An Improved Reversed Miller Compensation Technique for Three-stage CMOS OTAs with Double Pole-zero Cancellation and almost Single-pole Frequency Response

Authors: Francesco Centurelli, Pietro Monsurrò, Giuseppe Scotti, Pasquale Tommasino and Alessandro Trifiletti

Authors are with the Department of Information, Electronics, and Communications Engineering (DIET), Sapienza Università di Roma, Rome, Italy. (e-mail: pietro.monsurro@uniroma1.it).

Key-words: Reversed nested Miller compensation, pole-zero cancellation, multipath, low-power amplifiers, operational transconductance amplifiers, current reuse, bandwidth efficiency.

ABSTRACT

This paper presents an improved reversed nested Miller compensation technique exploiting a single additional feed-forward stage to obtain double pole-zero cancellation, and ideally single-pole behavior, in a three-stage Miller amplifier. The approach allows designing a three-stage operational transconductance amplifier (OTA) with one dominant pole and two (ideally) mutually cancelling pole-zero doublets. We demonstrate the robustness of the proposed cancellation technique, showing that it is not significantly influenced by process and temperature variations. The proposed design equations allow setting the unity gain frequency of the amplifier and the complex poles' resonance frequency and quality factor.

We introduce the notion of bandwidth efficiency to quantify the OTA performance with respect to a telescopic cascode OTA for given load capacitance and power consumption constraints, and demonstrate analytically that the proposed approach allows a bandwidth efficiency which can ideally approach 100%.

A CMOS implementation of the proposed compensation technique is provided, in which a current reuse scheme is used to reduce the total current consumption. The OTA has been designed using a 130nm CMOS process by STMicroelectronics, and achieves a DC gain larger than 120dB, with almost single-pole frequency response. Monte Carlo simulations have been performed to show the robustness of the proposed approach to PVT variations and mismatches.

1. INTRODUCTION

Battery-operated or energy harvested systems such as biomedical implantable devices or sensor nodes for Internet of Things (IoT) applications require an aggressive reduction of both supply voltage and power consumption of the analog interface circuits. The conventional cascoding techniques, usually adopted to increase the gain of operational transconductance amplifiers (OTAs), are not suitable to achieve high DC gain with supply voltages below 1V, unless special techniques such as body biasing [1] or low-voltage gain-boosting [2] are exploited. Cascading, as opposed to cascoding, allows achieving low-voltage and high-gain OTAs, but requires particular care in frequency compensation to guarantee the stability of multi-stage OTAs.

Three-stage reversed Nested Miller amplifiers [3-12] are used in high-gain, low-voltage applications where there is not enough voltage headroom for gain-boosting cascode OTAs. Several techniques to remove the positive zeros which plague OTAs exploiting Miller compensation have been proposed in the technical literature, such as series resistors [5-6], current buffers [5-7], voltage buffers [5], or mixtures of these techniques [6, 8], also including multipath [9-12].

In this work we present an improved reversed nested Miller compensation (RNMC) technique for low-voltage three-stage CMOS OTAs, in which a single feed-forward stage is used to perform double pole-zero cancellation, hence removing the two high-frequency poles of the OTA with two zeros. The proposed approach thus allows achieving a frequency response with one dominant pole and two (ideally) mutually cancelling pole-zero doublets (PZD) at the unity-gain frequency (ω_u), resulting in (almost) single-pole frequency behavior for a three-stage amplifier: this is important for settling time performance. The technique is based on a single additional feed-forward stage, which allows cancelling both high-frequency poles with two high-frequency zeros, leaving the frequency response essentially dominated by the first pole, with negligible effect of the high-frequency poles (and zeros) produced by the parasitic capacitances of MOS devices.

When imperfectly cancelled (because of unavoidable parameter variations), the PZDs create additional exponential transient responses which could degrade the settling performance of the OTA. In our approach both PZDs are placed at high frequency, near the ω_u, so that the impact of imperfect cancellation (due to parasitic effects or mismatches) on the settling performance of the OTA is limited. Furthermore, our sizing strategy allows choosing the quality factor of the two PZDs, in case the two zeros and the two poles form a complex conjugate pair. A simple solution is proposed for placing two real coincident zeros and poles at exactly the unity-gain frequency of the OTA, resulting in 90° phase margin, infinite gain margin, and single-pole transient behavior.

This approach can be applied to any three-stage amplifier, either for low-power, low-voltage applications, or even for very fast OTAs: in the latter case, the impact of parasitic capacitances would be more relevant, but the two main parasitic poles would still be almost cancelled, ensuring higher unity-gain frequency and phase and gain margins. We show that the technique is robust to parasitic capacitances, and process and temperature variations, because the equations allowing double pole-zero cancellation depend on the same parameters for the poles and for the zeros.

To quantify the advantages provided by the proposed compensation technique, we introduce the concept of bandwidth efficiency, η_{BW}, defined as the ratio between the ω_u of the three-stage OTA and the ω_u of a reference telescopic cascode OTA with the same capacitive load and power consumption (ω_{MAX}). We show that with a known load capacitor, it is always possible to design the 3-stage OTA with almost single-pole behavior, owing to double pole-zero cancellation. Ideally, there is no bandwidth penalty ($\eta_{BW} \rightarrow 100\%$) if the feed-forward stage is much larger than the three main stages, and an arbitrary large load capacitor can be driven with relatively small compensation capacitors.

The proposed compensation technique has been adopted to design a three-stage OTA in a commercial 130nm CMOS technology from STMicroelectronics. Transistor-level simulations have been carried out in CADENCE Virtuoso environment to test the small-signal and large-signal (transient) behavior of the three-stage OTA. Extensive Monte Carlo simulations have been performed to check the robustness of the designed OTA to PVT and mismatch variations. Compared to the reference telescopic cascode OTA, the designed three-stage OTA achieves a bandwidth efficiency of 47%, with a DC gain of more than 120dB.

In the following, Section 2 describes the proposed compensation technique, Section 3 details the resulting design methodology of a three-stage OTA as well as the CMOS implementation of the

OTA as a single-ended amplifier with rail to rail output swing. Section 4 presents the detailed design and summarizes the simulation results. Conclusions are reported in Section 5.

2. PROPOSED COMPENSATION TECHNIQUE

The block scheme of a three-stage OTA with feed-forward and RNMC is reported in Fig. 1, where g_i is the transconductance of stage *i*, g_{FF} the transconductance of the feed-forward stage, C_{C1} and C_{C2} are the compensation capacitors, C_L is the load capacitance and C_X and C_Y represent the total parasitic capacitances at node *X* and *Y* respectively. According to Fig. 1 we exploit only one feed-forward stage, which is sized to cancel both high-frequency poles with two corresponding zeros. Neglecting $C_X, C_Y \ll C_{C1}, C_{C2}, C_L$, the frequency response of the stage is:

$$\frac{v_o(s)}{v_i(s)} = \frac{g_1}{sC_{c1}} \frac{1+s\frac{C_{C2}(g_2 g_{FF}-g_1 g_3)}{g_1 g_2 g_3}+s^2 C_{C1}C_{C2}\frac{g_{FF}-g_1}{g_1 g_2 g_3}}{1+s\frac{C_{C1}g_2-C_{C1}g_3+C_L g_2}{C_{C1}g_2 g_3}C_{C2}+s^2\frac{C_{C2}C_L}{g_2 g_3}} \tag{1}$$

It is possible to cancel the two zeroes with the two high-frequency poles by imposing the condition:

$$C_{C1} = \frac{C_L g_1}{g_{FF}-g_1} \tag{2}$$

This condition cancels both the second-order and the first-order terms of the complex poles and zeroes, hence leaving only the dominant pole (in the origin, as output resistances have been neglected), whose unity gain frequency is ω_u.

