

Problem Set 4 Solutions

February 17, 2015

1

1.1

(1 points) Two devices of width W and length L are connected in parallel as pictured in Figure 1a. Calculate the Id-Vg characteristic for the pair of devices. Could they be replaced by a single device?

Solution: Each of the transistors in this configuration sees the same V_{ds} and V_{gs} , so we can write:

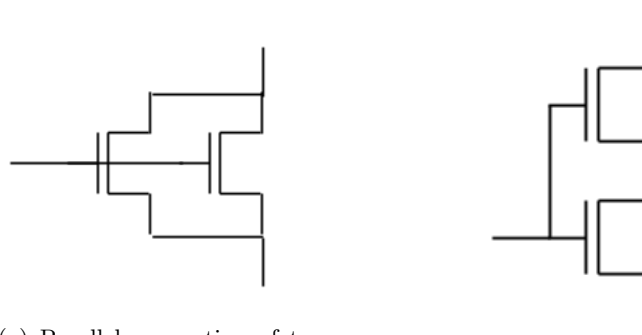
$$I_{right} = I_{left} = \mu C_{ox} \frac{W}{L} \cdot \begin{cases} 0 & V_{gs} < V_T \\ (V_{gs} - V_T)V_{ds} - V_{ds}^2/2 & V_{gs} > V_T \text{ and } V_{gd} > V_T \\ \frac{1}{2}(V_{gs} - V_T)^2 & V_{gs} > V_T \text{ and } V_{gd} < V_T \end{cases} \quad (1)$$

The total current is given by $I_{tot} = I_{right} + I_{left}$. This is the same as a transistor with a width of $W_{tot} = W_{right} + W_{left}$, so in this case we could replace the structure with one transistor of width $2W$ and length L . This holds down to even very small scale transistors.

1.2

(4 points) Two devices of width W and length L are connected in series as pictured in Figure 1b. Calculate the Id-Vg characteristic for this pair of devices. Could they be replaced by a single device?

Solution: We start by defining some terminology: the voltage at the source of the bottom resistor is V_s , the voltage at the drain of the top transistor is V_d , the voltage at the shared gate is V_g and the voltage at the node between the two transistors is V_x . Like a normal transistor, we can write the differences between these two voltages in a compact notation by relying on having two subscripts. For example: $V_{gx} = V_g - V_x$. We'll call the current flowing through both drain-source pairs I .



(a) Parallel connection of transistors.

(b) Series connection of transistors.

Figure 1: Interesting transistor connections.

We know that V_x must be smaller than V_d because there will be a voltage drop across the top transistor. Consequently, $V_{gx} > V_{gd}$. This lets us make some statements about what regions of operation the transistors are in. If the top transistor is in triode (not saturated) then $V_{gd} > V_T$. We know $V_{gx} > V_{gd}$, so the bottom transistor must be in triode as well. If the bottom transistor is in saturation, then $V_{gx} < V_T$, so the top transistor must be in saturation as well because $V_{gd} < V_{gx} < V_T$. We ignore the cutoff state because if either transistor is cutoff then both are because the current through them is zero.

This leaves us with three allowed combinations of states: top-saturated, bottom-saturated; top-triode, bottom-triode; and top-saturated, bottom-triode. Any other state combination would invalidate the inequalities we figured out in the previous paragraph.

Let's examine top-saturated, bottom-saturated first. We know the current through both transistors has to be the same, so:

$$\frac{1}{2}\mu C_{ox} \frac{W}{L} (V_{gx} - V_T)^2 = \frac{1}{2}\mu C_{ox} \frac{W}{L} (V_{gs} - V_T)^2 \quad (2)$$

$$(V_{gx} - V_T)^2 = (V_{gs} - V_T)^2 \quad (3)$$

$$V_{gx} - V_T = V_{gs} - V_T \quad (4)$$

$$V_x = V_s \quad (5)$$

$$\rightarrow I = 0! \quad (6)$$

It turns out that the top and bottom transistors can't both be saturated. If we assume they are then no drain-to-source voltage falls across the bottom transistor and no current flows through it, which would suggest that it operates in cutoff. That logical incompatibility means this arrangement can't occur. Great, one less state to worry about.

