



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



Neuromorphic Engineering 1 (NE1), Fall Semester 2014

Summary of the lectures 2012, 2013 & 2014

David Bontrager
Joachim Ott
Dora Sumisławska
Doris Ling
Roland Schlatter
Benjamin Ellenberger

Thursday 4th December, 2014

Professor:
Prof. Dr. Tobi Delbrück
Institute of Neuroinformatics, UZH

Contents

Contents	i
1 Introduction	1
2 On capacitance, MOSFETs, and regions	16
3 Subthreshold behavior	18
4 Above threshold behavior	24
5 Logarithmic I/V converter	27
6 Current source	28
7 Current Mirror	29
8 Source Follower	31
9 Differential Pair	33
10 Current Conveyor	36
11 Current Correlator	37
12 Bump Circuit	39
13 Transconductance Amplifier	41
14 Wide Range Transconductance Amplifier	43
15 Current Divider	44
16 Winner-take-all circuit (WTA)	45
17 Photodiodes, photoreceptor	50
18 Silicon Neurons	57
19 Important Values to Remember	59
20 Exam Questions	60
21 Lab Questions	61
22 Glossary	65
23 TODO	66
Bibliography	67

1 Introduction

Semiconductors

Semiconductors and other materials are arranged in regular structures, so called crystals. The outermost electrons of the atoms, the valence electrons, are distributed in a way that defines and holds together the structure of the crystal.

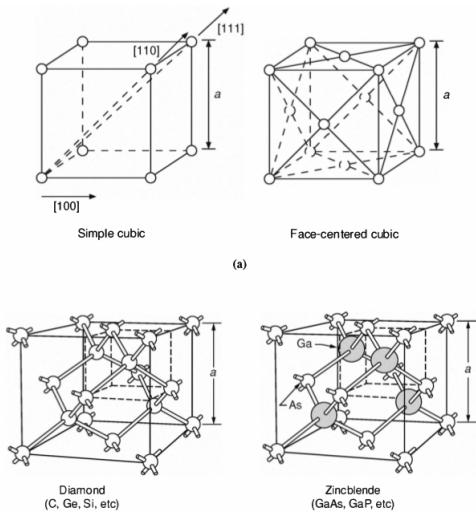


Figure 1: Crystal structure of important semiconductors. a) Possibilities of arrangement. b) Arrangement of silicon (Si) and Gallium arsenide (GaAs). Pure silicon crystallizes in a diamond crystal, Gallium arsenide crystallizes in a so called Zincblende structure.

Energy Band Diagrams

We classify crystals and other solids according to their electrical conductivity into **insulators**, **semiconductors**, and **conductors or metals** in the order of increasing conductivity. Semiconductors are at an intermediate conductivity level, a special property making them suitable for transistors because their conductivity is tunable. It can be modulated by varying the electrical boundary conditions and by introducing atoms of foreign elements into the crystal structure (so called **impurity doping**).

The conductivity of a solid is dependent on the amount of mobile electrons in the solid. Electrons can either be bound to an ion or can be mobile within the solid. Electrical currents are then carried by the motion of the mobile valence electrons. In order to be mobile, the electrons must acquire a minimum energy to break free. The electrons in solids can only be in certain

ranges, so called **energy bands**, and must avoid the forbidden ranges, so called **energy gaps or bandgaps**.

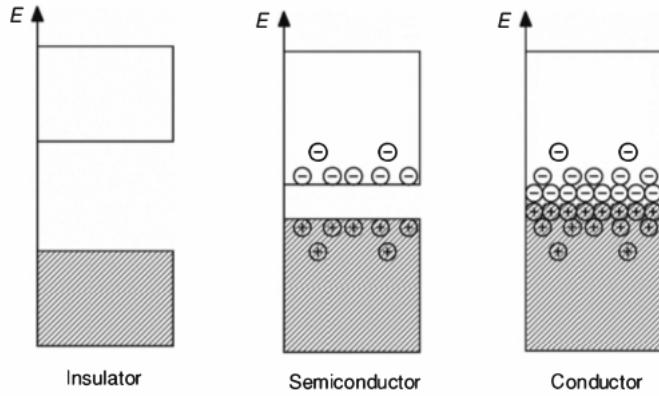


Figure 2: The schematic energy bands of an insulator, a semiconductor and a conductor. The hatched areas symbolize the states that are occupied by electrons at zero temperature. At non-zero temperature, some electrons occupy higher energy states leaving "holes" in the unoccupied lower-energy states. For insulators, the energy bands are either completely filled or empty, as all electrons are bound. A semiconductor has almost full or almost empty bands and the current can only flow under certain conditions. A conductor has partly filled bands within which electrons can move.

Valence bands in semiconductors are almost filled, conductance bands almost empty. A concept simplifying the view on semiconductors is that unoccupied states in a valence band are denoted as holes, being positively charged particles.

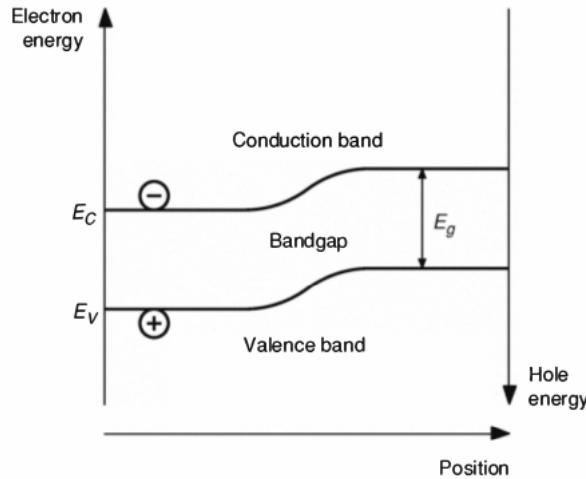


Figure 3: Valence and conduction band of a semiconductor. The energy is plotted as a function of position in one dimension. The energy of the band gap E_g denotes the difference between the lowest conduction band energy and the highest valence band energy. The band edges tell us the minimum amount of energy an electron has to acquire or lose to bridge the bandgap.

Carrier concentrations at Thermal equilibrium

Thermal energy makes some electrons move from the lower-energy states to the higher-energy states. The occupancy of energy states is described by the **Fermi-Dirac distribution** (E_F denotes the **Fermi level**, the energy at which the occupation probability is 0.5, k = Boltzmann constant, T =absolute Temperature).

$$F(E) = \frac{1}{1 + e^{\frac{E-E_F}{kT}}}$$

Typical impurity doping concentration, E_F is well separated from the valence and conduction band edges ($|E - E_F| \gg kT$), thus we can use the Boltzmann distribution

$$F(E) = e^{-\frac{E-E_F}{kT}}$$

This probability distribution is the reason for the exponential characteristics of diodes and transistors. The fermi level in an undoped semiconductor is usually halfway between the valence and conduction band.

Impurity doping

The conductivity of a semiconductor is increased significantly by impurity doping. By replacing several atoms of the semiconductor crystal structure

by a different element (a donor impurity if it has a valence electron more, an acceptor impurity if it has a valence electron less than the semiconductor atom), we get more unbound electrons/holes as the impurity atom itself is electrically neutral and the electron/hole can move freely.

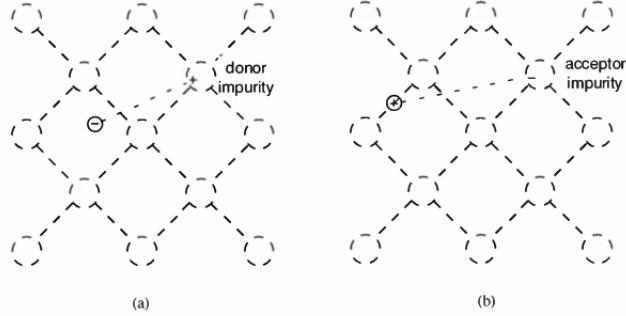


Figure 4: Semiconductor with a) donor and b) acceptor impurity doping. The excess charge carrier is only loosely bound to the ion cores (dashed circles).

The introduction of donors moves the Fermi level from near the center of the bandgap further towards the conduction-band edge, whereas the introduction of acceptors moves it closer to the valence-band edge. Higher acceptor doping semiconductors are called **n-type**, higher donor doping semiconductors are called **p-type**. Too strongly doped semiconductors have a Fermi level within the valence or conductance band and therefore are similar to metals. Common doping elements are phosphorus (P), arsenic (As) as donors and boron (B) as acceptor.

The mobile charge carriers that are dominant in the semiconductor in thermal equilibrium are the majority carriers, the sparser ones are the minority carriers.

Current densities

By historical definition the electron is assigned the negative elementary charge $-q$, where we define q to be positive. However, the direction of the current density is defined as the direction of positive charge flow. For each carrier type the current flow is due to two basic mechanisms, namely diffusion and drift.

Diffusion is a term borrowed from gas dynamics. It describes the process by which a net particle flow is directed from a region of higher particle density to a region of lower particle density along the density gradient. As we shall see, diffusion determines the current flow in diodes and, within the operating range mainly considered here, in transistors. Diffusion also governs the

ion flows in biological neurons.

Drift currents are caused by electric fields.

p-n Junction Diode

The fundamental semiconductor device is the p-n junction diode. It is built from an n-type semiconductor and a p-type semiconductor, which, when brought into physical contact, give rise to a net electron flow from n-type to p-type region and a net hole flow from p-type to n-type region. The diffusing minority carriers recombine with majority carriers in the vicinity of the junction and therefore gets depleted of mobile charge carriers, hence its name **depletion region or space-charge region**.

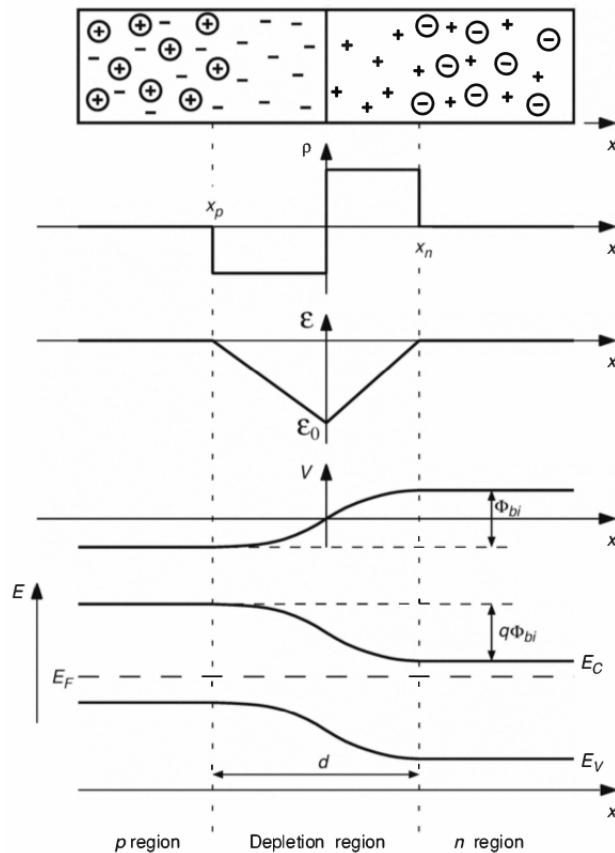


Figure 5: Characteristics of an abrupt p-n junction in thermal equilibrium with space-charge distribution ρ , electric field distribution \mathcal{E} , potential distribution V . and energy-band diagram E .

The depletion region on the n-type side of the junction is depleted of donor electrons and has surplus protons, the p-type side of the junction is

depleted of acceptor holes (filled with the donor electrons) and so have surplus electrons. This builds up an electric field pulling the majority carriers back into the other direction. In thermal equilibrium, diffusion and drift currents balance each other out and no net current flow is observed. The offset between the energy bands bordering the depletion region is an electrical potential called **built-in potential or diffusion potential** Φ_{bi} .

Forward and Reverse Bias

In a p-n junction, a net current flow through the diode is generated by applying a potential difference to it. A positive voltage difference applied to the p-type region relative to the n-type region is called a **forward bias** and the corresponding current is called **forward current**, a positive voltage difference applied to the n-type region relative to the p-type is called a **reverse bias**, and the corresponding current **reverse current**. In a steady state, the total current density must be constant throughout the diode. The electron current in the n-type region is transformed into a hole current in the p-type region. The applied voltage V appears across the depletion region as a change in the built-in voltage and thus modifies the width of the depletion region.

A **forward bias** shrinks the step across the junction. The positive potential applied to the p-type material repels the holes, while the negative potential applied to the n-type material repels the electrons. With increasing forward-bias voltage, the depletion zone eventually becomes thin enough that the zone's electric field cannot counteract charge carrier motion across the p-n junction, as a consequence reducing electrical resistance. The electrons that cross the p-n junction into the p-type material (or holes that cross into the n-type material) will diffuse in the near-neutral region. The forward current in the p-n junction when it is forward-biased involves electrons from the n-type material moving across the junction and combining with holes in the p-type material. Electrons can then proceed further onward by jumping from hole to hole, so the holes can be said to be moving to the opposite direction in this process.

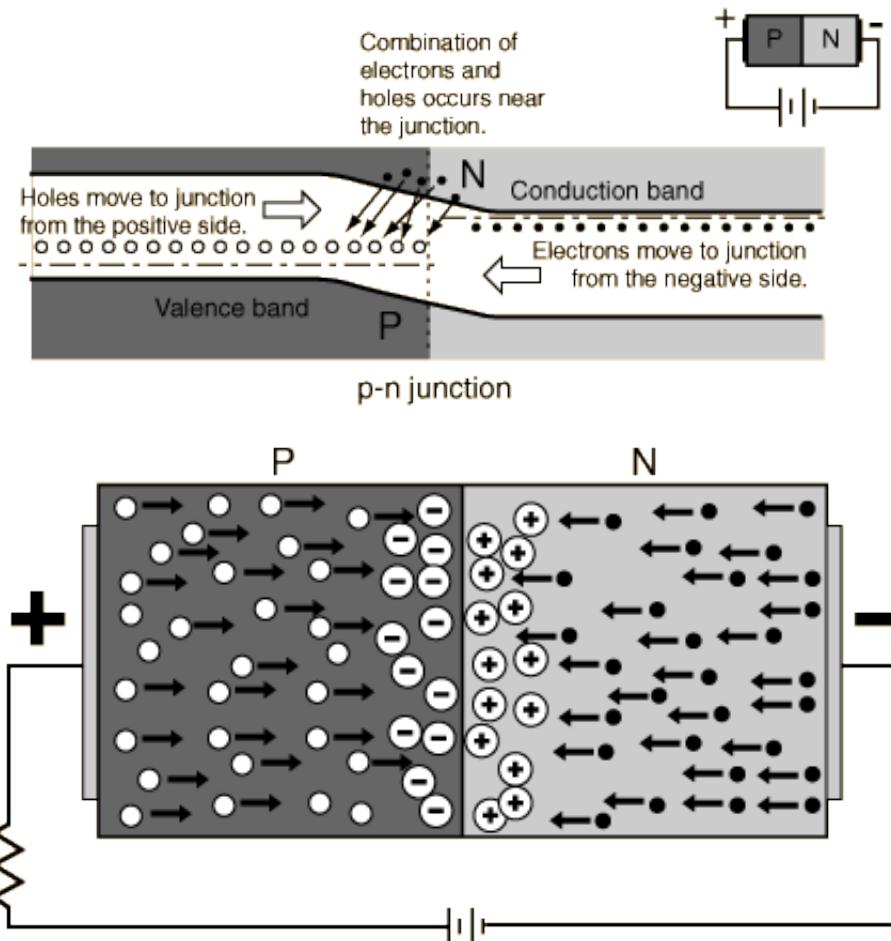


Figure 6: Forward biasing of an p-n junction.

