Note 4: Frequency Response and the Cascode

Rohit J. Braganza

Abstract Remember those parasitic capacitors? We will now take them into account. The reader may remember we listed four noticeable parasitics: the two fringing capacitances, C_{gs} and C_{gd}, and the two substrate capacitances, C_{sb} and C_{db}. Say we wish to know the frequency response of the LT1008 below, and maybe sketch a Bode plot. As before, we have two options: 1, we could draw all four capacitors for each transistor and do some quick Nodal analysis with 172 capacitors, some resistors, and some jω's here and there. Or there's the other option...

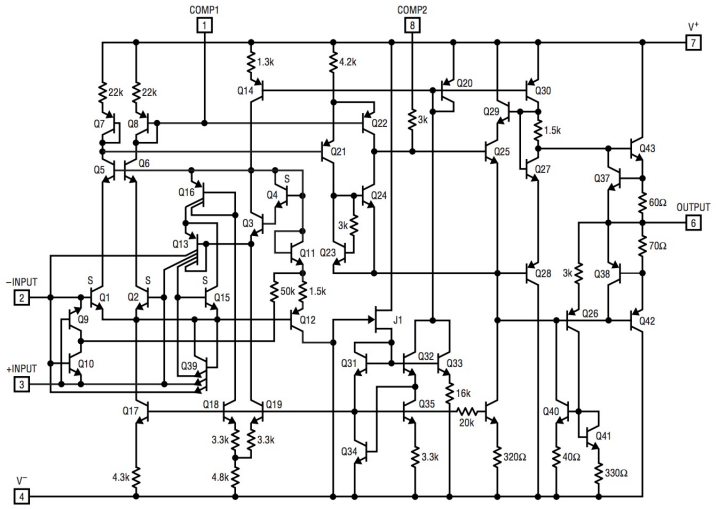

Fig. 1 The LT1008 returns.[1] [1]

Rohit Julian Braganza
email: rbraganza@berkeley.edu

1 Living in a RC World: The Open-Circuit Time Constant Method

Ideally every circuit would look something like that seen in **Fig. 2**: the simple RC circuit, perfect for the lazy- whoops I meant hard working- analog designer!

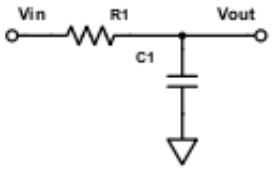

Fig. 2 The simple RC circuit we can all solve. $\tau = RC$. Doesn't get much better than that.

We will use the Open Circuit Time Constant method (OCTC) to convert our complicated circuits down to this nice simple circuit we can all solve.

Open Circuit Time Constant Method:

Step 1	Zero all sources, and short all AC-Coupling capacitors (caps of very large magnitude, in the signal path).
Step 2	Replace capacitor of interest with a test voltage or current. Open all other capacitors.
Step 3	Find the equivalent resistance seen by the capacitor, $R_{eq} = V_{test}/I_{test}$. The capacitor's time constant is then given by $\tau = R_{eq}C$
Step 4	Repeat for all other Caps!

Fig 3. gives a simple example.

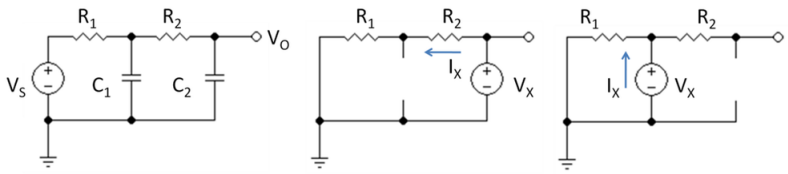

Fig. 3 Using OCTC for the leftmost circuit. The middle picture results in $\tau_2 = (R_1 + R_2)C_2$. The rightmost: $\tau_1 = R_1C_1$

What information exactly does the OCTC give us? It gives us an estimate of the poles in a circuit, as well as the 3dB point of a circuit. More precisely, the 3dB point is at: $\omega_{3dB} \approx \frac{1}{\sum_i \tau_i}$, where τ_i represents the i-th time constant. This approximation

[1] It's still probably a bit intimidating, but that's okay- we haven't covered differential pairs yet.

works well when one time constant is larger than all the others; in the worst case, for a two pole system, it can give a corner frequency twice as low as the actual, when the time constants are equal to each other. A thorough explanation of why this method works would take up the entire notesheet- the reader is instead highly recommended to read [2] and [3]. We will instead focus on how we can apply this method to our analysis.

