Three-Stage CMOS OTA for Large Capacitive Loads With Efficient Frequency Compensation Scheme

Alfio Dario Grasso, *Member, IEEE*, Gaetano Palumbo, *Senior Member, IEEE*, and Salvatore Pennisi, *Senior Member, IEEE*

***Abstract*—A simple compensation strategy, which employs passive components only, is adopted to design a three-stage operational transconductance amplifier (OTA) suitable for driving high capacitive loads. Compared to the classical nested Miller compensation technique, the new solution exploits two additional resistors and allows a reduction in the values of the compensation capacitors of about an order of magnitude. The OTA was fabricated using 0.35-μm CMOS technology and exhibits a 1.4-MHz gain–bandwidth with a load of 500 pF.**

***Index Terms*—Frequency compensation, nested Miller, operational transconductance amplifier (OTA).**

I. Introduction

With the scaling down of device sizes and supply voltages, single-stage amplifiers based on the cascading technique are no longer suitable for achieving high dc gains under low supply voltages. Currently, amplifiers exhibiting a dc gain in excess of 100 dB can be profitably implemented by cascading several simple transconductance gain stages. However, the design of such amplifiers is a challenging task since the increased number of high impedance nodes (and, in turn, of low-frequency poles) may result in instability. When three stages are used to implement the operational transconductance amplifier (OTA) and the last stage is the only inverting one, the nested Miller compensation (NMC) topology can be used [1]–[3]. This technique employs two compensation capacitors that exploit the Miller effect to split low-frequency poles and achieve the desired phase margin and transient response. However, this solution results in bandwidth reduction and high power consumption, especially in CMOS technology [4].

In recent years, many compensation techniques improving the basic NMC topology have been proposed [4]–[14]. The most efficient of these exploits additional transconductance stages to implement active compensation networks [7], [8], [11]–[14]. In this manner, optimized bandwidth is achieved at the expense of increased circuit complexity and dissipation.

In this brief, a simple compensation strategy, which uses passive components only, is exploited to design a three-stage OTA suitable for driving high capacitive loads. The circuit was fabricated and experimentally tested.

Comparison with other approaches shows that the adopted one achieves overall better performance.

Manuscript received April 13, 2006; revised May 8, 2006. This paper was recommended by Associate Editor A. Demosthenous.

The authors are with the Dipartimento di Ingegneria Elettrica, Elettronica e dei Sistemi (DIEES), University of Catania, 08028 Barcelona, Spain (e-mail: agrasso@diees.unict.it; gpalumbo@diees.unict.it; spennisi@diees.unict.it).

Digital Object Identifier 10.1109/TCSII.2006.882231

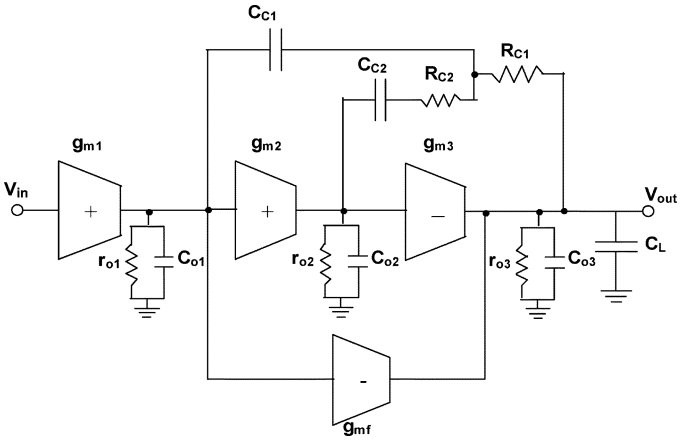

Fig. 1. OTA small-signal block diagram with the adopted compensation technique.

II. Adopted Compensation Technique

The OTA block diagram including the adopted compensation approach is shown in Fig. 1. It was originally presented in [6] along with a simplified analysis that will be better formalized here. Compensation exploits capacitors C_{C1} and C_{C2}, resistors R_{C1} and R_{C2}, and the additional active transconductance stage g_{mf} that, as will be clear later, can be implemented without entailing any extra transistor. In Fig. 3, g_{mi}, r_{oi}, and C_{oi} denote the ith stage transconductance, output resistance, and output capacitance, respectively, whereas g_{mf} is the transconductance of the feedforward transistor, and C_L is the load capacitor.