A **reverse bias** expands the step across the junction. To reverse-bias the p-n junction, the p side is made more negative, making it "uphill" for electrons moving across the junction. This will cause a transient current to flow as both electrons and holes are pulled away from the junction. When the potential formed by the widened depletion layer equals the applied voltage, the current will cease except for the small thermal current.

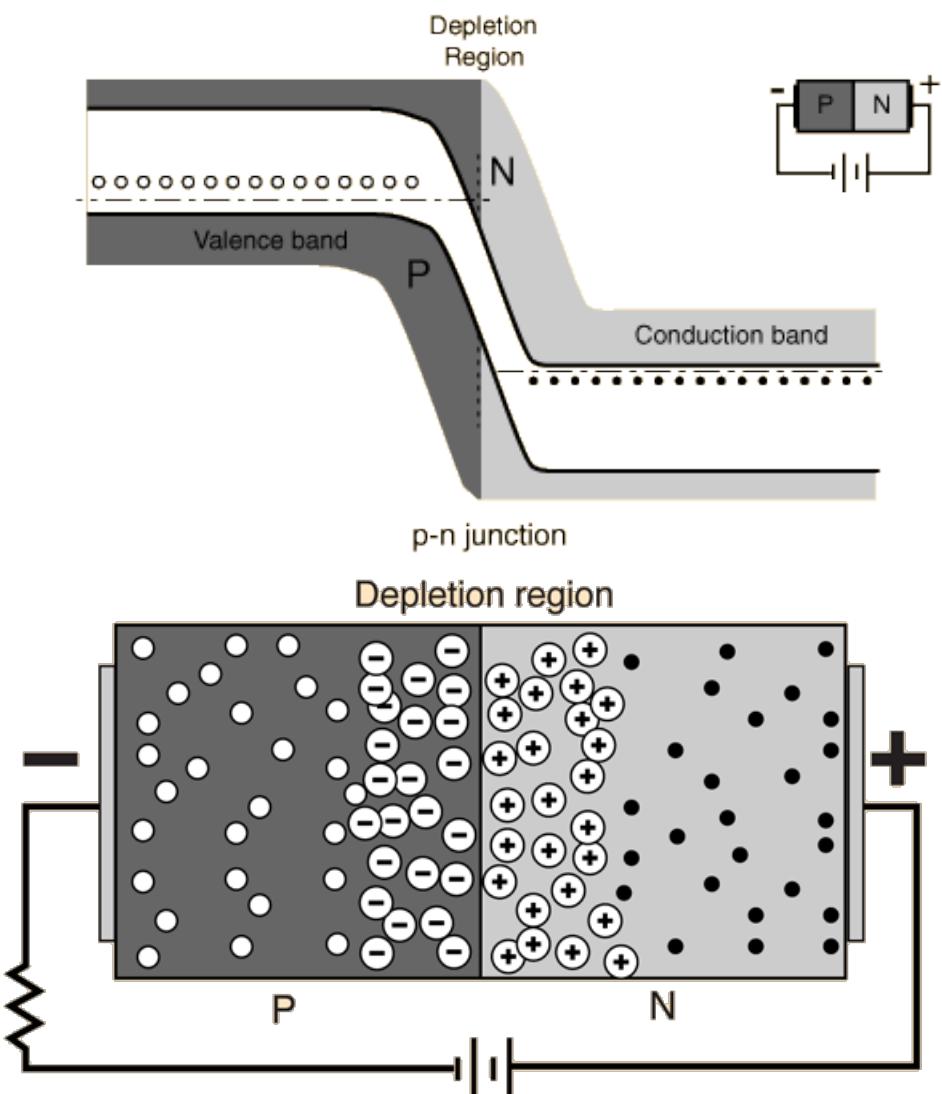


Figure 7: Reverse biasing of an p-n junction.

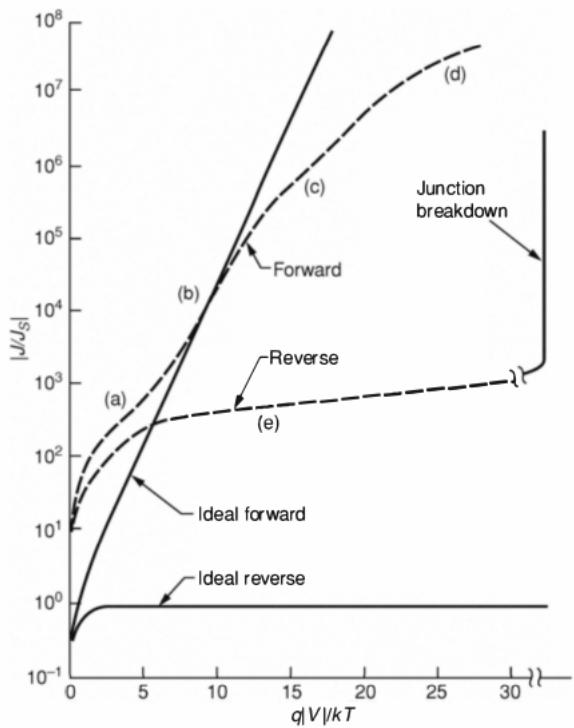


Figure 8: Biasing of an p-n junction.

For large reverse biases, a phenomenon called **junction breakdown** occurs that expresses itself in a sudden increase of reverse current at a certain reverse voltage. For silicon with typical impurity doping concentrations this effect is due to impact ionization: The generation of electron-hole pairs by collision with an electron or hole that has acquired sufficient kinetic energy in the electric field of the depletion region. A charge carrier may create multiple electron-hole pairs during its transition through the depletion region. The generated carriers can in turn create electron-hole pairs if they acquire sufficient energy, and so on. This effect is known as avalanche multiplication.

MOS Transistor Structure

MIS

Metal-Insulator-Semiconductor, consists of a conductor and a semiconductor, separated by thin insulation layer.

MOS

Metal-Oxide-Silicon, a version of MIS, in which silicon dioxide (SiO_2) is used as oxide.

CMOS

Complementary Metal Oxide Silicon, a MOS process in which both nFET and pFET are fabricated on the same substrate.

MOSFET

Metal-Oxide-Semiconductor Field-Effect Transistor, built using MOS and a p-n junction diode.

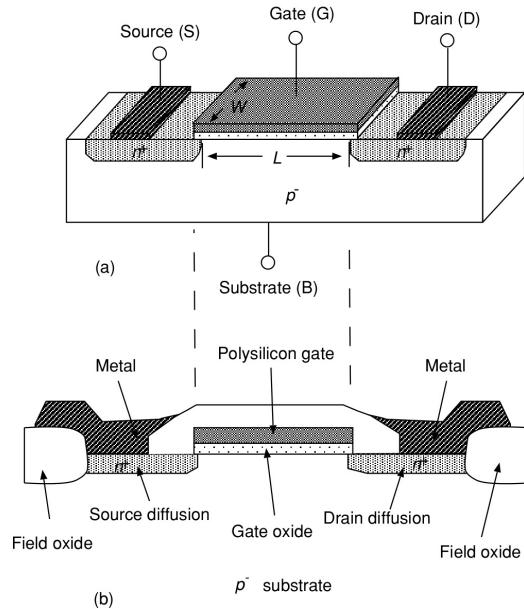


Figure 9: Structure of an n-type MOSFET in a pbody. The MOSFET has four terminals; the drain (D), the source (S), the gate (G), and the bulk (B). (a) Pictorial view of the MOSFET. (b) A more realistic picture of a cross-section of a fabricated MOSFET. Note that the gate oxide is much thinner than the field oxide. [1]

MOS Transistor Terminals

We can view the transistor as having four terminals:

1. Gate (G)
2. Source (S)
3. Drain (D)
4. Bulk (B)

MOS Transistor Types

n-type

Because the n+ source and drain regions can supply a lot of electrons to the channel, this device is called an n-channel MOSFET (nFET,n-type MOSFET, NMOS) [1]

p-type

In p-channel MOSFET (pFET,p-type MOSFET, PMOS) , the charge in the channel is carried by holes supplied from the source and drain regions.

MOS Transistor in Substrate

Most CMOS processes use a p-type starting substrate. The nFETs rest in the common p^- substrate, and the pFETs rest in n-wells within the substrate as shown in Fig. 10

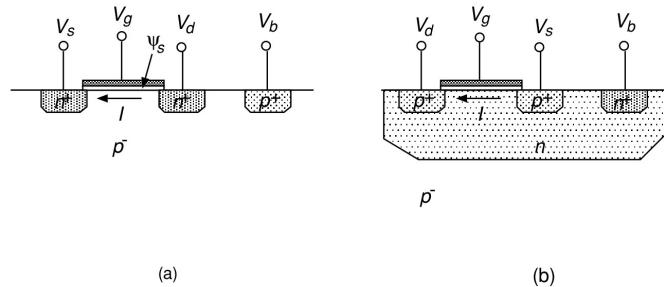


Figure 10: Physical structure of (a) an nFET and (b) a pFET in a common p^- substrate. The pFET rests in a n-well within the substrate. [1]

MOS Transistor Biasing

nFET

To ensure only a small leakage current between the n^+ regions to the p-substrate, the junctions have to be reverse biased. To do this, the drain voltage V_d and the source voltage V_s of the nFET should be greater than or equal to the bulk voltage V_b :

$$V_{sb} = V_s - V_b \geq 0 \quad (1)$$

$$V_{db} = V_d - V_b \geq 0 \quad (2)$$

The location of drain and source is defined by the carrier flow:[1]

- Electrons are supplied to the channel by the source, and removed by the drain.
- Holes are supplied to the channel by the source, and removed by the drain.

The n^+ region biased at the higher voltage is called the drain, and the other n^+ region is called the source. Because electrons are negatively charged, the direction of positive current flow, I , is from drain to source, even though the carriers flow from source to drain. [1]

pFET

In pFET , the p^+ regions should be biased negative relative to the bulk to reverse-bias the pn junctions.

$$V_{sb} \leq 0 \quad (3)$$

$$V_{db} \leq 0 \quad (4)$$

MOS Transistor Channel

The region underneath the gate and between the source and drain regions is called the channel. The channel has a width W , and a length L . The channel is insulated from the gate above by a layer of silicon dioxide. The gate is made of heavily doped (low resistivity) polycrystalline silicon. [1]

MIS Operation Domains

Depending on the charge of the gate, we either attract or repel majority carriers on the surface of the channel semiconductor.

Important: Relevant here is the type of substrate (p or n) the channel is made of, don't confuse it with nFET or pFET!

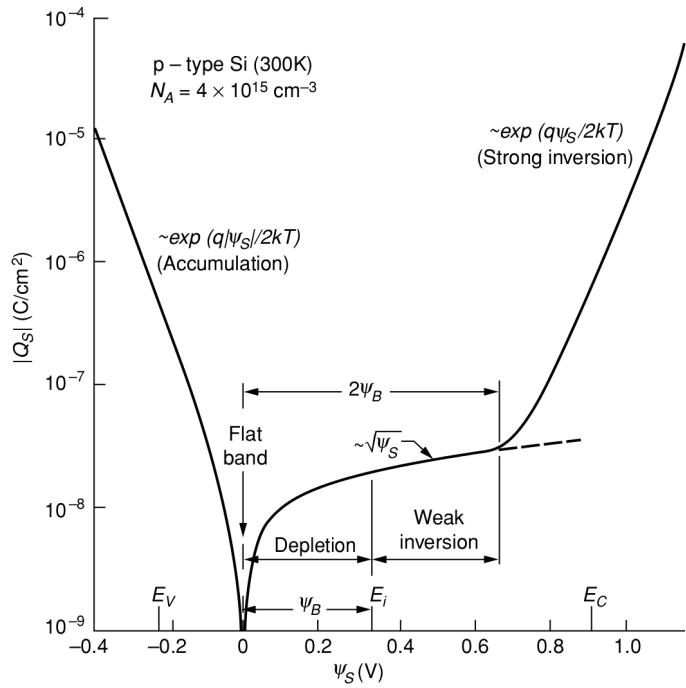


Figure 11: Dependence of the area charge Q_s on the surface potential for p-type silicon with acceptor density $N_A = 4 \times 10^{15} \text{ cm}^{-3}$ at room temperature. Figure adapted from S. M. Sze (1981), Physics of Semiconductor Devices, 2nd Edition. ©1981 by John Wiley & Sons, Inc. [1]

Accumulated Transistor Channel

Accumulation happens when a charge on the gate attracts a lot of majority carriers on the surface of the semiconductor underneath it.

For a p-substrate channel (default), as shown in figure 12, a negative voltage on the gate means mobile electrons on the gate. These electrons (negatively charged) attract positive charges on the semiconductor surface underneath it, leading to accumulation of them. Since positive charges (holes) are the majority carrier in p-substrate, we call this situation accumulation.

In a n-substrate channel, positive charge on the gate leads to accumulation of electrons (majority carriers in n-substrate) on the channel surface.