Even if the two PZDs are formally cancelled, it is better to avoid placing them at low frequencies, because, in the (unavoidable) case of imperfect cancellation, they create small transients with large time constants, which increase the settling time significantly. For this reason, we force as a second condition that the resonance frequency ω_0 of the two PZDs is the same as the unity-gain frequency ω_u, resulting in the condition:

$$C_{C2} = \frac{C_L g_2 g_3}{(g_{FF}-g_1)^2} \tag{3}$$

The double PZD have resonance frequency ω_0 and quality factor Q:

$$\begin{cases} \omega_u = \omega_0 = \frac{g_{FF}-g_1}{C_L} \\ Q = \frac{g_1(g_{FF}-g_1)}{(g_2-g_3)g_1+(g_{FF}-g_1)g_2} \end{cases} \tag{4}$$

We can assume that $g_2 = g_3$ to further simplify these conditions, leaving $Q = \frac{g_1}{g_2}$.

Within the limit of $g_{FF} \gg g_i$, $C_{C1} \approx C_L g_1/g_{FF}$ (1) can be much smaller than C_{C1}, and $C_{C2} \approx C_L g_2^2/g_{FF}^2$ (2) even smaller. Hence, for a given load capacitor, it is possible to achieve double pole-zero cancellation with arbitrary small compensation capacitors. The resulting unity-gain frequency (and resonance frequency of the high-frequency poles and zeros) would be $\omega_u \approx g_{FF}/C_L$, hence not influenced by the three main amplifier stages. By using all the power budget for the feed-forward stage, bandwidth can thus be maximized.

Due to parasitic effects, PVT variations and mismatches, PZDs are never exactly cancelled, so that pole-zero cancellation is not fully effective. The two poles and the two zeros "almost" cancel each other, but residual PZDs remain, creating additional transient functions which can be slower than that caused by the main dominant pole. Hence, it is advisable to choose the resonant frequency and the quality factor of the poles (and of the zeros, as they nominally cancel the poles) to avoid excessive resonances or low-frequency PZDs.

Conditions for double PZD cancellation (2) and PZD placement at the unity gain frequency (3) rely on the compensation capacitors chosen as a function of the load capacitor. This is possible in integrated circuits if they are implemented with the same capacitor type (such as MIM or Fringe), but it is not in general possible for operational amplifiers loaded by external capacitors: it is not possible to perform double pole-zero cancellation with an unspecified load capacitance. In many applications, however, the load capacitor is fixed and known a priori: when driving capacitors inside the integrated circuit, in applications such as analog or switched capacitor filters, Sample & Hold amplifiers, analog to digital converters (ADCs), OTAs are designed for a known load capacitor. In general, pole-zero cancellation, either single or double, cannot be achieved for more than one single value of the load capacitance, because C_L only affects the poles, and not the zeroes. This is true for all pole-zero cancellation schemes.

Equations (2-3) show that for $g_{FF} \gg g_1, g_2, g_3$, a large load capacitor C_L can be driven with relatively small capacitors C_{C1} and C_{C2}: the former is inversely proportional to g_{FF}, the latter to g_{FF}^2. For instance, for $Q = 1/2$ (two real coincident poles), hence $g_2 = g_3 = 2g_1$, and $g_{FF} = 10g_1$, C_{C1} will be 9 times lower than C_L, and C_{C2} about 20 times lower. Therefore, in principle, the proposed approach allows the usage of compensation capacitors much smaller than the load capacitor even if there are practical limitations when setting $g_{FF} \gg g_i$. In fact, a large g_{FF} would increase power consumption, whereas a small g_i would cause robustness issues under process and mismatch variations, due to the low device sizes (or bandwidth limitations, if current densities are low).

A. Effect of parasitic capacitances C_X and C_Y

If we include the parasitic capacitances C_X and C_Y in the analysis of the block scheme in Fig. 1, the frequency response becomes:

$$\frac{v_o(s)}{v_i(s)} = \frac{g_1}{sC_{C1}} \frac{1+\frac{C_{C2}(g_2 g_{FF}-g_1 g_3)}{g_1 g_2 g_3}s+\frac{C_{C1}C_{C2}g_{FF}-C_{C1}C_Y g_1-C_{C1}C_{C2}g_1+C_{C2}C_X g_{FF}+C_{C1}C_Y g_{FF}+C_{C2}C_Y g_{FF}+C_X C_Y g_{FF}}{g_1 g_2 g_3}s^2}{1+\frac{C_{C1}C_{C2}g_2-C_{C1}C_{C2}g_3+C_{C2}C_L g_2}{C_{C1}g_2 g_3}s+\frac{(C_{C1}C_{C2}C_L+C_{C1}C_{C2}C_X+C_{C1}C_{C2}C_Y+C_{C2}C_L C_X+C_{C1}C_L C_Y+C_{C2}C_L C_Y+C_{C1}C_X C_Y+C_L C_X C_Y)}{C_{C1}g_2 g_3}s^2} \quad (5)$$

The impact of the two capacitances is negligible when $C_X, C_Y \ll C_C, C_L$, which is a common hypothesis in the literature due to the complexity of the frequency response of three-stage amplifiers [3-12]. However, it can be shown that the condition (2) is sufficient to obtain double PZD cancellation also in (5), so that our double cancellation technique is robust against parasitic capacitances toward the ground.

It is still possible to force the condition $\omega_u = \omega_0$, but the resulting value of C_{C2} is different from (3) and will depend on the value of the parasitic capacitances. This implies that forcing the exact condition $\omega_u = \omega_0$ will not in general be easy: however, this condition is not required, because the impulse response of the filter with imperfectly cancelled PZDs will show fast transients with time constants around ω_0^{-1} also in this case.

B. Bandwidth efficiency

We introduce the notion of bandwidth efficiency to quantify the speed (in terms of unity-gain frequency) of an OTA given a power consumption budget and a load capacitor. The idea is to be able to choose amplifier topologies which provide the largest bandwidth for the same power consumption and load capacitance, or similar (or slightly lower) bandwidth but with additional benefits, such as higher phase margin and/or larger DC gain.

