Rail to Rail Folded Cascode Opamp Employing Class AB Output Stage

Imre Knausz, Analog IC Design, EE726

Table of Contents

Introduction

This paper will describe a high performance opamp architecture suitable for driving moderate off chip capacitances. This opamp is suitable for driving LCD panels. Below is a high level model of the opamp and it's load.

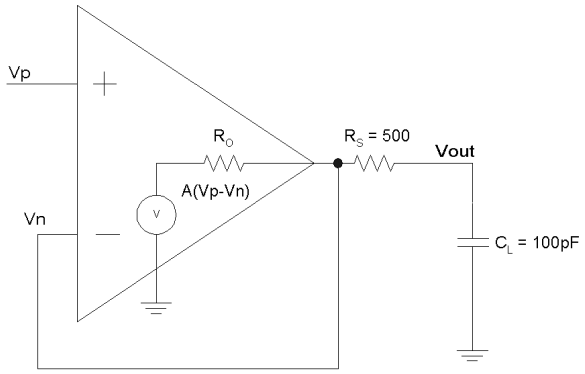

Figure 1: Ideal view of opamp and load.

Because the circuit can be used to drive LCDs, this amplifier has to be as low power as possible to enable efficient use of the battery in a portable application. The opamp will be designed towards the following specifications and simulated to verify adequate circuit operation.

- Power Consumption < 50µA (<250µW across supply voltage)
- Power Supply = 3.3 to 5V
- Low offset = ±2mV (1σ)
- Settling time (1% of final value) = 250ns
- Slew rate = 25 V/µs (C_L = 100pF, RS = 500Ω)
- Rail to rail input
- Rail to rail output (actually ±200mV from any rail)

Opamp Topology

The following is a schematic diagram of the opamp that was constructed:

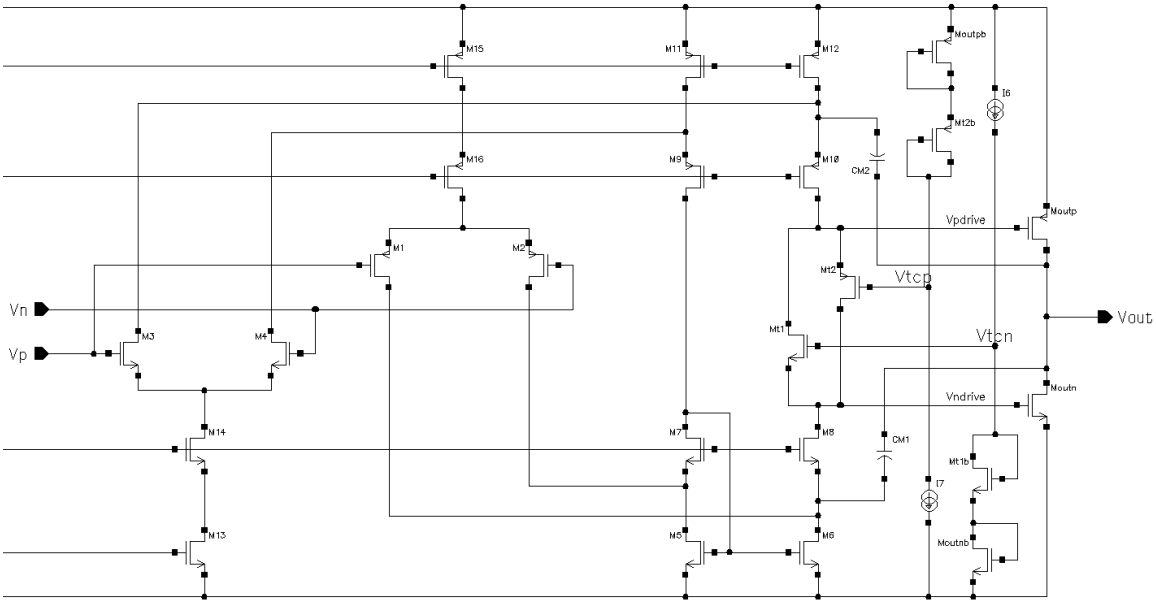

Figure 2: Simplified opamp schematic

First Stage

The first stage of the opamp is a folded cascode with a dual polarity front end. The folded cascode is configured the same as a standard folded cascode with a PMOS input, except this opamp has an extra NMOS input pair pulling current out of current mirror M11 and M12. The output of this stage is the drain of the M10 device and the drain of M8. Mt1 and Mt2 provide a function for the output stage as will be described later. These two devices form a low impedance path from the drain of M10 to M8. This is a low impedance path because we can consider Mt1 & Mt2 as common gate amplifiers and the input impedance (i.e. impedance looking into the source) of such a configuration is $1/g_m$.

