Optimum Split Ratio for Folded Cascode OTA Bias Current: A Qualitative and Quantitative Study

Conference Paper · December 2019

DOI: 10.1109/ICM48031.2019.9021755

1 author:

Hesham Omran
Ain Shams University
47 PUBLICATIONS **471** CITATIONS

SEE PROFILE

Optimum Split Ratio for Folded Cascode OTA Bias Current: A Qualitative and Quantitative Study

Hesham Omran
Integrated Circuits Laboratory (ICL), Faculty of Engineering, Ain Shams University, Cairo 11517, Egypt
Email: hesham.omran@eng.asu.edu.eg

***Abstract*—This paper presents a systematic study on the effect of varying the bias current split ratio (S) in the ubiquitous folded cascode amplifier. A unity gain capacitive feedback amplifier is used as a design example to show the effect of S on the amplifier parameters such as loop gain, closed-loop bandwidth, phase margin, and thermal noise. Moreover, the effect of S on the slew rate and the overall settling time of the amplifier is investigated. Analytical equations are derived and verified by simulation results using 180nm CMOS technology. Results show that the common practice of splitting the current equally does not yield optimum results even in the case of slewing. The optimum choice of S can result in 3dB increase in DC gain, 40% increase in bandwidth, 33% reduction in noise density, and 56% improvement in settling time.**

***Keywords*—folded cascode OTA, analog circuit optimization, systematic analog design, gm/ID methodology**

I. Introduction

The folded cascode operational transconductance amplifier (OTA) is the workhorse amplifier architecture in analog integrated circuit (IC) design [1]. The telescopic casocde OTA provides better efficiency (higher speed for the same bias current), higher gain, and lower noise [1], [2]. However, the folded cascode OTA is far more popular due to its distinct flexible common-mode input advantage. The schematic of a fully-differential folded cascode OTA is shown in Fig. 1, where the PMOS common-source (CS) input stage (M1) is followed by an NMOS common-gate (CG) stage (M3). The degraded efficiency of the folded cascode is due to using two separate bias currents for the CS and CG stages. On the other hand, the bias current is shared in the telescopic cascode. The gain of the folded cascode is lower than its telescopic counterpart because M2 appears in parallel with M1; thus, the output impedance is reduced. Moreover, M2 (which carries the largest current and thus has the largest g_m) contributes additional noise component which results in higher noise compared to the telescopic cascode. The flexible common-mode input advantage of the folded cascode is evident by noting that: 1) the OTA can achieve both wide common-mode input range and wide output swing simultaneously (they are not oppositely coupled as in the telescopic cascode), 2) the common-mode voltage at the input and the output can be set to the same value, which is a very useful property especially for cascaded amplifier stages, and 3) the common-mode input can be set to zero (or V_{DD} for NMOS input pair), which is another useful property as it may alleviate the need for an additional reference voltage.

This work was supported by the Information Technology Industry Development Agency (ITIDA) under Grant PRP2018.R24.7.

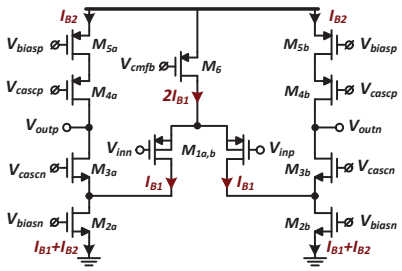

Figure 1: Schematic of fully-differential folded cascode OTA. A common-mode feedback (CMFB) circuit (not shown) senses the CM output and adjusts the biasing of M6.

