Design of a High-speed CMOS Fully Differential Op-amp

By Xiyao Zhang

Abstract

A high gain (100dB), high-speed (400MHz) and wide output swing (>1.2V) CMOS fully differential Operational Amplifier (op amp) is designed using 180 nm technology and its various parameters are simulated by Spectre®. Two generations of op amp are implemented and both of them are described in this report, but with emphasis on the final version. We summarize the lessons learned from the initial design and discuss how it helps the final op-amp to exceed several specifications and reduce the power consumption and noise at the same time. In addition, nulling resistor to cancel the second pole is implemented by a MOSFET operating in deep triode region and Common Mode Feedback (CMFB) is used to increase Common Mode Rejection Ratio (CMRR), improve Input Common Mode Range (ICMR) and define output common mode level.

I. Introduction to CMOS Fully Differential Op Amps:

Among all the analog circuits, Operational Amplifier (op amp) is perhaps the most fundamental and commonly used one, which is usually an integral part of other analog and mixed signal systems. CMOS op-amps often have a lower transconductance for a given current, lower gain, slower speed, higher input-referred offset and input-referred noise voltage than their BJT counterparts[1]. However, analog and digital circuit are now often integrated together -- signals that originate in analog form often need to be interfaced to the digital circuits in many complex systems, such as A/D and D/A converters. With CMOS technology becoming dominant because of its smaller size and less power dissipation, the integration requirement of mixed signal systems provide a strong economic incentive to use CMOS op-amps.

Unlike conventional signal-ended op amps, fully differential op-amps have a differential input and produce a differential output. They have some advantages over the signal-ended op amps and are widely used in modern CMOS integrated circuits. For example, by generating “complementary” outputs on both differential arms, fully differential op amps provide roughly twice output swing as their single-ended counterparts. And achieving a large swing becomes more and more important in the modern CMOS circuit design as the supply voltage decreases. Besides, as a balanced circuit with perfectly symmetric components on each side, fully differential op-amps are less susceptible to common-mode noise and even-order nonlinearities. A disadvantage of this topology is that it usually needs a common-mode feedback circuit to control the common-mode output voltage, which increases the design complexity.

In the following sections of this report, we discuss the design process of a fully differential op amp satisfied certain specifications and show related simulation results. The topology choice prior to the actual design is explained in Section II. Section III describes the initial version of this fully differential op amp, summarizes its pros and cons, as well as the lesson we learned. The final version of our op amps is discussed and a complete table summarizing the simulation results of both versions of op amp
vs. design specifications is included in Section IV. Section V is on the implementation of nulling resistor Rz by a NMOS operating in deep triode region and common mode feedback (CMFB). Finally, we conclude this report and make suggestions on how to further enhance this op amp’s performance in Section VI.

II. Design Considerations and Topology Choice:

A). Topology: Two-stage vs. Gain Boosting

Before starting the op amp design, we reviewed various topologies in Chapter 9 of Ref [2] and their advantages and disadvantages listed in Table 9.1. According to our design specifications, this op-amp should have a high gain (90 dB/30 K), large output swing (>1.2 V) and wide bandwidth (200 MHz), a single stage telescopic or folded cascode can hardly meet the specs. These two topologies are applicable only if we use a very small tail current to boost the gain and implement wide swing cascode configuration to achieve large output swing. The advantage of single stage cascode topology is that it doesn’t have a second pole to control and satisfaction of the phase margin requirement become easier. However, small tail current greatly limits the slew rate \( SR = \frac{I_{SS}}{C_L} \), which is an important portion of the settling time (10 ns in specs). Beside, the wide output swing topology need an additional current branch and at least four more transistors, which increase both power consumption and design complexity.

