**Abstract:** A general improved recycling folded cascode (IRFC) is analyzed and implemented. Analysis and comparisons among IRFC, original recycling folded cascode (RFC) and conventional folded cascode (FC) are made, and it is shown that with the flexible structure of IRFC, significant enhancement in transconductance, slew rate and noise can be achieved. Prototype amplifiers were fabricated in 0.13 μm technology, and measurement shows that with the same power and area consumption, IRFC has a 3 times enhancement in gain-bandwidth (GBW) and slew rate over conventional FC, and the enhancement is 1.5 times when compared with RFC.

1. Introduction

Operational transconductance amplifier (OTA) is still a necessary building block for many applications, such as baseband circuits [1] and data conversions [2], and consumes considerable power. A power-efficient OTA should have large transconductance while consume a small amount of power. In [3], Class AB output stage is adopted to enhance GBW, but it requires a large capacitor to emulate a battery. In [4], a slew rate enhancing method is proposed. In [5], recycling folded cascode (RFC) topology was proposed to enhance the power efficiency of OTA. In RFC, current mirrors are formed by splitting transistors to enhance transconductance and slew rate. However, AC and DC currents basically share the same path in RFC, which makes the DC currents from the cascoding branch necessary for current mirrors. This not only limits the enhancement of performance, but also makes the modification unsuitable for other structures of OTAs, such as telescopic OTA. In this paper, detailed analysis of IRFC and comparisons between IRFC and FC are given in Section 2. We introduce design parameters in IRFC to make it more flexible and more general than RFC, and we prove IRFC to be a more general type of folded-cascode. Next, optimization of transconductance and noise under constraint is discussed in Section 3. Simulation and measurement results are given in Section 4 to verify the effectiveness of the structure and our analysis. Finally, the conclusion is drawn in Section 5.

2. Analysis of IRFC

The topology and DC current of FC OTA and IRFC OTA are shown in Fig. 1(a) and (b), respectively. In IRFC, the original input pair M1 and M2 in FC is divided into M1a, M1b, M2a and M2b, in which the ratio of sizes of transistors is M1a:M1b=M2a:M2b=p:(1-p), where 0<p≤1. Also, M3 and M4 in FC are split into M3a, M3b, M3c and M4a, M4b, M4c (Fig. 1(b)), in which the ratio of sizes of transistors is M3a:M3b:M3c=M4a:M4b:M4c=(1+p):α(1-p):β(1-p), where α+β=1. M3a: M3b and M4a:M4b are current mirrors with a ratio of (1+p):α(1-p), which magnify AC currents to improve transconductance. On the other hand, in order to keep correct DC current, M3c and M4c form DC path. M3c and M4c are biased at constant gate voltage, so ideally only DC current flow into them. To further increase the AC impedance of DC path, cascode transistors M11b and M12b are inserted. Accordingly, M11a and M12a are inserted in AC path to improve the matching between DC and AC paths, in which the sizes of transistors M11a:M11b=M12a:M12b=α:β. Another benefit of transistors M11a and M12a is that they enhance the matching in current mirrors [5]. This separation of AC and DC paths is similar to current bleeding technique in mixers to enhance conversion gain [6], but here it not only improves the performances but also grants more flexibility in OTA design, as will be discussed below.

To make our discussion more general, we consider the limiting case of p→1, or, the case of FC. In this case, there is virtually no current mirror, so the ratio of current mirror (1+p):α(1-p)=0, i.e., we can express the boundary condition as:

$$\lim_{p \to 1} \alpha = \infty, \quad \lim_{p \to 1} \alpha \cdot (1-p) = \infty.$$  (1)

Equation (1) can also be intuitively understood by imagining that when p=1, the two branches of current α(1-p)I_h and β(1-p)I_h have the same infinity magnitude but the opposite direction, making the total current amount to zero. With (1), all equations in our analysis are effective for FC OTA. On the other hand, RFC is another special case of IRFC with p=0.5 and α=1, so IRFC is a more general type of folded cascode. For clear comparison, we will use the parameters of components of original FC and design parameter p and α in the following expressions to describe the performance of OTAs.
2.1 Effective Transconductance and Gain