Due to its widespread use in many analog IC design applications, the design and optimization of folded cascode OTA and its enhanced variants was extensively studied in both classical and recent literature [3]–[10]. However, there is an important design parameter for the folded cascode OTA that was not fully studied in the literature, namely, how the total OTA bias current is split between the input and output branches (i.e., between the CS and the CG amplifiers). For simplicity, it is usually assumed that the current is split equally [1], [4], [9]. Practically, some designers may put more current in the input stage aiming at higher gain-bandwidth product (GBW) and less noise. On the other hand, conservative designers tend to put 10% to 20% more current in the output branch to avoid starving M3 at the event of slewing. However, the optimum split ratio has not been systematically investigated before to the best of the author's knowledge. In this paper, a qualitative and quantitative study of this optimization problem is presented. The effect of the split ratio on the OTA DC and AC parameters is investigated. Moreover, the effect of the split ratio on the settling time under both linear settling and slewing conditions is explored.

II. Design Description

The total bias current of the folded OTA shown in Fig. 1 is given by

$$I_B = 2I_{B1} + 2I_{B2} \tag{1}$$

We define the split ratio (S) as the ratio between the CS and CG bias currents

$$S = \frac{I_{B1}}{I_{B2}} \tag{2}$$

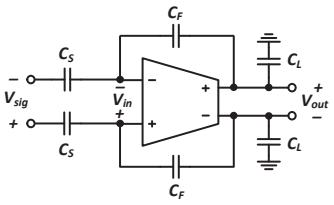

Figure 2: Capacitive feedback amplifier.

Table I: OTA design parameters.

	$L(\mu m)$	$g_m/I_D\,(S/A)$
M1	0.2	18
M2	1	14
M3	1	14
M4	1	12
M5	1	12

Thus, the bias currents can be written as

$$I_{B1} = \frac{S}{1+S}\cdot\frac{I_B}{2} \quad and \quad I_{B2} = \frac{1}{1+S}\cdot\frac{I_B}{2} \tag{3}$$

For the purpose of studying the proposed optimization problem in a real-life closed loop circuit, we use the capacitive feedback amplifier shown in Fig. 2 with $C_S = C_F = 1pF$. The load capacitance (C_L) is also set to $1pF$ and the total OTA bias current is set to $0.5mA$. The OTA is designed in a $180nm$ CMOS technology using g_m/I_D methodology. The design parameters of the OTA are shown in Table I. A relatively long channel length is used for M_{2-5} to boost the DC gain. A long channel length for M_1 would increase the OTA input capacitance and thus reduce the loop gain rather than increase it. Therefore, a short length is selected for M_1. A relatively large g_m/I_D is selected for the input pair to maximize the transconductance while avoiding excessive input capacitance. A moderate g_m/I_D is selected for the output branch as a compromise between output swing, noise, and stability. A slightly smaller g_m/I_D (larger overdrive) is used for $M_{4,5}$ to avoid very large PMOS devices. The device widths are calculated using the g_m/I_D starter kit [11]. Since it is required to study the effect of the current split ratio without disturbing the bias point of the OTA, the device widths are defined as

$$W_1 = W_1^*\cdot\frac{2S}{1+S} \quad and \quad W_{3-5} = W_{3-5}^*\cdot\frac{2}{1+S} \tag{4}$$

where W^* is the width at $S=1$. The width of M_2 is unchanged since it always carries the same total current.

III. Effect of S on the Amplifier DC and AC Parameters

Modeling the OTA as a single transconductor stage, it can be shown that the loop-gain (LG) of the amplifier in Fig. 2 is given by

$$LG \approx \frac{\beta A_{vo,OL}}{1+sR_{out}C_{out,tot}} = \frac{\beta G_m R_{out}}{1+sR_{out}C_{out,tot}} \tag{5}$$

where β is the feedback factor, $A_{vo,OL}$ is the OTA open-loop DC gain, G_m is the OTA transconductance, R_{out} is the OTA output resistance, and $C_{out,tot}$ is the total capacitance at the output node. G_m can be written as

$$G_m = \alpha g_{m1} \tag{6}$$

where α accounts for current division at the folding node

$$\alpha = \frac{g_{m3}+g_{mb3}+\frac{1}{r_{o3}}}{\frac{1}{r_{o1}}+\frac{1}{r_{o2}}+g_{m3}+g_{mb3}+\frac{1}{r_{o3}}} \tag{7}$$

and R_{out} is given by

$$R_{out} \approx \left[r_{o3}\left(g_{m3}+g_{mb3}\right)\left(r_{o2}||r_{o1}\right)\right]||\left[r_{o4}\left(g_{m4}+g_{mb4}\right)r_{o5}\right] \tag{8}$$