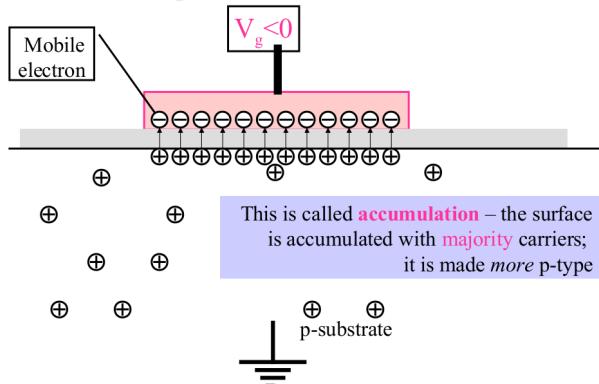


Figure 12: Accumulated Transistor Channel [7]

Flat-Band Transistor Channel

In an ideal MIS diode, with no bias applied, the work function of the metal and the semiconductor are the same. The Fermi levels line up and the energy bands in the semiconductor are flat. → Flat-Band Condition [1]

With $V_g = V_{fb} = 0$, the majority carrier density is constant and equal to the dopant density.

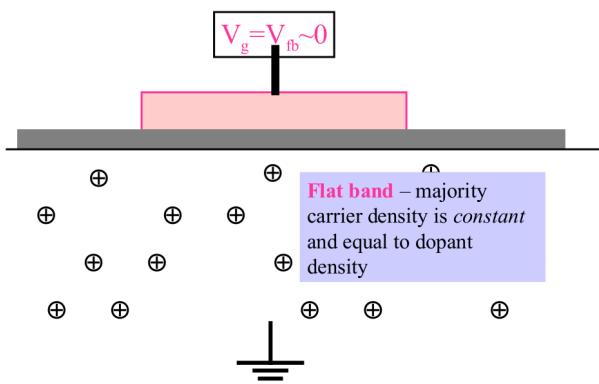


Figure 13: Flat-Band Transistor Channel [7]

Depleted Transistor Channel

For a p-substrate channel, if we put a positive subthreshold voltage on the gate (positive charges), we repulse positive majority carriers on the semiconductor surface. → Depletion of majority carriers.

For a n-substrate channel, the same happens if we put a subthreshold negative charge on the gate. → Depletion of electrons (majority carrier in n substrate) on the semiconductor surface.

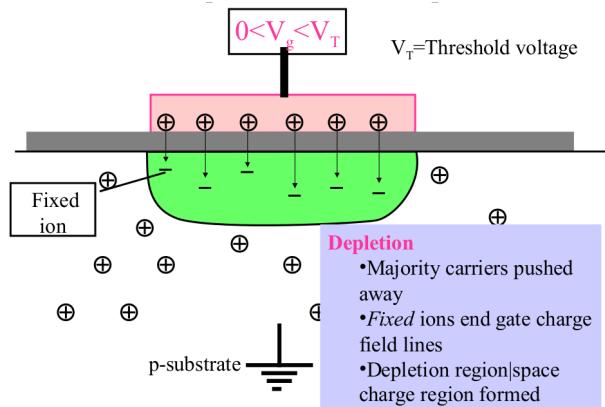


Figure 14: Depleted Transistor Channel [7]

Inverted Transistor Channel

For a depleted p-substrate channel, if we now further increase the positive charge on the gate, we will reach a point where we start attracting negative minority carriers on the semiconductor surface. The surface becomes inverted. (p-type \rightarrow n-type and vice versa in n substrate) \rightarrow Inversion

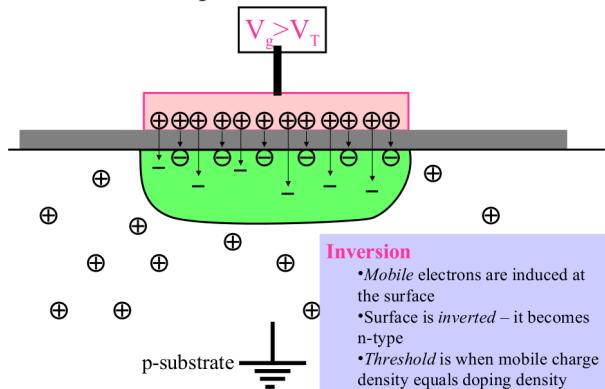


Figure 15: Inverted Transistor Channel [7]

2 On capacitance, MOSFETs, and regions

We start with nomenclature. There are two basic kinds of MOSFETs¹: nFETs and pFETs. These are sometimes also called *n*-channel and *p*-channel transistors (respectively), and – in an *n*-well device – native and well transistors (respectively). The nFET has an *n*-type source and drain sitting in a *p*-type substrate, and the pFET is vice-versa.

We have devices now – things to plug in. You tie the nFET’s drain to a higher voltage than its *source* (which will *source* electrons), and the pFET is vice-versa (because it will *source* holes). “Source” and “drain” are so named because they respectively *source* and *drain* the majority mobile carrier of the given transistor type.

You don’t want current crossing the *pn* junctions where the drain and source sit in the substrate, so we reverse bias those junctions². That means we stick the nFET bulk to ground (usually hardwired on the chip), and keep $V_s, V_d > 0$. Look at that – we’ve plugged it in!

We know about the sub- and above threshold regions. Moving between these two regions is a function of V_g , specifically the voltage between the gate and the source ($V_{gs} = V_g - V_s$). Both the sub- and the above threshold regions are divided into the triode³ and saturation regimes. For both sides of the threshold, we move in and out of saturation by changing V_{ds} – the drain-source voltage.

This is important: changing gate-source voltage from low to high moves us from subthreshold to above threshold for any $V_{ds} > 0$. Similarly, changing drain-source voltage from low to high moves us from ohmic regime to saturation regime for any $V_{gs} > 0$.

Surface voltage ψ_s deserves another mention. We’ve built a transistor, plugged it in, and biased the gate relative to the bulk⁴. Because of the positive gate voltage, the depletion region does not only surround the drain and source, but also extends between them. Having this channel of depletion region

¹Metal-Oxide-Semiconductor [what it’s made of] Field Effect Transistor [how it works]

²Put *n* at a higher voltage than *p*

³Interesting but irrelevant: “triode” comes from the original thought behind the transistor as a 3-terminal diode, thus *tri*-ode vs. *di*-ode. The entire transistor history on Wikipedia is a really interesting read.

⁴Really, we care about all voltages *relative to the bulk*, but bulk is almost always at ground so it’s an implicit relation

means that there is a built-in voltage across it⁵. So here we have two voltages in series – one across the insulating oxide, one across the depletion region. Current isn't flowing across either of these gradients, so what does that mean? It means static, accumulated charges are hanging out on either side of the gradients. In other words: they are capacitances. The relationship between the capacitances is described by κ , which we call the *capacitive coupling ratio* from gate to channel. It describes how effectively a change in gate voltage can change the surface voltage.

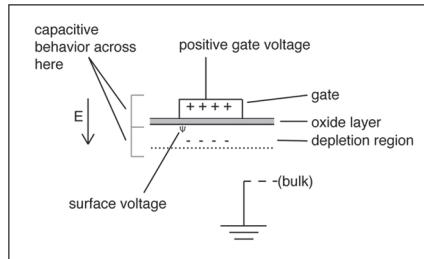


Figure 16: Cross-section of the transistor gate, oxide layer, and depletion region. When $V_{gate-bulk} > 0$, both the oxide insulator and the depletion region (channel) act as capacitors. See Figure 2.18, p.42 in the textbook for a proper version of this figure.

MOSFET Operation Domains

The operation domain of an nFET is set by the relative values of the four terminals of the transistor.

In general, these operation domains (e.g. accumulation) are equivalent to the ones of the MIS structure.

We can map the regimes to a MOSFET mode:

When the MOSFET is in this domain:	We say the MOSFET operates in:	Which can be further divided into (depending on V_{gs}):
Accumulation	Cutoff	
Depletion	Cutoff	
Weak Inversion	Subthreshold	Triode Mode Saturation Mode
Strong Inversion	Above Threshold	Triode Mode Saturation Mode

⁵Not across it as in a *drain-source* voltage, but across it as in perpendicular to the plane of the insulating oxide

3 Subthreshold behavior

Depending on how we want a transistor to behave, we will want to keep it in subthreshold. A few of the biggest reasons to do so are low power consumption (nano- or microamps), less dominant Early voltage effects, and the usefulness of logarithmic behavior (can cover many orders of magnitude of current for less than one order of magnitude of voltage). Most circuits in this class involve subthreshold transistors, so these are extremely important concepts to understand. In this chapter I will discuss only nFET transistors. To do the equations for pFET devices, you simply multiply all voltages by -1^6 .

On subs, threshes, and holds

Yeah it's a bad section title. Don't worry about it. We mentioned κ above. It's pretty darn important, so let's look at the mathematical definition:

$$\kappa = \frac{C_{ox}}{C_{ox} + C_{dep}} = \frac{\partial \psi_s}{\partial V_g} \quad (5)$$

where C_{ox} is the capacitance across the insulating oxide and C_{dep} is the capacitance across the depletion layer, as illustrated in Figure 16. ψ_s is a function of a number of terms, but this is what it boils down to and why it is important⁷. Typical values for κ are between 0.4 and 0.9, though it clearly must - mathematically - be less than 1 (look at (5))

Okay, so we're subthreshold. V_{gs} is low but non-zero. Let's say the source and drain, both of which are *n*-type material sitting in the *p*-type substrate, are also both at 0V relative to ground. Because of the low gate voltage, the electrons have an energy barrier to overcome⁸ if they want to travel between

⁶We did not previously mention this, but pFET bulks are tied to V_{dd} , not ground, so all voltages are relative to that (that is, the relevant V_g for a pFET would be calculated as $V_{dd} - V_g$). The same goes for drain and source voltages. Generally, what the bulk is tied to is hard-wired into the device, though sometimes we make a device where there is a fourth lead controlling this.

⁷p.44 in the textbook has details on how the partial derivative comes into play, but the lecture slides do a better job of making this point

⁸That's not the best way to put it. The built-in potential from the *pn* junctions creates the energy barrier. This voltage will be such that the *n*-type material (drain and source in an nFET) is at a higher voltage than the *p*-type material (the substrate in an nFET), so due to the electric field electrons want to flow from *p*-type to *n*. If we then increase V_g , the positive charge on the gate pushes the *substrate's* mobile carriers (holes) away from the silicon oxide, leaving a negatively-charged depletion region immediately under the surface. This is the channel where we see inversion - this is how we form the inversion layer. Broadening this channel (increasing V_g) makes it take less energy for an electron to move from the drain or source into the channel, and *that* is what's happening with the energy barrier.

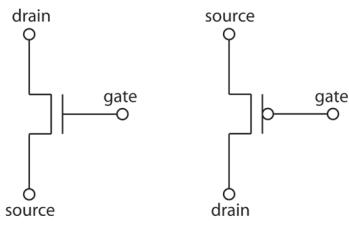


Figure 17: Circuit diagrams for an nFET (left) and a pFET (right). Terminals are labeled such that, in a standard circuit diagram, V_{dd} is above the components and *ground* is below them. However, nothing in the device forces it to be this way. They are physically symmetric and can work with current flowing in either direction.

drain or source and the channel. It's difficult - the natural state is quite resistive. Raising the gate voltage (but keeping it under 0.7V, the approximate threshold voltage) means electrons can more easily move through the channel because the energy barrier is lower. But they have no reason to move. It is not more favorable energy-wise to be at the source compared to being at the drain, because drain-source voltage is still zero.

We raise the drain voltage, though only a little. Tens of millivolts. The drain side now has a slightly higher energy barrier for its electrons to cross if they want to enter the channel⁹. The mobile carrier concentration at either end of the channel is a function of this barrier height. If we change the barrier height at one end and not the other, carrier concentration will decrease on the side with the higher barrier, setting up a concentration gradient. What do we know happens across gradients? Yes! Diffusion! *This is important to note – the electric field across the channel is still not the main driving force in current flow. Electrons are moving primarily due to the concentration gradient.* Sure, it's only tens or hundreds of nanoamps, but it's current!

Current flows with even very small drain-source voltages. For a small range of voltages, it stays in the *triode/ohmic/linear* region. The name comes from the fact that current has an approximately *linear* dependence on drain-source voltage *when the transistor is above threshold*¹⁰. In fact it is also approximately linear while subthreshold for V_{ds} below U_T ¹¹, but then the exponential starts kicking in. To be clear, it only looks linear instead of exponential here *because*

⁹This energy barrier stuff takes a little thinking to get used to, but soon you don't worry about this because you only care about the "what," not the "why." They spend a reasonable amount of time on it in class, so pay attention and ask questions and it should make sense

¹⁰The "ohmic" part of the name is simply projecting Ohm's law on the linear regime, describing the fact that the transistor acts as linear (ohmic) resistor in that range. That is, current increases linearly with voltage, just like $V = IR$ says.

¹¹about 25 mV, remember?

the exponential itself behaves linearly over a small range. It's a math thing. Once you get above $4U_T$ (about 100mV), you enter the saturation region, where current is nearly constant across any higher drain-source voltage for a given gate voltage. This means that operating in subthreshold saturation region is an easy way to get a constant current source if you can set your gate and source voltages and you are unsure about what your drain voltage will be (as long as it stays above $4U_T$, the drain voltage can move around without affecting the output current).

We could go through all the derivations for getting the drain-source current equations, but we'll leave that to the textbook. You start with current as a function of electron diffusion current density, channel width, and channel depth¹². The diffusion current density is a function of electron mobility and the concentration gradient across the channel. So we calculus-up our equations, and end up with a few constants prefixing a couple exponential terms raised to the power of various voltages and our old friend κ . It looks like this:

$$I = I_0 e^{(\kappa V_g - V_s)/U_T} - I_0 e^{(\kappa V_g - V_d)/U_T} \quad (6)$$

Now we see why it flattens out once V_{ds} goes above $4U_T$ - the second term becomes negligible and thus the output does not change for higher V_{ds} . The form for (6) comes from the fact that the total current (I) is the difference between the forward current (the first term) and the reverse current (the second term). A different form is

$$I = I_0 e^{(\kappa V_g - V_s)/U_T} (1 - e^{-V_{ds}/U_T}) \quad (7)$$

We've already said it, so we might as well come out with the equation for a transistor in subthreshold *saturation*. Remember, this applies when $V_{ds} > 4U_T$ (in saturation) and $V_{gs} < 0.7V$ (subthreshold).

$$I = I_0 e^{(\kappa V_g - V_s)/U_T} \quad (8)$$

or, since we often tie V_s to ground:

$$I = I_0 e^{\kappa V_g / U_T} \quad (9)$$

Spend some time looking at the plots in the textbook and the lecture slides - that will help you get a sense for what is different and what changes when you play with the inputs on a transistor¹³. *An important point: as long as they're in subthreshold, transistors always go between saturation and triode regimes at $V_{ds} = 4U_T$, regardless of V_{gs} . However, an above threshold transistor's transition V_{ds} to go between triode and saturation regimes does change for*

¹²Equation 3.2.5, p. 55

¹³Figs 3.6-3.8, pp. 58-60, and Fig 3.10, p. 63

different values of V_{gs} .¹⁴

The constant “ I_0 ” is just a bunch of other constants smushed together for convenience and because we’re rather embarrassed by them. No self-respecting physicist wants half a dozen consecutive constants sullying their equations, so you clean it up with a compound constant¹⁵. The only part that we need to remember for later is that it includes the transistor dimension ratio $\frac{W}{L}$ for channel width (W) and length (L).