The effective transconductance of IRFC can be found by calculating the short-circuit current at the output with regard to input [7]. Suppose the transconductance of folded cascode OTA is $g_{m_{FC}}$ (which is the transconductance of M1/M2, $g_{m1,2}$), then the transconductance of IRFC OTA with the same power consumption can be expressed as:

$$G_{m_{IRFC}} = (p + \frac{1 + p}{\alpha})g_{m_{FC}} \quad (2)$$

From (2), it can be seen that we may choose proper parameter $p$ and $\alpha$ to enhance the transconductance of IRFC, which means larger GBW, higher speed and better power efficiency of OTA. Compared with RFC which can achieve 2 times enhancement over FC, IRFC can have better enhancement with proper $p$ and $\alpha$. In reality, the enhancement of transconductance is slightly lower than (2) due to the finite output impedance of M3c and M4c. The low frequency gain of IRFC is decided by its effective transconductance and low frequency output impedance. Besides larger transconductance, IRFC has higher output impedance because compared with M4 in FC, less current flows through M4a [5]. With enhanced effective transconductance and higher output impedance, IRFC has a higher DC gain than FC.

2.2 Poles and Zero

The dominant pole ($\omega_{p1}$) in IRFC is at output node and one non-dominant pole ($\omega_{p2}$) of IRFC is decided by the parasites at the source node of M9/M10, like the case in FC OTA. In addition, the current mirrors of IRFC introduce a pole-zero pair ($\omega_{p3}$ and $\omega_{z3}$). The location of poles and zero in S-plane is shown in Fig.2. The pole-zero pair of IRFC can be expressed as:

$$\omega_{p3} = \frac{\alpha (1-p)g_{m3,4}}{1 + p + \alpha (1-p)}C_{gs3,4} \quad (3)$$

$$\omega_{z3} = \frac{[(1-p)(1+p)+ap(1-p)]g_{m3,4}}{p(1+p)+ap(1-p)}C_{gs3,4}, \quad (4)$$

where $g_{m3,4}$ and $C_{gs3,4}$ are the transconductance and gate-source capacitance of transistor M3/M4 in FC OTA. By comparing the location of the pole and zero in S-plane, we have:
\[
\frac{\omega_{z3}}{\omega_{p3}} = \frac{(1-p)(1+p) + ap(1-p)}{ap(1-p)} = 1 + \frac{1+p}{ap} \tag{5}
\]

For FC, \( p = 1 \). Combining (1) and (5), we have \( \omega_{z3}/\omega_{p3}=1 \), i.e., the zero and the pole cancel each other. Actually, phase margin can be a main factor limiting the selection of parameter \( p \) and \( \alpha \), as discussed in the next section.

2.3 Slew Rate

The slew rate (SR) of IRFC OTA can be calculated as follow. Suppose \( V_p \) goes high, then M1a and M1b turn off, and consequently M4b, M4a and M2a turn off. The current flowing through M2b is \( 2i_b \). Part of this current flows through DC path (M11b and M3c), while the rest goes into M3b and is magnified by current mirror. With the magnified current, the slew rate is also enhanced. The slew rate can be expressed as:

\[
SR_{IRFC} = \frac{(1+p)[2-(1-\alpha)(1-p)]I_b}{\alpha(1-p)C_{\alpha}}. \tag{6}
\]

Equation (6) is also valid for FC and RFC. In practice, the slew rate is limited by non-idealities and thus cannot reach the theoretical value. First, the transient response of current mirror to large signal is different from the case of small signal, so the current ratio is no longer accurate [5]. In addition, for low supply voltage in advanced technologies, the transistors M5 and M6 are likely to enter linear region when output current is large, and thus limit the slew rate.

2.4 Noise

Using the method in [8], the input-referred noise of IRFC OTA is expressed in (7):

\[
e_{\text{in}}^2 = \frac{8k_BT_f}{g_m(p+(1+p)/\alpha)^2} \left[ p + \frac{(1+p)^2}{\alpha^2(1-p)} + \frac{g_mL_1}{g_mL_1} \left( 1 + \frac{p}{2} - \frac{1}{2\alpha^2(1-p)} \right) + \frac{g_{m5}}{g_{m6}} \right] + \frac{K_{FP}}{\mu_pC_{\alpha}W_L^2} \left[ p + \frac{(1+p)^2}{\alpha^2(1-p)} + \frac{K_{FN}}{K_{FP}} \left( 1 + \frac{p}{2} - \frac{1}{2\alpha^2(1-p)} \right) \left( \frac{L_1}{L_3} \right)^2 \left( \frac{L_2}{L_4} \right)^2 \right], \tag{7}
\]

where \( L_1, L_3 \) and \( L_5 \) are the gate length of M1, M3 and M5 in FC, respectively. Equation (7) is also valid for FC and RFC.