Assuming for each stage a dc gain $A_{Vi} = g_{mi}r_{oi}$, that is much greater than 1, the open-loop transfer function of the OTA is expressed as that shown in (1) at the bottom of the next page, where $A_0 = g_{m1}r_{o1}g_{m2}r_{o2}g_{m3}r_{o3}$ is the dc gain, and $\omega_{P1} \approx 1/(r_{o1}g_{m2}r_{o2}g_{m3}r_{o3}C_{C1})$ is the dominant pole due to the Miller effect on C_{C1}. Therefore, the gain–bandwidth product is, as usual, $\omega_{\mathrm{GBW}} = g_{m1}/C_{C1}$. The function exhibits two other poles, which depend on C_{C2} and C_L, and an additional high-frequency pole, which is proportional to the parasitic capacitance C_{o2}. Moreover, two zeros are also included.

By inspection of (1), it can be seen that, as the main advantage of the adopted approach, the zeros can be made both negative and, in particular, their values can be adjusted to exactly cancel out the two higher poles. Indeed, by setting

$$R_{C1} = \frac{1}{g_{m3}} \tag{2}$$

and equating the coefficients of the second-order polynomials

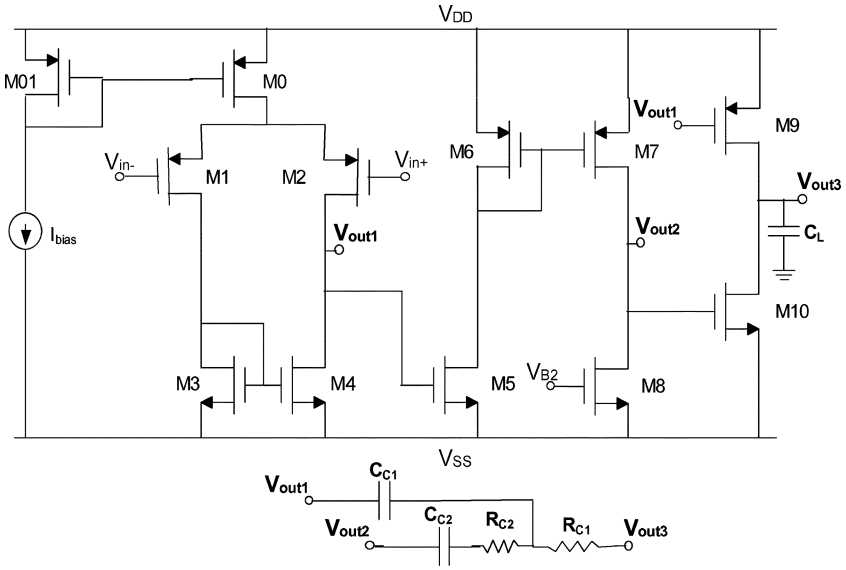

Fig. 2. Schematic diagram of the OTA implemented.

at the nominator and denominator of (1), we get

$$R_{C2} = \frac{C_L}{g_{m3}C_{C2}} - \frac{g_{mf}}{g_{m3}g_{m2}} \tag{3}$$

$$C_{C1} = \frac{g_{m3} - g_{m2}}{g_{m2}} C_{C2}. \tag{4}$$

The feedforward transconductance g_{mf} can be set equal to g_{m3} to obtain a symmetrical push–pull output stage that enhances slew rate performance. To ensure positive values of R_{C2} and C_{C1}, we must satisfy the constraints $C_{C2} < C_L(g_{m2}/g_{m3})$ and $g_{m3} > g_{m2}$. Note that, in principle, small values of C_{C2} can be utilized. This causes a decrease of C_{C1} from (4) (thereby improving ω_{GBW} and slew rate or, alternatively, enabling the use of a lower g_{m1} for a given ω_{GBW}) but increases R_{C2} from (3). Therefore, a tradeoff between power and area consumption must be achieved when setting C_{C2}. Of course, the compensation capacitors must be greater than the parasitic capacitances at high impedance nodes for our derivations to be valid. Moreover, it is worth noting that unlike conventional compensation strategies where C_{C1} is proportional to C_L, now C_{C1} is proportional to C_{C2}. This suggests that the compensation strategy is well suited for high capacitive load.