Putting more current in the input pair $(S>1)$ will have a minor effect on R_{out} because the increase in r_{o3-5} is counteracted by the decrease in $g_{m3,4}$. However, it will clearly increase G_m. As a result, there will be an overall increase in $A_{vo,OL}$. Fig. 3a shows that using $S=4$ improves the DC gain by around $3dB$, while using $S=0.25$ gives around $7dB$ reduction.

In order to investigate the effect of β, it is worth noting that the equivalent resistance at the folding node $\left(R_{fold}\right)$ has different values at low and high frequencies. At low frequencies, the drain of M_3 is connected to a high-impedance node; thus, R_{fold} is given by

$$R_{fold,LF} \approx r_{o1}||r_{o2}||\frac{1+\frac{r_{o4}(g_{m4}+g_{mb4})r_{o5}}{r_{o3}}}{g_{m3}+g_{mb3}+\frac{1}{r_{o3}}} \tag{9}$$

On the other hand, the dominant output pole creates a low impedance node at high frequencies

$$R_{fold,HF} \approx r_{o1}||r_{o2}||\frac{1}{g_{m3}}||\frac{1}{g_{mb3}}||r_{o3} \tag{10}$$

Noting that $R_{fold,LF} \gg R_{fold,HF}$, Miller effect across C_{gd1} yields two values for the OTA input capacitance (C_{in})

$$C_{in,LF/HF} = C_{gs1}+C_{gb1}+C_{gd1}\left(1-g_{m1}R_{fold,LF/HF}\right) \tag{11}$$

Consequently, the feedback factor can be written as

$$\beta_{LF/HF} = \frac{C_F}{C_S+C_F+C_{in,LF/HF}} \tag{12}$$

Using $S>1$ means larger width for M_1; thus, C_{in} increases. Consequently, β_{LF} will have an inverse relation to S as shown in Fig. 3b. The decrease in β_{LF} counteracts the increase in $A_{vo,OL}$, causing an overall negligible change in the DC loop gain $(LG_o = \beta_{LF}A_{vo,OL})$ for $S>1$ as shown in Fig. 3c. However, for $S<1$, LG_o decreases as the degradation in $A_{vo,OL}$ is sharper than the increase in β_{LF}. As a result, the static gain error $(\varepsilon_s \approx 1/LG_o)$ will be as shown in Fig. 3d.

The amplifier closed-loop bandwidth (BW_{CL}) is equal to the LG unity gain frequency $(f_{u,LG})$, which can be written using (5) as

$$f_{u,LG} = \frac{1}{2\pi}\frac{\beta_{HF}G_m}{C_{out,tot}} \tag{13}$$

The total output capacitance is

$$C_{out,tot} = C_{out}+C_L+C_F||\left(C_S+C_{in,HF}\right) \tag{14}$$

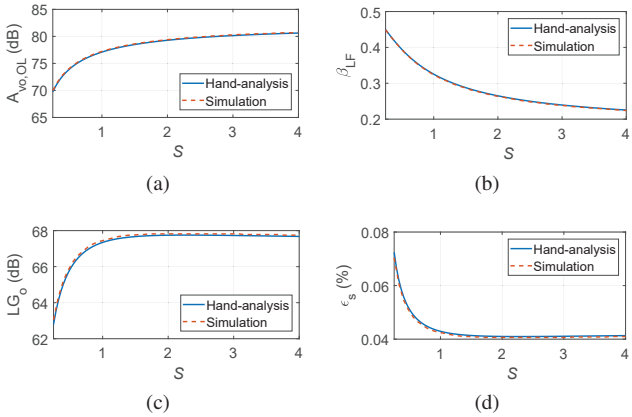

Figure 3: Amplifier DC parameters vs the split ratio (S).

