Physics 364, Fall 2014, reading due 2014-10-26.

Email your answers to ashmansk@hep.upenn.edu by 11pm on Sunday

Course materials and schedule are at http://positron.hep.upenn.edu/p364

Assignment: (a) First read (or at least skim) Eggleston's chapter 5 (Field Effect Transistors, pages 133–149). You can **skip** the section on JFETs (5.2.1, pages 134-136), as we will only study MOSFETs. You can also ignore the common-gate amplifier. And don't worry about depletion-mode devices, as we will only work with enhancement-mode MOSFETs (both n- and p-channel). (b) Then read through my notes (starting on the next page), which directly relate to what we will do in Labs 17 and 18. (c) Then email me your answers to the questions below.

1. What are the names of the three terminals of an n-channel MOSFET corresponding to (respectively) the base, collector, and emitter of an NPN bipolar transistor? (Be sure to say which one corresponds to which one.) Between which two terminals (and in which direction) does electric current flow?

2. Can you make an analogy between the BJT parameter that we have been calling "little r_e" and the FET parameter known as "transconductance" g_m? (To check that you have the correct analogy, consider the units of r_e and the units of g_m.)

3. What is the property of the gate of a field-effect transistor that makes FETs amazingly useful for the input stage of an opamp?

4. Is there anything from this reading assignment that you found confusing and would like me to try to clarify? If you didn't find anything confusing, what topic did you find most interesting?

5. How much time did it take you to complete this assignment? Also, I continue to welcome suggestions for ways in which I might adapt the course to help you to learn most efficiently.

FYI, there is a detailed description of the internal workings of a 741 opamp in the textbook by Sedra & Smith. They spend about half of a chapter analyzing the 741. In case you are interested in seeing this analysis, I put a PDF of the relevant pages up on the (password-protected) Canvas site for this course.

`https://canvas.upenn.edu/courses/1253913/files/49121448/`

(The filename is "`sedra_smith_741_analysis.pdf`" under "Files" on Canvas.)

Having now studied Bipolar Junction Transistors for two weeks, this week we will study Field Effect Transistors. (Actually, we'll do PID controllers on Oct 27/28, then FETs on Oct 29/30 and Nov 3/4.) Because FETs are the basis of all modern digital electronics, but are also commonly used in analog circuits, FETs make a nice segue between analog and digital electronics. I'm going to start by quoting/paraphrasing the opening of the FET chapter from Horowitz & Hill:

> Both BJTs and FETs are 3-terminal devices in which the conduction between two electrodes is controlled by a voltage applied to a third **control electrode** (called the **base** for a BJT and called the **gate** for a FET). In an NPN BJT, the collector→base junction is reverse-biased, so no current normally flows between collector and base. Forward-biasing the base→emitter junction by ≈ 0.6 V causes electrons from the emitter to enter the base region, where they are strongly attracted to the collector; although some base current results, most of these electrons ("minority carriers" in the p-type base) are captured by the collector. This results in a collector current, controlled by a (much smaller) base current. The collector current I_C is proportional to the rate of electron injection into the base region, which is an exponential function of the $B \to E$ potential difference V_{BE}. You can think of a BJT either as a current amplifier (with roughly constant gain β) or as a voltage-controlled current source: $I_C \propto \exp(V_{BE}/25\text{ mV})$, where 25 mV really means kT/e.
>
> In a FET, conduction in a **channel** is controlled by an electric field, which is produced by applying a voltage to the **gate** electrode. [The electric field either attracts charge carriers into the channel from the surrounding silicon, increasing the channel's conductivity, or else repels charge carriers from the channel into the surrounding silicon, decreasing the channel's conductivity.] A key advantage of a FET is that the gate draws no DC current. As with BJTs, there are two polarities: n-channel FETs (conduction by electrons) and p-channel FETs (conduction by holes), analogous to NPN and PNP transistors, respectively.