After pole-zero cancellation, (1) is reduced to a two-pole transfer function in which the second pole depends upon the parasitic capacitance C_{o2}. However, we can assume a phase margin almost equal to 90° if the parasitic pole is much higher than the gain–bandwidth product. This condition is ensured by

$$C_{C2} \gg \frac{g_{m1}g_{m2}}{(g_{m3} - g_{m2})^2} C_{o2}. \tag{5}$$

Since the described approach is based on a double pole-zero cancellation, which cannot be obtained exactly, the effect of mismatches must be investigated. The first consideration is that mismatches are more critical if the second-order polynomials in (1) exhibit low damping factors ξ. Using (2) and (3), the damping factor is

$$\xi = \frac{1}{2}\left(\frac{C_L}{C_{C2}} + \frac{g_{m3}}{g_{m2}} - 1\right)\left[\frac{C_L}{C_{C2}}\left(\frac{g_{m3}}{g_{m2}} - 1\right)\right]^{-\frac{1}{2}} \tag{6}$$

which is always greater than 1 if conditions ensuring positive values of R_{C2} and C_{C1} are met. This means that the poles and zeros are all real (and also negative).

In addition, it is shown in the Appendix that, by following our design equations, the cancellation of the lower frequency doublet exhibits a low sensitivity to process variations. As a result, the higher frequency doublet (which is more sensitive to mismatches) should be placed well beyond the gain–bandwidth product, to reduce the effect of the phase margin. This is achieved by the condition (usually met)

$$\omega_{\mathrm{GBW}} < \frac{g_{m3}}{C_L}\left(\frac{g_{m3} + g_{m2}}{g_{m3} - g_{m2}}\right)\left(\frac{C_L}{C_{C2}} - 1\right). \tag{7}$$

III. Implementation and Experimental Results

Fig. 2 shows the simplified schematic of the OTA adopting the compensation network discussed above. The input stage is made up of transistors M0–M4 that implement a pMOS differential pair with a current mirror load. The second noninverting stage is made up of transistors M5–M8. The last stage is implemented through common source transistors M9 and M10.

$$A_v(s) = A_0 \frac{1 + s\left[R_{C1}C_{C1} + \left(R_{C2} - \frac{1}{g_{m3}} + \frac{g_{mf}}{g_{m2}g_{m3}} + R_{C1}\right)C_{C2}\right] + s^2 \frac{(1+g_{m2}R_{C2})g_{m3}R_{C1} - 1 + g_{mf}R_{C1}}{g_{m2}g_{m3}} C_{C1}C_{C2}}{\left(1 + \frac{s}{\omega_{P1}}\right)\left[1 + s\left(R_{C2} + \frac{1}{g_{m2}} - \frac{1}{g_{m3}} + \frac{g_{mf}}{g_{m2}g_{m3}}\right)C_{C2} + s^2 \frac{1 - g_{m2}R_{C1}}{g_{m2}g_{m3}} C_{C2}C_L\right]\left(1 + s\frac{R_{C1}}{1 - g_{m2}R_{C1}} C_{o2}\right)} \tag{1}$$

TABLE I
TRANSISTOR DIMENSIONS

Transistor	Aspect ratio
M0	2x(40/1)
M01	40/1
M1, M2	4x(40/1)
M3, M4	40/0.7
M5	2x(40/0.7)
M6, M7	50/0.7
M8	40/1
M9	4x(40/0.7)
M10	8x(40/0.7)

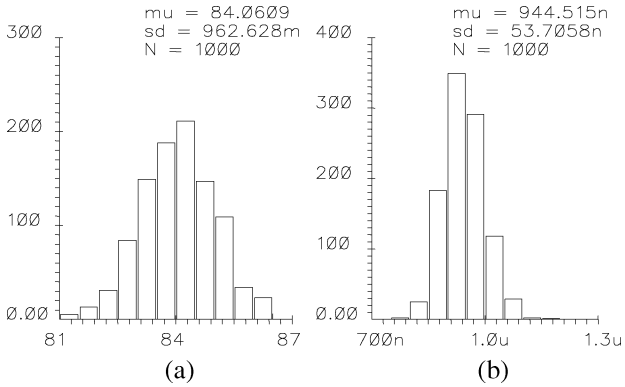

Fig. 3. Statistical distribution of (a) phase margin and (b) 1% settling time resulting from 1000 Monte Carlo post-layout simulations due to local mismatches.