To define the bandwidth efficiency, we denote as g_T the transconductance of a single transistor consuming the whole current budget I_{TOT}. Then we consider the telescopic cascode as the most

efficient OTA. In fact, for the same capacitive load C_L and the same total bias current $I_{TOT} = I_B$ (and hence transconductance), it achieves a single-pole (neglecting cascoding effects) frequency response with unity-gain frequency $\omega_{MAX} = g_T/C_L$. All other amplifiers are less efficient, because they have more current branches to bias, hence yielding a lower unity-gain frequency because of the pole proportional to C_L being driven with a lower transconductance.
Hence we define as bandwidth efficiency the ratio between the ω_u of a generic amplifier and the unity-gain frequency of a telescopic cascode OTA with the same capacitive load and power consumption:

$$\eta_{BW} \equiv \frac{\omega_u}{\omega_{MAX}} = \frac{\omega_u C_L}{g_T} \tag{6}$$

Three-stage OTAs are typically used when high DC-gain is needed, but conventional three-stage Miller OTAs have a complex frequency response resulting in a limited bandwidth efficiency.
Unlike conventional three-stage Miller OTAs, the proposed compensation technique allows to design three-stage OTAs with high DC gain and with a frequency response which is ideally single-pole, resulting in a bandwidth efficiency which can ideally approach 100%, as shown in the next Section.
This can be seen within the limit of $g_{FF} > g_i$, because all the power budget would be used to obtain a large g_{FF}, yielding a unity-gain frequency $\omega_u = g_{FF}/C_L$ (2-4). A large feed-forward stage would drive the load capacitance, yielding a large bandwidth, while the three main stages, with relatively small compensation capacitors, would yield very high DC gain.

3. DESIGN OF A 3-STAGE CMOS OTA WITH PROPOSED RNMC TECHNIQUE

A straightforward CMOS implementation of a three-stage OTA suitable for the compensation technique discussed in the previous section is reported in Fig. 2. The first gain stage is implemented through the differential pair M_1, M_2 loaded with the n-channel current mirror M_3, M_4. The second stage is implemented by the common source NMOS transistor M_6, whereas the third stage (non-inverting) is composed by M_8, M_9, and M_{10}. The feed-forward stage is implemented by the input differential pair M_1, M_2, the current mirror M_3, M_{14}, and the additional inverting current mirror composed of devices M_{12} and M_{13}. Biasing current sources are implemented through transistors M_5, M_7, M_{11} and M_{15} respectively, with biasing voltage V_B produced by a biasing network (not shown).
The transconductance g_1 of the first stage is equal to the transconductance $g_{m1,2}$ of the differential pair M_1-M_2, the transconductance g_2 of the second stage is equal to the transconductance g_{m6} of M_6, whereas the transconductance of the third stage is set by g_{m8} and the current mirror ratio $\frac{g_{m10}}{g_{m9}}$.
Finally, the transconductance g_{FF} of the feed-forward stage is set by the transconductance $g_{m1,2}$ of the differential pair M_1-M_2 and the current gains of the current mirrors M_3-M_{14} and M_{13}-M_{12}, divided by two because it only uses one branch of the input differential pair.
The compensation strategy discussed in section 2 can be applied to the CMOS three-stage OTA presented in Fig. 3. As an example, if we want to size the OTA for real coincident ($Q = 1/2$) poles and zeros (see Fig. 2) we have the following design equations:

$$g_1 = g_{m1,2} \tag{7a}$$

$$g_2 = g_{m6} = 2g_{m1,2} \tag{7b}$$

$$g_3 = g_{m8} \cdot \frac{g_{m10}}{g_{m9}} = 2g_{m1,2} \tag{7c}$$

$$g_{FF} = \frac{g_{m1,2}}{2} \cdot \frac{g_{m14}}{g_{m3}} \cdot \frac{g_{m12}}{g_{m13}} = 5g_{m1,2} \tag{7d}$$

$$C_{C1} = C_{C2} = \frac{C_L}{4} \tag{7e}$$

In this way, we obtain double pole-zero cancellation (2) with resonance frequencies at the unity-gain frequency of the OTA (3), with two real coincident PZDs.

A. Bandwidth efficiency of the proposed three-stage OTA

In this section we compute the bandwidth efficiency of the three-stage OTA reported in Fig. 2. Devices M_1-M_4, $M_{8,9}$ and $M_{13\text{-}14}$ have unitary size (*1x* in Fig. 2). The ratio g_2/g_1 is set by the size of M_6 (*2x* in Fig. 2), with respect to the size of $M_{1,2}$, whereas g_3/g_1 is set by $\frac{g_{m8}}{g_{m1,2}} \cdot \frac{g_{m10}}{g_{m9}}$ and g_{FF}/g_1 by $\frac{g_{m14}}{g_{m3}} \cdot \frac{g_{m12}}{g_{m13}}$. Considering that the bias current of a branch with *1x* transistors is equal to I_B, the total current consumption I_{TOT} of the circuit can be derived by inspection as follows:

$$I_{TOT} = \left(2 + \frac{g_2}{g_1} + 1 + \frac{2g_{FF}}{g_1} + 2\right) I_B = \left(5 + \frac{g_2}{g_1} + \frac{2g_{FF}}{g_1}\right) I_B. \tag{8}$$

Because g_1 is proportional to I_B, g_T is equal to $\frac{5g_1+g_2+2g_{FF}}{2}$, hence:

$$\omega_{MAX} = \frac{5g_1+g_2+2g_{FF}}{2C_L}. \tag{9}$$

If we now assume double PZD cancellation (2) and $\omega_0 = \omega_u$ (3), we obtain:

$$\eta_{BW} = \frac{2g_{FF}-2g_1}{2g_{FF}+5g_1+g_2} \tag{10}$$

Hence, the bandwidth efficiency can grow up to 100% for $g_{FF} \gg g_1, g_2$. This implies that the feed-forward stage becomes much larger than the main path of the three-stage amplifier: the main path provides low-frequency gain, and the feed-forward path provides bandwidth by driving the capacitive load. Under these conditions, ideally, bandwidth would be the same as for a single-stage amplifier, but with a DC gain three times larger (in dB). By using a large feed-forward stage in parallel with a three-stage amplifier with relatively small compensation capacitors, it is possible to drive a large capacitive load with high bandwidth efficiency, high gain, and almost single-pole frequency response.

Under the assumption $g_2 = g_3$, which is innocuous for bandwidth efficiency, and with double pole-zero cancellation and placing of the pole-zero doublets at the unity-gain frequency, we have:

$$\eta_{BW}(Q) = \frac{\frac{g_{FF}}{g_1}-1}{\frac{g_{FF}}{g_1}+\frac{5}{2}+\frac{1}{2Q}} \tag{11}$$

For our choice of real coincident poles $Q = \frac{1}{2}$, and $g_{FF} = 5g_1$:

$$\eta_{BW}\left(\frac{1}{2}\right) = \frac{8}{17} \approx 47\% \tag{12}$$

A better efficiency of 67% could be achieved for $g_{FF} = 10g_1$, and the maximum of 100% would be achieved asymptotically. However, this would limit bandwidth, because the current mirror M12-M13 in Figs 2-3 would have a pole at too low a frequency if M12 were too large.