At some point we will want to calculate κ . One easy way is, for an nFET, to tie the drain to V_{dd} , tie the source to ground through an ammeter (like the Keithley 236 SMU), and sweep the gate voltage from 0V to 1.2V or a similar range. Plotting the output current on a semilog plot (log of measured current on y-axis, gate voltage on x-axis) gives a curve where, in the linear regime, the slope is κ/U_T .

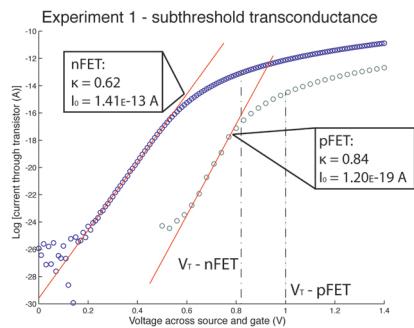


Figure 18: Reasonable-looking results from lab 2 - measuring κ and threshold voltage as described in class (slope of red lines is κ/U_T)

On conductance, Early voltage, and mismatch

Not all transistors behave exactly the same. In fact, it’s not easy to find two that do. We call this *device mismatch*. It’s primarily a function of the randomness involved in the doping procedure, which is a diffusion process. Because of this, we cannot ensure that any two transistors have the same response characteristics without testing them and finding two that are close enough. We simply cannot control the process that well. There is also uncertainty with dimension tolerances¹⁶.

¹⁴In all of these cases, we keep V_{gs} constant while sweeping V_{ds} , and then repeat for a different value of V_{gs}

¹⁵If you want the details: Equation 3.2.7, p. 56

¹⁶In any manufacturing process, a dimension will be produced to a tolerance of $\pm\epsilon$, so this inherent variation in physical dimensions is going to change the behavior of the transistor (I_0 is a function of width and length, remember?)

This becomes a problem if you design a circuit that is very sensitive to device response characteristics and relies on symmetry between certain devices within the circuit. There are ways to mediate or minimize the problem, but basically it manifests itself as deviation from an ideal behavior (lateral translation of output vs input plots, threshold variation, etc).

Another non-ideal behavior of transistors produces what we call the *Early voltage*¹⁷. We said earlier that a transistor does not increase its output current for higher *drain-source* voltages after passing the saturation voltage. This is not actually true, especially for short length MOSFETs. In the above threshold saturation regime, increasing drain voltage beyond the saturation point extends the pinchoff region¹⁸ further into the channel away from the drain, decreasing the effective length of the transistor and thus increasing its output current. To calculate it, you simply measure the current through a transistor while sweeping the drain voltage. Extrapolate a line from where the saturation region has flattened out and see where it hits the x-axis. This will be the negative Early voltage¹⁹. The slope of this line is also the *drain conductance* of the transistor at this point, represented as

$$g_{ds} = \frac{\partial I_{ds}}{\partial V_{ds}} = \frac{I}{V_e} \quad (10)$$

As transistor length increases, so does Early voltage. The ideal transistor would have an Early voltage of $V_e = \infty$ (horizontal slope, no x-intercept). More imperfections are discussed in the textbook section 3.5 (p. 75), but these are the most important two. It's good to be aware of the other effects (and you have to discuss them in one or two of the lab writeups), but they make fewer appearances throughout the course. The important thing is to understand how mismatch can affect your circuit and what kind of change in behavior you would expect to see for mismatch showing up in any given transistor.

While *conductance* is on our minds, let's spend another minute on it. We do know what it is in a general sense - it's R^{-1} , right? The inverse of the resistance? If $V = IR$, then $R = V/I$ and $g = I/V$. This is all just a very general statement from our old friend Ohm's Law with no transistor-specific properties. These relationships are a little more involved in transistors than in normal, linear resistors, so we take a more dynamic view and call the conductance the partial derivative of the drain current in terms of the voltage

¹⁷Named after a person, not a temporal relation

¹⁸See p. 67, "Saturation Region" section for a description of the *pinchoff region*. Really, all you need to remember about the Early voltage is 1) what it does to your output, 2) how to calculate it, and 3) that it's more pronounced in short transistors

¹⁹See p.76, Fig 3.16

of whichever terminal we're interested in. We tend not to move the source around too much, so mostly we talk about the drain conductance, as shown in (10), and the gate conductance, most often called the *transconductance*. We call it the transconductance because the current we are concerned with does not go through the terminal whose voltage we are changing²⁰.

$$g_m = \frac{\partial I}{\partial V_g} \quad (11)$$

To get conductance expressions, you simply differentiate your relevant current equation, like (8), in terms of the relevant voltage.

That pretty well covers the basics of subthreshold behavior, though probably not everything. I didn't plan this thing out all that well. There's also some stuff about κ not actually being constant, but you'll see that in the labs.

²⁰Remember, it's the gate, and no current goes in or out of the gate

4 Above threshold behavior

Above threshold we have a different phenomenon. We established that sub-threshold current moves around because of diffusion - moving across concentration gradients. Once we get above threshold, the transistor goes into *strong inversion* and the electric field from drain to source takes over and *drift* is the primary motive force. Once you get the basics of chapters 2 and 3, the rest of the course is extrapolating and applying it. So really focus on learning the material in these two chapters - you will be expected to know it well and be able to use that knowledge on the fly during lab sessions.

On current equations, κ , and conductances

First off, κ . Once the transistor goes into strong inversion²¹, surface potential ψ_s is pretty well coupled with gate potential, so we set κ to 1²². We will no longer have κ in our equations, but another little guy - β - gets introduced as a modified version of I_0 . We are now above threshold, but we still have the ohmic/linear/triode regime and the saturation regime.

So current is now tumbling down the electric field between the source and drain, being pushed along by our old friend the electric field. This means the current will flow according to the rules that govern charges moving in a field, with the slight caveat that it happens inside matter, not in the ubiquitous vacuum we studied in undergrad. So what are the main points of these rules? Well, we have I_{ds} as a product of carrier mobility, charge density, and electric field. When we expand the terms we get

$$I_{ds} = \mu C_{ox} \frac{W}{L} (V_{gs} - V_T) V_{ds} = \beta (V_{gs} - V_T) V_{ds} \quad (12)$$

where V_T is the threshold voltage and μ is electron mobility (previously hidden inside I_0). And look at that! I_{ds} has a linear dependence on both V_{gs} and V_{ds} . No wonder they call it the *linear* regime. So as long as you hold one of those two voltages constant and sweep the other, you can easily measure β as the slope of the line divided by the voltage that is being held constant (either drain or gate). If you recall, when a transistor is subthreshold, it goes from triode to saturation regime at the same drain voltage ($4U_T$), regardless of gate voltage. Above threshold we don't get that. That saturation V_{ds} now depends on V_{gs} - for higher V_{gs} you get a higher saturation voltage V_{dsat} ²³.

²¹This is where the gate voltage is high enough to attract minority carriers so their concentration in the channel is higher than the dopant concentration

²²A little more technically: any change in ψ_s is due to a change in inversion charge (mobile electron concentration), not depletion charge (fixed ions), and κ is limited to describing static charges, being based on capacitances and all

²³ $V_{dsat} = V_{gs} - V_T$

Say we put V_{ds} higher than V_{dsat} , so we're in the saturation regime of above threshold. The current equation changes to

$$I_{ds} = \frac{\beta}{2}(V_{gs} - V_T)^2 \quad (13)$$

with the same β as before. Remember this, you will likely get asked this at some point: *subthreshold current is an exponential function of voltage, but above threshold it's linear in the linear region and quadratic in the saturation regime.* Also, try to remember which voltages it's linear or quadratic or exponential with. Here if we plot the square root of the current against V_{gs} , the slope of the line is a nice, friendly $\sqrt{\beta/2}$. All of this is shown pretty clearly in the slides.

So, κ has not completely disappeared. I know, I know, we said it's no longer relevant to the surface potential. That is true. Now, however, it has a hand in controlling the threshold voltage, along with V_s as $V_T = V_{T0} + \frac{V_s}{\kappa}$. The class does not really spend much time on that, though.

Another word on current dependencies

The full above threshold equation, which contains all the information we need to derive both (12) and (13) is

$$I = \frac{\beta}{2} [((V_g - V_{T0}) - V_s)^2 - ((V_g - V_{T0}) - V_d)^2] \quad (14)$$

which we can look at as the difference between forward (the half with V_s) and reverse (the half with V_d) currents. We have to do some algebra to get this in the form of (12), but essentially the current is product of inversion charge in the channel (which is linear in $V_g - V_{T0}$) and the electric field across the channel (which is linear in V_{ds}).

In saturation, V_d is high enough that any electron that gets near the drain is instantly sucked up into it, so that end of the channel actual goes sub-threshold²⁴ because the concentration of mobile carriers is so low (they all get pulled into the drain by the electric field arising from the voltage between the channel and the drain, where they are whisked away to V_{dd} and the greener pastures that await them there).

Overall, current is a proportional to inversion charge concentration and the electric field²⁵. Integrating this relationship over the length of the channel

²⁴This is the *pinchoff* region

²⁵ $I = J_{drift} Width_{channel} Depth_{channel} = \mu W Q_i \mathcal{E} \propto Q_i \frac{d}{dz} Q_i = \frac{d}{dz} Q_i^2$

gives $I \propto Q_i^2 = (Q_s^2 - Q_d^2)^{26}$. Saturation says $Q_d \rightarrow 0^{27}$, so current is only a function of Q_s and we drop off the second half of (14) to get (13).

I found it very helpful to sit down with the equations and trace dependencies back to their roots²⁸. For example, you look at (14) to see how current depends on voltages. You then look in the book and find that the voltages are functions of inversion charge. You look farther back to see what inversion charges are functions of, and so on. Without knowing what phenomena give rise to a given term, it's hard to build an intuition for any of this. But maybe that's just me.

I'll only mention conductances here to say that we calculate them the same way as before - differentiate your current in terms of whatever your relevant voltage is. And there you have it. Study your labs.

²⁶Inversion charge concentrations at source and drain ends of channel

²⁷Remember? all electrons that make it to that end immediately bugger off through the drain

²⁸Letting you do this is where the book really earns its keep

5 Logarithmic I/V converter

We can call this a circuit, but this is simply what a transistor does in sub-threshold - converts the log of the current to a voltage²⁹. We can think of the transistor as having three variable parameters: V_g , V_s , and I_{ds} ³⁰. The circuit is just a single transistor where current into the drain is our input and the output is either the gate or source voltage. The equations are simply algebraic manipulations of the subthreshold saturation equation (8).

$$V_s = \kappa V_g - U_T \log \left(\frac{I}{I_0} \right) \quad (15)$$

$$V_g = \frac{1}{\kappa} \left(V_s + U_T \log \left(\frac{I}{I_0} \right) \right) \quad (16)$$

In each case, you fix the voltage that is not being measured at a constant value. *Important note:* due to the infinite impedance between gate and channel, V_g (determined - by definition - by charge on the gate) cannot be directly influenced by the input current. For this version to work, there must be feedback between the gate and the drain, which can be accomplished by shorting the two together (see Figure 19). This is a *diode-connected* transistor, called such because it functionally has only two terminals, like a diode.

Doing this has two effects. First, it ensures that the transistor stays in the saturation regime by maintaining a reverse bias between drain and channel. Second, if we force a certain current through, this will set the drain voltage and then necessarily the gate voltage (because the two are shorted together). See? Positive feedback from drain to gate (and therefore negative feedback from gate to drain).

²⁹Or vice versa - that's a current source

³⁰We ignore V_d for now

6 Current source

This is another iteration of “we call this a circuit but really it is simply what transistors do.” In this case, V_s , V_d , and V_g are set to constant values and I_{ds} is our output. In practice, we tie V_s to ground, put V_g at some value less than 0.7V (to keep it subthreshold), and V_d is put above 100mv to keep it in saturation. If we use a long transistor we can ignore Early voltage effects.

Technically this is a current *sink*, not a current *source* because the current is pulled into the drain from above³¹. To have a current source that will go higher in the circuit than whatever is consuming your current, you simply use a pFET instead of an nFET, tie the source to V_{dd} , and set gate voltage such that $V_g > V_{dd} - 0.7$. In either case the equation to calculate output current is by definition the subthreshold saturation current $I_{ds} = I_0 \exp((\kappa V_g - V_s)/U_T)$.

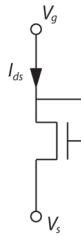


Figure 19: Diode-connected nFET

³¹“Above” refers to “closer to V_{dd} ”

7 Current Mirror

The current mirror circuit is built by connecting the gate of a diode-connected transistor to the gate of another transistor. Both transistors have a fixed source voltage, are in saturation, and act as a current source.

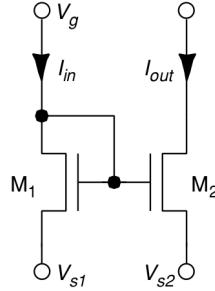


Figure 20: nFET Current Mirror [1]

- M1: Diode connected transistor
- Input current I_{in} sets common gate voltage V_g , hence also sets I_{out}

Scaling of I_{in} vs I_{out}

- By choosing different source potentials V_{s1} and V_{s2}
Leads to exponential scaling: $I_{out} = e^{V_{s1}-V_{s2}/U_T} \cdot I_{in}$
- By choosing different transistor sizes
The current scales linearly with transistor size difference
Important: Fixed transistor width vs. length ratio

Above Threshold Operation

Input vs. output currents depends mainly on drain voltage of M_1 and M_2 , because of larger Early effect. The dependence on the source voltage is much weaker.