Therefore, we should choose from the two popular high gain topologies, i.e. two-stage and gain boosting. Compared to gain boosting, two-stage topology has a larger output swing and lower power dissipation and input referred noise. Since the supply voltage of 180 nm standard CMOS technology is merely 1.8 V, the output swing of 1.2 V in this design should be one of the main considerations in choosing topologies. Although power and noise is not our primary concern, it always better to have them reduced. So two-stage topology seems more favorable for this particular op amp design. Gain boosting is also possible to achieve larger swing level by using wide swing configuration, but again it is extra power and complexity.[3]

On the other hand, the disadvantage of two-stage is its narrower bandwidth, and more important, its second pole which might cause stability problems. Normally the second pole must be pushed far enough (for example, twice of gain bandwidth product, GBP) to raise the phase margin to 60 degrees. This is achieved by adding compensation capacitor and increasing the transconductance of second stage input transistor. However, since the load capacitance is well-defined (4 pF) in our application, we should also be able to just move the first zero of the op amp from right half plane (RHP) to left half plane (LHP) and cancel the second pole, without worrying too much about load dependency of the nulling resistor.[2]

Based on these analysis analyses, two-stage topology with first-stage high gain and second-stage high output swing is chosen. Gain boosting is the back-up plan, in case that controlling of the 2nd pole becomes infeasible.

B). Output Stage: Differential vs. Single-Ended
Basically, there are two choices for the output stages: differential or single-ended (using no tail transistor, real ground instead virtual ground). Apparently differential topology helps reduce the electromagnetic interference (EMI) noise and increase CMRR. For single-ended setup, the differential mode gain \((A_{v,\text{dm}})\) and common mode gain \((A_{v,\text{cm}})\) is identical and its CMRR is essentially 0 dB. To raise the CMRR in an op amp with single-ended output stage, it is necessary to add CMFB to the first stage to compensate the large \(A_{v,\text{cm}}\) of the output stage. The disadvantage of differential configuration is smaller output swing because the tail device takes a part of the total allowable overdrive voltage. Besides, if the tail transistors are too big, they may act as a large capacitor being charged and discharged when the input is a step function, thereby causing overshooting and jeopardizing the settling time of our op amps.

Hence, we choose differential output stage for its simplicity of design and will try to size the tail devices as small as possible to avoid overshooting. When the settling time becomes an issue, we’ll use single ended output stage with CMFB to solve the problem.

**III. The First Generation Design and Lessons Learned:**

**A). First Generation Design**

With the circuit topology chosen to be two-stage with differential output stage, we design an initial generation of fully differential op amp, as shown in Figure 1(a). Stage 1 of this op amp is a differential folded cascode amplifier and stage 2 is a differential common source (CS) amplifier. The DC gain allocation is 3K for the first stage and 10 for the second stage: this is achieved by using longer transistors \((L = 4 \times \lambda)\) and small tail current of the first stage (100 uA). Overdrive voltage for each transistor is set initially as 200 mV and later some adjustments are made to reach the optimum operating point for 2nd stage output swing. The final operating points and node voltages is shown in Figure 1(b). Both compensation capacitor (800 pF) and nulling resistor (1.5 kohms) are added to raise phase margin up to 55 degrees. Since it is the first generation op amp, instead of going into the details of the design flow, we put emphasis on the lessons acquired from designing this op amp in the next part of this section.

**B). Lessons**

Simulation results (Figure 2) show that the first generation op amp meet all the specifications except the settling time of 10 ns. There are mainly three lessons we learned from designing this circuit:

1). Sizing of the 2nd stage input devices and tail MOSFETs:

Since the only unsatisfied spec is settling time, it is the most important thing to be improved. A slow settling may result from different reasons, such as ringing or overdamping, limited current (no enough first stage tail current to charge up the miller capacitor) or overshooting. In our op amp, phase margin is set as 55 degrees, which causes neither ringing nor overdamping. For the rest two reasons, small first stage current generally limits the slew rate and overshooting makes the stabilization of output worse, which both contribute to the settling time.
To increase the slew rate, we could either raise the first-stage current $I_{\text{st}}$ or reduce the total miller capacitance $C_{\text{miller}} (SR = I_{\text{st}} / C_{\text{miller}})$. The overall miller capacitance present between the input and output of second stage CS amplifier can be roughly calculated by

$$C_{\text{miller}} \approx (1 + A_{v2}) \times (C_{gd,15/16} + C_c).$$  \hspace{1cm} (1)

Although the compensation capacitor $C_c$ is only 800 fF, $C_{gd,15/16}$ is directly proportional to the size of transistor M15,16 and is quite large in this first generation op-amp. Assuming the gain is fixed, we should size M15,16 down to raise the slew rate. The tradeoff here is larger 2nd stage transistors usually provide large output swing and meanwhile sustain a high tail current. A higher tail current can push the second pole further from the origin and improve the frequency response of the op amp.