B. Effect of feed-forward path on DC gain

Assuming ideal current mirrors and denoting as r_{oi} the output resistance of the generic transistor M_i in Fig. 2, the DC gain of the proposed three-stage OTA can be easily computed as follows:

$$A_0 \cong g_{m1,2}(r_{o2}//r_{o4})\, g_{m6}(r_{o6}//r_{o7}) g_{m8} \frac{g_{m10}}{g_{m9}} (r_{o10}//r_{o11}//r_{o12}). \quad (13)$$

The presence of the feed-forward path reduces the gain of the third stage, which sees the relatively large load of the feed-forward stage, whose output resistance $r_{o11}//r_{o12}$ is low owing to the sizing of M_{11} and M_{12}. The gain penalty is a factor 5 with the sizing in (7), because M_{12} is five times larger than M_{10}, reducing the output resistance (together with M_{11}, which is four times larger than M_{10}), whereas the transconductance remains the same with or without feed-forward.

C. Effect of process and mismatch variations

In this section we discuss the robustness of the proposed compensation technique with respect to process variations. In particular, we focus on the pole-zero cancellation under process variations and show that the double pole-zero cancellation, and the quality factors and resonance frequency of these poles and zeros, depend essentially on device matching (small within die random variations), whereas they are not dependent on the much larger process (die to die) variations.
At this purpose we consider all the devices used in the OTA in Fig. 2, which is made up of NMOS transistors, PMOS transistors and integrated capacitors. Then we focus on the parameter variations of these devices and on the effect of these variations on the pole-zero cancellation. The parameters of NMOS transistors, PMOS transistors and integrated capacitors can vary independently across dies, resulting in large variations for performance figures depending on them, whereas performance figures depending on the ratio between the parameters of devices of the same type will be affected only by mismatch (within die) variations, and will exhibit much lower variability.
Let's analyze the relationships between the parameters of the block scheme in Fig. 1 and the parameters of the CMOS circuit in Fig. 2. The transconductance g_1 in Fig. 1 is given by:

$$g_1 = \frac{g_{m1,2}}{2}\left(1 + \frac{g_{m4}}{g_{m3}}\right) \quad (14)$$

The term $\frac{g_{m4}}{g_{m3}}$ is the ratio of the transconductances of two NMOS transistors and is affected only by mismatch variations, hence we assume that the transconductance g_1 of the first stage changes as $g_{m1,2}$ (i.e. as the transconductance of a PMOS transistor).
The transconductance g_2 in Fig. 1 is given by:

$$g_2 = g_{m6} \quad (15)$$

The transconductance g_2 changes as g_{m6} (i.e. the transconductance of a NMOS transistor).
The transconductance g_3 in Fig. 1 is given by:

$$g_3 = g_{m8} \cdot \frac{g_{m10}}{g_{m9}} \quad (16)$$

The term $\frac{g_{m10}}{g_{m9}}$ is the ratio of the transconductances of two PMOS transistors and is affected only by mismatch variations, hence we assume that the transconductance g_3 of the third stage changes essentially as g_{m8} (i.e. as that of a NMOS transistor).
Finally, the transconductance g_{FF} in Fig. 1 is given by:

$$g_{FF} = \frac{g_{m1,2}}{2} \cdot \frac{g_{m14}}{g_{m4}} \cdot \frac{g_{m12}}{g_{m13}} \quad (17)$$

The terms $\frac{g_{m14}}{g_{m4}}$ and $\frac{g_{m12}}{g_{m13}}$ are ratios of the transconductances of two NMOS transistors and are affected only by mismatch variations, hence the transconductance g_{FF} of the feed-forward stage changes essentially as $g_{m1,2}$ (i.e. as that of a PMOS transistor).

Taking into account the above considerations and denoting with δ_N, δ_P and δ_C the process variations of NMOS transistors, PMOS transistors and integrated capacitors respectively, the frequency response in (1) can be rewritten as:

$$\frac{v_o(s)}{v_i(s)} = \frac{g_1(1+\delta_P)}{sC_{c1}(1+\delta_C)} \frac{1+s\frac{1+\delta_C}{1+\delta_N}\frac{C_{C2}(g_2g_{FF}-g_1g_3)}{g_1g_2g_3}+s^2\frac{(1+\delta_C)^2}{(1+\delta_N)^2}C_{C1}C_{C2}\frac{g_{FF}-g_1}{g_1g_2g_3}}{1+s\frac{1+\delta_C}{1+\delta_N}\frac{C_{C1}g_2-C_{C1}g_3+C_Lg_2}{C_{C1}g_2g_3}C_{C2}+s^2\frac{(1+\delta_C)^2}{(1+\delta_N)^2}\frac{C_{C2}C_L}{g_2g_3}} = \frac{\omega_u(1+\delta_P)}{s(1+\delta_C)} \frac{1+\frac{(1+\delta_C)s}{(1+\delta_N)Q\omega_0}+\frac{(1+\delta_C)^2s^2}{(1+\delta_N)^2\omega_0^2}}{1+\frac{(1+\delta_C)s}{(1+\delta_N)Q\omega_0}+\frac{(1+\delta_C)^2s^2}{(1+\delta_N)^2\omega_0^2}} \quad (18)$$

Hence, the relative positions of the zeros and poles in the two PZDs are not influenced by process variations, because they both depend on the variations of integrated capacitors and NMOS transistors. On the other hand, the unity-gain frequency will change with PMOS devices and capacitors, hence the condition $\omega_0 = \omega_u$ will in general not hold precisely under process variations. This condition is not necessary: it only ensures that in case of imperfect cancellation, which is unavoidable, the two pole-zero doublets will provide two additional poles with fast transients, and relatively small amplitude (zero under ideal double cancelation).

4. SIMULATION RESULTS AND COMPARISONS

The proposed compensation technique has been applied to design a three-stage OTA with a DC gain in excess of 120dB based on the circuit presented in Fig. 2 and implemented in a commercial 130nm CMOS process from STMicroelectronics. To guarantee high DC gain in spite of the low intrinsic gain of short-channel MOS transistors, all the current mirrors have been implemented as high-swing cascode current mirror (HSCCM) stages, resulting in the detailed schematic reported in Fig. 3. A supply voltage $V_{DD} = 1V$ and a loading capacitance $C_L = 1pF$ have been assumed as constraints for OTA design. The compensation capacitors have been set to $C_{C1} = C_{C2} = C_L/4 = 250fF$ as explained in (7). Bias currents and transistor dimensions are reported in Table I.

A summary of the performance figures of the three-stage OTA simulated in open-loop configuration and in the typical PVT corner is reported in Table II. The amplifier exhibits an excellent DC gain A_0 and unity-gain frequency f_u, good phase margin m_φ and gain margin m_G with good common mode rejection (CMRR) and power supply rejection (*PSRR*).

The open-loop frequency response of the amplifier (gain and phase) is reported in Fig. 4.