Next we'll look at top-saturated, bottom-triode. We'll skip the step where we cancel $\mu C_{ox} W/L$ off both equations.

$$(V_{gs} - V_T)V_{xs} - V_{xs}^2/2 = (V_{gx} - V_T)^2/2 \quad (7)$$

$$2V_{gs}V_{xs} - 2V_TV_{xs} - V_{xs}^2 = V_{gx}^2 - 2V_{gx}V_T + V_T^2 \quad (8)$$

$$(9)$$

At this point, we need to figure out how to get V_{gx} out of our equations

$$V_{gx} = V_g - V_x = V_g - V_x + V_s - V_s = V_{gs} - V_{xs} \quad (10)$$

Great! Let's sub that in:

$$2V_{gs}V_{xs} - 2V_TV_{xs} - V_{xs}^2 = (V_{gs} - V_{xs})^2 - 2(V_{gs} - V_{xs})V_T + V_T^2 \quad (11)$$

$$2V_{gs}V_{xs} - 2V_TV_{xs} - V_{xs}^2 = V_{gs}^2 - 2V_{gs}V_{xs} + V_{xs}^2 - 2V_{gs}V_T + 2V_{xs}V_T + V_T^2 \quad (12)$$

$$0 = 2V_{xs}^2 + V_{xs}(-2V_{gs} - 2V_{gs} + 2V_T + 2V_T) + V_{gs}^2 - 2V_{gs}V_T + V_T^2 \quad (13)$$

$$0 = V_{xs}^2 - 2V_{gs}V_{xs} + \Delta V^2/2 \quad (14)$$

$$\text{where } \Delta V = V_{gs} - V_T \quad (15)$$

A quadratic equation later we have

$$V_{xs} = \Delta V \pm \sqrt{\Delta V^2 - \Delta V^2/2} \quad (16)$$

$$= \Delta V \left(1 \pm \frac{1}{\sqrt{2}} \right) \quad (17)$$

We can substitute this back into our current expression for the bottom transistor to find current:

$$I = \mu C_{ox} W/L \cdot (\Delta V V_{xs} - V_{xs}^2/2) \quad (18)$$

$$I = \mu C_{ox} W/L \cdot \Delta V^2 \cdot (1 \pm 1/\sqrt{2} - (1 \pm 1/\sqrt{2})^2/2) \quad (19)$$

$$I = \mu C_{ox} W/L \cdot \Delta V^2 \cdot (1 \pm 1/\sqrt{2} - (1 \pm 2/\sqrt{2} + 1/2)/2) \quad (20)$$

$$I = \mu C_{ox} W/L \cdot \Delta V^2 \cdot (1/4) \quad (21)$$

$$I = I_{one\ transistor}/2 \quad (22)$$

This suggests we can replace this structure with a single transistor of twice the length.

Saying that conclusively requires us to look at the last state: top-triode, bottom-triode. We skip the $\mu C_{ox} W/L$ step again and use $V_{gx} = V_{gs} - V_{xs}$ and $V_{dx} = V_{ds} - V_{xs}$:

$$(V_{gs} - V_T)V_{xs} - V_{xs}^2/2 = (V_{gx} - V_T)V_{dx} - V_{dx}^2/2 \quad (23)$$

$$V_{gs}V_{xs} - V_TV_{sx} - V_{xs}^2/2 = (V_{gs} - V_{sx} - V_T)(V_{ds} - V_{xs}) - (V_{ds} - V_{xs})^2/2 \quad (24)$$

$$V_{gs}V_{xs} - V_TV_{xs} - V_{xs}^2/2 = V_{gs}V_{ds} - V_{xs}V_{ds} - V_TV_{ds} - V_{gs}V_{xs} + V_{xs}^2/2 + V_TV_{xs} - V_{ds}^2/2 + V_{ds}V_{xs} - V_{xs}^2/2 \quad (25)$$

$$0 = V_{xs}^2 + V_{xs}(-V_{gs} - V_{gs} + V_T + V_T + V_{ds} - V_{ds}) + V_{ds}(V_{gs} - V_T) - V_{ds}^2/2 \quad (26)$$

$$0 = V_{xs}^2 - 2V_{xs}\Delta V + \Delta V V_{ds} - V_{ds}^2/2 \quad (27)$$

Applying the quadratic formula again:

$$V_{xs} = \Delta V \pm \sqrt{\Delta V^2 - (\Delta V V_{ds} - V_{ds}^2/2)} \quad (28)$$

$$= \Delta V \left(1 \pm \sqrt{1 - \frac{(\Delta V V_{ds} - V_{ds}^2/2)}{\Delta V^2}} \right) \quad (29)$$