It is worth noting that the gate of M9 is driven by the output of the first stage to implement a pseudo class AB output stage able to drive the load capacitor C_L with a current much higher than the output branch quiescent current. As a result, the slew rate is determined by the maximum available current from the first stage charging C_{C1}. Moreover, with this connection, transistor M9 also implements the feedforward transconductor required for the compensation, which hence does not require any extra transistor. The OTA was designed using a triple-metal double-poly 0.35-μm n-well CMOS process supplied by AMS and accessed through EUROPRACTICE. The transistor aspect ratios are reported in Table I. The supply voltage is 1.5 V, and I_{bias} is set equal to 10 μA. The total current dissipation is hence 150 μA. The stage transconductances are $g_{m1} = 296$ A/V, $g_{m2} = 478$ A/V, and $g_{m3} = g_{mf} = 1.23$ mA/V. Consequently, to obtain a gain–bandwidth product of 1.5 MHz with a capacitive load of 500 pF, we get $C_{C1} = 30$ pF, and from design equations, we set $R_{C1} = 800\ \Omega$, $R_{C2} = 18\ k\Omega$, and $C_{C2} = 20$ pF. All these values are obtained by strictly following the design equations. Of course, a further optimization can be obtained through computer simulations. In our case, the value of R_{C2} was only slightly increased to 20 $k\Omega$ to slightly reduce peaking in the step response. The damping factor was evaluated according to (6) and resulted in a value equal to 2.1. Besides, it can be seen that (7) is also satisfied.

The robustness of the compensation technique against mismatches was addressed in the Appendix. To further confirm that the double pole-zero cancellation is rather insensitive to mismatches, we performed post-layout Monte Carlo simulations. Fig. 3(a) and (b) illustrates the phase margin and 1% settling time distribution for 1000 iterations, taking into account the local variations (mismatches) of circuit parameters. The mean value of the phase margin is about 84° with a standard deviation of about 1°. The mean value of the settling time is 945 ns with a standard deviation of 53 ns. If lot-to-lot variations are considered, larger standard deviations respectively equal to 5° and 83 ns are found.

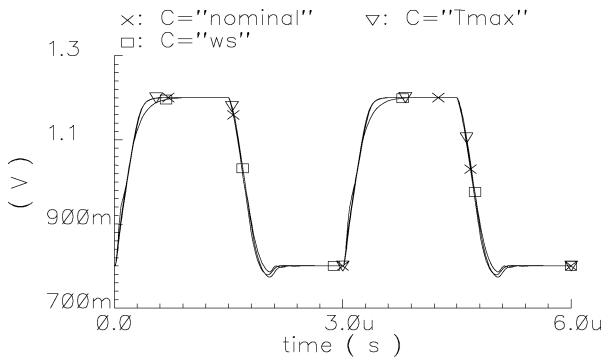

Fig. 4. Corner simulation step response (nominal: nominal models; ws: worst speed; Tmax: maximum temperature 85 °C).

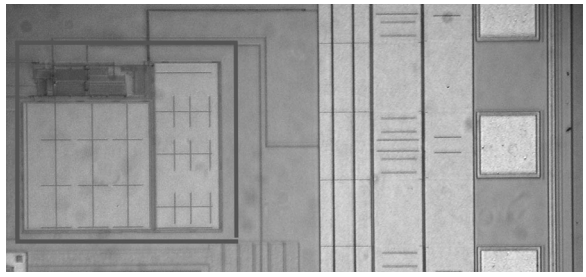

Fig. 5. Chip microphotograph of the OTA.

TABLE II
OTA MAIN PERFORMANCE PARAMETERS

PARAMETER	VALUE
Technology	0.35 µm CMOS
Area	0.075 mm^2
Power Supply	1.5 V
Total Bias Current	150 µA
Loading Capacitance	500 pF
Input Offset Voltage	2.3 mV
DC Gain	113 dB (simulated)
Gain-bandwidth Product	1.4 MHz
Phase Margin	75°
CMRR @ DC	78 dB (simulated)
Positive (Negative) Slew Rate †	2.2 (-1.8) V/µs
Positive (Negative) Settling Time at 1%	810 (740) ns
HD2/HD3 @ 100 kHz, 400 mVpp †	-54.25/-60.36 dB
HD2/HD3 @ 100 kHz, 500 mVpp †	-48.20/-56.71 dB

† in unity gain configuration

Additional corner simulations were executed. The phase margin was 96° and 80° in the worst speed and worst temperature (85 °C) case, respectively. Fig. 4 illustrates the response of the OTA in unity gain to a 400-mV$_{\text{p-p}}$ input step. It can be noted that there is no sensible variation in the settling time in the worst temperature case, while the settling time increased to about 170 ns when worst speed models were considered.