The OTA has then been tested in unity-gain buffer configuration. The closed-loop frequency response is reported in Fig. 5. The amplifier is well compensated, with good phase and gain margin, and high-frequency poles (and zeros) are due to parasitic C_{gs} and C_{gd} not accounted for in (1).

Table III shows a summary of the closed loop simulation results, where the Gain-Bandwidth product GBW is estimated as the -3dB bandwidth in unity gain configuration, settling times $T_{settle}^{1\%}$ are measured from a square waveform of 1µs period, distortions from a single-tone sinusoidal waveform at 1MHz. Distortions are dominated by the second harmonic (*HD2*), and are lower than the signal-to-noise ratio (*SNR*), so that the optimal output swing where *SNR* and total harmonic distortion *THD* are comparable is higher than $\pm 0.2V$: however, since distortions vary with mismatches more than noise, a lower signal swing is preferred. With this choice, *SNR* dominates over *THD*.

The transient response to a $0.4 \pm 0.2V$ input square wave is reported in Fig. 6 showing a 134ns and 103ns rise and fall time respectively.

As a further evaluation, closed-loop simulations have been carried out in a transimpedance configuration with a feedback resistor $R_f = 200k\Omega$.

Transient simulations in the transimpedance configuration have been carried out with an input current $I_{in} = \pm 1.5\mu A$ resulting in an output swing V_{swing} of $0.5 \pm 0.3V$. Fig. 7 shows the closed-

loop step response, and Fig. 8 the closed-loop sinusoidal response at 1MHz (in time and in frequency) for the transimpedance configuration. The rising edge of the output voltage is limited by slew rate at about $8V/\mu s$, while the rate on the falling edge exceeds $12V/\mu s$. The rising edge is limited by slew-rate, because the input differential pair has 4μA tail current and is loaded by two 250fF capacitors, yielding a $8V/\mu s$ slew rate limitation.

A. Temperature and supply voltage variations

Table IV reports the results of temperature variations from -30 to 120°C (with 30°C steps), and supply voltage varying from 0.9 to 1.1V. The amplifier performance is stable, with a slight variation of the bandwidth mostly due to the variation of the transconductances with temperature. Most variations are monotonic and only the extreme values are reported.

B. Process variations and mismatches

Monte Carlo simulations have been carried out in Cadence ADE XL environment by using accurate statistical models provided by the IC manufacturer. Table V reports the results of Monte Carlo simulations (process only), whereas Table VI reports Monte Carlo simulations for mismatches, showing good robustness against both process and mismatch parameter variations as expected. Table VII shows the results for combined process and mismatch variations. Process variations can be used as a substitute for corner simulations, if the number of simulations ($N = 100$) is sufficient, because the device parameters will change throughout the distribution of possible values. Furthermore, we have added the minima and maxima of the values obtained by Monte Carlo simulations (process only, and combined process and mismatch), to better determine the extreme points under corner simulations.

C. Comparison with three-stage CMOS OTAs from the literature

Table VIII compares the main performance of our design with other three-stage amplifiers taken from the literature with V_{DD} equal or less than 1V. The comparison is carried out considering also the well-known figures of merit defined as:

$$FOM_S = GBW \cdot C_L/P_{TOT} \tag{19}$$

$$FOM_L = SR \cdot C_L/P_{TOT} \tag{20}$$

where C_L is the loading capacitance and P_{TOT} the overall power consumption. η_{BW} has also been computed for comparison, using Eq. (6c).

Our proposed OTA has the highest gain (due to cascading of the three stages) and highest phase margin (theoretically it should be 90°, but cascading creates additional poles, not accounted for in the model). It has the second best small-signal FOM and IFOM after [18], which however has less than half the gain. Its large-signal FOM is significantly lower than [18], but only marginally lower than [15] and [17], which also have less than half the gain. The proposed OTA can be sized to operate with a much larger load capacitor, provided that the compensation capacitors are scaled: this would not change the FOMs because it would trade-off bandwidth for load capacitance, but would increase the phase margin (which would be closer to the theoretical 90° of a single-pole amplifier), because the parasitic poles of the cascode stages would be less relevant.

5. CONCLUSIONS

An improved compensation strategy for three-stage CMOS OTAs based on a single feed-forward path has been presented. The technique allows double pole-zero cancellation, placing of the pole-

zero doublets, and controlling of their quality factor. Thanks to double pole-zero cancellation, the two high-frequency poles in three-stage OTAs can be removed, leaving a single-pole transient response in the ideal case. In the case of imperfect cancellation, the two pole-zero doublets are placed at high frequency to ensure fast transient. Hence, by using a single feed-forward stage, it is possible to almost remove two poles in three-stage OTAs.

Compared with a telescopic cascode OTA, the proposed approach achieves a bandwidth efficiency which can ideally approach 100 % when setting $g_{FF} \gg g_i$, hence, a large feed-forward stage in parallel with the three-stage reversed nested Miller amplifier can achieve large gain, large bandwidth, and almost single-pole transient response. Given the value of the compensation capacitor, it is thus possible to design an OTA with high efficiency and almost single-pole behavior with relatively low compensation capacitors, in the limit case $g_{FF} \gg g_i$.

The proposed topology and compensation strategy have been used to design an OTA. It allows achieving a DC gain in excess of 120dB in a short-channel 130nm technology. The amplifier further employs current reuse to implement the feed-forward stage: by summing a PMOS current and an NMOS current, the output node implements feed-forward without needing extra bias current. The performance of the designed amplifier has been shown to be stable under temperature and process variations, both by means of a theoretical analysis and by means of parametric and Monte Carlo simulations.

REFERENCES

1. Monsurrò, P., Scotti, G., Trifiletti, A., Pennisi, S., "Sub-1V CMOS OTA with body-driven gain boosting", (2008) European Conference on Circuit Theory and Design 2007, ECCTD 2007, art. no. 4529651, pp. 535-538.
2. Cellucci, D., Centurelli, F., Di Stefano, V., Monsurrò, P., Pennisi, S., Scotti, G., Trifiletti, A., "0.6-V CMOS Cascode OTA with Complementary Gate-Driven Gain Boosting and Forward Body Bias", (2019), International Journal of Circuit Theory and Design (in press).
3. Palumbo, G., Pennisi. S., “Feedback amplifiers”, Springer, 2002.
4. Eschauzier, R., Huijsing, J., “Frequency compensation techniques for low-power operational amplifiers”, Springer, 1995.
5. Mita, R., Palumbo, G., Pennisi, S., "Design guidelines for reversed nested Miller compensation in three-stage amplifiers", (2003) IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, 50 (5), pp. 227-233.
6. Centurelli, F., Monsurrò, P., Pennisi, S., Scotti, G., Trifiletti, A., "Design solutions for sample-and-hold circuits in CMOS nanometer technologies", (2009) IEEE Transactions on Circuits and Systems II: Express Briefs, 56 (6), pp. 459-463.
7. Garimella, A., Rashid, M.W., Furth, P.M., "Reverse nested miller compensation using current buffers in a three-stage LDO", (2010) IEEE Transactions on Circuits and Systems II: Express Briefs, 57 (4), art. no. 5438730, pp. 250-254.
8. Ho, K.-P., Chan, C.-F., Choy, C.-S., Pun, K.-P., "Reversed Nested Miller Compensation with Voltage Buffer and Nulling Resistor", (2003) IEEE Journal of Solid-State Circuits, 38 (10), pp. 1735-1738.