The input-referred noise of FC OTA is:

\[
e_{\text{in}}^2 = \frac{8k_BT_f}{g_mL_1} \left[ 1 + \frac{g_{mL}}{g_mL_1} \right] + \frac{K_{FP}}{\mu_pC_{\alpha}W_L^2} \left[ 1 + 2 \frac{K_{FN}}{K_{FP}} \left( \frac{L_1}{L_3} \right)^2 \left( \frac{L_2}{L_4} \right)^2 \right] \tag{8}
\]

Equation (8) can be obtained by setting \( p=1 \) in (7). Comparing (7) and (8), we can observe that the noise currents of transistors M1b, M2b, M3b and M4b are amplified by current mirrors, while the gain of OTA increases at the same time, which suppresses noise of transistors. The two factors affect the input-referred noise in opposite direction, so we can find optimized \( p \) and \( \alpha \) to minimize noise, as discussed in the next section.

2.5 Offset

Using the method in [5], we can calculate the input offset variance of IRFC as:

\[
\sigma^2(V_{OS,IRFC}) = \frac{2A_{\text{VTP}}^2}{W_iL_1} \left[ p + \frac{(1+p)^2}{\alpha^2(1-p)} \right] \left[ 1 + \frac{\mu_k}{\mu_p}A_{\text{VTPN}}^2 \left( \frac{L_1}{L_3} \right)^2 \left( \frac{L_2}{L_4} \right)^2 \right]. \tag{9}
\]

where \( A_{\text{VTP}} \) and \( A_{\text{VTN}} \) is the area proportionality constant for threshold voltage of PMOS and NMOS, respectively.

For FC OTA, the input offset variance is

\[
\sigma^2(V_{OS,FC}) = \frac{2A_{\text{VTP}}^2}{W_iL_1} \left[ 1 + 2 \frac{\mu_k}{\mu_p}A_{\text{VTPN}}^2 \left( \frac{L_1}{L_3} \right)^2 \left( \frac{L_2}{L_4} \right)^2 \right]. \tag{10}
\]

Like in the case of noise, the current mirrors boost the drain-current variance and the overall transconductance at the same
time, making optimization possible.

2.6 Applicability
In IRFC, the AC and DC current have different paths, so the DC current from the cascode branch is no longer necessary to sustain the correct DC current of current mirrors, which is another advantage over RFC. With the separated AC and DC path, the modification can be applied to various OTAs with differential input and active load, such as differential telescopic OTA (Fig. 3).

![Fig. 3 Improved recycling differential telescopic OTA](image)

3. Design Considerations

3.1 Available Transconductance Enhancement
The enhancement of transconductance is limited by phase margin. Consider (3), (4) and (5), with decreased $\alpha$ and fixed $p$, $\omega z_3/\omega p_3$ becomes larger, which means that the zero can no longer compensate the pole. In addition, when $\alpha$ decreases, the decreased transconductance of M3b/M4b pushes $\omega p_3$ to lower frequency meanwhile the GBW of the OTA increases, thus the phase margin of OTA is degraded. By setting $p=0.5$, we calculated phase margin versus $\alpha$ and compared it with simulation results (Fig. 4). In calculation, we assumed $C_L=100C_{gs3,4}$ and $g_{m1,2}=g_{m3,4}$. Since calculation ignores the effect of non-dominant pole to GBW, there is some difference between calculation and simulation; however, the trends of calculation and simulation fit well. From Fig.4, we can see that $\alpha$ should be greater than 0.4 to achieve enough phase margin and thus avoid ringing in transient response. Generally, the transconductance enhancement should be smaller than 3.5 to guarantee enough phase margin (about 70°). In fact, when designing IRFC, we have one more dimension of freedom, since we can make trade-off between phase margin and other performances, while in other single stage OTAs, phase margin is generally constant (about 90°).