1.1 Frequency Response of the Common Source- sans C_{gd}

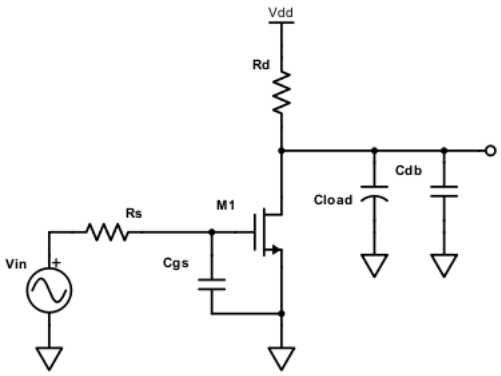

Fig. 4 The common source with parasitics, except for C_{gd}. The load is modeled as a capacitor, C_{load}. The substrate is assumed to be grounded.

Without C_{gd}, and with the help of the OCTC it is easy to determine the poles the circuit. First notice C_{sb} is shorted out (not shown). Looking at C_{gs}, this capacitor "sees" a resistance of R_s.

$$\therefore \exists \text{ A pole at: } s = 1/(R_S(C_{gs})) \tag{1}$$

$$\text{Or, equivalently}^2\text{, a time constant at: } \tau = R_S(C_{gs}) \tag{2}$$

Likewise, at the drain, we see the combined capacitance of $C_{load} + C_{db}$, and a resistance of $R_d \parallel r_o$.

$$\therefore \exists \text{ An output time constant at: } \tau = (R_d \parallel r_o)(C_{db} + C_{load}) \tag{3}$$

[2] This is not exactly true. The pole gives the time constant only in a single pole circuit. However, for now, we pretend that the time response mirrors a one pole system even though we have many poles. This works surprisingly well for most op-amps.

1.2 The big, bad & ugly C_{gd}- the Miller Cap

Things get a little bit more complicated when we consider C_{gd}: looking at **Fig. 5**, notice that the OCTC would not seem to help much at all- the test source we would replace for the capacitor would not be grounded (always a pain, and reserved mainly for first level circuits classes only), and not to mention that there is the gain cell as well. To solve this issue, we use a handy theorem: *Miller's Theorem.*

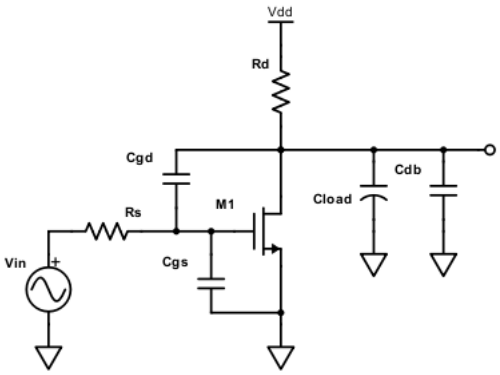

Fig. 5 Now we include all parasitics, including C_{gd}. Finding the pole for C_{gd} with the OCTC would be extremely complicated. Notice how the transistor creates a current that depends on the source- the capacitor is essentially in feedback with an amplifier.

Miller's Theorem Consider the circuit below:

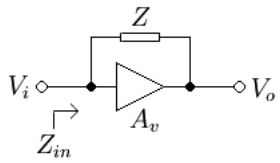

Fig. 6 A generic gain block with gain A_v and an impedance Z in feedback.

Let's find the input impedance:

$$Z_{in} = \frac{V_{in}}{I_{in}} = \frac{V_{in}Z}{V_{in} - A_v V_{in}} = \frac{Z}{1 - A_v} \tag{4}$$

And likewise, the output impedance:

$$Z_{out} = \frac{Z}{1 - \frac{1}{A_v}} \tag{5}$$

Miller's Theorem for Capacitors
We know for a capacitor: $Z = 1/sC$

$$\implies \quad Z_{in} = \frac{1}{s(1-A_v)C} \tag{6}$$

$$Z_{out} = \frac{1}{s(1-\frac{1}{A_v})C} \tag{7}$$

Notice that for a capacitor, if $A_v < -1$, looking into the input we see a *larger capacitance, by a factor of* $(1-A_v)$*!* From the output however, so long as $|A_v| \gg 1$, we essentially see the same capacitance.

Let's now see how we can apply the Miller Theorem to our C_{gd} problem. We first note that we can consider the transistor as a gain element.