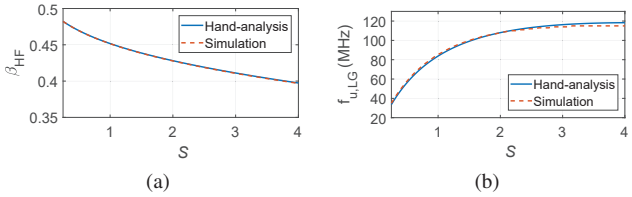

Figure 4: Amplifier AC parameters vs the split ratio (S).

where C_{out} is the OTA self-loading due to the drain capacitance of $M_{3,4}$.

The variation of β_{HF} vs S is less severe than the case of β_{LF} because Miller effect is significantly reduced at high frequencies as shown in Fig. 4a. This decrease in β_{HF} is opposed by a higher increase in G_m and a slight decrease in $C_{out,tot}$. Thus, there is an overall significant increase in $f_{u,LG}$, which is the main advantage of using $S>1$. For $S=4$ there is a 40% increase in bandwidth. On the other hand, the bandwidth degrades considerably for $S<1$.

The stability of the OTA depends on the ratio between $f_{u,LG}$ and the non-dominant pole (f_{pnd}). The non-dominant pole is due to the folding node and is given by

$$f_{pnd} = \frac{1}{2\pi R_{fold,HF} C_{fold}} \tag{15}$$

where C_{fold} is the capacitance at the folding node. The OTA phase margin (PM) is given by

$$PM \approx 90 - \tan^{-1}\left(\frac{f_{pnd}}{f_{u,LG}}\right) \tag{16}$$

As S increases, g_{m3} decreases causing an increase in $R_{fold,HF}$ and degradation in f_{pnd} as shown in Fig. 5a. This degradation, accompanied by the increase in $f_{u,LG}$, causes a decrease in PM. As shown in Fig. 5b, simulations indicate that PM decreases from 81^o at $S=1$ to 71^o at $S=4$. There is a slight deviation between hand-analysis and simulations at large S due to the nature of the approximate expression in (16).

The noise density referred to the amplifier input (V_{sig}) can be calculated by multiplying the noise density at V_{in} by the

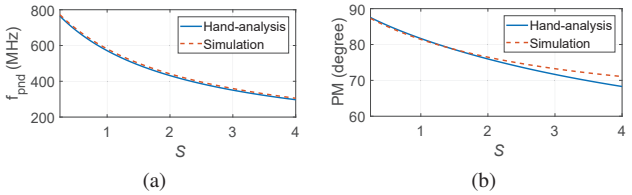

Figure 5: The non-dominant pole (f_{pnd}) and the phase margin (PM) vs the split ratio (S).

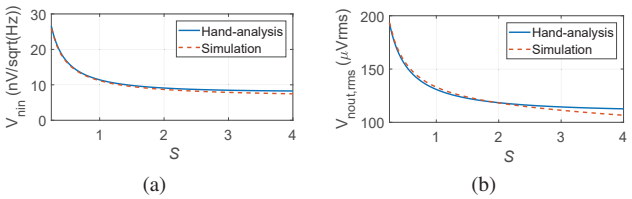

Figure 6: (a) Input-referred noise density and (b) rms output noise vs the split ratio (S).

amplifier non-inverting gain (i.e., referring it to V_{out}) then dividing it back by the closed loop inverting gain (i.e., referring it back to V_{sig})

$$V_{nin}^2(f) \approx \frac{8kT\gamma_{eff}}{g_{m1}} \cdot \left(\frac{1/\beta_{HF}}{C_S/C_F}\right)^2 \tag{17}$$

where γ_{eff} is the effective noise coefficient

$$\gamma_{eff} \approx \gamma_1 + \gamma_2\frac{g_{m2}}{G_m} + \gamma_5\frac{g_{m5}}{G_m} \tag{18}$$

Fig. 6a shows that using $S=4$ reduces the input-referred noise density by 33% due to the increase in g_{m1} despite the decrease in β_{HF}. On the other hand, for $S=0.25$, there is more than two-fold increase in noise density.