A downside of using such a configuration is that the g_m of the input stage is not constant through the operating range. The input circuitry can be separated into three different operating ranges. While the common mode range is in or near the center, both differential pairs are active and the effective g_m of the input stage is $g_{m\text{-}n} + g_{m\text{-}p}$. If the input common mode voltage drops below $V_{GS\text{-}N} + 2*V_{ON\text{-}N}$, the NMOS differential pair is off and only the PMOS input pair is active and thus the total g_m equals $g_{m\text{-}p}$. Similarly, the PMOS differential pair turns off with high common mode voltages and the g_m once again drops to about half of it's maximum value (assuming the NMOS and PMOS devices were sized to have the same g_m). The main problem this variation of g_m causes is that it can move the unity gain frequency around and this makes it difficult to optimally compensate the opamp. In this application, we will not employ a constant g_m circuit because the load will cause difficulties in compensating the opamp which we will solve later in the compensation section.

For completeness, I will mention an effective method for correcting this g_m variation. The easiest method is to boost the current though the active differential pair (when the other differential pair is off). To double the g_m of remaining differential pair, we need to quadruple the tail current since g_m is proportional to $\sqrt{I_D}$. We can do this by injecting an additional $3*I_{tail}$ into the active pair. This assumes the input pair is in saturation, if the input pair is operating in subthreshold (weak inversion) region to save power, then only 1x extra tail current needs to be injected because in this mode of operation, the transistor's g_m is directly proportional to it's drain current.

Class AB Output Stage

Three basic choices are available for output stages – class A, class B and class AB. Class A output stages are normally used in classic '7 transistor' opamps. Their problem is that they waste power and the output slew rates are asymmetric. Slewing in one direction is done by the output device (usually a common source device), but slewing in the other direction is performed by an active load which is just a

current source. The current source device's gate is tied to a constant bias voltage, so the current output of the device is limited. The second choice is a class B output stage which is basically two common source output stages with the output of one stage functioning as a load for the other and vice-versa. The gates of the output devices are driven by the first stage of the opamp. The problem with this approach is that PMOS and NMOS devices require a different DC bias voltage on the gate. Because of this, class B output stages require some kind of feedback to avoid this unacceptable crossover distortion.

The third option is to employ a class AB output stage [1, p. 150]. A diagram below details operation.

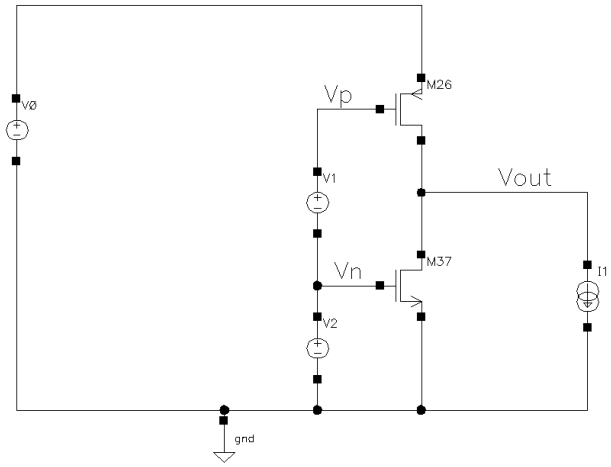

Figure 3: Simplified view of class AB output stage.

Voltage source V1 represents the floating voltage source which is required to properly bias the output devices. This scheme is called class AB because it is a cross of class A and B. The output devices are biased with a current, just like class A, but when the amplifier needs to sink or source a large amount of current, it can drive the gate of the NMOS or PMOS to either rail to overdrive the NMOS or PMOS device. Because of the floating voltage source, even when one device if being overdriven, the other device still has it's quiescent current flowing through it. The advantage of this is that neither device is completely turned off and the amplifier can respond quicker to transients, reducing crossover distortion. The simulation result below illustrates the biasing scheme:

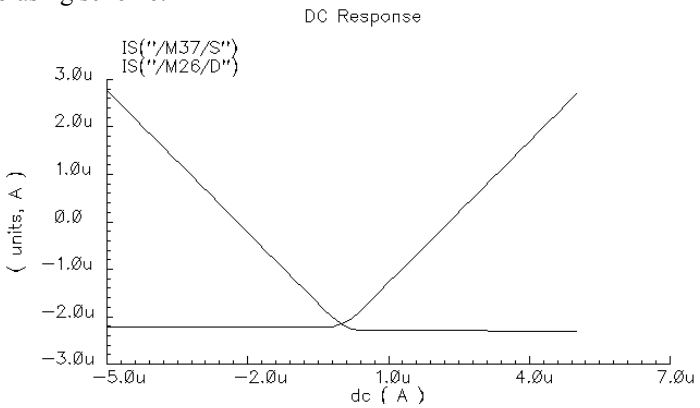

Figure 4: Output current versus output device currents

The following is a zoomed in view of our implementation:

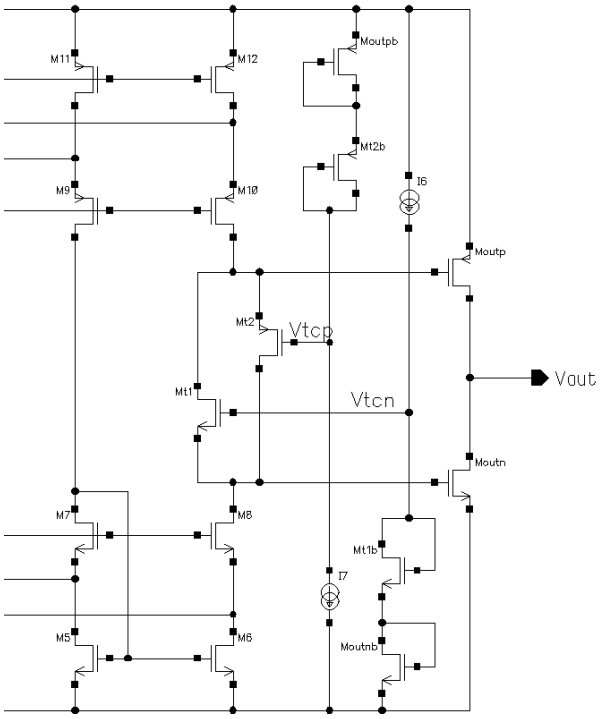

Figure 5: Zoomed in view of the class AB implementation.

Mt1 & Mt2 create the floating voltage source. Their bias voltages are created by the diode stacked devices Moutnb & Mt1b and Moutpb & Mt2b respectively. This configuration forms a translinear loop. The NMOS side of the loop is formed by diode stacked devices connecting to the gate of Mt1 down to the gate of Moutn. In terms of quiescent biasing, we can write a voltage loop equation to obtain:

$$V_{gs-outnb} + V_{gs-t1b} = V_{gs-t1} + V_{gs-outn} \tag{1}$$

The threshold voltages all cancel out of the equation and we are left with:

$$V_{ON-outnb} + V_{ON-t1b} = V_{ON-t1} + V_{ON-outn} \tag{2}$$

which reduces to:

$$\sqrt{\frac{I_{D-outnb}}{\left(\frac{W}{L}\right)_{outnb}}} + \sqrt{\frac{I_{D-t1b}}{\left(\frac{W}{L}\right)_{t1b}}} = \sqrt{\frac{I_{D-t1}}{\left(\frac{W}{L}\right)_{t1}}} + \sqrt{\frac{I_{D-outn}}{\left(\frac{W}{L}\right)_{outn}}} \tag{3}$$

We will use equation 3 to size the devices because we will likely have larger currents in the output than in the folded cascode devices. We might also want to minimize the bias currents through Mt1b & Moutnb.

To understand how the "floating voltage source" function of this biasing scheme works, we will examine the large signal characteristics of the output stage [2, p. 1507]. Consider M5 as one current source and M12 as the other. In-phase input signals coming from the differential pairs into the drains of M5 & M12 will control the floating voltage source. Consider the case where current ΔI is pulled. Mt1's current in increased by ΔI whereas Mt2's drain current decreases by the same amount. This will cause the gate voltages of the output devices to move down. This is because the gate to source voltages Mt1 & Mt2 increase due to the increased current, but their gate voltages remain constant. This will start to reduce the current through Moutn and increase the current though Moutp, causing the opamp to source current. This

trend will continue until all of the current flows through Mt1. At this point some minimum current will flow through Moutn which is set up by the sizes of the devices in our translinear loop.

Also important to note is the fact that this output stage has some component of quiescent power that is supply voltage dependent. This is mainly due to the fact that the output devices don't have the same drain to source voltages as their diode connected bias counterparts. This will result in channel length modulation and one can expect 10-20% more current in the output stage than calculated.