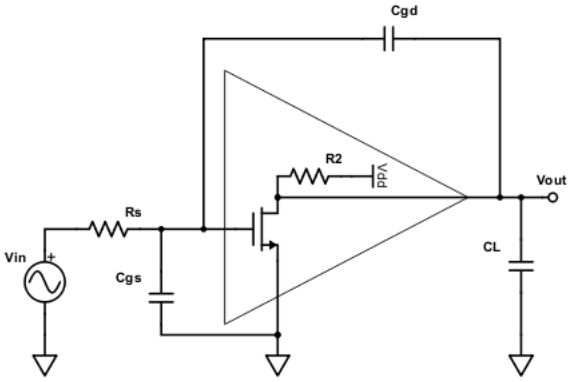

Fig. 7 Figure 6- but with a triangle! Note we left out the substrate caps for clarity.

We can now apply Miller's theorem to C_{gd}, as seen in **Fig. 8**.

Remember how the common source has an inverting gain? Well, unfortunately, from the Miller's theorem, this makes the input capacitance much, much greater. We know the gain of the amp to be $-g_m(R_2 \parallel r_o)$, and using (5) and **Fig. 8** this means we have an input pole at:

$$\text{input } \tau = R_s(C_{gs} + (1 + g_m(R_2 \parallel r_o))C_{gd}) \tag{8}$$

For the output pole, we make the assumption $A_v \ll -1$, and so C_{gd} appears unchanged.

$$\therefore \ \exists \text{ An output } \tau = (R_2 \parallel r_o)(C_{gd} + C_{db} + C_{load}) \tag{9}$$

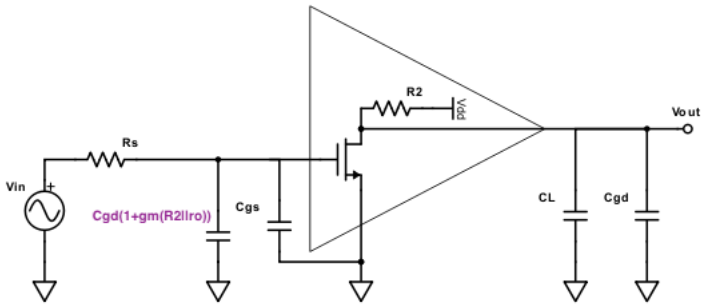

Fig. 8 The equivalent C_{gd} as seen from the input and output; we assume the gain is much greater than 1 for the output impedance. Note the Miller multiplied input cap, highlighted in purple.

1.3 A Need for Speed: The Dominant Pole Approximation

It is good to look back and wonder what exactly we have accomplished here. Using the OCTC and Miller's theorem, we will be able to approximate the locations of the poles of a circuit. From this, we can determine which one is *dominant*, also known as the lowest frequency pole. This in turn gives us a rough estimate of the bandwidth of a circuit: we will have the full low frequency gain until we hit this 3dB point, after which we will get the characteristic 20dB/decade loss.

Why do we often refer to a circuit's frequency response as its "speed?" Well, remember, for a basic RC circuit, the pole location is the inverse of what we call its "time constant"- essentially a circuit's speed is limited by the amount of time we need to actually charge up the parasitic capacitors. The dominant pole therefore gives some indication of a circuit's speed since it represents the node that will take the longest to charge[3]. Put it this way: if it weren't for these capacitors, all the voltages would be changing near the speed of light, and everything would be much faster. Unfortunately, nothing is ever easy.

2 Beware the Miller Cap

We will do an example to observe the power behind the OCTC and Miller theorem- and to show why we should *always* watch out for that pesky miller cap.

Let's look at the multistage example in **Fig. 9**. Let's see- 4 FETs, 4 parasitics each- that's 16, plus the load, or 17 capacitors total! You might be be a little worried at this point... and you should be, but not for the reason you might expect. For hidden in this little circuit there lies an extremely dangerous node...

[3] In reality, amplifiers are rarely used in an open loop configuration, and instead used in feedback. The effects of feedback on speed will be looked into later, but the first pole is still a good indicator of a circuits speed capability.