With bidirectional input, the current mirror can be used as half wave rectifier, because the input transistor has high impedance in the opposite direction.

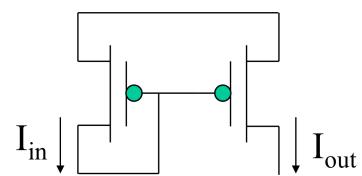


Figure 21: pFET Current Mirror [8]

8 Source Follower

The source follower

- Transforms a weakly-driven voltage signal into a more strongly-driven voltage signal
- Is a two transistor circuit consisting of two MOSFET in series
- Consists of a fixed current source that is connected to the source of another MOSFET in saturation
- Linearly transforms a voltage V_{in} at a high impedance input terminal M_1 into a voltage at a lower impedance output terminal V_{out}

M_1 source is equal to V_{out}

V_{out} is able to drive larger loads

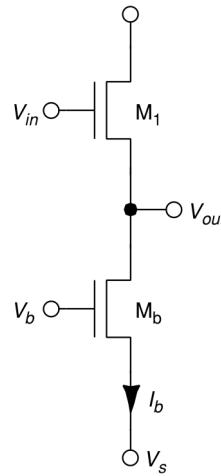


Figure 22: Source Follower Circuit [1]

Proper Operation Requirements

Saturation of the biasing transistor: $V_{out} > V_s + 4U_T$
that is $V_{in} > K_n^{-1}(\kappa_b V_b + 4U_T)$

Subthreshold Operation

In subthreshold, the output voltage changes according to:

$$V_{out} = \kappa_n V_{in} - U_T \log\left(\frac{I_b}{I_{n0}}\right) = \kappa_n V_{in} - \kappa_b V_b + V_s \quad (17)$$

κ_b is the subthreshold slope factor of M_b

V_{out} is linearly related to V_{in}

$$1 > \kappa_n > 0$$

V_{out} has a fixed offset that depends on bias current I_b

Gain

V_{out} follows V_{in} with gain of K_n

Important: K_n is not constant due to the body effect of M_n

9 Differential Pair

The Differential Pair circuit, as shown in Figure 23, is composed of two source followers with a common Fixed Current Source, M_b , with the bias voltage, V_b . The transistors, M_1 and M_2 , have variable voltage inputs, V_1 and V_2 , respectively, and share a common source, node V, which also acts as the drain of M_b .

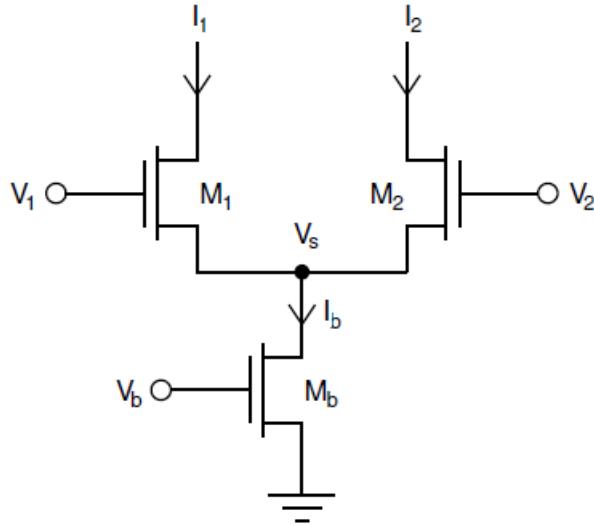


Figure 23: Differential Pair Circuit

The output current is proportional to the difference in the input voltages according to:

$$I_1 = I_b \frac{e^{\kappa V_1}}{e^{\kappa V_1} + e^{\kappa V_2}} = \frac{I_b}{1 + e^{\kappa(V_2 - V_1)}} \geq 0 \quad (18)$$

$$I_2 = \frac{I_b}{1 + e^{\kappa(V_1 - V_2)}} \geq 0 \quad (19)$$

Before the currents I_1 and I_2 are input, the Differential pair is off because the drain of M_b (node V) is off. Once I_1 and I_2 are on and in saturation, they will charge node V to turn on M_b and put I_b into saturation. The transistor with the lower input voltage (V_1 or V_2) will act as a drain choke and allow less current through its drain. The losing transistor will see its source voltage (node V) increase and thus fall out of saturation.

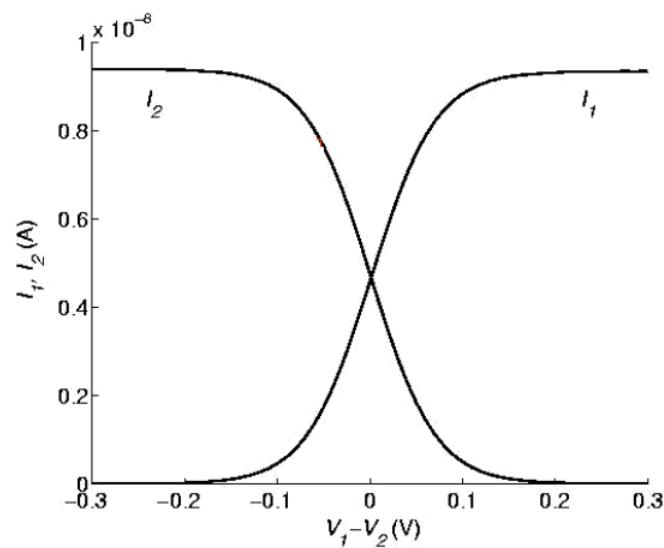


Figure 24: I-V characteristics of Differential pair

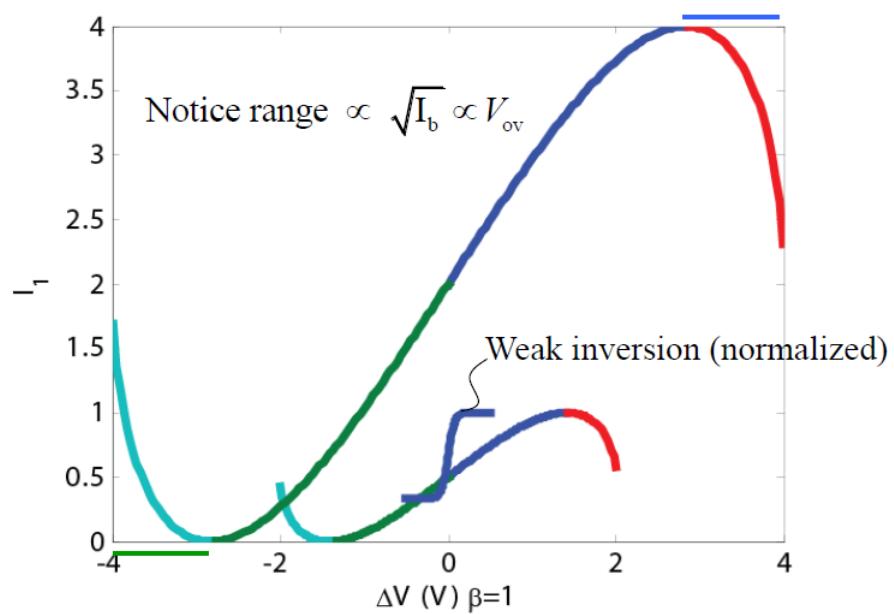


Figure 25: I-V characteristics of Differential pair with reference to input voltage difference in weak & strong inversion

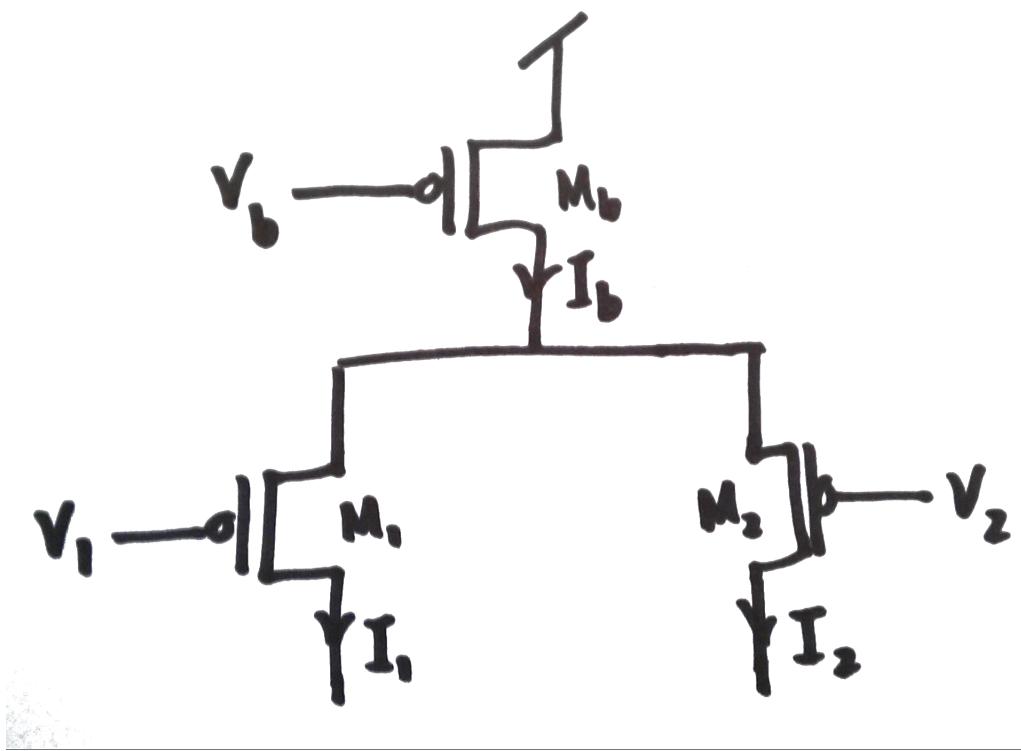


Figure 26: pFET Differential pair → flipped w bias on top connected to Vdd, and M1, M2 sinking current

10 Current Conveyor

The Current Conveyor, also commonly known as a Buffered Current Mirror, consists of two transistors. where the M_1 gate is tied to the M_2 drain, and the M_2 gate is tied to the M_1 source as shown in Figure 27.

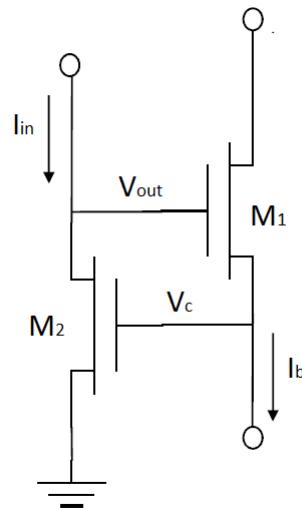


Figure 27: Current Conveyor circuit

The Current Conveyor acts like a source follower (M_1) whose output terminal is tied to the gate of a current sink (M_2). The input current, I_{in} , will charge V_{out} to turn on M_1 whose current will charge V_c , turning on M_2 until M_2 sinks I_{in} . The current being sunk by M_2 follows V_c according to:

$$I = e^{\kappa * V_c} \geq 0 \quad (20)$$

11 Current Correlator

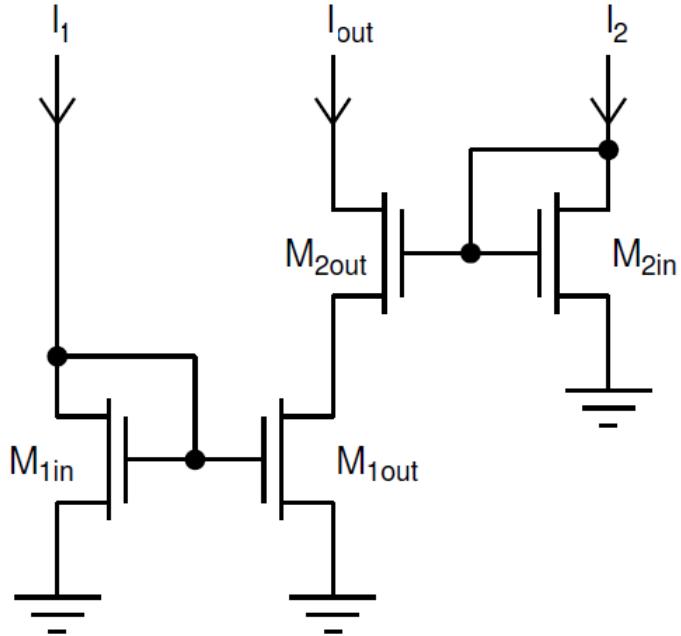


Figure 28: Current correlator circuit

The Current Correlator is comprised of two stacked current mirrors that have the output of one mirror (M_2) lead into the drain of the other mirror (M_1). Currents, I_1 and I_2 , are sunk through the diode connected half of both mirrors, pulling V_1 and V_2 , respectively, so that they follow:

$$I_{out} = I_o e^{\kappa V_1} \cdot (1 - e^V) \geq 0 \quad (21)$$

If $V_1 > V_2$, M_1 will sink I_1 and pull up the common node voltage V (of the M_1 drain voltage and M_2 source voltage), reducing the current through M_2 . Thus I_{out} settles at a self normalized product between I_1 and I_2 .

$$I_{out} = \frac{I_1 \cdot I_2}{I_1 + I_2} \geq 0 \quad (22)$$

However, if V_1 is off, I_2 will charge common node V until M_2 shuts off once $V > V_{dM2} + 4U_t$. Likewise, if M_2 is off, the drain of M_1 will be off.

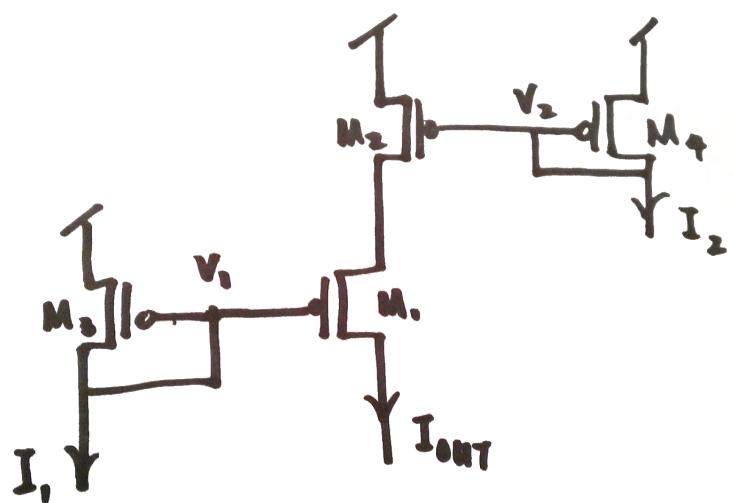


Figure 29: pFET Current correlator → two stacked pFET mirrors with 3 current sinks

12 Bump Circuit

The Bump-Antibump circuit is comprised of a Differential Pair with the gates of the input transistors tied to the gates of two stacked transistors.