Opamp Gain

The gain expression is relatively easy to find since we can consider each stag individually. The gain of the first stage can be found using the standard folded cascode gain expression, the only difference being that we have two sets of input devices. We can combine the effects of both by using superposition resulting in:

$$A_{V1} = \left(g_{m1,2} + g_{m3,4}\right) \cdot \left[\left(g_{m8} r_{o8} \left(r_{o6} \parallel r_{o3}\right)\right) \parallel \left(g_{m10} r_{o10} r_{o12}\right)\right] \tag{4}$$

Next, we can find the gain of the output stage using superposition. If we consider the output stage as two common source amplifiers (using each other as their load), then it is easy to see that their gain is:

$$A_{V2} = \left(g_{m-outn} + g_{m-outp}\right) \cdot \left(r_{o-outn} \parallel r_{o-outp}\right) \tag{5}$$

Thus the total gain is:

$$A_{V1} = \left(g_{m1,2} + g_{m3,4}\right)\left(g_{m-outn} + g_{m-outp}\right) \cdot \left[\left(g_{m8} r_{o8} \left(r_{o6} \parallel r_{o3}\right)\right) \parallel \left(g_{m10} r_{o10} r_{o12}\right)\right]\left(r_{o-outn} \parallel r_{o-outp}\right) \tag{6}$$

Compensation

This amplifier makes use of cascoded Miller compensation. This variation of Miller compensation allows us to use a smaller compensation capacitor since our load capacitor is relatively large. Normal Miller compensation give us an output pole of:

$$\omega_{out} = \frac{g_{m0}}{C_L} \tag{7}$$

The schematic below shows how to use cascode Miller compensation:

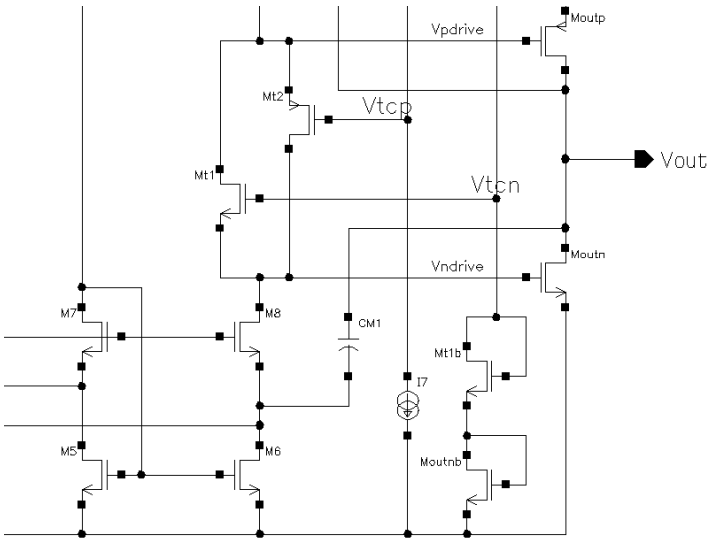

Figure 6: Cascode Miller compensation schematic.

Above we can see the class AB devices Mt1 & Mt2 and the current mirror M5-8. To employ cascode Miller compensation, we hook up one end of the compensation capacitor to the output and the other end of the capacitor to the drain of the current mirror device M6. To see what this does for us, we can write this simplified small signal model of half of the output stage:

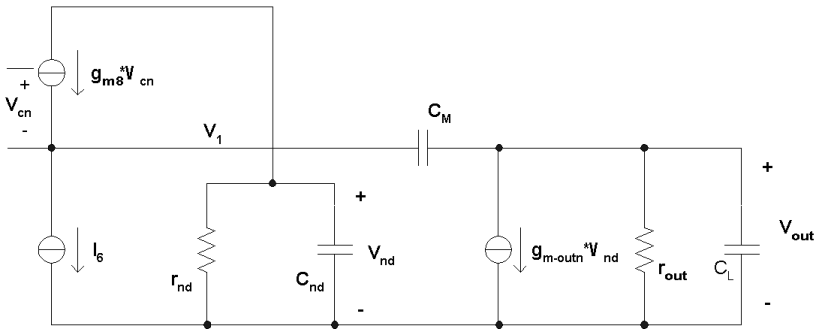

Figure 7: Small signal model of output stage

r_{nd} represents the total resistance at the V_{ndrive} node and r_{out} represents the resistance at the output node. Current source M6 is represented by current source I6 and the cascode device M8 is driven by it's gate to source voltage, V_{cn}. We consider I6 to be the input signal current. Also, since V1 is such a comparatively low impedance node, we can consider it to be a virtual AC ground. Therefore, we will write an expression for Vout:

$$V_{out} = -g_{m-outn} \cdot V_{nd} \cdot Z_{out} = -g_{m-outn} \cdot V_{nd} \cdot \frac{\frac{1}{s(C+C_L)} r_{out}}{\frac{1}{s(C+C_L)} + r_{out}} \tag{8}$$