9. Grasso, A.D., Marano, D., Palumbo, G., Pennisi, S., "Improved Reversed Nested Miller Frequency Compensation Technique with Voltage Buffer and Resistor", (2007) IEEE Transactions on Circuits and Systems II: Express Briefs, 54 (5), pp. 382-386.
10. Grasso, A.D., Palumbo, G., Pennisi, S., "Advances in reversed nested Miller compensation", (2007) IEEE Transactions on Circuits and Systems I: Regular Papers, 54 (7), pp. 1459-1470.
11. Marano, D., Palumbo, G., Pennisi, S., "Step-response optimisation techniques for low-power, high-load, three-stage operational amplifiers driving large capacitive loads", (2010) IET Circuits, Devices and Systems, 4 (2), art. no. ICDSB400000400000200008700000l, pp. 87-98.
12. Surkanti, P.R., Furth, P.M., "Converting a three-stage pseudoclass-AB amplifier to a true-class-AB amplifier", (2012) IEEE Transactions on Circuits and Systems II: Express Briefs, 59 (4), art. no. 6170880, pp. 229-233.
13. T. Lehmann and M. Cassia, "1-V power supply CMOS cascode amplifier," (2001) IEEE J. Solid-State Circuits, 36 (7), pp. 1082–1086.
14. G. Raikos and S. Vlassis, "Low-voltage bulk-driven input stage with improved transconductance," (2011) Int. J. Circuit Theor. Appl., 39, (3), pp. 327–339.
15. M. R. Valero Bernal, S. Celma, N. Medrano, and B. Calvo, "An ultra low-power low-voltage class-AB fully differential OpAmp for long life autonomous portable equipment," (2012) IEEE Trans. Circuits Syst. II, Exp. Briefs, 59, (10), pp. 643–647.
16. L. Zuo and S. K. Islam, "Low-voltage bulk-driven operational amplifier with improved transconductance," (2013) IEEE Trans. Circuits Syst I, Reg. Papers, 60, (8), pp. 2084–2091.
17. O. Abdelfattah, G. W. Roberts, I. Shih, and Y.-C. Shih, "An ultra-lowvoltage CMOS process-insensitive self-biased OTA with rail-to-rail input range," (2015) IEEE Trans. Circuits Syst. I, Reg. Papers, 62, (10), pp. 2380–2390.
18. E. Cabrera-Bernal, S. Pennisi, A. D. Grasso, A. Torralba, R. G. Carvajal, "0.7-V Three-Stage Class-AB CMOS Operational Transconductance Amplifier," (2016) IEEE Trans. Circuits Syst I: Reg. Papers 63, (11), pp. 1807 – 1815.
19. W. Qu, S. Singh, Y. Lee, Y.-S. Son, and G.-H. Cho, "Design-oriented analysis for Miller compensation and its application to multistage amplifier design," (2017) IEEE J. Solid-State Circuits, 52, (2), pp. 517–527.

TABLE 1. DEVICE SIZES AND CURRENTS OF THE PROPOSED OTA

MOS	Width/Length (μm/μm)	Bias Current (μA)
M_1,M_2	4.5/0.4	1
M_5	2x1.5/0.4	2
M_3,M_4	0.5/0.4	1
M_{3A},M_{4A}	1.5/0.4	1
M_6	2x0.5/0.4	2
M_{6A}	2x1.5/0.4	2
M_7	2x1.5/0.4	2
M_{7A}	2x4.5/0.4	2
M_8	0.5/0.4	1
M_{8A}	1.5/0.4	1
M_9	1.5/0.4	1
M_{9A}	4.5/0.4	1
M_{10}	2x1.5/0.4	2
M_{10A}	2x4.5/0.4	2
M_{11}	8x1.5/0.4	8
M_{11A}	8x4.5/0.4	8
M_{12}	10x0.5/0.4	10
M_{12A}	10x1.5/0.4	10
M_{13},M_{14}	0.5/0.4	1
M_{13A},M_{14A}	1.5/0.4	1
M_{15}	2x1.5/0.4	2
M_{15A}	2x4.5/0.4	2

TABLE II. OPEN-LOOP SIMULATION RESULTS IN TYP27 CONDITIONS

Name	Value	Unit
I_{TOT}	16.7	uA
V_{DD}	1	V
P_{TOT}	16.7	uW
A_0	126.6	dB
f_u	13.7	MHz
m_φ	73	deg
m_G	19	dB
CMRR	68	dB
PSRR+	68	dB
PSRR-	74	dB

TABLE III. CLOSED-LOOP SIMULATION RESULTS IN TYP27 CONDITIONS

Name	Value	Unit	Notes
GBW	24.8	MHz	
V_{swing}	0.6	V	Peak-peak, around 0.5V
$T^{1\%}_{settle,+}$	117	ns	600mVpp output
$T^{1\%}_{settle,-}$	105	ns	600mVpp output
SR+	9	V/µs	600mVpp output
SR-	11	V/µs	600mVpp output
V_{noise}	0.86	mV	Integrated from 1Hz to 1GHz
SNR	47.8	dB	Assuming output sinusoid 0.6Vpp
HD2	-46	dB	@1MHz
HD3	-52	dB	@1MHz
HD4	-60	dB	@1MHz
THD	-44.8	dB	@1MHz
DR	46.2	dB	Assuming HD3 dominates

TABLE IV. CLOSED-LOOP AND OPEN-LOOP SIMULATION RESULTS UNDER TEMPERATURE AND SUPPLY VOLTAGE VARIATIONS

Name	Min Value	Max Value	Min Value	Max Value	Unit
V_{DD}	1	1	0.9	1.1	V
T	-30	120	27	27	°C
I_{TOT}	16.5	16.9	17.1	16.4	µA
GBW	17.0	20.6	24.8	25.1	MHz
V_{swing}	625	565	585	595	mV
$T^{1\%}_{settle}$	115	119	148	95	ns
V_{noise}	850	920	860	860	µV
A_0	122	127	125.8	127.1	dB
f_u	12.2	14.5	12.6	14.1	MHz
m_φ	72.8	73.4	73.0	73.3	°
m_G	18.7	19.7	19.8	19.0	dB
CMRR	64	70	60.6	71.7	dB
PSRR+	63	75	82.6	63.9	dB
PSRR-	71	82	90.3	70.7	dB
HD2	-42	-50	-50.8	-44.1	dB
HD3	-49	-55	-50.1	-50.5	dB

TABLE V. CLOSED-LOOP SIMULATION RESULTS UNDER MISMATCH VARIATIONS

Name	Mean	Std	Unit
I_{tot}	16.8	2.4	µA
V_{swing}	590	20	mV
HD2	-47.3	9.8	dB
HD3	-52.9	7.1	dB
V_{noise}	960	350	µV
GBW	19.8	3.7	MHz