Finally:

$$I = \mu C_{ox} W/L \cdot (\Delta V V_{xs} - V_{xs}^2/2) \quad (30)$$

$$I = \mu C_{ox} W/L \cdot \Delta V^2 \left[1 \pm \sqrt{1 - \frac{(\Delta V V_{ds} - V_{ds}^2/2)}{\Delta V^2}} - \left(1 \pm 2\sqrt{1 - \frac{(\Delta V V_{ds} - V_{ds}^2/2)}{\Delta V^2}} + 1 - \frac{(\Delta V V_{ds} - V_{ds}^2/2)}{\Delta V^2} \right) /2 \right] \quad (31)$$

$$I = \mu C_{ox} W/L \cdot \Delta V^2 \cdot \frac{(\Delta V V_{ds} - V_{ds}^2/2)}{2\Delta V^2} \quad (32)$$

$$I = \frac{1}{2} (\mu C_{ox} W/L \cdot (\Delta V V_{ds} - V_{ds}^2/2)) = I_{one\ transistor}/2 \quad (33)$$

Thus we can replace this structure with a single transistor of length $2L$. Remember that fact and you'll never have to go through this math again.

2 points for reasoning about regions of operation, 1 point for correctly equating current / equation setup, 1 point for evaluation.

2

2.1

(4 points) A NAND2 gate is pictured in Figure 2. Calculate the leakage current in the gate for each value of A and B .

The gate is sized to match the pull-down delay of an inverter with width W . $V_{dd} = 1V$, $V_{T,nmos} = 0.4V$, $V_{T,pmos} = 0.4V$, and the DIBL constant $\eta = 0.3$. Take I_0 as a constant.

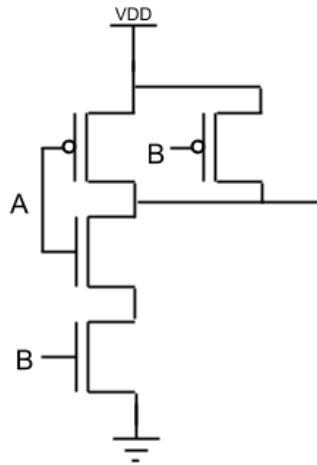


Figure 2: A NAND2 gate.

Solution: There are three possible leakage cases: two PMOS leaking (when both inputs are 1), one NMOS leaking (when inputs are 01 or 10) or a stack of two NMOS leaking (when both inputs are zero). When a transistor is on we can treat it as a short compared to the leakage current, and this allows us to figure out the voltage across the leaking transistors.

In the case where one NMOS is leaking, the output is pulled to V_{dd} by the PMOS that is on. If $A = 0$ and $B = 1$, then the node between the NMOS transistors – which we will call node X – is at ground because the lower NMOS will pull it to zero. If $A = 1$ and $B = 0$, then node X is at V_{dd} because it is pulled up by the active, upper, NMOS device. Either way, the NMOS device that is off will see a V_{ds} of V_{dd} and a V_{gs} of $0V$. Leakage current through that device can be calculated by the formula from class:

$$I_{leak,01} = I_{leak,10} = I_0 \exp\left(\frac{q}{nkT}(V_{gs} - V_T + \eta V_{ds})\right) \left(1 - \exp\left(-\frac{q}{kt}V_{ds}\right)\right) \quad (34)$$

$$= I_0 \exp\left(\frac{q}{kT}(-V_T + \eta V_{dd})\right) \left(1 - \exp\left(-\frac{q}{kt}V_{dd}\right)\right) \quad (35)$$

$$= I_0 \exp\left(\frac{1}{26\text{mV}}(-0.4\text{V} + 0.3 \cdot 1\text{V})\right) \left(1 - \exp\left(-\frac{1}{26\text{mV}}1\text{V}\right)\right) \quad (36)$$

$$= 0.02I_0 \quad (37)$$

Notice that I substitute $q/kT = 26\text{mV}$, this is a common substitution which assumes $T \approx 300\text{K}$ or around room temperature. q/kT is a measure of thermal energy as expressed in volts, it is the random energy of an electron. It is referred to as the thermal voltage and I'll sometimes abbreviate it ϕ_{th} . I use that notation to distinguish it from the threshold voltage, V_T or V_{th} .

