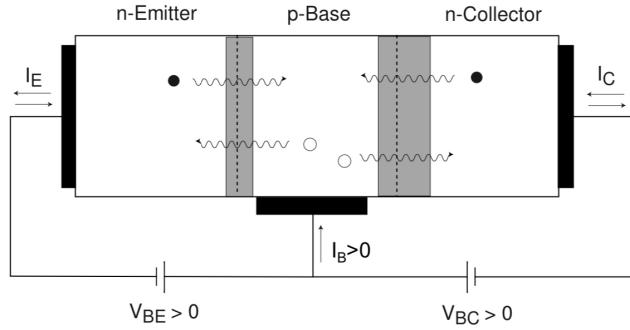


Figure 30: Cut Off Region

Saturation Region: $V_{BE} > 0, V_{BC} > 0$



Minority Carrier profiles (*not to scale*):

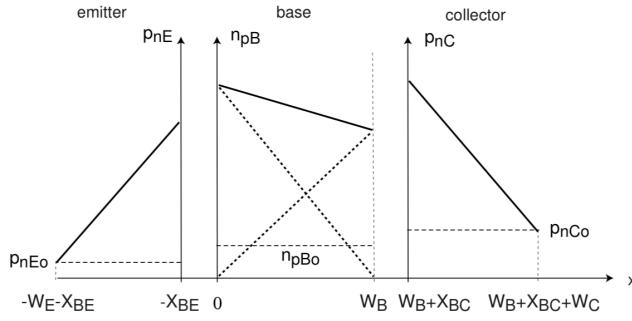


Figure 31: Saturation Region

As you might have noticed this is essentially the superposition of the forward and backward active cases so naturally their current equations reflect that:

$$I_C = I_S \left(e^{\frac{V_{BE}}{V_{th}}} - e^{\frac{V_{BC}}{V_{th}}} \right) - \frac{I_S}{\beta_R} \left(e^{\frac{V_{BC}}{V_{th}}} - 1 \right) \quad (32)$$

$$I_B = \frac{I_S}{\beta_F} \left[e^{\frac{V_{BE}}{V_{th}}} - 1 \right] + \frac{I_S}{\beta_R} \left[e^{\frac{V_{BC}}{V_{th}}} - 1 \right] \quad (33)$$

$$I_E = -I_S \left(e^{\frac{V_{BE}}{V_{th}}} - e^{\frac{V_{BC}}{V_{th}}} \right) - \frac{I_S}{\beta_F} \left(e^{\frac{V_{BE}}{V_{th}}} - 1 \right) \quad (34)$$

7 Single Stage Bipolar/MOS transistor amplifiers(8)

7.1 General Amplifier Concepts(8.1)

References

- [1] Roger Thomas Howe and Charles Giona Sodini. *Microelectronics: an integrated approach*. Prentice Hall electronic and VLSI series. Upper Saddle River, N.J: Prentice Hall, 1997. ISBN: 978-0-13-588518-5.