TABLE VI. CLOSED-LOOP SIMULATION RESULTS UNDER PROCESS VARIATIONS

Name	Mean	Std	Min	Max	Unit
I_{tot}	16.7	0.1	16.6	16.8	µA
V_{swing}	595	40	510	660	mV
HD2	-46.0	0.7	-47.8	-44.9	dB
HD3	-52.0	0.5	-53.1	-51.1	dB
V_{noise}	850	20	800	930	µV
GBW	24.7	0.8	22.4	25.8	MHz

TABLE VII. CLOSED-LOOP SIMULATION RESULTS UNDER PROCESS AND MISMATCH VARIATIONS

Name	Mean	Std	Min	Max	Unit
I_{tot}	16.8	2.4	13.8	21.4	µA
V_{swing}	590	40.8	500	665	mV
HD2	-44.6	8.3	-60.4	-31.1	dB
HD3	-51.1	7.0	-70.0	-39.2	dB
V_{noise}	840	66	710	960	µV
GBW	24.6	2.4	19.0	28.4	MHz

TABLE VIII. PERFORMANCE COMPARISON WITH STATE OF THE ART

	This work	[13]	[14]	[15]	[16]	[17]	[18]	[19]
Technology (µm)	0.13	0.5	0.18	0.18	0.35	0.065	0.18	0.18
Supply Voltage (V)	1	1	1	0.8	1	0.5	0.7	1.2
Loading Capacitance (pF)	1	20	1	8	15	3	20	18
DC gain (dB)	127	69	64	51	88	46	57.5	100
Power (µW)	16.7	40	130	1.2	197	183	25.41	69.6
GBW (MHz)	19.7	2	2	0.057	11.67	38	3	1.18
PM (°)	73	57	45	60	66	57	60	59.6
SR (V/µs)	10	0.5	0.7	0.14	1.95	43	2.8	0.22
Compensation strategy*	RNMC + FF	D. P.	M + N. R	D.P	M + N. R	RNMC + DFC	RNMC +CB	RNMC +AZ
FOM_S (MHz pF/mW)	1485	1000	15	380	889	623	2361	366
$IFOM_S$ (MHz pF/mA)	1485	1000	15	475	889	1246	3372	305
FOM_L (V pF/µs mW)	598	250	5	934	148	705	2204	68

*RNMC + FF: Reversed nested Miller compensation + Feed Forward;
D. P.: Dominant pole compensation;
M + N. R.: Miller Compensation with Nulling Resistor;
RNMC + DFC: Miller Compensation with Damping Factor Control
NMC+CB: Nested Miller Compensation with Current Buffer
RNMC: Reversed nested Miller compensation with active zero.

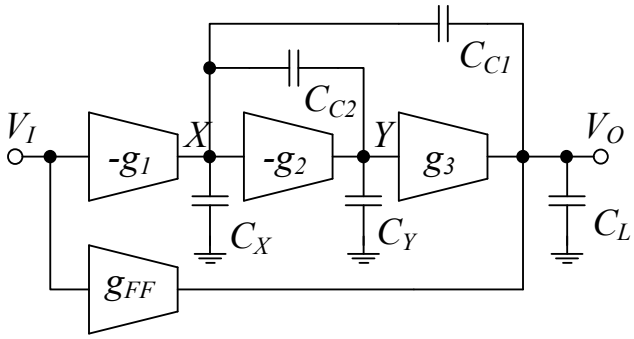

Fig. 1. Block scheme of the proposed three-stage RNMC OTA with a single feed-forward stage, to compute its frequency response. The input is differential (not shown).

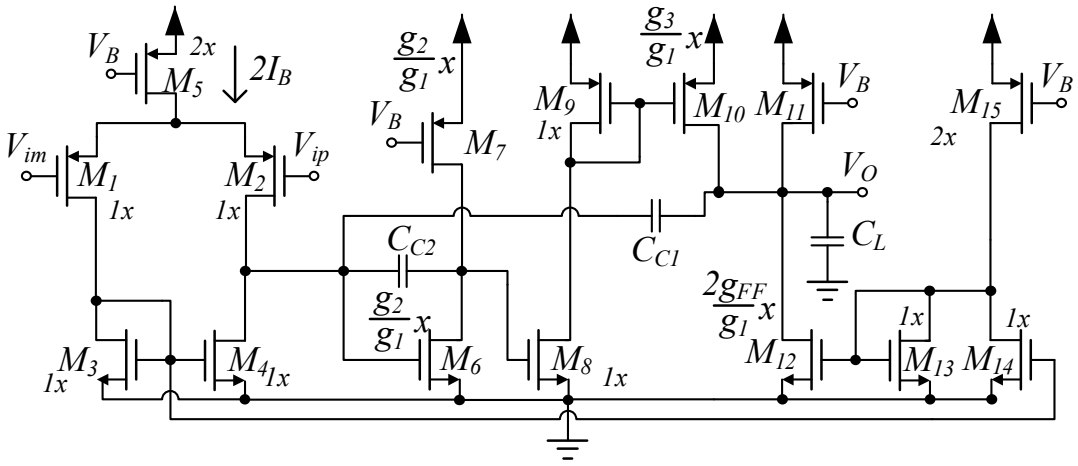

Fig. 2. Schematic of the proposed three-stage RNMC OTA with feed-forward stage.

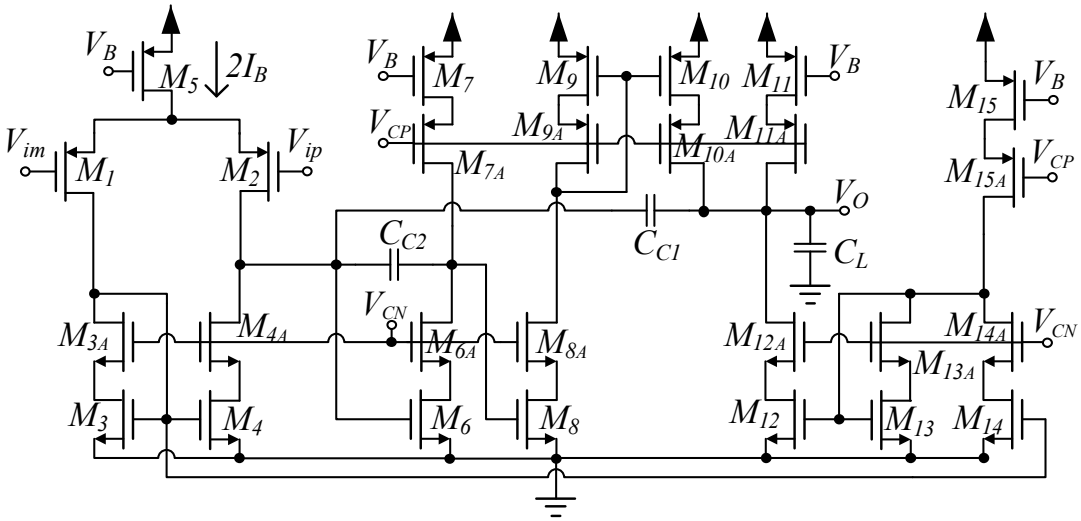

Fig. 3. Detailed schematic of the proposed three-stage RNMC OTA with feed-forward path.