Also notice that the $\exp(1\text{V}/26\text{mV})$ term is very small and has almost no effect on the equation. In fact, its only effect is to shut off the leakage at very low levels of V_{ds} . That term can be safely ignored when analyzing transistors with even moderate V_{ds} across them. However, it is quite significant for the PMOS which is turned off in our problem, which exhibits no leakage because $V_{ds} = 0$ so that $\exp(V_{ds}/26\text{mV}) = 1$ and the parenthetical part of the leakage expression cancels to zero.

In the case that two PMOS devices are leaking, $A = 1$ and $B = 1$, we can apply the same formula. V_{gs} for both PMOS devices is zero – recall that the source of a PMOS is the highest voltage it is attached to – and $V_{ds} = V_{dd}$. Each device will see the same V_{gs} and V_{ds} , so we get a leakage current

from each device.

$$I_{leak,11} = 2I_0 \exp\left(\frac{q}{nkT}(V_{gs} - V_T + \eta V_{ds})\right) \left(1 - \exp\left(-\frac{q}{kt}V_{ds}\right)\right) \quad (38)$$

$$= 2I_0 \exp\left(\frac{1}{26\text{mV}}(-0.4\text{V} + 0.3 \cdot 1\text{V})\right) \left(1 - \exp\left(-\frac{1}{26\text{mV}}1\text{V}\right)\right) \quad (39)$$

$$= 0.04I_0 \quad (40)$$

The leakage current through two NMOS devices in series will depend on the voltage at the node between them. To find that voltage, we use the approximation we observed earlier: that $\exp(V_{ds}/\phi_{th})$ is very small. We can equate the currents through both devices under this approximation:

$$I_0 \exp((-V_x - V_T + \eta(V_{dd} - V_x))/\phi_{th}) \approx I_0 \exp((-V_T + \eta V_x)/\phi_{th}) \quad (41)$$

$$-V_x - V_T + \eta(V_{dd} - V_x) \approx -V_T + \eta V_x \quad (42)$$

$$V_x \approx V_{dd} \frac{\eta}{2\eta + 1} \quad (43)$$

$$V_x \approx 0.19 \quad (44)$$

We need to check if this approximation is good. If $V_x = 0.19\text{V}$, then $\exp(V_{ds}/\phi_{th}) = \exp(-7.2) = 0.007$. The approximation is good.

Now that we have a value for V_x we can calculate the current through the stack of NMOS devices

$$I_{leak,11} = I_0 \exp\left(\frac{1}{26\text{mV}}(-0.4\text{V} + 0.3 \cdot 0.19\text{V})\right) \left(1 - \exp\left(-\frac{1}{26\text{mV}}0.19\text{V}\right)\right) = 1.86 \times 10^{-6} I_0 \quad (45)$$

This is a dramatic reduction in leakage from our other states. It is because stacking the two transistors resulted in a negative V_{gs} for the upper device and a small V_{ds} for the lower device. This “stack effect” is well known and it is a commonly used tool to combat leakage.

2 points for identifying correct cases and V_{gs} , V_{ds} . 1 point for getting the middle node of the double stacked NMOS. 1 point for evaluation.

2.2

(2 point) If inputs A and B each have a 50% chance of being one each cycle, what is the average leakage energy per cycle through the gate? The logic is being run at a frequency f_s .

Solution: We have a 1/4 chance of being in each of our states, so the average leakage current is $I_{avg} = 0.25(I_{leak,11} + I_{leak,10} + I_{leak,01} + I_{leak,00}) = 0.02I_0$. We will leak for an amount of time $\approx 1/f_s$ each cycle. That means the average energy per cycle is $E = 0.02V_{dd}I_0/f_s$.

Notice that the leakage energy depends on how fast we can make the cycle time of the processor. Increasing V_T will reduce the leakage current as we saw in the first part of this problem, but it also reduces the *on* current of transistors (a function of $V_{dd} - V_T$) which decreases f_s . This suggests that there is an optimum V_T which will minimize the power of a digital system. We’ll look at that in class soon.