The rms output noise can be calculated by multiplying the noise density with the closed loop gain and the equivalent noise bandwidth

$$V_{nout,rms}^2 \approx V_{nin}^2(f) \cdot \left(\frac{C_S}{C_F}\right)^2 \cdot \frac{\pi f_{u,LG}}{2} \approx \frac{2kT\gamma_{eff}\alpha}{\beta_{HF}C_{out,tot}} \tag{19}$$

Fig. 6b shows that the rms noise decreases by 20% at $S=4$. This improvement is justified by noting that the increase in G_m for $S>1$ decreases γ_{eff} by reducing the noise contribution of $M_{2,5}$. The percent improvement in rms noise is less than the improvement in noise density because the closed loop bandwidth of the amplifier has been extended. In all the previous results, there is an excellent match between simulation results and hand-analysis expectations, by virtue of using the precomputed lookup tables methodology presented in [9].

IV. Effect of S on the Amplifier Slew Rate and Settling Time

The results presented in Sec. III are all in favor of using $S>1$, except for the small degradation in PM. However, in order to obtain a complete picture, the effect of S on the OTA slew rate must be studied. For symmetrical slewing, the OTA

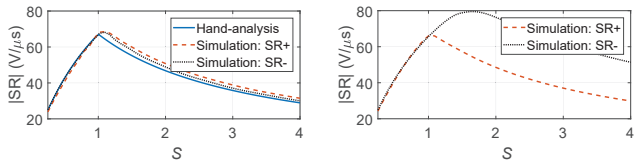

Figure 7: Positive and negative slew rate vs S for (a) fast CMFB loop and (b) slow CMFB loop.

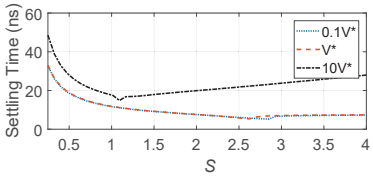

Figure 8: Settling time vs S for an input step of $0.1V^*$,V^*,and $10V^*$.

slew rate is limited by the smaller bias current out of I_{B1} and I_{B2} and can be roughly written as

$$SR \approx \frac{min(I_{B1}, I_{B2})}{C_{out,tot}} \tag{20}$$

Fig. 7 shows the simulated positive and negative slew rate ($SR+$ and $SR-$ respectively). If a fast common-mode feedback (CMFB) loop is used, $SR+$ and $SR-$ are kept almost equal as shown in Fig. 7a, and the best SR is in the vicinity of $S = 1$ as expected from (20). However, Fig. 7b shows that for a slow CMFB loop, the OTA will suffer from asymmetrical slewing for $S > 1$, and the common-mode output will suffer from large transients.

Since the settling time in general is divided into linear settling and slewing parts, the increased closed loop bandwidth for $S > 1$ is counteracted by the decrease in SR and may lead to an overall increase in settling time. The 0.1% settling time from transient simulation results (for the case of fast CMFB loop) is plotted vs S in Fig. 8 for a small, medium, and large input steps ($0.1V^*$,V^*,and $10V^*$, respectively, where $V^* = \frac{2}{g_m/I_D}$ resembles the overdrive voltage of the long-channel model). The optimum value of S shifts to the left as the input step increases. For small and medium input steps the linear settling dominates and the overall settling time significantly improves for $S > 1$. On the other hand, for a large input step the linear settling improvement is counteracted by the degradation in the slew rate. However, the optimum point is still located at $S > 1$. Using the optimum S can result in a 56% improvement in settling time for small and medium input steps and 30% improvement in the case of slewing.