On the other hand, a super large tail NMOS M18 results in overshooting in this op amps, which dramatically hamper the circuit from stabilizing. As discussed in Section II, this tail NMOS has to be sized down, even at the expense of reducing the second stage tail current and assigning more overdrive voltage across it.

2). Design sequence of the 1st and 2nd stage operation points:

A natural thought of designing op amps is determining the first stage and then the second stage. However, from the author’s personal experience, a reverse order seems more reasonable for this design problem. Consider the output stage of our op amp, each output should swing at least 1.2 V under a 1.8 V power supply. Hence, the output common mode (CM) voltage should be round 1 V and the drain to source voltage ($V_{ds}$) across the tail NMOS need to be right above its overdrive voltage (i.e. at the triode edge, maybe 50 mV above it) to enable the maximum swing. And $V_{ds18}$ can be pushed down only by lowering the input CM level of the second stage. Because of these tight constrains, it is very difficult to tweak the transistor sizes to achieve optimum operating point of the 2nd stage after fixing its input CM level. Alternatively, we could design the 2nd stage first and then choose the 1st parameters accordingly.

### IV. The Second Generation Design and Summary Table of Simulation Results:

#### A). Second Generation Design

In this section, let us discuss the design procedure and some hand calculations of the final version op-amp. Figure 3(a) is its schematic of telescopic cascode plus common source two-stage topology and Figure 3(b) shows the DC operating points of this circuit. From the experience of our previous work, we learned that the second stage should be designed first to meet the output swing specification with a limited budget of transistor sizes. Following is the design flow with circuit parameters determined by necessary hand estimations and simulations:

1) Compensation Capacitor: Usually the capacitance of compensation cap can be estimated as $C_c \approx 0.22C_L [3]$. Plug in 4 pF as the load capacitance, the compensation capacitance needs to be
around 1 pF (which seems reasonable according to the previous work).

2) First stage tail current: As mentioned several times in this paper, slew rate (SR) of our op amp is an important parameter for it to meet 10 ns settling time spec. Assume the output takes one third of the settling time \((\approx 3\,\text{ns})\) to reach the desired value (say 1.2 V), the slew rate can be estimated as \(1.2V/3\,\text{ns} = 400\,\text{V}/\mu\text{s}\). Since \(SR = I_{SS1}/C_C\), we need \(I_{ss1}\) to be at least 400 uA.

3) Output Common Mode Voltage: The output CM level are usually set to a certain voltage to allow the maximum output swing. Assuming 200 mV overdrive is assigned to 2\textsuperscript{nd} stage tail NMOS M\(_5\), output CM voltage is given by

\[
V_{out,CM} \approx \frac{V_{DD} - V_{DS5,\text{min}}(sat)}{2} + V_{DS5,\text{min}}(sat),
\]

where \(V_{DS5,\text{min}}(sat)\) is the minimum voltage across M\(_5\) to keep it in saturation. According to equation (2), \(V_{out,CM}\) should be about 1 V.

4) Transistor sizing: We size the 2\textsuperscript{nd} stage transistors based on leaving large enough headroom. The 1\textsuperscript{st} stage transistors are sized long enough to provide high DC gain. Transistor size shouldn’t be too large for both stages, not only because of the slew rate and overshooting problems explained in Section III, but also because we want to keep the third pole P\(_3\) as far as possible. P\(_3\) mainly come from the internal capacitance at the first stage output and will determine the phase margin after P\(_2\) being canceled by Z\(_1\).

5) Nulling resistor R\(_z\): When the load capacitance of a two-stage CMOS op amp is well-defined, the right half plane zero can often be moved to the left half plane and canceled the second pole by adding a nulling resistor R\(_z\) in serial with compensation capacitor. This occurs if

\[
R_z = \frac{C_L + C_E + C_C}{g_{m1,2}C_C},
\]

where \(C_E\) is the internal capacitance at the output node of stage 1. The actual value of R\(_z\) is found by simulation sweep in this design.