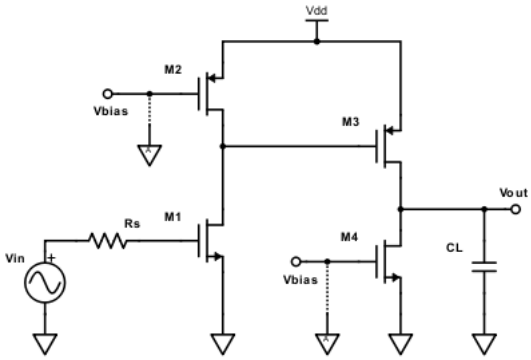

Fig. 9 An example. The dashed grounds on M2 and M4 are just friendly reminders that the gates are AC grounded.

Well, let's get started! Note that there are three nodes in this circuit: the input, the output, and the middle node, which we will label as Node A. We will therefore have to find the equivalent capacitance and resistance seen at each node, taking special care to recognize the miller caps. It's important to note that not all C_{gd} capacitors are "miller caps"- only those in the signal path, and with a gain. For instance, C_{gd2} and C_{gd4} are *not* miller multiplied because they are not in the signal path.

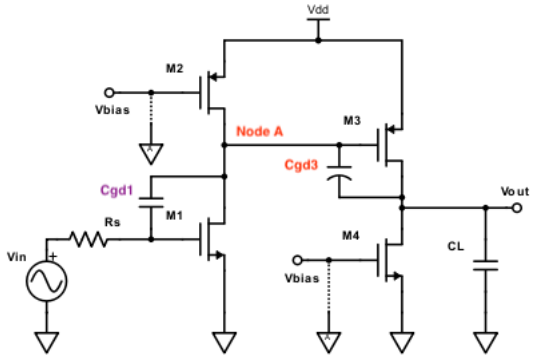

Fig. 10 Figure 9, with the middle node and the miller caps highlighted.

Let's find each time constant. The input pole is similar to that of (7), albeit with increased gain due to an active load. Assume $g_m r_o \gg 1$.

$$\therefore \quad \tau_{input} = R_s(C_{gs1} + C_{gd1}(1 + g_{m1}(r_{o1} \parallel r_{o2})))$$

Moving on our list: To Node A! Note that C_{gd3} is an input-referred Miller multiplied cap[4].

$$\tau_A = (r_{o1} \parallel r_{o2})(C_{gd1} + C_{db1} + C_{gd2} + C_{db2} + C_{gd3}(1 + g_{m3}(r_{o3} \parallel r_{o4})))$$

[4] If needed, draw the capacitors in yourself to see where they come from.

And the grand finale:

$$\tau_{output} = (r_{o3} \parallel r_{o4})(C_L + C_{gd4} + C_{db4} + C_{gd3} + C_{db3})$$

Now, which is the dominant pole (and what is the danger we were told to look out for)? First things first: an important skill to learn is how to recognize and remove the fluff. We have a ton of caps here- which ones matter more than the others? Basically what we need to look for are the miller caps: they are the ones multiplied by the gain of each stage- they can be way more than ten times greater than the others! Re-writing the above time constants into a simplified form:

$$\tau_{input} \approx R_s(C_{gd1}(1 + g_{m1}(r_{o1} \parallel r_{o2})))$$
$$\tau_A \approx (r_{o1} \parallel r_{o2})(C_{gd3}(1 + g_{m3}(r_{o3} \parallel r_{o4})))$$
$$\tau_{output} \approx (r_{o3} \parallel r_{o4})C_L$$

Let's first look at the output node. It has a high output impedance, on the order of r_o, but it has no miller multiplied capacitor; unless the load is outrageously large (over say 10-100pF), this likely will not be a problem.

What about the input node? Notice that although it contains a miller multiplied C_{gd1}, its resistance is only R_s, ideally low, in the order of 50Ω. The input τ is therefore often the smallest.

Now look carefully at Node A. Notice that it has *both a high impedance and a miller multiplied capacitor*. It will therefore produce a very low frequency pole! This is what we must be aware of. Notice that the circuit given here provides a fantastic low frequency gain $A_v \propto (g_m r_o)^2$, but due to Node A it will have a slower response. For instance, say $r_o = 100k\Omega$, $A_{v3} = 40$, and $C_{gd} = 0.5pF$; this would result in a pole at 800kHz- not exactly high speed[5]. Beware the miller cap on a high impedance node!

$$\therefore \quad \tau_A > \tau_{output} > \tau_{input}$$

3 The Cascode: Taming the Miller Cap

What if there was a way to have high gain without sacrificing speed? Well, you guessed it- there is! Enter the Cascode.