Prototypes of the proposed circuit were then fabricated and experimentally characterized. The die micrograph is depicted in Fig. 5. It can be seen that the compensation capacitors occupy about 85% of the overall area, which is equal to 0.075 mm^2.

TABLE III
PERFORMANCE COMPARISON

	MNMC [7]	NGCC [8]	NMCFNR [9]	DFCFC [11]	AFFC [12]	DLPC [13]	ACBC$_F$ [14]	**This work**[a]
C_L (pF)	100	20	100	100	100	120	500	500
GBW (MHz)	100	0.61	1.8	2.6	5.5	7	1.9	1.4
SR (V/μs)[a]	35	2.5	0.79	1.32	1.41	3.3	1	2
Power (mW@V_{DD})	76@8	0.68@2	0.406@2	0.42@2	0.250@1.5	0.33@1.5	0.324@2	0.225@1.5
FOM_S (MHz·pF/mW)	132	18	443	619	2200	2545	2932	**3111**
FOM_L (V/μs·pF/mW)	46	74	195	314	564	1200	1543	**4444**

a. average value

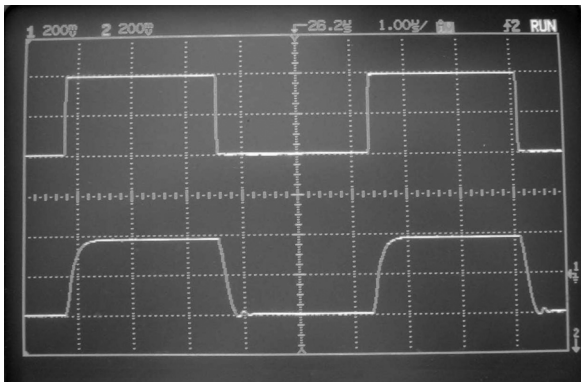

Fig. 6. Measured transient response to a 400-mV$_{p\text{-}p}$ input step (Xdiv: 1 μs, Ydiv: 200 mV).

Fig. 6 shows the measured transient response to a 400-mV$_{p\text{-}p}$ input step with an off-chip capacitive load of 500 pF. We measured a positive and a negative slew rate equal to 2.2 and -1.8 V/μs, respectively. The positive and negative settling times at 1% of the final value were 810 and 740 ns, respectively.

To measure the gain–bandwidth product and the phase margin, we used the indirect method proposed in [16] that allows us to determine the open-loop parameters of an amplifier by measuring its closed-loop 3-dB cutoff frequency and the corresponding phase. In particular, we measured the magnitude and phase of the OTA in unity–gain configuration, and we get an estimated value of the gain–bandwidth product and the phase margin equal to 1.4 MHz and 80°, respectively. The measured results are summarized in Table II.

To provide a performance comparison between the adopted compensation technique and other reported compensation approaches, two figures of merit [FOM$_S$ $=$ $(\omega_{\rm GBW} \cdot C_L)/(2\pi \cdot$ Power) and FOM$_L$ $=$ $(SR \cdot C_L)$Power] are commonly used [11]–[14]. The higher the value of these figures of merit, the better is the amplifier performance. The figures of merit along with the main performance parameters for several three-stage amplifiers reported in the literature are summarized in Table III. It is apparent that the proposed OTA exhibits the highest values of both figures of merit.

IV. CONCLUSION

A compensation technique suitable for heavy capacitive loads was used in a three-stage OTA. It was implemented using the transistors of the basic OTA topology, thus optimizing power consumption and entailing a double pole-zero cancellation. Under the usual conditions, this cancellation is reliable, thus avoiding an *ad hoc* bias circuit to allow compensation resistors tracking the MOS transconductances. To confirm this feature, several theoretical analysis and statistical simulations were provided, which, however, are not necessary during the design process. Prototypes were fabricated using a standard CMOS 0.35-μm technology and show better small-signal and large-signal performances than other previously reported compensation techniques.