![Fig. 4 Calculated and simulated phase margin vs. $\alpha$](image)
3.2 Noise Consideration

In [5], RFC can consume less power to achieve the same GBW as FC but at the cost of noise, since no noise optimization was made for RFC. With the flexible structure of IRFC, however, noise performance can be optimized. To simplify (7), we assume that $L_1 = L_5$ and $g_{m5} = g_{m1}$. With the assumptions above, we have:

$$e_{nf, IRFC}^2 = \frac{1 + p + \frac{(1 + p)^2}{\alpha^2 (1 - p)}}{2(p + (1 + p)/\alpha)^2}, e_{nf, FC}^2$$  \hspace{1cm} (11)

With (11), we can find optimized $p$ and $\alpha$ for noise. By finding derivative of (11) regarding $\alpha$ and setting it to zero, we obtain:

$$\alpha = \frac{p}{1 - p}  \hspace{1cm} (12)$$

And the optimized noise power density for a given $p$ is:

$$e_{nf, IRFC, opt}^2 = \frac{1 + p}{2} e_{nf, FC}^2$$  \hspace{1cm} (13)

From (13) we can see that the optimized noise is proportional to $p$. This can be explained by the fact that with small $p$, the ratio of current mirror $M3a:M3b$ and $M4a:M4b$ is also small, so the noise currents of M1b and M2b are not significantly amplified; in addition, the overall transconductance is mainly decided by $\alpha$, so it is possible to have small $p$ and large transconductance at the same time, and then the input-referred noise is optimized.

Considering the stability, $\alpha$ should not be too small. Suppose we limit the transconductance enhancement under 3.5 to guarantee phase margin of 70°, then the optimized $\alpha$ with stability constraint is:

$$\alpha = \max \left[ \frac{1 + p}{3.5 - p}, \frac{p}{1 - p} \right]  \hspace{1cm} (14)$$

We define the optimized normalized noise as the ratio of the input-referred noise of noise-optimized IRFC and conventional FC, and simulated it for different $p$ (with stability constraint) and compared it with (13), as shown in Fig. 5. From Fig. 5 we can see that for $p<0.2$, $\alpha$ is constrained by stability condition and the optimized noise is larger than (13). For $0.2<p<0.5$, the simulated results fit well with (13). For $p>0.5$, the optimized value of $\alpha$ in (12) is greater than 1 and cannot be reached, and thus the noise of IRFC for $p>0.5$ cannot reach its lower bound in (13). The global optimized noise of IRFC can be obtained with $p=0.1$; however, setting $p=0.1$ means $M1b:M1a=9:1$ and this large ratio can induce matching problem [8]. Thus, a more practical value of $p$ is 0.2, when $M1b:M1a=4:1$, and the input-referred noise reduction compared to FC is 39% (2.1 dB). This also implies the potential to design a noise-optimized IRFC having the similar (or better) noise performance and transconductance with FC but consuming far less power. In addition, since the expression of input offset in (9) is similar to (7), we can expect that the offset of IRFC has similar characteristic with noise.

As for designs that do not meet the mentioned assumptions, optimized noise can still be found, though the expression can be more complicated. The trend of optimized noise versus $p$ should be similar, since the most influential element in optimized noise is still the ratio of current mirrors.
4. Experimental and Simulation Results

In this work, the IRFC OTA \((\rho=\alpha=0.5)\) was implemented in SMIC 0.13 \(\mu\)m technology with nominal VDD of 1.2 V to verify the effectiveness of our analysis. For comparison, an FC OTA and an RFC OTA were also implemented. The RFC and IRFC were modified based on the same FC OTA, so three OTAs have the same area (17×67 \(\mu\)m²) and power consumption (360 \(\mu\)W). The OTAs were configured in unit-gain amplifier with a feedback factor of 0.5 (with on-chip RC network where \(R_1=490\) k\(\Omega\) and \(C_1=1\) pF). The micrograph of the chip and test schematic are shown in Fig. 6 (a) and (b). An off-chip buffer is inserted to drive 50 \(\Omega\) load of network analyzer. The on-chip load capacitance is 4.5 pF, and the parasitic capacitance of pad and PCB trace is about 0.5 pF, so the effective load is approximately 5.5 pF (including the loading effect of feedback capacitance \(C_1\)) for all OTAs.

The measured BWs of FC, RFC and IRFC are 11.7 MHz, 22.9 MHz and 33.5 MHz, respectively (Fig. 7). Thus the GBWs of FC, RFC and IRFC are their close-loop BWs divided by feedback ratio, i.e., 23.4 MHz, 45.8 MHz and 67 MHz, respectively. Compared with FC, the GBW of IRFC is enhanced almost 3 times; while the enhancement is about 1.5 times when compared with RFC. The enhancement is smaller than the (2) because of the finite output impedance, as discussed in
Section II. To measure the settling time, a small signal step, 75 mVpp at 5 MHz, was applied to the amplifiers (Fig. 8). The measured 1% settling times of FC, RFC and IRFC are 64.9 ns, 32.6 ns and 21.4 ns, respectively. To measure slew rate, a large signal step (600 mVpp at 1 MHz) was applied. The transient response to the large signal is shown in Fig. 9, and the measured slew rates of FC, RFC and IRFC are 6.7V/μs, 17 V/μs and 20.7V/μs, respectively; that is, the slew rate of IRFC is enhanced 3 times over the FC and 1.2 times over the RFC. In addition, in transient responses in Fig. 8 and Fig. 9, no signs of ringing are observed, so we can conclude that the IRFC with \( p=\alpha=0.5 \) does not induce stability problem, which was also confirmed by phase margin simulation. Also, with greater transconductance and output impedance, gain enhancement of IRFC over RFC and FC was observed in simulation.