Let's first find the low frequency small signal gain. First step: G_m. Note M1 produces a current $g_m v_{in}$, which immediately becomes divided by the r_o of M1 and the input impedance of M2, on the way to the output.

$$G_m = g_m \frac{r_o}{\frac{1}{g_{m2}} + r_o} \tag{10}$$

[5] When you calculate the pole, don't forget the 2pi factor to convert to Hz!

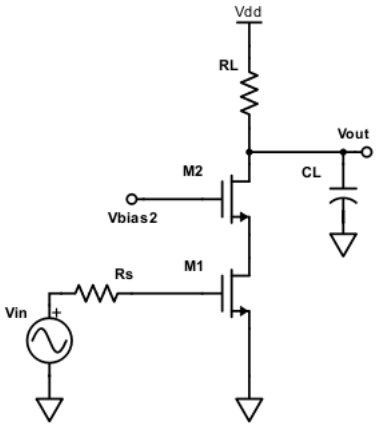

Fig. 11 The Cascode: a common source (M1) and common gate (M2) working in tandem.

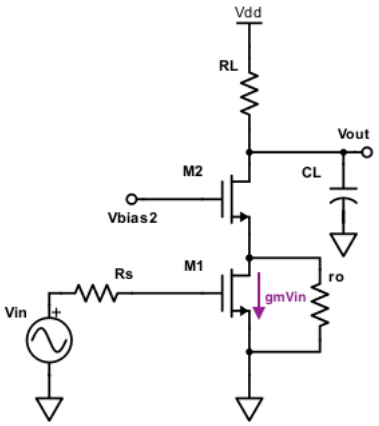

Fig. 12 Note that if we were to find G_m, M1 creates a current $g_m v_{in}$ that flows directly to the output, if $r_o \gg 1/g_{m2}$.

And since $r_o \gg 1/g_m$ almost always:

$$G_m \simeq g_m \tag{11}$$

For the output impedance, we have an interesting case: we have RL in parallel with the drain of M2, except notice that the source of M2 is connected with the drain of M1! Remember the full equation?

$$R_{out,drain} = r_o + (1 + g_m r_o) R_s \tag{12}$$

Notice now we have an "R_s," in this case the drain of M1! Therefore $R_{s,2} = r_{o1}$, and

$$R_{out,drainM2} = r_{o2} + (1 + g_{m2} r_{o2}) r_{o1} \approx g_{m2} r_{o2} r_{o1} \tag{13}$$

This is in parallel with R_L of course.

$$\therefore \quad R_{out} = R_L \parallel g_{m2} r_{o2} r_{o1} \implies \quad A_v = gm(R_L \parallel g_{m2} r_{o2} r_{o1}) \tag{14}$$

Notice something interesting? The output impedance of this circuit could be *potentially higher than r_O by a factor of $g_m r_o$*! Unfortunately using the resistor R_L spoils all of that, so we instead use the circuit seen in **Fig 13.**, where we we use the stacked active load instead. Assuming the g_m and r_o of each device are the same, we can see that looking both up and down from the output we have the same resistance. Therefore:

$$A_v = \frac{1}{2}(g_m r_o)^2 \tag{15}$$

We have almost squared our gain, for no cost in power (we can use the same current and supply as a common source)! Now, you may wonder what exactly is stopping

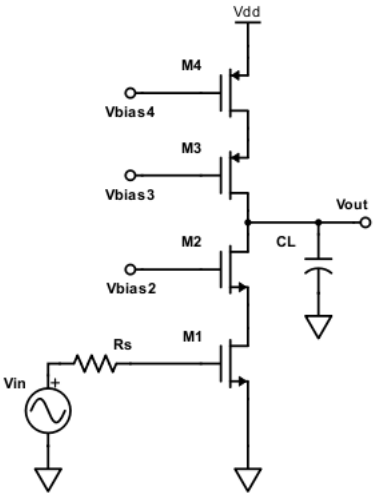

Fig. 13 The actively loaded cascode, made for maximum gain.

us from stacking more and more transistors to get more and more gain! Well, unfortunately voltage headroom issues and speed will come into play.