APPENDIX

In this section, we shall investigate the effects of an imperfect (double) pole-zero cancellation in (1) caused by process tolerances. Indeed, the presence of a pole-zero doublet in the loop gain can lead to a slower settling [3], [15] or even instability.

With this purpose, consider the generic transfer function

$$A(s) = \frac{\omega_{\rm GBW}}{s} \cdot \frac{1 + b_1 s + b_2 s^2}{1 + a_1 s + a_2 s^2} \tag{A1}$$

which represents an approximation of (1).

Let us model the effect of mismatch on coefficients b_i through parameters $\delta_{1,2}$ as

$$\begin{aligned} b_1 &= a_1(1+\delta_1) = \frac{2\xi}{\omega_n}(1+\delta_1) \\ b_2 &= a_2(1+\delta_2) = \frac{1}{\omega_n^2}(1+\delta_2) \end{aligned} \tag{A2}$$

where we introduced the damping factor ξ and the poles frequency ω_n in the rightmost expressions. We have already demonstrated that ξ is always greater than 1. This ensures that the poles (zeros) are real (and negative) and are located at a frequency approximately equal to $1/a_1 = \omega_n/2\xi$ and $a_1/a2 = 2\xi\omega_n$. By imposing $a_1/a_2 > \omega_{\rm GBW}$ (i.e., the second doublet is at the right of the gain–bandwidth product), we derived (7). More specifically, the evaluation of ω_n from (1) indicates that ω_n can be around or greater than $\omega_{\rm GBW}$. Assuming pessimistically $\omega_n \cong \omega_{\rm GBW}$, the lower frequency (higher frequency) doublet is about 2ξ times lower (greater) than $\omega_{\rm GBW}$. Therefore, the mismatch of parameters a_1 and b_1 (i.e., of the lowest-frequency doublet) could, in principle, worsen significantly the settling time and even modify the actual unity-gain frequency.

Consider now the (pessimistic) case in which the mismatches are all uncorrelated and assume that the values of "all" the circuit parameters (transconductances, compensation resistors, and capacitors) are uniformly distributed within $\pm 20\%$ of their nominal values. In these conditions, we evaluated the variation of

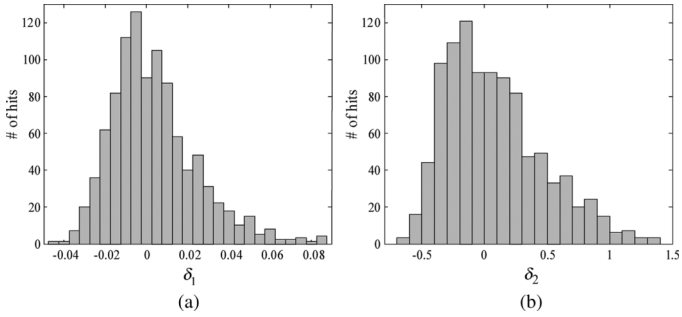

Fig. 7. Distribution of the mismatch between the coefficients of the second-order polynomials in (1).

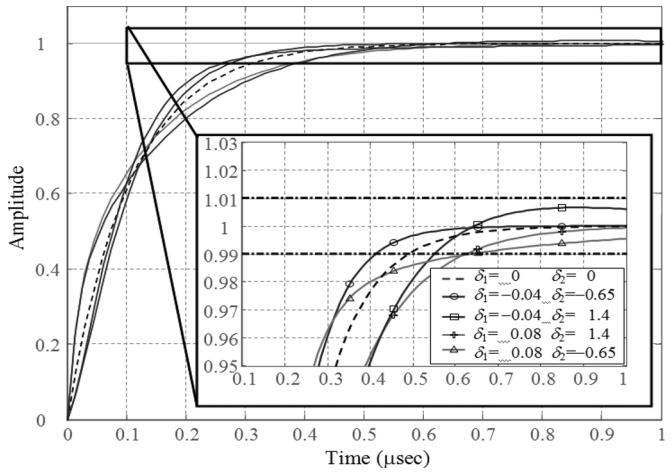

Fig. 8. Worst-case step responses of (A1) in unitary feedback (the inset shows the 1% settling range).