In a BJT, you wiggle the base voltage in order to control the flow of current between the collector and the emitter. Analogously, in a FET, you wiggle the **gate** voltage in

order to control the flow of current between **drain** and the **source**.[1]

Because of FETs' negligible gate current, FET-based amplifiers can have input resistances of order 10^{14} Ω. So a FET-based opamp has both a gigantic input resistance and $I_{\text{bias}} \approx 0$ (typically measured in picoamps, vs. about 100 nA for the BJT-based 741 opamp). This lack of gate current also accounts for the low power consumption of FET-based digital circuits (microprocessors, etc.) compared with older BJT-based circuits: FET-based logic gates consume negligible power except while switching between the LOW and HIGH binary output states.[2] FETs can also make an excellent "analog switch" (like an ON/OFF switch whose setting is controlled electronically, rather than with your fingers): this allows charge to be stored on a capacitor for a long time in a computer's DRAM memory or for a short time in the "sample & hold" circuit that holds steady the input to an analog-to-digital converter. We will see some applications of FET-based analog switches in Lab 18.

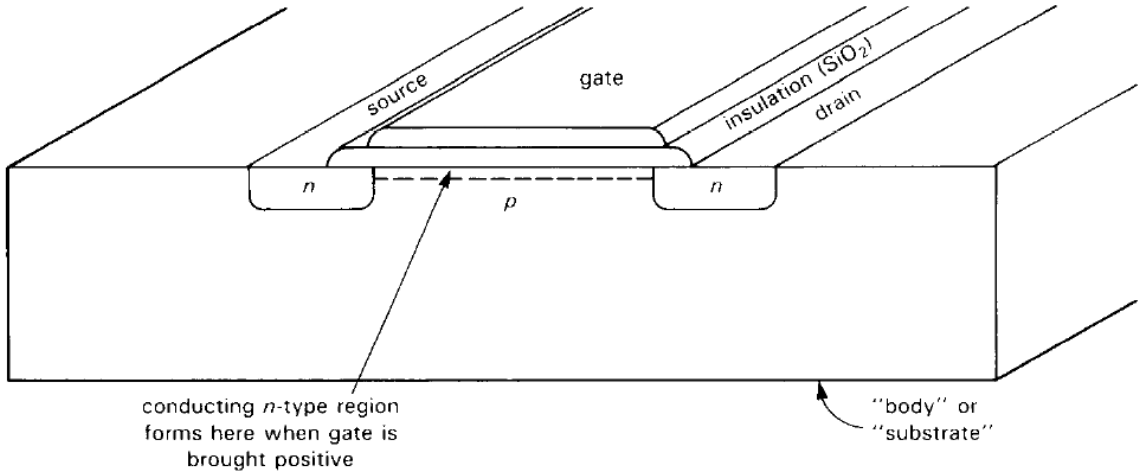


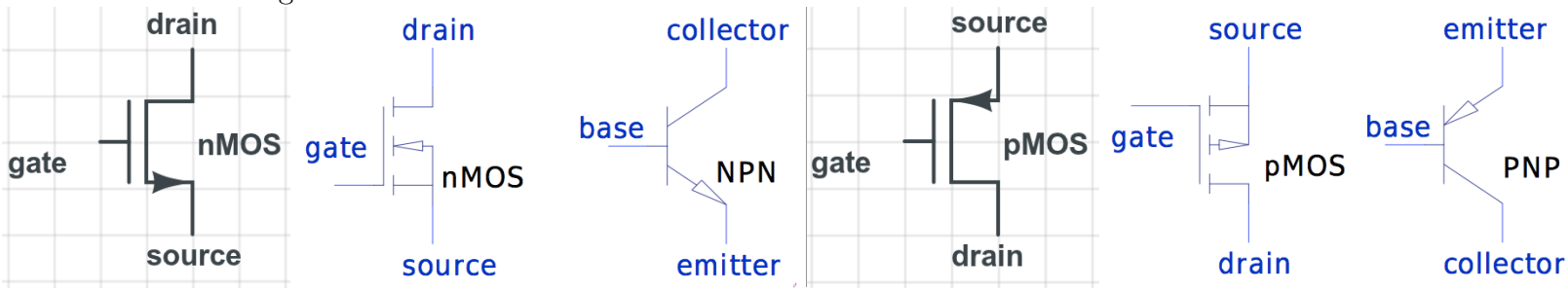
Figure 3.5. An *n*-channel MOSFET.