6) Power vs. Frequency Response: Generally speaking, there is a trade off between a circuit’s performance and its power dissipation. In our design, we need to balance the power and frequency performance of the op amp. The second generation design consumes 40% less power than the first generation. It realizes DC gain of 100 dB, bandwidth of 400 MHz and 55 degrees phase margin at 400 MHz, as well as output swing larger than 1.2 V. It is possible to degenerate these performance parameters to cut the power dissipation further down. However, since the power is already decreased compared to the previous design, we choose to keep the performance.
7) Input Referred Noise: The second generation design also has better performance than the first generation. The reason is that the input referred noise (dominated by flicker noise) is normally proportional to W*L of the first stage input transistors and this is confirmed by our noise simulation. Compared to the initial version, the final design has a larger input device area and the telescopic topology itself can also improve the noise performance. The simulation result also shows which transistor contributes most to the noise, as listed in the summary table. We’ll return to this point later in the conclusion section.

B). Summary Table of Simulation Results vs. Design Specifications

Table 1 is a complete table summarizing the simulation results of both 1st and 2nd generation op amps vs. design specification of this project. All results are simulated with a 4 pF capacitive load and with 1.8 V power supply.

<table>
<thead>
<tr>
<th>Specifications</th>
<th>1st Generation</th>
<th>2nd Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Frequency Gain</td>
<td>90 dB</td>
<td>96 dB</td>
</tr>
<tr>
<td>Unity Gain Frequency</td>
<td>200 MHz</td>
<td>200 MHz</td>
</tr>
<tr>
<td>Phase Margin</td>
<td>55 degrees</td>
<td>55 degrees</td>
</tr>
<tr>
<td>Settling Time</td>
<td>10 ns</td>
<td>28 nS</td>
</tr>
<tr>
<td>Output Swing</td>
<td>±0.6 V</td>
<td>±0.6 V</td>
</tr>
<tr>
<td>CMRR</td>
<td>&gt;50 dB</td>
<td>63.5 dB</td>
</tr>
<tr>
<td>Input Referred Noise</td>
<td>No Spec</td>
<td>6.7 nV / √Hz</td>
</tr>
<tr>
<td>Noise Contributors</td>
<td>No Spec</td>
<td>M5/6 22.67%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M9/10 16.08%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M1/2 11.22%</td>
</tr>
<tr>
<td>Power Dissipation</td>
<td>No Spec</td>
<td>14.7 mW</td>
</tr>
</tbody>
</table>

Table 1 Simulation Results vs. Project Specifications

V. Implementation of Nulling Resistor and Common Mode Feedback:

A). Nulling Resistor

As mentioned in the previous sections, the nulling resistor is load dependent and has to be an exact value (given by equation 3) to cancel the second pole. In reality, the absolute value of a CMOS resistor varies from run to run. The temperature coefficient of a resistor is also different from the transistors in CMOS standard process, thereby causing error. An alternative way of implementing Rz by a MOS transistor operating in triode region is unsusceptible to process and temperature variations.
The method of realizing $R_z$ by a transistor is given in Figure 10.31 of the textbook. In our op-amp, NMOS rather than PMOS biasing is used because the second stage input device is a NMOS transistor. We choose the biasing current and size of the lower diode-connected transistor M21 in Figure 5(a) to set drain voltage of M21 to the gate voltage of M1. And then use equation (10.30) in the textbook to determine the width of the triode MOSFET M22. The simulation result of Figure 6(a) shows that the phase margin constrain can be well satisfied by substituting $R_z$ with this triode device.

B). Common Mode Feedback

Although our op amp satisfied the CMRR constrain, we still implemented common mode feedback (CMFB) of both stages to make the design complete, and it also helps reduce the sensitivity of low frequency gain to common mode input voltage. Another advantage of CMFB is that it gives us the flexibility of setting the common mode output voltage $V_{o,cm}$, and guarantee the maximum output swing with process variations.

The CMFB circuit is also given in the textbook. We just use two large resistors (100 kohms each) to take out the CM output voltage. Although the absolute value of a CMOS resistor is not accurate, we can usually guarantee the two resistors have the same value by properly laying them out. The design flow of our CMFB is summarized in the following steps: (from the author’s own experience)

1) Set the optimum operating point by changing $V_{in,cm}$ and determine the best $V_{o,cm}$;

2) Put a test current source in and test the relationship between the tail current and $V_{o,cm}$;

3) Referred to this relationship, cut the ORINGINAL tailcurrent so that the variation of the amount of current drain by this test source can adjust the $V_{o,cm}$ over a large enough range, and include the optimum output CM level in this range.