Which then simplifies to:

$$V_{out} = -g_{m-outn} \cdot V_{nd} \cdot \frac{r_{out}}{1 + s \cdot r_{out}(C+C_L)} \tag{9}$$

to find an expression for $V_{nd,}$ we can write a node equation at node V1:

$$-g_{m8} \cdot V_{cn} + I_6 + (V_1 - V_{out}) \cdot sC = 0 \tag{10}$$

since the AC current through M8 is the same as that through rnd & Cnd,

$$\frac{V_{nd}}{z_{nd}} + I_6 + (V_1 - V_{out}) \cdot sC = 0 \tag{11}$$

Since V1 is at a virtual ground, the above becomes:

$$\frac{V_{nd}}{z_{nd}} + I_6 - V_{out} \cdot sC = 0 \tag{12}$$

substituting for z1 and solving for Vcn we get:

$$V_{nd} = (-I_6 + V_{out} \cdot sC) \frac{r_{nd}}{1 + sC_{nd} r_{nd}} \tag{13}$$

Plugging this expression into (9) and multiplying through we get:

$$V_{out} = \frac{I_6 r_{nd} g_{m-outn} r_{out} - V_{out} s C_M r_{nd} g_{m-outn} r_{out}}{1 + s C_{nd} r_{nd} + s r_o (C + C_L) + s^2 C_{nd} r_{nd} r_{out} (C + C_L)} \tag{14}$$

Now we rearrange to solve for the transfer function:

$$\frac{V_{out}}{I_6} = \frac{r_{nd} g_{m-outn} r_{out}}{1 + s(C_{nd} r_{nd} + r_o (C + C_L) + C_M r_{nd} g_{m-outn} r_{out}) + s^2 C_{nd} r_{nd} r_{out} (C + C_L)} \tag{15}$$

Assuming the poles are far apart, i.e. p1 >> p2, we can write them as [3, p. 618]:

$$p_1 \cong \frac{-1}{C_{nd} r_{nd} + r_o (C + C_L) + C_M r_{nd} g_{m-outn} r_{out}} \cong \frac{-1}{C_M r_{nd} g_{m-outn} r_{out}} \tag{16}$$

$$p_2 \cong \frac{C_{nd} r_{nd} + r_o (C + C_L) + C_M r_{nd} g_{m-outn} r_{out}}{C_{nd} r_{nd} r_{out} (C + C_L)} \cong \frac{-C_M}{C_{nd}} \cdot \frac{g_{m-outn}}{C_L} \tag{17}$$

This is a very positive result. P_2 has bee increased by a factor of C_M/C_{nd} which pushes out the unity gain frequency of the opamp and increases it's performance. Alternately, to save space or to increase the slew rate, we can decrease C_M and still maintain the same unity gain frequency we would get with conventional Miller compensation. Another very positive result of this analysis is that no RHP zero results from this form of compensation!

Offset Calculations

The majority of the offset of this amplifier results from 4 main components – the offset of each differential input pair, the offset of the current source ($M_{9\text{-}12}$ figure 1) and the current mirror ($M_{5\text{-}8}$ figure 1). The standard deviation is equal to the square root of the sums of the variances of the components:

$$\sigma_{OS} = \sqrt{\sigma_{os-thn}^2 + \sigma_{os-thp}^2 + \sigma_{os-cs}^2 + \sigma_{os-cm}^2} \tag{18}$$

The standard deviations of threshold voltages are normally available for each process, however, we don't have access to this AMI 0.5u data. I will take a guess at what the fab data might look like based on other similar processes. Threshold voltage is linearly proportional to the inverse of the square root of the gate area [4]. The following chart illustrates the concept.