![Fig. 7 Measured close-loop bandwidth](image)

![Fig. 8 Measured settling behavior](image)

![Fig. 9 Measured large-signal response](image)

The noise performances of fabricated IRFC, RFC and FC are characterized by simulation. In addition, a noise-optimized IRFC (IRFC2) with \( p=0.2 \) and \( \alpha=3/8 \) was also simulated to verify our analysis in Section III. The spectral density is shown in Fig. 10. The integrated input-referred noise (1 Hz to 100 MHz) of IRFC, RFC, FC and IRFC2 is 98.5 \( \mu \)Vrms, 95.5 \( \mu \)Vrms, 110.0 \( \mu \)Vrms and 107.7 \( \mu \)Vrms, respectively. The noise-optimized IRFC has the same noise performance as FC, but it consumes about 40% less current and has about 200% GBW and slew rate.
The measurement and simulation results of OTAs were summarized in Table I. To compare the overall performance, two commonly used figures of merits (FoM) are calculated. The expressions of FoMs are:

\[
\text{FoM}_1 = \frac{\text{GBW} \cdot C_L}{I_D} \quad \text{FoM}_2 = \frac{\text{SR} \cdot C_L}{I_D}
\]

Compared with FC, IRFC and IRFC2 can achieve about 3 times and enhancement in both FoMs; while the enhancement over RFC is 1.5 times over FoM 1 and 1.2 times over FoM 2. The enhancement in FoMs shows the significant performance improvement of IRFC.

### Table 1 Performance Summary

<table>
<thead>
<tr>
<th>Parameter</th>
<th>FC</th>
<th>RFC</th>
<th>IRFC</th>
<th>IRFC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supply voltage (V)</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Current (μA)</td>
<td>300</td>
<td>300</td>
<td>300</td>
<td>185*</td>
</tr>
<tr>
<td>Area (μm×μm)</td>
<td>17×67</td>
<td>17×67</td>
<td>17×67</td>
<td>N/A</td>
</tr>
<tr>
<td>Capacitive Load (pF)</td>
<td>5.5</td>
<td>5.5</td>
<td>5.5</td>
<td>5.5</td>
</tr>
<tr>
<td>DC Gain (dB)*</td>
<td>53.61</td>
<td>60.2</td>
<td>64.9</td>
<td>66.1</td>
</tr>
<tr>
<td>GBW (MHz)</td>
<td>23.4</td>
<td>45.8</td>
<td>67</td>
<td>44.9*</td>
</tr>
<tr>
<td>Open Loop Phase Margin (deg)*</td>
<td>87.8</td>
<td>83.6</td>
<td>72.7</td>
<td>80.1</td>
</tr>
<tr>
<td>Slew Rate (V/μs)</td>
<td>6.7</td>
<td>17</td>
<td>20.7</td>
<td>12.7*</td>
</tr>
<tr>
<td>FoM1 (MHzpF/μA)</td>
<td>429</td>
<td>839.7</td>
<td>1228.3</td>
<td>1335*</td>
</tr>
<tr>
<td>FoM2 ((V/μs)pF/μA)</td>
<td>122.8</td>
<td>311.7</td>
<td>379.5</td>
<td>377.6*</td>
</tr>
</tbody>
</table>

*: Simulation Result

### 5. Conclusion

A novel transconductance-enhanced FC OTA, IRFC, is analyzed and implemented. Analysis shows that the IRFC is a more general type of FC OTA, and expressions of IRFC are proved to be effective for both RFC and FC OTA. It has been demonstrated by measurement results that the modification can achieve almost 3 times enhancement in FoMs over conventional FC OTA while maintaining enough stability. In addition, simulation shows that noise-optimized IRFC can achieve similar noise performance with FC but with 40% less current and 2x GBW and slew rate. The theoretical results are in good agreement with experimental and simulation results.

### References