4) Replace this test current source by "$V_{cmc}$" (common mode control voltage) biased NMOS find out the relationship between this $V_{cmc}$ and $V_{o,cm}$;

5) According to step 3, find out the $V_{cmc}$ that can bias $V_{o,cm}$ to the desired level;

6) Make a simple op-amp, set its common mode input level ($V_{cm}=V_{in}=V_{in}$) as $V_{o,cm}$. Size this op-amp so that its common mode output ($V_{cmc+}$ and $V_{cmc-}$, as shown in the schematic) is $V_{cmc}$ acquired from step 4. Another choice is to size the additional tail NMOS to achieve the same thing. Make sure the feedback is negative.

7) Use two 100k resistors to take $V_{o,cm}$ and feed into the CMFB op-amp. And connect $V_{cmc}$ to the gate of the additional tail NMOS for CMFB.

Figure 6(b)-(d) shows the effect of CMFB on the sensitivity of DC gain vs. $V_{in,cm}$. With no CMFB, the response curve is quite sharp, indicating that the DC gain depends on input CM voltage a lot. With one CMFB to the first stage, we see a much flatter response. If CMFB is applied to both stages of our op amp, the characteristic is further improved and the op amp is useful almost over the entire input common mode range (ICMR depends on biasing).

In terms of controlling $V_{o,cm}$, we could change $V_{o,cm}$ linearly from 500 mV to 1.2 V, as shown in Figure 6(e). This makes the adjustments of DC operating point at the output end (to the middle of
output swing) much easier and unsusceptible to process variations.

Finally, CMFB also improves the common mode rejection ratio (CMRR) by introducing negative feedback for the common mode signals. Figure 6(f) is the common mode gain after adding CMFB, which drops below zero dB.

VI. Conclusions and Comments on Further Enhancing the Performance of This Op-Amp:

To conclude this report, we design a high-speed CMOS fully differential op amp using 180 nm technology and several of its performance metrics exceed the project specification. Two generations of op amp are described in this report, with emphasis on the final version. We learned a couple of lessons, which are very valuable. The hand-on experiences on analog circuit design were accumulated and design method was learned from these experiences. Power consumption and noise are improved from the first design to the second. Nulling resistor by a MOSFET operating in deep triode region and Common Mode Feedback (CMFB) is implemented. We attached all the simulation result supporting our conclusions in this report.

For the future works, we should try to use a folded cascode topology instead of telescopic cascode for the first stage if the input common mode range is specified or crucial. Simulation results show that the noise of this circuit is mainly come from its input transistors. Therefore, input referred noise can be reduced by sizing the input transistor larger. A more advanced trick is let the input device work in the subthreshold region to save power and reduce noise at the same time.[4] But the cost of reducing noise is the increasing of the internal capacitance and the moving of the third pole closer to the origin. This can cause stability problems. For CMFB, if we could replace the two sampling resistors by triode transistor, it will reduce the area of the layout. Finally, in this design, we have not considered the reference, for a real op amp, either a PTAT current or bandgap reference generator should be included.

Reference:

Figure 1(a) Schematic of the First Generation Op Amp

Figure 1(b) DC Operating Points of the First Generation Op Amp
Figure 2 Simulation Results of the First Generation Op Amp

- DC Gain, Unity Gain Bandwidth and Phase Margin of 1st Generation Design

- Common Mode Gain of 1st Generation Design

- Output Swing of 1st Generation Design

- Settling Time of 1st Generation Design

- Noise Response of 1st Generation Design
Figure 3(a) Schematic of the Second Generation Op Amp

Figure 3(b) DC Operating Points of the Second Generation Op Amp
Figure 4 Simulation Results of the Second Generation Op Amp

- DC Gain, Unity Gain Bandwidth and Phase Margin of 2nd Generation Design
- Common Mode Gain of 2nd Generation Design
- Small Signal Gain of 2nd Generation Design
- Settling Time of 2nd Generation Design
- Noise Response of 2nd Generation Design
Figure 5(a) Complete Schematic of the Second Generation Op Amp with Rz and CMFB Implemented
Figure 5(b) DC Operating Points of the Second Generation Op Amp with Rz and CMFB Implemented
Figure 6 Simulation Results of the Second Generation Op Amp with Rz and CMFB Implemented