As if it weren't already confusing enough for you to have to replace the words *base*, *collector*, and *emitter* (which you only learned a couple of weeks ago) with the new corresponding words **gate**, **drain**, and **source**, there is the additional complication that FETs come in many varieties. Drain and source are at opposite ends of a thin semiconductor **channel** (see figure above, from Horowitz & Hill, where the dashed line indicates the channel); the gate is an electrode parallel to the channel. In a **MOSFET**, the gate is a thin layer of metal, separated from the channel by an insulator (hence Metal + Oxide + Semiconductor). In a JFET (which we will not study), the gate is a semiconductor with opposite (e.g. p vs. n) doping from the channel. In enhancement-mode devices, the channel normally contains very few charge carriers,

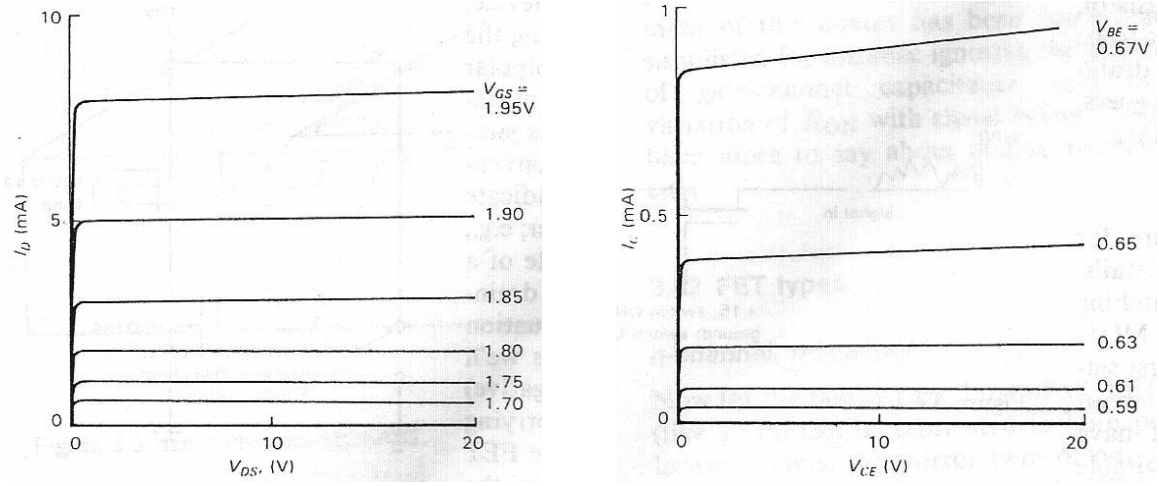
[1] In an n-channel FET (analogous to NPN transistor), electrons flow from source to drain, so conventional current flows from drain to source. My mnemonic is to imagine electrons flowing out of a spigot (source) and down the drain; the electric current is opposite the flow of electrons.

[2] If you don't yet know what a logic gate is, don't worry. We will introduce them in the next week or two.

and hence does not conduct; applying a voltage to the gate electrode attracts carriers into the channel, thus making the channel conductive. In depletion-mode devices (which we will not study), the channel is doped such that it normally contains many charge carriers, and hence allows current to flow between drain and source; applying a voltage to the gate electode repels carriers from the channel, thus making the channel highly resistive to current flow. In this course, we will study only n-channel and p-channel enhancement-mode MOSFETs, which are analogous to NPN and PNP transistors, respectively. The figure below shows (from left to right) the schematic symbols for an n-channel MOSFET (two common variants of the same symbol), then its bipolar analogue the NPN transistor, then a p-channel MOSFET (two common variants of the same symbol), then its bipolar analogue the PNP transistor. Just as the arrow indicates where the emitter is on a BJT, the horizontal line coming in to the gate indicates where the source is on a MOSFET.



Let's look in more detail at the n-channel MOSFET, which resembles the familiar NPN transistor. The four figures below show side-by-side characteristic curves for nMOS and NPN transistors. Normally $V_{DS} > 0$, i.e. the drain (analogous to collector) is kept more positive than the source (analogous to emitter). Once the gate voltage is raised sufficiently far above the source, current I_D can flow from drain to source.