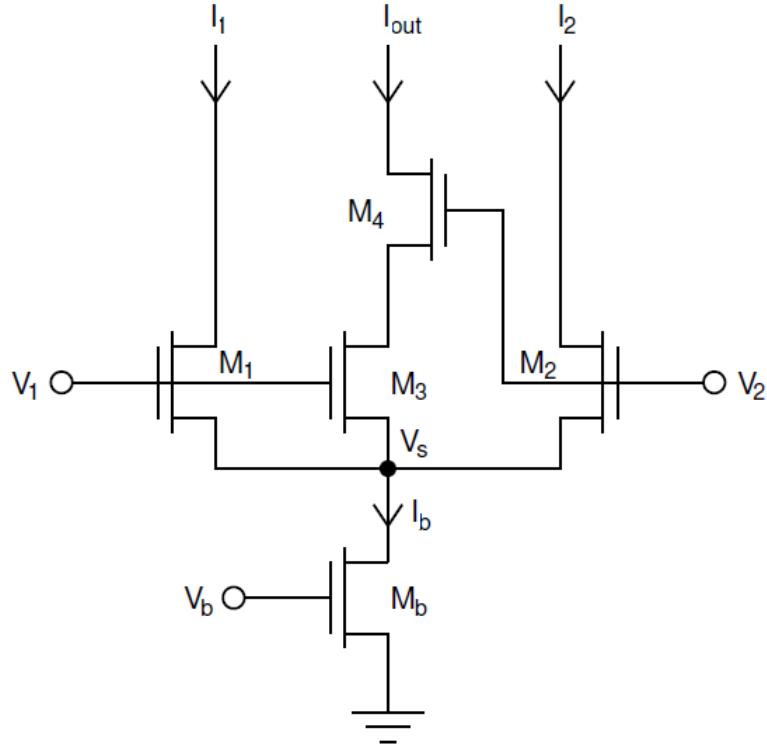


Figure 30: Bump circuit

If $V_1 \gg V_2$ or $V_2 \gg V_1$, then the circuit behaves as a Differential Pair with I_{out} as a function of $V_2 - V_1$.

$$I_{out} = \frac{I_b}{1 + \frac{4}{S} \cosh^2 \frac{\kappa d V}{2 U_T}} \geq 0 \quad (23)$$

where the transistor geometry factor,

$$S = \frac{(W/L)_{middle}}{(W/L)_{outer}} \quad (24)$$

If the difference between V_1 and V_2 is small, the stacked transistors will have similarly biased gates, and I_{out} will reflect the sum of the two currents according to:

$$I_1 + I_2 = I_b + I_{out} = \frac{I_B}{1 + \frac{S}{4} \cosh^{-2} \frac{\kappa dV}{2U_T}} \geq 0 \quad (25)$$

If the limit of $I_1 + I_2$ is taken as $\delta V \rightarrow 0$, the sum reaches an absolute minimum at

$$I_1 + I_{2\min} = \frac{I_b}{1 + \frac{S}{4}} \geq 0 \quad (26)$$

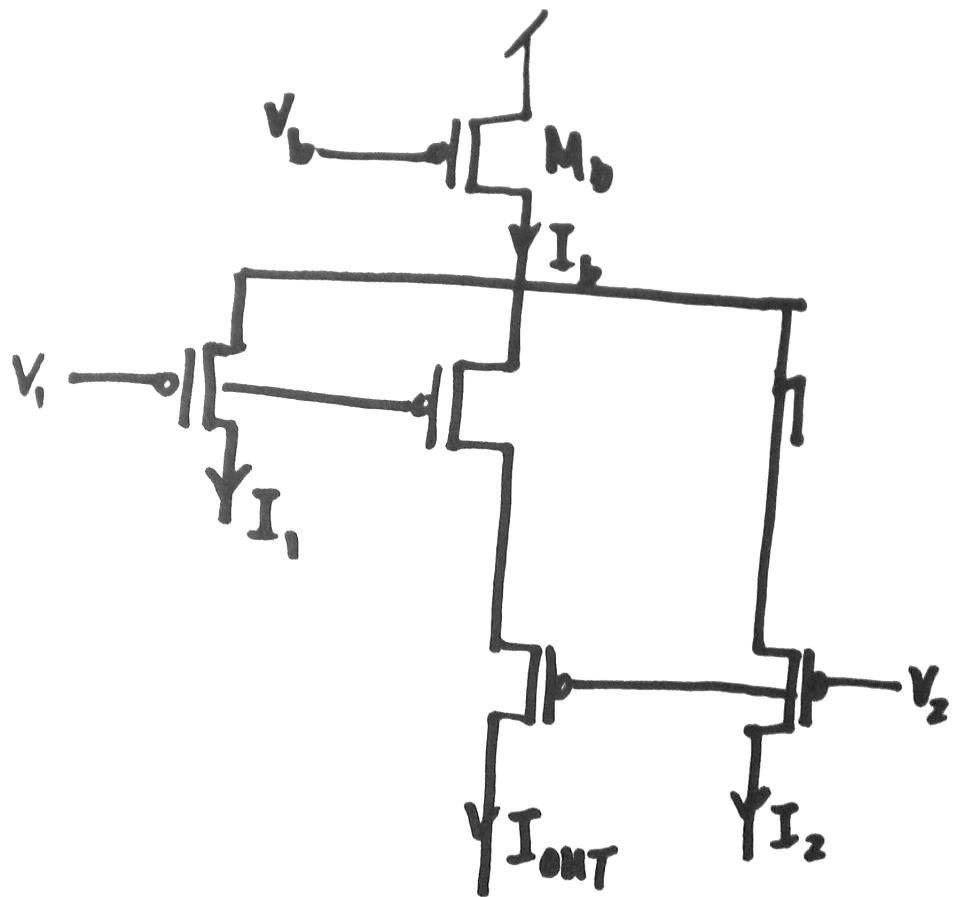


Figure 31: pFET Bump circuit

13 Transconductance Amplifier

The simple transconductance amplifier is very easy to understand if you know how the differential pair and the current mirror works, so check back if you don't.

The bottom part of the transconductance amplifier (Fig. 32a) is a differential pair with a current mirror from one of the two outputs to the other one. What now happens is that the current I_1 get mirrored on top of I_2 . We assume that V_1 and V_2 are not equal, and so aren't I_1 and I_2 . The difference between the two currents then flows in or out of the amplifier in a way that Kirchhoff's current law is fulfilled:

$$I_{out} = I_1 - I_2 = I_b \tanh\left(\kappa \frac{V_1 - V_2}{2U_T}\right)$$

Or for a small differential input ($|V_1 - V_2| < 200mV$):

$$I_{out} \approx g_m (V_1 - V_2) \text{ with } g_m \approx \frac{\kappa I_b}{2U_T}$$

Since amplifiers are usually used in voltage and not in current mode, we focus on its behavior when the output current I_{out} is fixed. Let's set $V_1 = V_2$ such that I_{out} is zero and keep it there. If we now increase for example V_2 , the corresponding transistor would drain more current, which can't be provided by the upper transistor since it only mirrors I_1 . In order to fulfill

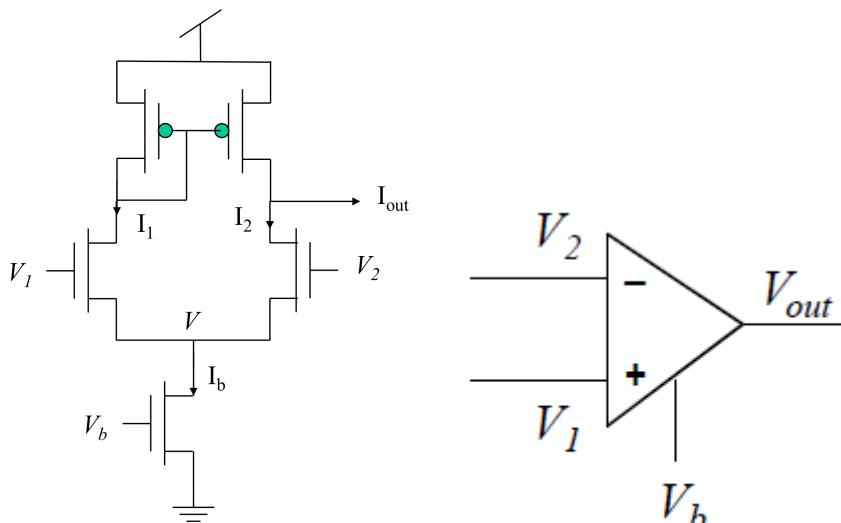


Figure 32: A simple transconductance amplifier.

all transistor equations and Kirchhoff's current law, V_{out} decreases such that the lower transistor is no longer in saturation. Since the transistors behave exponential, even a small change of the input voltages leads to a big change in the output voltage.

For this amplifying behavior, it is very important that all transistors operate in saturation region. We already know the common node voltage is $V = \kappa(\max(V_1, V_2) - V_b)$ if $|V_1 - V_2| > 4U_T$. Therefore, the saturation condition for the bias transistor is $\max(V_1, V_2) > V_b + \frac{4U_T}{\kappa}$. The saturation conditions for the upper and bottom transistors at the output are $V_{dd} - V_{out} > 4U_T$ and $V_{out} - V_s > 4U_T$. We see that V_{out} is very restricted (especially towards ground) at the output with

$$\kappa(\max(V_1, V_2) - V_b) + 4U_T < V_{out} < V_{dd} - 4U_T$$

Outside of this range, the amplifier no longer behaves like an amplifier. The output voltage simply won't change anymore.

Usually, we use amplifiers as depicted in Fig. 32b, where

$$V_{out} = A(V_1 - V_2)$$

and we call $A = \frac{g_m}{g_d}$ the amplifiers gain. In subthreshold operation the gain is $A \approx \frac{\kappa V_E}{2U_T}$ and above threshold it is $A \approx \sqrt{\frac{\beta}{I_b}} V_E$.

14 Wide Range Transconductance Amplifier

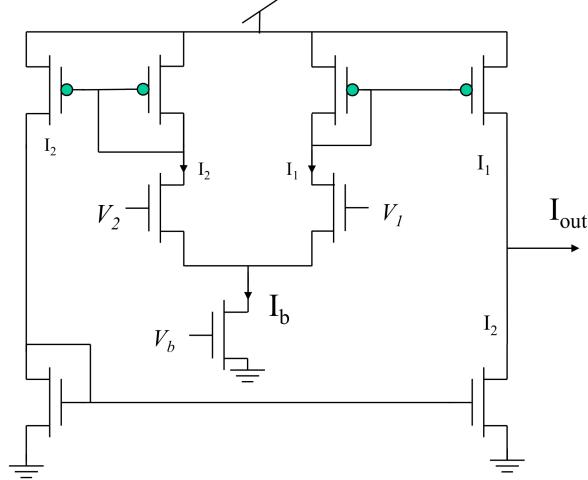


Figure 33: Wide Range Transconductance Amplifier [8]

A wide-output range transconductance amplifier behaves very similar to a simple transconductance amplifier. Even though the circuit looks very complicated, it only consists of a differential pair and three current mirrors. The goal of these mirrors is to remove the restrictions on the output voltage of the amplifier.

Each branch of the differential pair is connected to a current mirror, but the currents I_1 and I_2 are still separated. After that, the same thing is done as in the simple transconductance amplifier, where one current (in Fig. 33 this would be I_2) is mirrored to the same node as the other one, and this node is connected to the output of the amplifier. The differential pair is now no longer connected to the output, which removes its restrictions from the allowed output condition. The new condition is $4U_T < V_{out} < V_{dd} - 4U_T$.

15 Current Divider

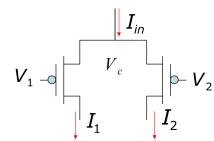


Figure 34: Current Divider [9]

16 Winner-take-all circuit (WTA)

How does the WTA circuit work?

The circuit of Fig. 35 is a continuous time, analog circuit that implements a WTA network, designed by Lazzaro et al. (1989).

It processes all the (continuous-time) input signals in parallel, using only two transistors per input cell, and one global transistor that is common to all cells. Collective computation and global connectivity is obtained using one single node common to all cells.

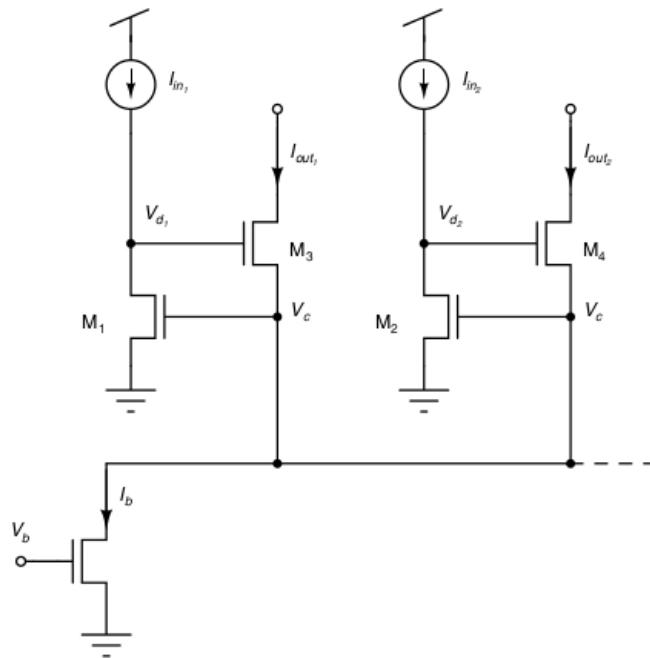


Figure 35: Two cells of a current mode WTA circuit [1].

The WTA network is modular and can be extended to n cells.

Each cell comprises a current-controlled conveyor as in figure 36. The inputs of the circuit are applied currents, output signals are encoded both by I_{out} currents, and the V_d voltages.