Frequency response Let's look at **Fig. 14** and determine its poles. Let's first do the output pole, assuming $R_L < r_o$, and $C_L \gg C_{gd}$:

$$\tau_1 = R_L C_L \tag{16}$$

Now for the more interesting point: what about C_{gd1}? Notice now that the drain of M1 is now a *low impedance node*, due to the source of M2. If $R_L \ll r_o$, then we know that this impedance is just $1/g_{m2}$; the gain across M1 therefore is g_{m1}/g_{m2}.

$$\therefore \quad \tau_{input} = ((1 + \frac{g_{m1}}{g_{m2}})C_{gd} + C_{gs})R_s \tag{17}$$

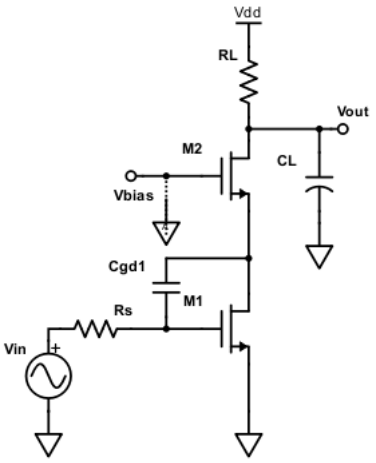

Fig. 14 Frequency Response of the Cascode. Notice that the miller cap C_{gd1} has now been interfaced with the *low impedance* source of M2.

Notice the difference? Because of the low impedance node, we have essentially removed the miller-effect: doubling C_{gd} instead of multiplying it by a factor of say 20.

Frequency Response of the Actively Loaded Cascode What about the actively loaded cascode? Do we still mitigate the miller capacitance?

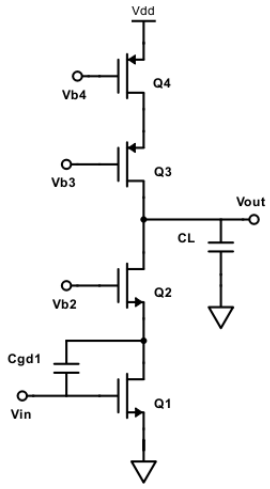

Fig. 15 Frequency Response of the actively loaded Cascode. Is the source of M2 still a low impedance node?

Let's first find the output pole. Assuming all the g_m's and r_o's are equivalent:

$$\tau_{out} \approx \frac{1}{2} g_m r_o^2 C_L \tag{18}$$

Notice that this can become extremely large if the load cap is large; the active cascode therefore should not be used as an output stage unless the designer knows the capacitance the amplifier will be driving.

Now what we've been waiting for: the Miller cap. The question really is, is the source of M2 still a low impedance node? To find the input impedance, remember we *open the output*, and apply a test source. Because the output is an open, we can no longer say our equivalent "R_L" is small- remember, looking into the drain of M3 we see approximately $g_{m3} r_{o3} r_{o4}$. As mentioned in the last note-sheet, the full formula for the source impedance:

$$R_{source} = \frac{1}{g_m}(1 + R_d / r_o) \tag{19}$$

Plugging in $g_{m3} r_{o3} r_{o4}$ for R_d and assuming all all r_o's and g_m's are equal, we see that:

$$R_{S2} \approx r_o \tag{20}$$

And so as a result we will still have the Miller effect take place, since the gain across M1 will be $\frac{1}{2} g_m r_o$.

$$\implies \quad C_{in} \approx (1 + \frac{1}{2} g_m r_o) C_{gd1} + C_{gs1} \tag{21}$$

Not all hope is lost however. Notice first this is still the same as the common source- but we have much much more gain! And, again the input capacitance usually is interfaced with a low impedance (if not, make sure to buffer the input to make it low!). It turns out however, that we can actually make the input capacitance lower than that of the simple common source- remember, ultimately the parasitic capacitances depend on the W and L of the transistor- we can sacrifice some gain (since we have plenty) and make the transistors smaller![6]

Why was the Source of M2 not a high impedance node when we found the transconductance? This is a key thing to note: remember when we find the input resistance, we open v_{out}; when we find the transconductance however we must *ground the output*. This essentially removes the PMOS pair (M3-M4) from the circuit under question, and as a result, looking into the source we see only $1/g_{m2}$.

Voltage Headroom And Swing There is one downside to the cascode: as stated before, we have now decreased our swing (**Fig. 16**). Notice now we have limited the output swing between $[V_{dd} - V_{ov4} - V_{ov3}, V_{ov1} + V_{ov2}]$. This means proper biasing must be ensured to keep the overdrives to a minimum (100mV) for maximum swing.