The figure above[3] compares (left) n-channel MOSFET characteristics with (right) NPN BJT characteristics. The left graph shows drain current I_D vs. drain-source voltage difference V_{DS}, for a range of different values of gate-source voltage difference

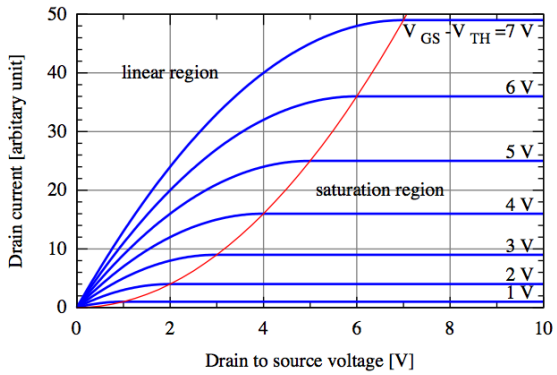
[3] from Horowitz & Hill figure 3.2.

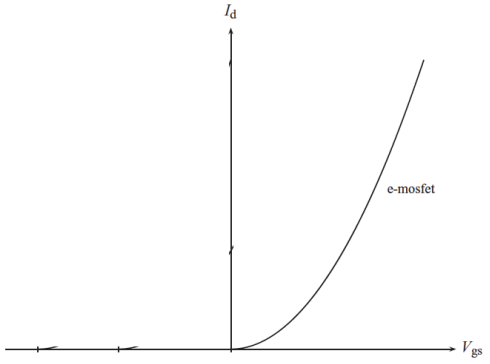
V_{GS}. The right graph shows collector current I_C vs. collector-emitter voltage difference V_{CE}, for a range of different values of base-emitter voltage difference V_{BE}. The emphasis in both graphs is on the "active" region that is used for amplification.

No appreciable drain current I_D will flow until V_{GS} exceeds a **threshold voltage** V_t, which can vary from MOSFET to MOSFET in the range 0.5 V $\lesssim V_t \lesssim$ 5 V. The left figure below graphs I_D vs. $V_{GS} - V_t$. The red line $V_{DS} = V_{GS} - V_t$ separates two modes of operation. The region to the left of the red line, where $V_{DS} < V_{GS} - V_t$, is called the **linear** or **ohmic** region, because a FET in this mode of operation is sometimes used as a programmable resistor. The drain current in the ohmic region follows $I_d = K \cdot (2(V_{GS} - V_t) - V_{DS})\, V_{DS}$, which is approximately linear in V_{DS} if $V_{DS} \ll 2(V_{GS} - V_t)$. To the right of the red line, the region $V_{DS} > V_{GS} - V_T$ is called the **saturation** or **active** region, which corresponds to the active region of the BJT.[4] The drain current in the active region follows

$$I_D = K \cdot (V_{GS} - V_t)^2$$

where coefficient K depends on FET geometry, charge-carrier mobility, temperature, etc. As shown below (right), I_D increases quadratically with V_{GS}, whereas for a BJT, I_C increases exponentially with V_{BE}.





The generally similar shapes of nMOS and NPN transistor characteristics makes it plausible that MOSFETs can be used in the same kinds of amplifier circuits as BJTs. This is indeed the case, particularly for applications that require extremely large input resistance. One word of caution, though, is that FET charactersitics vary much more from device to device than BJT characteristics, so FET-based amplifier circuits can be more difficult to design. As a result, FET-based amplifiers tend to appear inside of integrated circuits (where the designer has more control over manufacturing

[4] Some regrettable terminology choices give FETs and BJTs confusingly different (and sometimes contradictory) vocabulary. **Active** mode means the same thing for BJTs or FETs. But **saturation** mode has opposite meanings for FETs and BJTs: in a FET, it is the large-V_{DS} region, while for a BJT, it is the small-V_{CE} region. Also, the BJT's active region (i.e. large V_{CE}) is sometimes also called its "linear" region, while **linear** region refers to the small-V_{DS} region for a FET. So it is probably best to use the words **active** and **ohmic** for the two FET regions, to avoid confusing yourself; these two regions correspond respectively to the "active" and "saturation" regions of a BJT.

variations) but are seldom designed using discrete transistors. In any case, we will look at the FET analogues of a few of our familiar BJT amplifier circuits mainly so that you can see that the analogous circuits are possible.