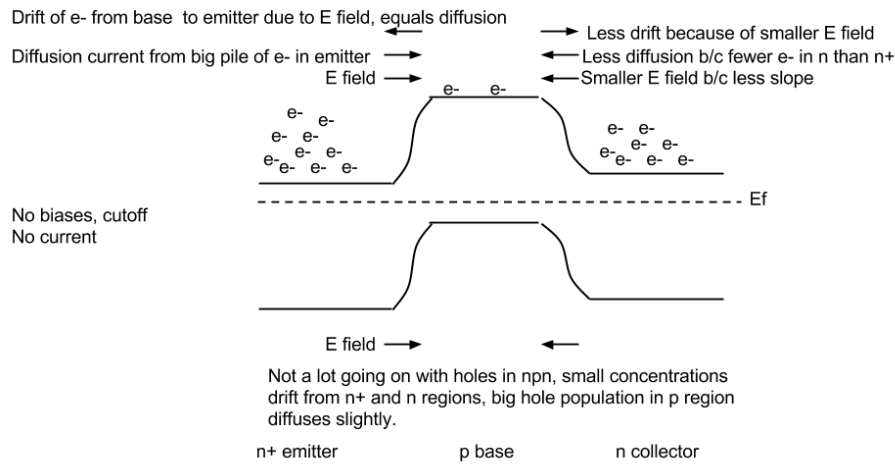
1 point for averaging the current, 1 point for calculating energy in a cycle.

3

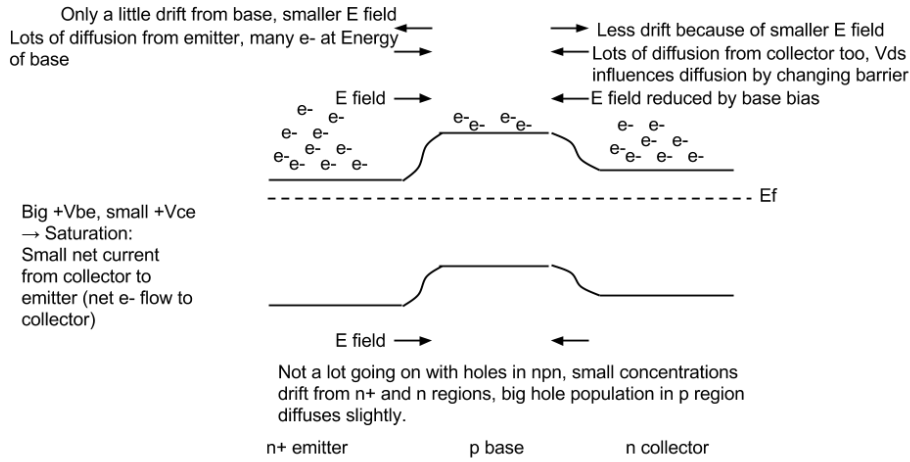
3.1

(3 point) Draw the band diagram for a NPN transistor in all three modes of operation and indicate the qualitative directions and magnitudes of carrier flow. As background, npn transistors consist of two back-to-back pn junctions with the p region being very small. It is doped as n_+ , p , n .

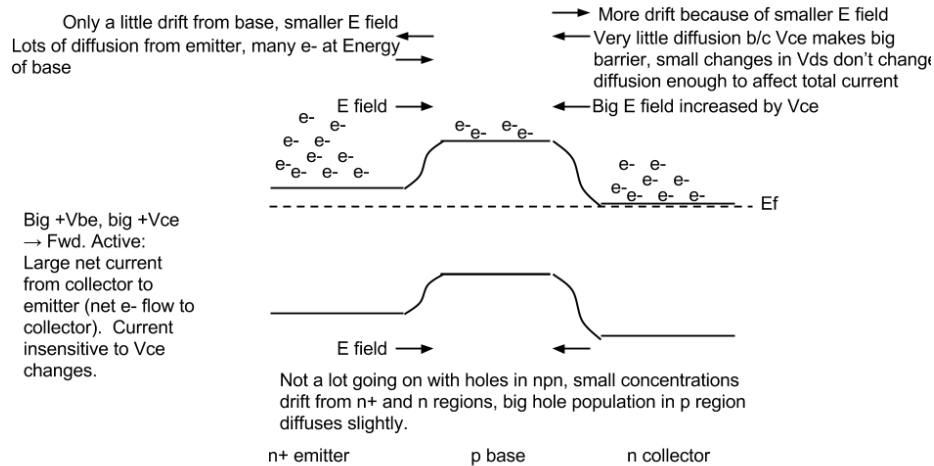
Solution: See Figure 3. Notice that a NPN transistor is a minority carrier device: current that is transmitted from collector to emitter using electrons as charge carriers, but they pass through a p doped region. Thus the charge carriers are the minority carriers in the conduction region. This



(a) Cutoff.