The application – current coping in which the current-conveyor circuit is used in WTA is showed in the figure 37. Before to copy a current we used the current-mirror. It is very efficient – as we need 1 transistor less. However, as the output of the circuit is connected in feed-back loop to the input, the sum of parasitic capacitances C_{gs} will significantly slow down the output. In the most drastic case the outputs will even not detect any change in input

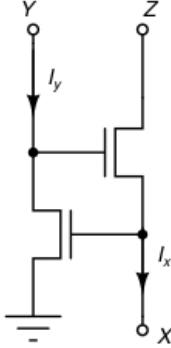


Figure 36: Current-controlled conveyor [1].

current signal – for high frequency time variant inputs (slew-rates limits output respond, higher time constant $\tau = RC$).

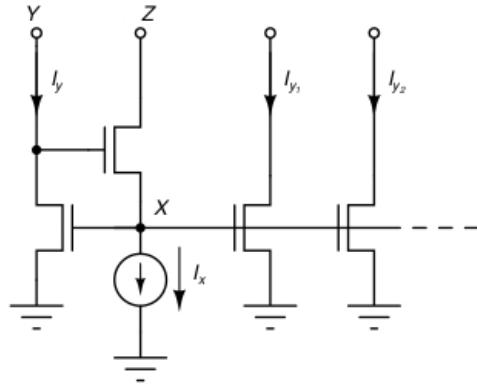


Figure 37: System application of current conveyor for the generation of multiple copies of an input current [1].

Transistors M_1 and M_2 discharge nodes V_d – implement inhibitory feedback. Transistors M_3 and M_4 implement an excitatory feed-forward path by charging node V_c .

The circuit selects the largest input current because cell provides j provides $I_{out}^j = I_b$ and so suppresses all other output voltages and currents (all other $V_d \approx 0$ and $I_{out} \approx 0$). Cell j wins the competition because its voltage V_d^j determines V_c (V_{gs} is set by fixed value of I_b). Since all cells have common V_c gate potentials it will be greater than V_d for all other cells than j – they will be shut-off [1].

Briefly: the circuit chooses one winner by selecting a cell with the highest input current (max() function). All other cells are suppressed (inhibited).

Hysteretic WTA: by feeding back the output of the circuit to its input,

the system gets more stable and less sensitive to probable new winner.

Can you reason through its behaviour?

We consider three cases (the network with two cells):

- I. inputs are equal;
 - II. one input much larger than the other;
 - III. two inputs differ by a very small amount (small-signal regime).
- I. Because gates of M_1 and M_2 are tied to the same common node V_c , the drain voltages of M_1 and M_2 must take the same value. As a result, the output transistors M_3 and M_4 will have equal V_{gs} value and in case both of them are in saturation, the output currents have to be identical. Upon Kirchhoff's current law:

$$I_{out_1} = I_{out_2} = \frac{I_b}{2} \quad (27)$$

- I. We can consider the case in which $I_{in_1} \gg I_{in_2}$. In this case, the drain voltage of M_1 (V_{d_1}) will be greater than the drain voltage of M_2 (V_{d_2}).

Because the two transistors M_1 and M_2 have a common gate voltage V_c , and both their sources are tied to ground, the forward current (dependent only on V_{gs}) for both them is equal. The need to suppress the value of forward current through M_2 arises. Therefore voltage V_{d_2} has to be decreased (lower current for the same gate-to-source voltage). We say that reverse current – dependent only on V_{ds} – will increase, limiting overall current through drain to I_{in_2} value.

Decrease of V_{d_2} causing M_2 to operate in its ohmic region (out of saturation), switches off M_4 . This implies $I_{out_2} = 0$.

Consequently, output transistor M_3 sources all the bias current

$$I_{out_1} = I_b, \quad (28)$$

with V_{d_1} satisfying the equation

$$I_0 e^{\kappa V_{d_1} - V_c} = I_b. \quad (29)$$

- I. To analyze the circuit in this regime, we will use small-signal analysis. We must consider the Early effect of the transistor operating in the saturation region (Eq. 30):

$$I_{ds} = I_{sat} \left(1 + \frac{V_{ds}}{V_e}\right) \quad (30)$$

, V_e is the Early voltage. By derivating this equation (small-signal analysis) we get: $\delta I_{ds} = I_{sat} \frac{\delta V_{ds}}{V_e}$.

Assume that the two input currents I_{in_1} and I_{in_2} are initially equal.

If we now increase the input current I_{in_1} by a small amount δI we would like to be able to obtain a difference in M_1 transistor's V_{d_1} , which is:

$$\delta V_d = \delta I_d \frac{V_e}{I_{sat}} \quad (31)$$

As V_{d_1} is also the gate voltage of transistor M_3 , the I_{out_1} will be amplified by an amount proportional to $e^{\delta V}$. The Kirchhoff's law requires that I_{out_2} decreases by the same amount. This reduction means the gate voltage V_{d_2} of M_2 must decrease by δV [1].

All three cases can be summarised by the graphic 38.

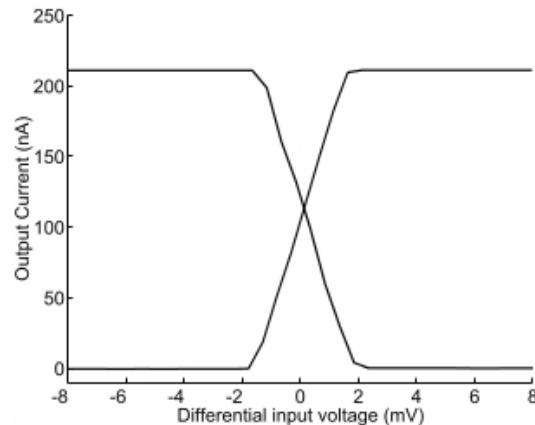


Figure 38: Responses of the two-cell WTA circuit. Current output (I_{out_1} and I_{out_2}). The bias voltage $V_b = 0.7V$ [1].

How does the bias current affect its performance?

With increasing bias current it will get harder to choose a winner – it takes larger input current difference to suppress all cells excluding the winner.

???

How can you adjust the gain of the circuit by layout of the transistors?

Since the gain in the small-signal regime of the competition of the WTA circuit is expressed by equation:

$$\frac{\delta V}{\delta I} = \frac{V_e}{I_{sat}} \quad (32)$$

the only transistors' layout variables are early Voltage V_e (dependent on the length of transistor) and I_0 coefficient of the saturation current (linearly dependent on W/L). Therefore, we can increase the gain by decreasing the width of the transistor or by increasing its length.

???

17 Photodiodes, photoreceptor

How does the I-V curve of a diode change in the presence of light?

The current-voltage characteristic of a photodiode has the same shape as that of a normal diode, but the curve is displaced along the current axis by the value of the photocurrent (see fig. 39).

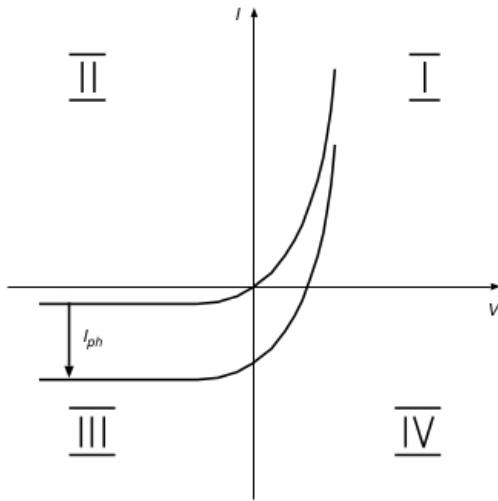


Figure 39: Steady-state current-voltage characteristics of a photodiode. The upper curve is the normal diode characteristic (dark characteristic). The lower curve shows the characteristic under illumination [1].

In the presence of an applied external bias the (negative) photocurrent is superimposed onto the diode current.

How does phototransduction occur in silicon?

Transduction means transformation of the energy from one form to another. Phototransduction means transforming from photon to electronic signals [5].

In a semiconductor, an incident photon and therefore its energy can be absorbed by an electron. A photon with an energy larger than or approximately equal to the bandgap energy can excite an electron from the valence band into the conduction band (corresponds to the generation of an electron-hole pair). Illumination of a semiconductor therefore increases the concentration of mobile charge carriers (electrons).

If the motion of the carriers is driven by diffusion, then the generation is balanced by recombination (creation of pairs).

However, if electron-hole pairs are generated in depletion region and subjected to its built-in electric field, pairs are likely to be separated. Some of the separated carriers contribute to an electrical output signal (reverse photo

current). The separated mobile charges are "swept home" as the minority charges dominate in depletion region.

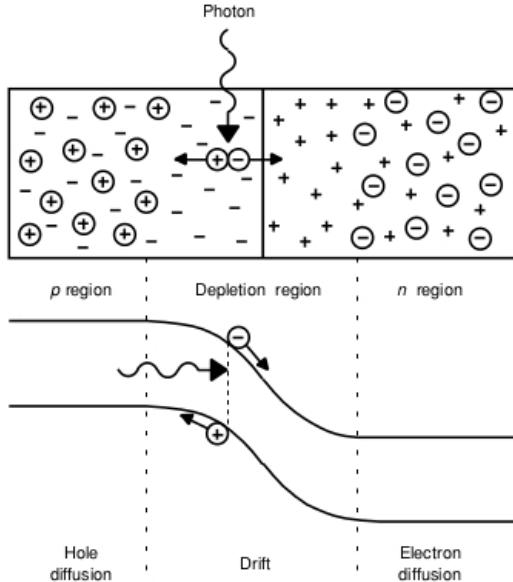


Figure 40: Principle of operation of a photodiode. Electron-hole pairs generated by incident photons in or within a diffusion length outside the depletion region become separated and contribute to a reverse generation current [1].

If the photodiode is open-circuited, that is, no external current is allowed to flow, then generated charge accumulates at the boundaries of the depletion region until a steady state is attained. In steady state, a forward diffusion current in the junction compensates for the photo-current [10].

If the photodiode terminals are short-circuited the photocurrent can be measured as a reverse diode current. In the presence of an applied external bias the photocurrent is superimposed onto the diode current [1].

How you can use adaptation in a feedback loop to cancel out circuit mismatch?

In the simple source-follower logarithmic photoreceptor circuit (see fig. 41) in real world application there are two important problems: mismatches and slow response (because the huge capacitance have to be charged and discharged – photodiode junction C and the parasitic C_{ds}, C_{gs} , by a very small photocurrent).

The differences between supposedly identical receptor outputs are as large as the typical signals variations produced by real scenes – circuit is

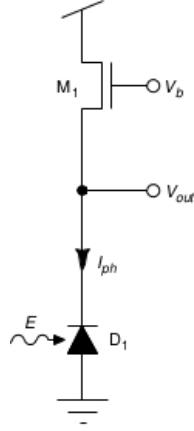


Figure 41: Photosensors with logarithmic irradiance-to-voltage conversion consisting of a photodiode and a MOSFET in source-follower configuration [1].

unusable. Hence the necessity for adaptation in a feedback loop, to deal with the circuit mismatch problem (see fig. 42).

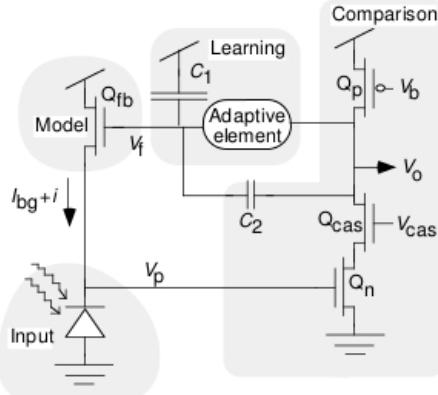


Figure 42: Adaptive receptor circuit [10].

Briefly: The feedback amplifier and the input fight to control the source voltage of Q_{fb} , but the feedback amplifier wins because it has much higher gain. The input voltage v_p moves enough so that the output voltage v_o moves enough so that v_f moves enough so that v_p is held nearly clamped.

The photocurrents are proportional to the areas of the corresponding photodiode junctions, whose spatial variations are the main source of photocurrent mismatches of identically designed photosensing elements resulting in *fixed-pattern noise*.

The FPN can be reduced by individual tuning of each photosensor. A more economic solution is the enhancement of transient signals with respect to the mismatched steady-state signals using a well-matched amplifier stage (capacitive-devider). Adaptation to the DC value can be provided by a resistive element [1].

Adaptive element (see fig. 43) has current-voltage relationship as shown in the fig. 44.

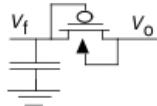


Figure 43: Expansive adaptive element with the capacitor that stores the adaptation state [10].

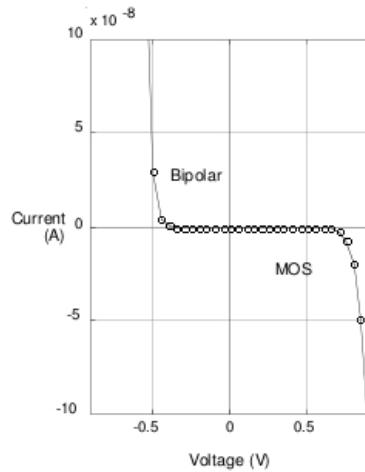


Figure 44: Measured current–voltage relationship for the adaptive element. The voltage scale changes logarithmically with the current scale [10].

Due to the use of an adaptive element with an expansive nonlinearity there is different respond to input:

- Large changes (high frequencies, huge changes in illumination, ambient lighting, background change – from shadow into sunlight) adapt rapidly. Low gain for static signals (including circuit mismatches) – the feedback is a short circuit across the adaptive element.
- Small signals around an adaptation point (small contrast variation) adapt slowly. The transient gain of the receptor is high, set by the capacitive-divider ratio as no charge flows through the adaptive ele-

ment, changes in v_o are coupled to v_f through the capacitive divider [10].

The amplitude of the response to the small contrast variation is almost invariant to the absolute intensity (as in retina in eye), owing to the logarithmic response property.

How you can build a fast logarithmic current-sense amplifier, by using feedback to make a virtual ground?

To make a virtual ground – to clamp voltage V_s to constant level.