There is one more piece of nomenclature that is used differently for BJTs and FETs. For BJTs, we defined "little r_e" to be $r_e = \mathrm{d}V_{BE}/\mathrm{d}I_C$, the dynamic resistance that behaves as if it were a little resistor placed inside the emitter. At the time, I pointed out in a footnote that engineers sometimes write $1/g_m$ instead of writing r_e. Well, for FETs, everyone writes $1/g_m$ instead of r_e. So we define the **transconductance** $g_m = \mathrm{d}I_D/\mathrm{d}V_{GS}$. In the active region, we can evaluate $g_m = 2K \cdot (V_{GS} - V_t)$.

You should look at Eggleston's analysis of the **source follower** (section 5.4.4), which is the FET analogue of the emitter follower. Notice that the input resistance is determined only by the resistors forming the biasing network, which can be very large, since there is no gate current. The voltage gain Eggleston finds is

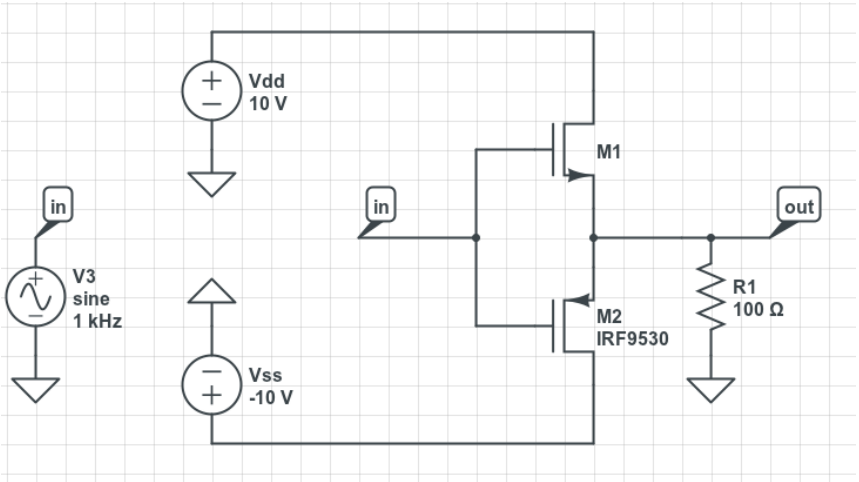
$$\frac{v_{\text{out}}}{v_{\text{in}}} = \frac{R_S \parallel R_{\text{load}}}{(1/g_m) + (R_S \parallel R_{\text{load}})}$$

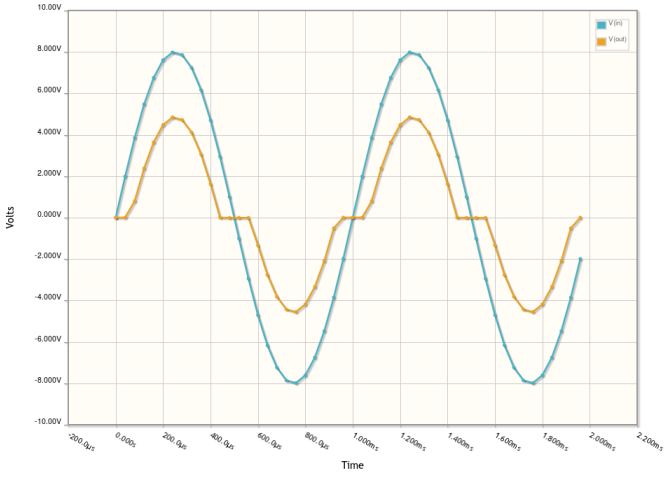
which approaches 1 in the limit $1/g_m \ll (R_S \parallel R_{\text{load}})$, i.e. the output voltage follows the input, but the output current can be much larger than the input current.