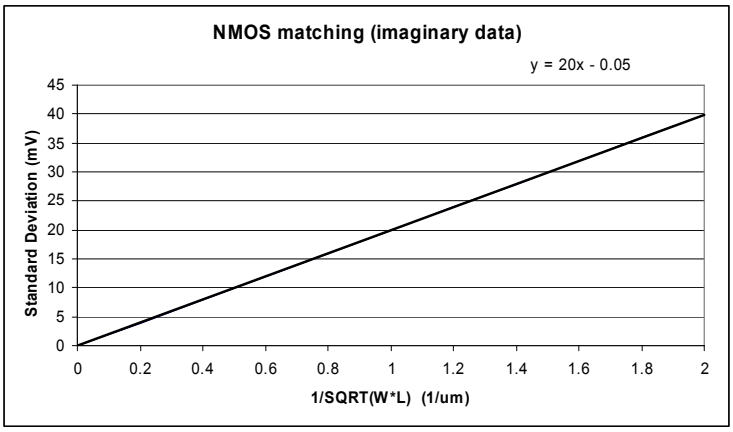

Figure 8: Sample matching data

For this paper we will use a slope of 20mV for NMOS devices and a slope of 25mV for PMOS devices.

Next we will calculate the input referred offset voltages of the current source. Let's define $\Delta I = I_1 - I_2$ which is the difference in current from one leg of the current source (or mirror) to the other. In the current mirror,

$$\Delta I = g_m \cdot \Delta V_{th} \tag{19}$$

In a feedback configuration, the opamp will try to offset this current mirror mismatch and this action will manifest itself as an input referred offset. The amp will have an offset voltage proportional to the current offset that it must account for and this constant of proportionality is g_m. Therefore,

$$V_{os} = \frac{g_{m-cs}}{g_{m-1,2}} \cdot \Delta V_{th} = \sigma^2_{os-cs} \tag{20}$$

The variance of the current mirror input referred offset voltage is found in the same way.

A very important observation is that the opamp's offset is not constant over the input common mode range. A the input common mode goes down, the NMOS differential pair will turn off and it's offset will no longer contribute to the total. Also, the g_m's of the current source and the current mirror will evaluate to different values since those devices now have different current. The converse is also true for high input common modes when the PMOS differential pair is off.

Implementation

The first thing that will be done is to find the amount of current we need in the folded cascode stage to slew the input of the second stage. We will assume that the output stage loads the first stage with about 0.25pF of capacitance. The slew rate is defined by the following expression:

$$SR = \frac{I_b}{C} \tag{21}$$

Our total slew rate budget is 250ns, so we will assume we can use half of the budget for internal slewing. We only need to slew about 3V in any direction before the output devices are turned on hard, therefore our internal slew rate spec should be 3/125ns = 24V/us. Given this, I_b comes out to be 6uA.

Device sizes in the folded cascode are not critical in this application (in terms of gain), we can pick devices as large as possible (but making sure devices are still in the inversion region). We will use a V_{ON} of 200mV to pick original sizes and increase the sizes if need be. The differential pair needs to be sizeable for matching.

The output stage needs to be sized to slew the load capacitance. The output stage has to slew 5V in 125ns, therefore, its required slew rate is 40V/us, and this will require a current of 4mA. Since the output devices are in the triode region, the devices can be sized using:

$$\left(\frac{W}{L}\right) = \frac{I_D}{K' \cdot (V_{GS} - V_{th}) V_{DS}} \tag{21}$$

Using this equation (with V_{GS}=4V and with V_{DS} equal to 200mV, to make sure we can slew quickly across our full range), $(W/L)_{outn}$ comes out to be 55. Tripling this gives us 165 for the PMOS device.

Equations 16 & 17 show that the poles are independent (with respect to each other, i.e. they both move by the same amount due to a change in bias current) of the bias current in the output stage. The devices will respond faster if they are biased in the active region.

Simulation Results

References:

1. Roubik Gregorian, *Introduction to CMOS Opamps and Comparators*. John Wiley & Sons, Inc., New York, 1999
2. Hogervorst *et al.*, "A Compact Power efficient 3V CMOS Rail-to-Rail Input/Output Operational Amplifier for VSLI Cell Libraries," *IEEE J. Solid-State Circ.*, vol. 29, pp. 1505-1512, Dec. 1990.
3. Paul R. Gray and Robert G. Meyer, *Analysis and Design of Analog Integrated Circuits*. John Wiley & Sons, Inc., New York, Third Edition, 1993
4. Marcel J. M. Pelgrom, "Matching Properties of MOS Transistors," *IEEE J. Solid-State Circ.*, vol. 24, No. 5, pp. 1433-1440, Oct. 1990.