[6] Of course there are downsides to smaller transistors besides less gain, such as an increase in offset voltages due to greater device variability. No one said this will be easy- it's an art-form.

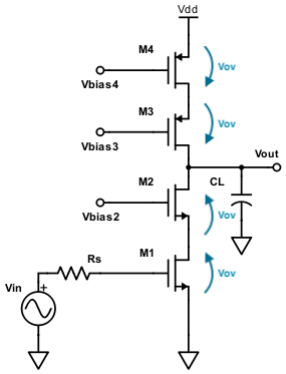

Fig. 16 Swing of the Actively Loaded Cascode

4 The Zero[7]

Question: Consider the common source, with only C_{gd}. What will happen to the transfer function as ω approaches infinity? *Answer*: Eventually C_{gd} will short the output to the input and the transfer function will approach one (feed-forward path). *Implication:* this implies that there must be a *zero* in the circuit-why? Because remember before the gain was inverting- now it is not! Therefore the must be some intermediate frequency where the gain was zero![8][5]

Where is this zero? It comes from the moment the feed-forward path begins to effect the transfer function. Consider **Fig. 17**. The zero occurs when the current

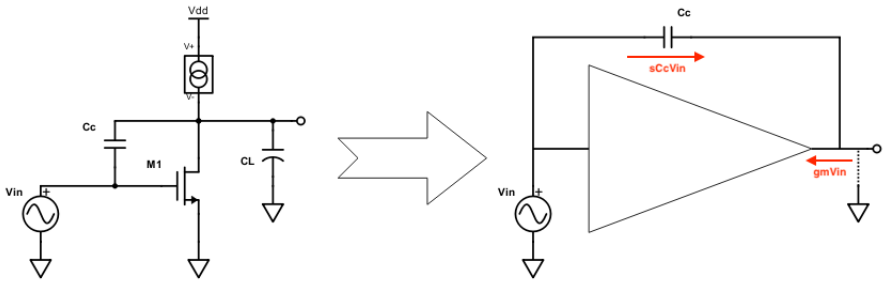

Fig. 17 Analysis of the right hand plane zero. We consider M1 as a generic gain block as before. Note the output is grounded for a transconductance calculation.

[7] This section can be skipped, or read just for interest, and be reviewed later when covering compensation. It is however a very important concept relating to feedback stability and transient responses, and is often asked about in job interviews. This of course is a rather advanced subject.

[8] Note that I stated: consider only C_{gd}; in reality C_{gs} and the substrate caps will short as well, which pulls the transfer function towards zero. The point being made here is however still valid.

through C_c (think C_{gd}) equals that coming from the transistor itself. This occurs if:

$$sC_cV_{in} = -g_mV_{in} \tag{22}$$

$$\implies \quad \text{Zero at: } s = \frac{+g_m}{C_c} \tag{23}$$

Notice that we have a *right-hand plane zero*, meaning it will cause the phase to continue to decrease with increasing frequencies. This can become problematic for stability reasons (which we will cover later).

5 What about Phase?

Yes, poles and zeros do indeed have a phase response, which turns out to be extremely important for concerns such as stability. For instance, if our amplifier has a phase of -180°, think: what would happen if we connect it in feedback? This leads to an important topic in circuits, known as *compensation*, which would take up an entire notesheet on its own (and probably more). It will be covered in more advanced notes/courses.

6 Summary

We covered a lot in this section. We found out how to use a simple method to approximate the poles in complicated circuit (to what turns out to be surprisingly good accuracy). We discovered that the miller effect can have a profound effect, and that the cascode offers a way to mitigate the miller effect. We also had a quick glance at zeros in circuits. So put away your algebraic fears- you don't have to do crazy nodal analysis anymore!

References

1. LT1008 Datasheet. http://cds.linear.com/docs/en/datasheet/1008fb.pdf
2. OCTC wikipedia example. https://en.wikipedia.org/wiki/Open-circuit_time_constant_method
3. Paul R. Gray, Hurst P J Lewis S H & Meyer RG (2001). Analysis and design of analog integrated circuits (Fourth ed.). New York: Wiley. p. 7.3.2 pp. 517520. ISBN 0-471-32168-0.
4. Miller's Theorem, Wikipedia. https://en.wikipedia.org/wiki/Miller_theorem
5. Great article on RHPZ in edn: Demystifying the RHPZ. https://www.edn.com/electronics-blogs/analog-bytes/4419561/Demystifying-the-RHPZ