Also look through Eggleston's description of the common-source amplifier (section 5.4.3), which is the FET analogue of the common-emitter amplifier. Eggleston's derivation finds a voltage gain (where v denotes a small ΔV at the input or output)

$$\frac{v_{\text{out}}}{v_{\text{in}}} = -\frac{(R_D \parallel R_{\text{load}})g_m}{1 + g_m R_S} = -\frac{R_D \parallel R_{\text{load}}}{R_S + (1/g_m)}.$$

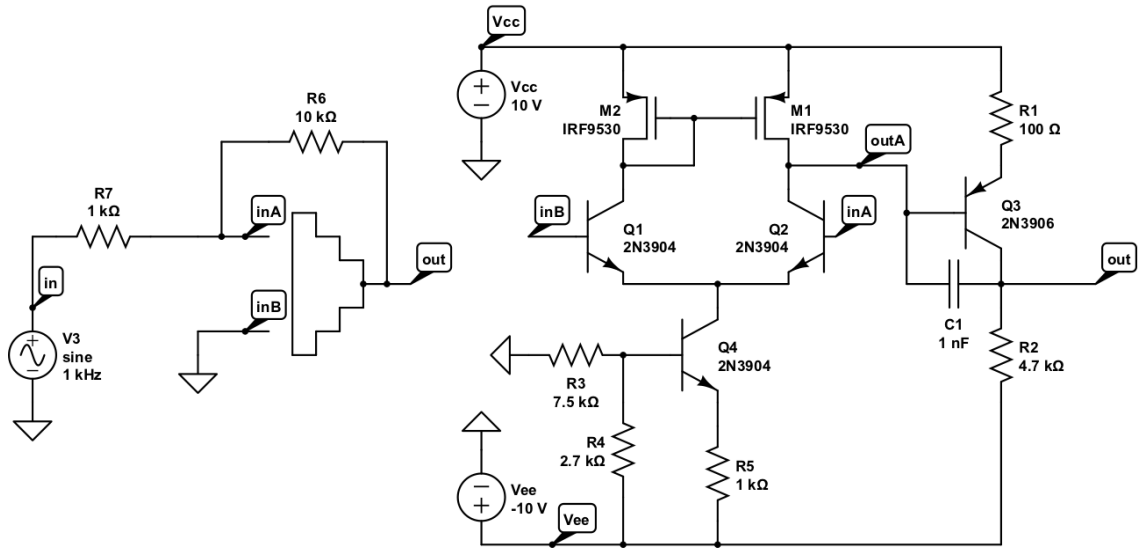
If we assume $R_{\text{load}} \gg R_D$ (so that the input resistance of the downstream load is much larger than the output resistance of our amplifier) and substitute $R_D \to R_C$, $R_S \to R_E$, and $(1/g_m) \to r_e$, we get the familar expression from the BJT common-emitter amplifier: $v_{\text{out}}/v_{\text{in}} = -R_C/(R_E + r_e)$. So this circuit is not so surprising.





To cover one example that is not in Eggleston's text, the figure above shows (left) the schematic for a CMOS (the "C" for **complementary**, i.e. one n-channel MOSFET (a.k.a. nMOS) and one p-channel MOSFET (a.k.a. pMOS)) push-pull follower, and (right) its input and output vs. time. For the default MOSFETs in CircuitLab, the threshold voltage V_t is 3 V for nMOS (and -3 V for pMOS), so the crossover distortion is much more dramatic than for the BJT push-pull. (Instead of an 0.7 V gap between $V_{\rm in}$ and $V_{\rm out}$, you have a 3 V gap.) Notice also that CircuitLab uses a variant of the MOSFET schematic symbols that are designed to resemble NPN and PNP transistors. In the schematic, M_1 is nMOS and M_2 is pMOS.

Most other BJT circuits have MOSFET analogues, though in general FETs are harder to use for making amplifiers but much easier to use for making circuits that just switch a current on and off. The circuit shown below is yet another example of a highly simplified opamp. The first stage is a differential amplifier made with two NPN transistors, just as we built in class last week. The two emitter resistors (beneath Q_1 and Q_2) have been removed, to increase the differential gain, at the cost of some nonlinearity (which is cured by using the opamp with negative feedback). There is an NPN current source in the tail of the differential pair, as we did in class, to make the common-mode gain as small as possible. The second stage, as with last week's opamp, is a PNP common-emitter amplifier. The capacitor between the input and output of the second stage is what gives the opamp's gain its $1/f$ frequency dependence, i.e. the "frequency compensation" that was briefly mentioned in `reading05`. In this opamp, I omitted the third stage (normally a push-pull follower), so this opamp has a 4.7 kΩ output resistance, whereas a real opamp's output resistance would be much smaller. (As long as the gain is high enough, negative feedback will make the opamp circuit's output impedance very low, even if the output impedance of the opamp itself is not so small, as we saw in Lab 7.)