(b) Saturation.



(c) Forward Active.

Figure 3: NPN band diagrams.

is the opposite of a MOSFET, where the conduction region is an inversion layer that has the same kind of carriers as the drain and source: i.e. MOSFETs are majority carrier devices.

1 point for getting bands in the right order, 1 point for identifying where diffusion/drift dominate, 1 point for overall current flow. Half credit if only one of the three modes of operation was incorrect in each category.

3.2

(3 points) Repeat for a PNP transistor.

Solution: See Figure 4.

1 point for getting bands in the right order, 1 point for identifying where diffusion/drift dominate, 1 point for overall current flow. Half credit if only one of the three modes of operation was incorrect in each category.

4

4.1

(1 point) Find an expression for electron concentration and hole concentration as a function of doping by using the law of mass action for electrons and holes ($np = n_i^2$) and the law of charge conservation for electrons and holes.

Solution: We're given the law of mass action in the text. Charge conservation comes from the fact that every positive hole is generated by a negative boron ion (acceptor), and every negative electron is generated by a positive phosphorous ion (donor). Over any given volume these will average to zero charge. So the sum of the positive charges $N_D + p$ has to equal the sum of the negative charges $N_A + n$.

$$np = n_i^2 \text{ and } N_D + p = N_A + n \quad (46)$$

$$p + N_A - N_D - n_i^2/p = 0 \quad (47)$$

$$p^2 + p(N_A - N_D) - n_i^2 = 0 \quad (48)$$

$$p = \frac{N_A - N_D}{2} + \sqrt{\frac{(N_A - N_D)^2}{4} + n_i^2} \quad (49)$$

$$n + N_D - N_A - n_i^2/n = 0 \quad (50)$$

$$n^2 + n(N_D - N_A) - n_i^2 = 0 \quad (51)$$

$$n = \frac{N_D - N_A}{2} + \sqrt{\frac{(N_D - N_A)^2}{4} + n_i^2} \quad (52)$$

4.2

(2 points) What do these expressions simplify to if $N_A \gg N_D$.

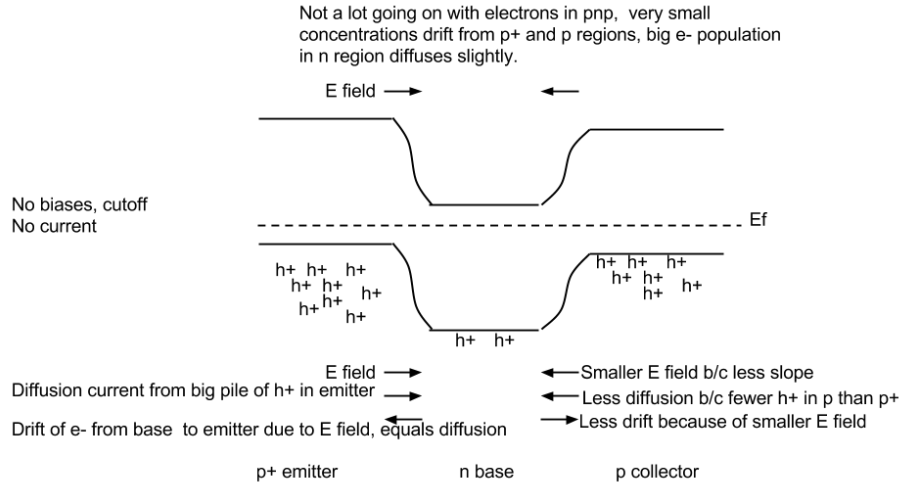
Solution: In terms of doping levels, $N_A \gg N_D$ translates to factors of 1000 or more. For instance, in class we looked at an example with $N_A = 10^{12}/\text{cm}^3$ and $N_D = 10^{15}/\text{cm}^3$. That means that $N_A - N_D \approx N_A$ is a really good approximation. With that in mind:

$$p = \frac{N_A - N_D}{2} + \sqrt{\frac{(N_A - N_D)^2}{4} + n_i^2} \quad (53)$$

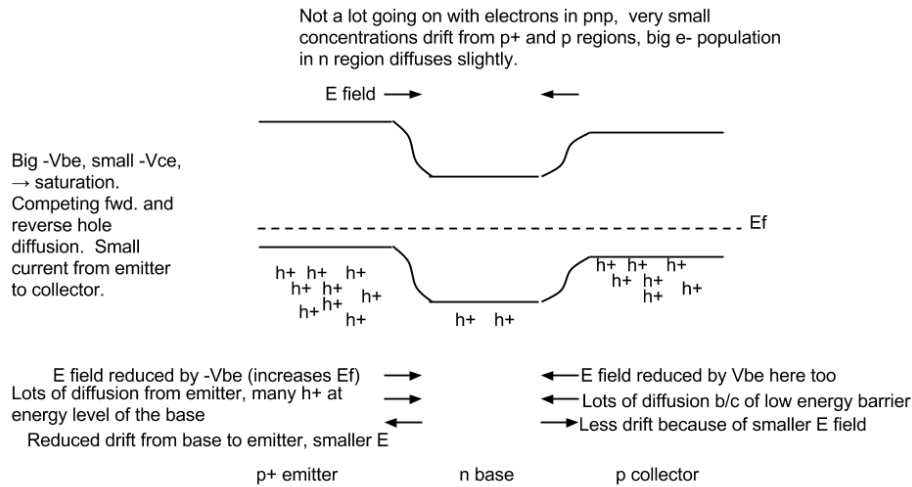
$$\approx \frac{N_A}{2} + \sqrt{\frac{N_A^2}{4}} \quad (54)$$

$$\approx N_A \quad (55)$$

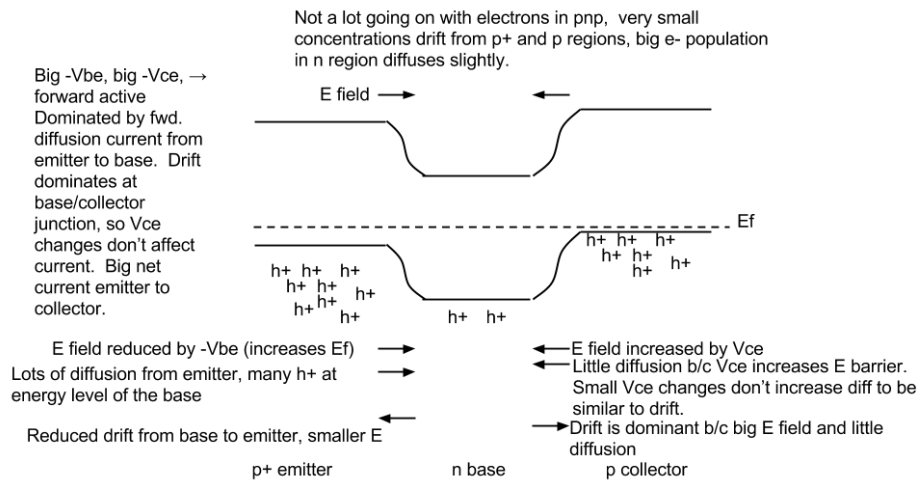
This makes good intuitive sense: the concentration of holes is set by the number of acceptors if we have a lot of acceptors. That's what dopants are designed to do.



(a) Cutoff.



(b) Saturation.



(c) Forward Active.

Figure 4: NPN band diagrams.

If we apply the exact same approximation to n , we find that $n \approx 0$. This isn't a super useful approximation, and more information is available if we do a series expansion:

$$n = \frac{N_D - N_A}{2} + \sqrt{\frac{(N_D - N_A)^2}{4} + n_i^2} \quad (56)$$

$$\approx -\frac{N_A}{2} + \left(\frac{N_A^2}{4} + n_i^2\right)^{1/2} \quad (57)$$

$$\approx -\frac{N_A}{2} + \frac{N_A}{2} \left(1 + \frac{4n_i^2}{N_A^2}\right)^{1/2} \quad (58)$$

$$\approx -\frac{N_A}{2} + \frac{N_A}{2} \left(1 + \frac{1}{2} \frac{4n_i^2}{N_A^2}\right) \quad (59)$$

$$\approx \frac{n_i^2}{N_A} \quad (60)$$

1 point for each n and p . Only take off half a point if the answer for n was 0 and the series expansion was skipped.