V_s node is output of the source-follower log receptor with high parasitic capacitance – requiring a lot of time to charge or discharge it – circuit is not responding to high frequencies. By decoupling output node from the V_s node we get higher bandwidth (grater passband of frequencies). Hence the necessity for active feedback, to deal with the problem of slow response [10]. For the circuit schematic see fig. 45

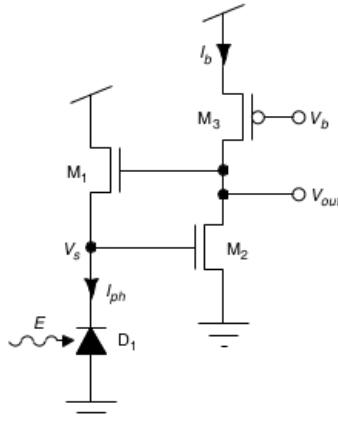


Figure 45: Logarithmic photosensor with feedback loop increasing the bandwidth by clamping the voltage V_s [10].

The voltage output signal then appears at the gate of the MOSFET M_1 , while the source and therefore the voltage across the photodiode is practically clamped. The source node drives V_s a two-transistor inverting amplifier with a pFET current source clamping V_s [1].

How you can use a capacitive divider in the feedback loop of an amplifier to set gain?

The mismatches can be reduced by a well-matched transient gain stage fabricated with a capacitive divider.

On short time scales, no charge flows through the adaptive element, but changes in v_o are coupled to v_f through the capacitive divider (see fig. 42). The transient gain of the receptor is thus set by the capacitive-divider ratio.

$$A_c = \frac{C_1 + C_2}{C_2}, \quad \delta V_{out} = A_c A \delta V_s \quad (33)$$

The larger C_1 is relative to C_2 , the larger the gain of the circuit [1].

How you can use a cascode configuration to increase effective drain resistance?

If we take a cascode with 2 nFET transistors and name the upper one M_c and the lower one M_2 and take variables g_{dc} , g_{sc} – drain and source, output, input transconductance of M_c , g_{d2} – drain, output transconductance of M_2 and g_0 – overall conductance of the cascode, V_x – common node between 2 transistors we can write:

$$g_0 = \frac{\delta}{V_0}, \quad (34)$$

$$\delta = g_0 V_0 = g_{d2} V_0 - g_{sc} V_x, \quad (35)$$

$$(g_{d2} + g_{sc}) V_x = g_{dc} V_0. \quad (36)$$

Excluding unknown V_x from both equation we get:

$$\delta = \frac{g_{d2} g_{dc}}{g_{d2} + g_{sc}} V_0 \approx \frac{g_{d2} g_{dc}}{g_{d2} + g_{sc}} V_0. \quad (37)$$

$$g_0 = g_{d2} \frac{g_{dc}}{g_{sc}} \approx g_{d2} \frac{U_T}{V_e}. \quad (38)$$

And finally we can derive the expression for the change resistance after connecting additional transistor in cascode configuration:

$$r_0 = r_{d2} \frac{V_e}{U_T}. \quad (39)$$

As we can see, the resistance will be increased by approximately $\frac{V_e}{U_T} \gg 1$.

What is the Miller effect?

The Miller effect occurs when a capacitor feeds back the output of an inverting, high gain amplifier back to the input. If the input needs to move a certain amount, it must charge not only its own side of the capacitor, but also the other side of the capacitor which moves A times as much in the opposite direction. Hence a small capacitance C looks like a capacitance $(A + 1)C$ to the input. Since $A \gg 1$, we usually ignore the 1. The Miller capacitors C_n (see fig. 42) from the gate to the drain of Q_n has a substantial effect on the time-response of the receptor. The Miller effect increases the relevant gate-drain and gate-source capacitances by a factor of A [10].

How can a cascode be used to reduce it?

The parasitic capacitance C_p from the source of M_1 onto the output node via the gate-to-drain capacitance of nFET M_2 gives rise to the Miller effect (see fig. 46).

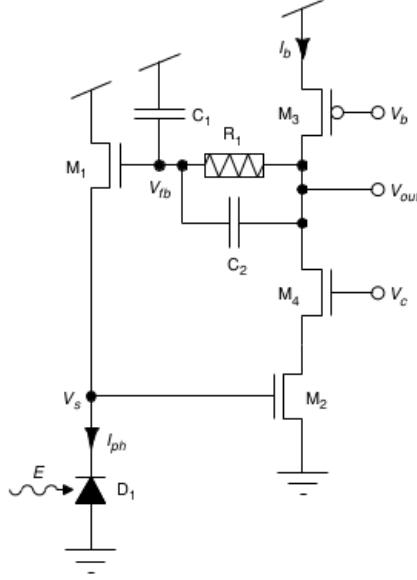


Figure 46: Adaptive logarithmic photosensor with cascode transistor increasing the bandwidth for small photocurrents [1].

Hereby, the apparent capacitance, as seen from the source of M_1 , is increased by C_p multiplied by the voltage gain from gate to source. For small photocurrents this effect leads to limited response band width ($I = C \frac{\delta V}{\delta t}$). The introduction of a cascode with a fixed gate voltage V_c clamps the voltage on the drain of M_2 , because the current through the amplifier is approximately constant, and largely nullifies the Miller effect (parasitic capacitance has only influence on feedback for changing voltage).

However, for certain biasing conditions the presence of the cascode can make the circuit unstable [1].

18 Silicon Neurons

What is a neuron and what are its components?

The basic anatomical unit in the nervous system is specialised cell called the neuron.

[2]

1. Synapse
2. Soma
3. Dendrite

What types of models are used to simulate neurons?

How does the spikegeneratng mechanism work?

What is an FI curve?

Can you draw the circuit schematic of the axon-hillock neuron?

The schematic of the axon-hillock neuron can vary upon the chosen amplification. General schematic is depicted on the figure 47.

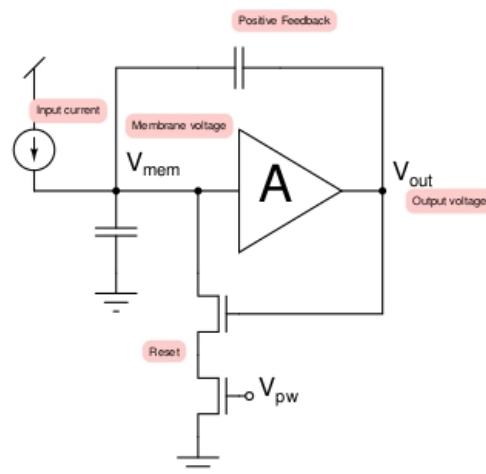


Figure 47: The Axon-Hillock circuit [3].

The specific amplification was chosen and is depicted on the Axon-hillock integrate and fire circuit (see fig. 48).

[6].

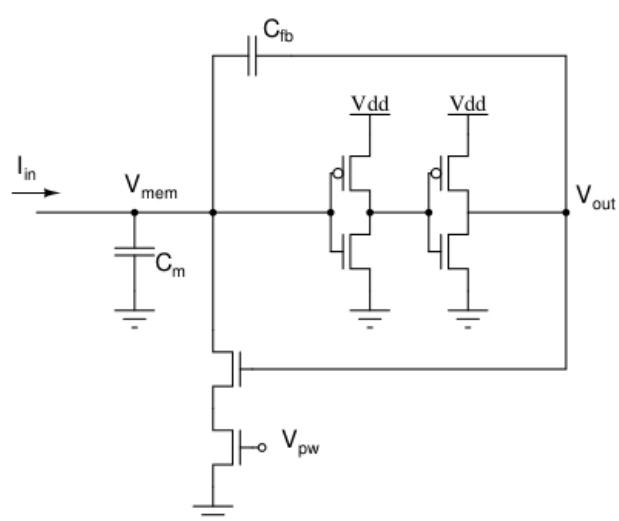


Figure 48: Axon-hillock integrate and fire circuit [6].

19 Important Values to Remember

Subthreshold slope factor κ

nFET: 0.4- 0.9

pFET: 1

Threshold Voltage V_t

Typical value nFETs: 0.7V

Early Voltage

20-30V

Dark Current

? pico Amps?

I_0

nFET: $1 \cdot 10^{-14}$ A

pFET: $1 \cdot 10^{-18}$ A

20 Exam Questions

- 2013

- Draw and argue through a photodiode.
- Explain the time domain ($\tau = R \cdot C$) of a photo diode.
- Draw and argue through a photodiode circuit with feedback and cascode.
- Draw and argue through a WTA circuit given all is in subthreshold, and I_2 current is bigger than I_1 .

Given current source 1 has voltage x, current source 2 has voltage y, and the bias voltage is z, what are the voltages of all the nodes in the circuit.

- What is U_T and what does it stand for (Knowing the formula $U_T = \frac{kT}{q}$ is NOT enough! What does it physically mean?)

- 2014

21 Lab Questions

Lab 1

- Remember: Experimental techniques and devices are never ideal.

Lab 2

- What does it mean for a MOS transistor channel to be accumulated, flat-band, depleted, inverted?
- Knowledge of how subthreshold transistor operation is a diffusion process and why it depends exponentially on the terminal voltages.
- What is the meaning of "saturation"?
- What is the triode or linear operating range? I_{ds} vs V_{gs} on log scale.
- Differences between n- and p-fets.
- Typical values of I_0 , κ and subthreshold operating range.
- What are wells and how should the wells be biased relative to the substrate?
- What is the "back gate" or "body effect"?
- How is the back gate related to κ ?
- How do you measure κ ?
- How to make a MOS capacitor and what is its C-V relationship.
- How a source follower works and how to compute the gain of a source follower.

Lab 3

- How transistors work above threshold.
- What is the linear or triode region and what is the saturation region?
- How they depend on gate and threshold voltage. What is the "over-drive"?
- What is the specific current? How the Early effect comes about.
- Typical values for Early voltage.
- How to sketch graphs of transistor current vs. gate voltage and drain-source voltage.

- How above-threshold transistors go into saturation and why the saturation voltage is equal to the gate overdrive.
- The above-threshold current equations.
- How above-threshold current depends on C_{ox} and mobility.
- How transconductance and drain resistance combine to generate voltage gain and what is the intrinsic voltage gain of a transistor.
- What effect does velocity saturation have on transistor operation, specifically, how does it change the relation between saturation current and gate voltage?
- What is DIBL (drain induced barrier lowering) and II (impact ionization)?
- What is the dominant source of mismatch?
- How does transistor mismatch scale with transistor size?
- What are typical values of transistor threshold voltage mismatch?

Lab 4

- Can you sketch a transamp, a wide range transamp, a current correlator, and a bump circuit in both n- and p-type varieties?
- How does a differential pair work?
- How does the common-node voltage change with the input voltages?
- How can you compute the differential tail currents from the subthreshold equations, and how do you obtain the result in terms of the differential input voltage?
- How does a current-correlator work?
- How does a bump circuit work?
- The I-V characteristics of a transconductance amplifier below threshold.
- What's the functional difference between simple and wide-output-range transamp?
- The subthreshold transconductance g_m .
- The relation between gain A , transistor drain conductances g_d , and transconductances g_m .

- Can you reason through all the node voltages in these circuits? I.e., if we draw the circuit and provide specific power supply and input voltages, can you reason to estimate all the other node voltages, at least to first order approximations, assuming $\kappa = 1$?

Lab 6

- How does the WTA circuit work?
- Can you reason through its behavior?
- How does the bias current affect its performance?
- How can you adjust the gain of the circuit through the sizing of the transistors?

Lab 7

- How to use an oscilloscope and a function generator.
- How to compute the time-constant of a low-pass filter and how to estimate it from the measurements.
- to change the time- constant of a follower-integrator circuit.
- In what way does the follower integrator behave nonlinearly for large signal input?

Lab 8

- How does the I-V curve of a diode change in the presence of light?
- How does phototrans- duction occur in silicon?

Lab 9

- How you can use adaptation in a feedback loop to highpass amplify signal but not mismatch.
- How you can build a fast logarithmic current-sense amplifier, by using feedback to make a virtual ground.
- How you can use a capacitive divider in the feedback loop of an amplifier to set a gain.
- How you can use a cascode configuration to increase effective drain resistance.
- What is the Miller effect, and how a cascode can be used to nullify it.

Lab 10

- The schematic for a synapse circuit.
- How the synaptic current changes as a function of the synaptic weight, the time constant, and the presynaptic frequency. Can

Lab 11

- What is a neuron and what are its components (synapse, soma, dendrite)?
- What types of models are used to simulate neurons?
- How does the spike-generating mechanism work?
- What is an FI curve?
- Can you draw the circuit schematic of the axon-hillock neuron?

Lab 12

- How do tunneling and injection mechanisms work?
- What is the shape of the energy band diagram in the channel and oxide during tunneling and injection?
- How are the memory cell circuits used to control tunneling and injection?

22 Glossary

Term	Brief explanation of the term
A very important term	The much longer explanations breaking over several lines here and more text to see the line breaking again and again and again and again.
Another line	Another explanation.

23 TODO

- **ADD THE EXAM QUESTIONS YOU GOT ASKED!!** (The guys next year will thank you!)
- Improve layout (currently ugly at some points...)
- Complete "Silicon Neurons" chapter
- Complete current divider circuit
- Add more values to "Important Values to Remember"
- Add more basic theory (e.g. definition and meaning of U_T)

Bibliography

- [1] Liu, Kramer, Delbrück, Indiveri, Douglas – *Analog VLSI: Circuits and Principles*, 2002 – Massachusetts Institute of Technology
- [2] Carver Mead – *Analog VLSI and Neural Systems*, 1989 – Addison-Wesley Publishing Company
- [3] Delbrück, Indiveri, Liu – *Neuromorphic Engineering I, Fall Semester 2013, Silicon neurons. Integrate and fire models. Analog VLSI implementations of adaptive neurons. The silicon neuron: detailed VLSI model with adaptation properties.*
- [4] Delbrück, Indiveri, Liu – *Neuromorphic Engineering I, Fall Semester 2013, Lab 4 Document*
- [5] Delbrück, Indiveri, Liu – *Neuromorphic Engineering I, Fall Semester 2013, Lab 8 Document*
- [6] Delbrück, Indiveri, Liu – *Neuromorphic Engineering I, Fall Semester 2013, Lab 11 Document*
- [7] Delbrück, Indiveri, Liu – *Neuromorphic Engineering I, Fall Semester 2013, Lecture 2 Document*
- [8] Delbrück, Indiveri, Liu – *Neuromorphic Engineering I, Fall Semester 2013, Lecture 4 Document*
- [9] Delbrück, Indiveri, Liu – *Neuromorphic Engineering I, Fall Semester 2013, Lecture 6 Document*
- [10] T. Delbrück, C.A. Mead – *Analog VLSI Phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits*, CNS Memo no.30, April 2. 1996 – California Institute of Technology