One big difference between this opamp and last week's is the current mirror (M_2 and

M_1) up above the collectors of Q_1 and Q_2. This current mirror is made with two p-channel MOSFETs, and replaces the resistors that were above the collectors of Q_1 and Q_2 in last week's opamp. (Since the output was taken only from the collector of Q_2, the resistor above Q_1 could be omitted, by the way. I think I mentioned that very quickly last week in class.) Remember that the differential amplifier's gain is proportional to the resistor placed above Q_2. One way to make that resistor appear very large is to replace it with a current source, since a current source has very large $R_{\text{out}} = |\mathrm{d}V_{\text{out}}/\mathrm{d}I_{\text{out}}|$. So in fact we could use a single PNP (or pMOS) transistor to make a current source above Q_2. The current source above Q_2 would make the gain very large because even a very small change in the current through Q_2 (in response to a wiggle in Q_2's base voltage) causes a very large change in Q_2's collector voltage (in proportion to the large dynamic resistance above Q_2's collector).

It turns out that using a *current mirror* above Q_1 and Q_2 instead of a single current source above Q_2 makes the differential gain a factor of two larger than using a single current source. Imagine V_{inA} wiggling up by ΔV and V_{inB} wiggling down by the same ΔV. The change in V_{inA} causes Q_2's collector current to increase by $\Delta I = \Delta V/r_e = g_m \Delta V$. The corresponding ΔV_{out} is then $-R_{\text{out}}\Delta I$, where R_{out} is the large dynamic resistance discussed above for transistor current sources.[5] If we were using independent current sources, rather than a current mirror, this would be the end of the story. But because M_1 mirrors the current in M_2, the corresponding ΔI from Q_1 is mirrored at M_2 and thus also appears at Q_2's collector. So the ΔI is doubled, thus doubling the corresponding ΔV_{out}. The fact that the mirror has a very large dynamic resistance (like a current source) gives us large differential gain. And the fact that a change in M_2's current is mirrored at M_1 gives us another factor of two. I put Sedra & Smith's 3-page writeup of this circuit (differential amplifier with current mirror in place of collector resistors) on the Canvas site for the course, in case you're interested.

`https://canvas.upenn.edu/courses/1253913/files/49125691`

(The filename is "`sedra_smith_diffamp_current_mirror.pdf`" under "Files" on the Canvas site.)

In addition to analog circuits that require very large input resistance, the real application of FETs is for switching things on and off. We will spend some time in Lab 18 working with a FET-based analog switch, which acts like an ordinary switch (i.e. you can open or close it to disable or enable the flow of current) that is controlled by an electrical signal rather than by your finger. This might be useful, for example, to allow a computer to zero the charge on the capacitor of your opamp-based integrator at the start of a new measurement cycle, or to allow a computer to select one of several different input signals to direct to a shared oscilloscope channel. MOSFETs are also commonly used to allow computers to switch on and off high-power devices such as motors, because a very large current can be switched on and off without needing to

[5] To do this right, you need to consider in parallel the output resistances of Q_2's collector and M_1's drain.

supply any current to the FET's gate.

Finally, the most widespread application of MOSFETs is in CMOS (complementary MOS, i.e. matched p-type and n-type pairs) digital logic circuits, which we will study extensively a week or two from now.

Please look through the file (from the Harvard course) describing MOSFET switches at this Canvas URL:

`https://canvas.upenn.edu/courses/1253913/files/49125748`

(The filename is "`mosfet_switches.pdf`" under "Files" on Canvas.) We will borrow one section of Tom Hayes's analog-switch lab, for Lab 18 on Nov 3/4, so you'll need to have read this `mosfet_switches.pdf` file by the end of next weekend.