The transient simulation results are depicted in Fig. 9, which again shows that $S = 4$ is superior for a small input step. However, it suffers from a slight overshoot and undershoot due to the decrease in PM (underdamped response). The slewing behavior is very clear for the case of $10V^*$ input step, at which $S = 1$ provides faster settling, although the optimum point is located at a slightly larger S as shown in Fig. 8.

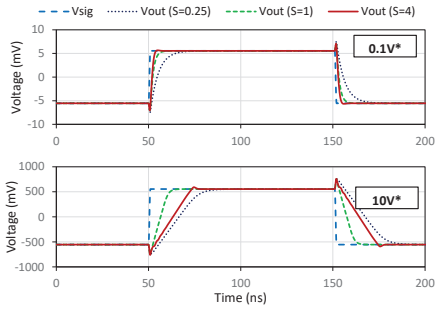

Figure 9: Transient simulation results for $S = 0.25, 1$, and 4 at an input step of $0.1V^*$ and $10V^*$.

V. Conclusion

An extensive study on the effect of the split ratio (S) of the folded cascode OTA bias current was presented. Results show that using $S > 1$ (i.e., putting more bias current in the CS stage) can improve the amplifier DC gain, bandwidth, and noise. For continuous-time analog signal processing and bias circuits, using $S \approx 2-3$ can lead to an overall significant improvement in bandwidth and settling time. However, for circuits that suffer from frequent large output excursions, e.g., switched capacitor circuits, using $S \approx 1.1-1.2$ may be optimum, but higher S may cause significant asymmetrical slewing. Aside from a small improvement in PM, the case of $S < 1$ does not seem to provide any favorable behavior.

References

[1] B. Razavi, *Design of analog CMOS integrated circuits*, 2nd ed. McGraw-Hill Education, 2017.

[2] O. A. Abu-El-ela, A. A. Aldaly, Y. Abd-El-azim, and H. A. Omran, "Design automation of a fully differential switched capacitor amplifier using a telescopic cascode OTA," in *2018 35th IEEE National Radio Science Conference (NRSC)*, March 2018, pp. 385–392.

[3] S. Mallya and J. H. Nevin, "Design procedures for a fully differential folded-cascode CMOS operational amplifier," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 6, pp. 1737–1740, Dec 1989.

[4] H. C. Yang, M. A. Abu-Dayeh, and D. J. Allstot, "Small-signal analysis and minimum settling time design of a one-stage folded-cascode CMOS operational amplifier," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 7, pp. 804–807, July 1991.

[5] A. Girardi, F. P. Cortes, and S. Bampi, "A tool for automatic design of analog circuits based on gm/ID methodology," in *2006 IEEE International Symposium on Circuits and Systems*, May 2006, pp. 4 pp.–.

[6] R. S. Assaad and J. Silva-Martinez, "The recycling folded cascode: A general enhancement of the folded cascode amplifier," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 9, pp. 2535–2542, Sept 2009.

[7] Po-Cheng, Hung-Ming, and Chien-Chih, "PAGE: Parallel agile genetic exploration towards utmost performance for analog circuit design," in *2013 Design, Automation Test in Europe Conference Exhibition (DATE)*, March 2013, pp. 1849–1854.

[8] J. Ou and P. M. Ferreira, "A gm/ID-based noise optimization for CMOS folded-cascode operational amplifier," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 61, no. 10, pp. 783–787, Oct 2014.

[9] P. Jespers and B. Murmann, *Systematic Design of Analog CMOS Circuits Using Pre-Computed Lookup Tables*. Cambridge University Press, 2017.

[10] M. P. Garde, A. Lopez-Martin, R. G. Carvajal, and J. RamÃrez-Angulo, "Folded cascode OTA with 5540 MHzpF/mA FoM," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018, pp. 1–5.

[11] B. Murmann. Gm/ID Starter Kit. [Online]. Available: https://web.stanford.edu/ murmann/gmid