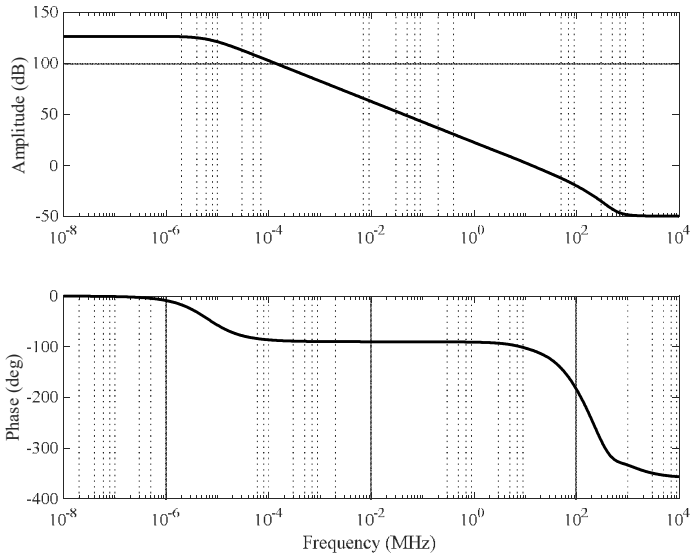

Fig. 4. Open-loop small-signal response (magnitude and phase).

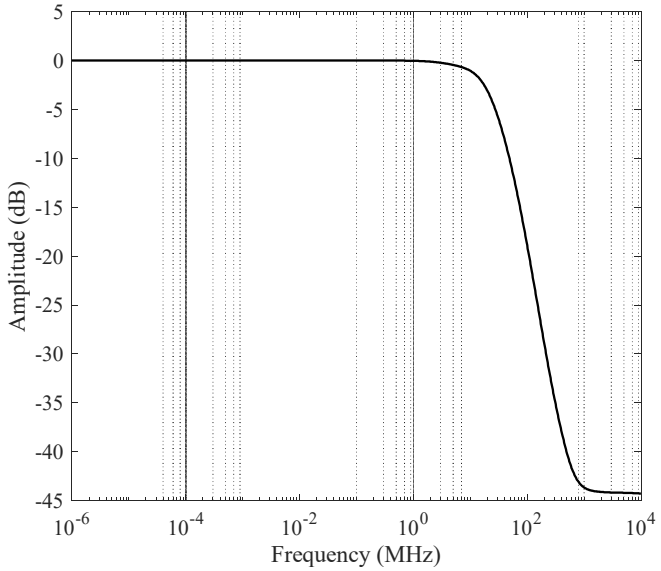

Fig. 5. Closed-loop small-signal response.

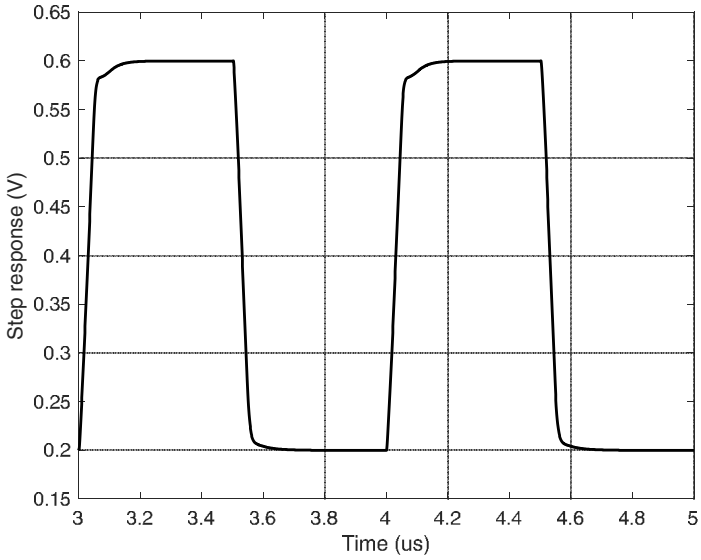

Fig. 6: Transient response of the OTA in unity gain buffer configuration with a $0.4 \pm 0.2V$ input square wave.

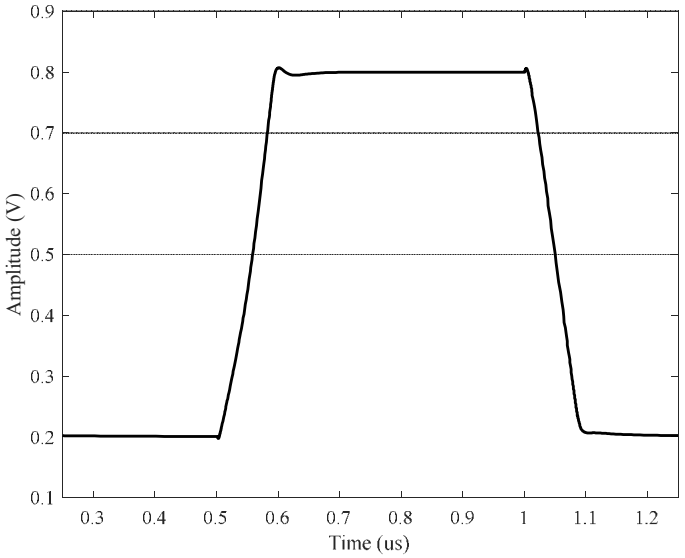

Fig. 7. Transient response of the OTA to +/-1.5μA input square wave in transimpedance configuration.

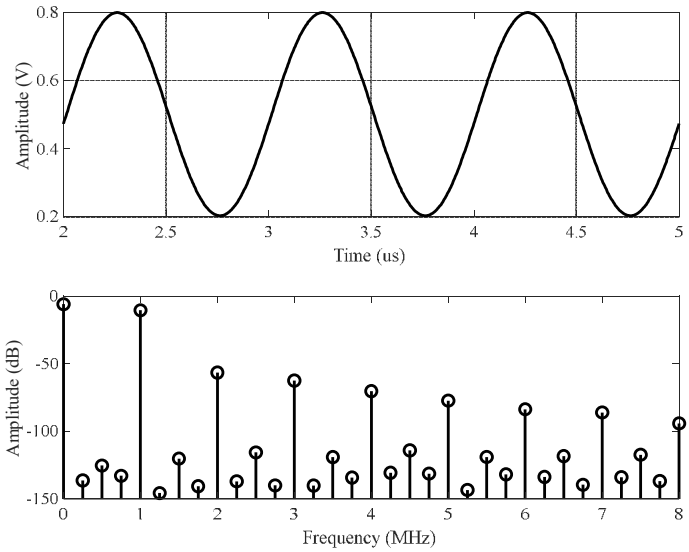

Fig. 8. Transient response to input sinusoid, in time (top) and frequency (bottom).