parameters δ_1 and δ_2. The distribution over 1000 iterations is shown in Fig. 7, where we set $C_L/C_{C2} = 25$, $g_{m3}/g_{m2} = 2.5$, $\omega_{\mathrm{GBW}}/\omega_n = 0.95$, and $\xi = 2$ (similar results were obtained choosing other values of the parameters provided that the design equations in Section II are satisfied). From Fig. 7(a), it is apparent that δ_1 is almost insensitive to parameter mismatches since its variation is approximately limited between -0.04 and 0.08 (with a mean value of 0.003). Therefore, the adopted compensation technique inherently attenuates the effects of mismatches on the imperfect cancellation of the lower pole-zero doublet, thus limiting the worsening of the settling time. On the other hand, Fig. 7(b) shows that the parameter δ_2 is much more affected by mismatches since its value approximately varies between -0.65 and 1.40 (with a mean value of 0.08). Nevertheless, as already stated, δ_2 is responsible for the imperfect cancellation of the highest pole-zero doublet located well beyond the gain–bandwidth product. The modification of the settling time (and phase margin) was then evaluated. Fig. 8 shows the step response of (A1) used in a unitary feedback loop for the nominal case and for the extreme values of δ_1 and δ_2. An acceptable maximum settling time variation was found (about 30% greater than the nominal case with $\delta_1 = \delta_2 = 0$). These results are similar to those given by the Monte Carlo simulation at transistor level. Moreover, a maximum phase margin variation of $\pm 10^\circ$ was found.

Finally, it can be seen that resistors R_{C1} and R_{C2} could be implemented through triode-biased MOS exploiting tuning techniques to further reduce sensitivity to process variations but comes at the expense of circuit complexity and power dissipation.

REFERENCES

[1] J. H. Huijsing and D. Linebarger, "Low-voltage operational amplifier with rail-to-rail input and output ranges," *IEEE J. Solid-State Circuits*, vol. SSC-20, no. 6, pp. 1144–1150, Dec. 1985.

[2] R. G. H. Eschauzier and J. H. Huijsing, *Frequency Compensation Techniques for Low-Power Operational Amplifiers*. Boston, MA: Kluwer, 1995.

[3] G. Palumbo and S. Pennisi, *Feedback Amplifiers: Theory and Design*. Boston, MA: Kluwer, 2002.

[4] K. N. Leung and P. K. T. Mok, "Analysis of multistage amplifier–frequency compensation," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 48, no. 9, pp. 1041–1056, Sep. 2001.

[5] ——, "Right-half plane zero removal technique for low-voltage low-power nested miller compensation," in *Proc. ICECS*, 1999, vol. II, pp. 599–602.

[6] G. Palumbo and S. Pennisi, "Design methodology and advances in nested-Miller compensation," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 7, pp. 893–903, Jul. 2002.

[7] R. G. H. Eschauzier and J. H. Huijsing, "A 100-MHz 100-dB operational amplifier with multipath nested miller compensation," *IEEE J. Solid-State Circuits*, vol. 27, no. 12, pp. 1709–1716, Dec. 1992.

[8] F. You, S. Embabi, and E. Sanchez-Sinencio, "Multistage amplifier topologies with nested Gm-C compensation," *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 2000–2011, Dec. 1997.

[9] K. N. Leung and P. K. T. Mok, "Nested Miller compensation in low-power CMOS design," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 48, no. 4, pp. 388–394, Apr. 2001.

[10] K. N. Leung, P. K. T. Mok, W. H. Ki, and J. K. O. Sin, "Three-stage large capacitive load amplifier with damping-factor-control frequency compensation," *IEEE J. Solid-State Circuits*, vol. 35, no. 2, pp. 221–230, Feb. 2000.

[11] H. Lee and P. K. T. Mok, "Active-feedback frequency-compensation technique for low-power multistage amplifiers," *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 511–520, Mar. 2003.

[12] ——, "Advances in active-feedback frequency compensation with power optimization and transient improvement," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process*, vol. 51, no. 9, pp. 1690–1696, Sep. 2004.

[13] H. Lee, K. N. Leung, and P. K. T. Mok, "A dual-path bandwidth extension amplifier topology with dual-loop parallel compensation," *IEEE J. Solid-State Circuits*, vol. 38, no. 10, pp. 1739–1744, Oct. 2003.

[14] X. Peng and W. Sansen, "AC boosting compensation scheme for low-power multistage amplifiers," *IEEE J. Solid-State Circuits*, vol. 39, no. 11, pp. 2074–2077, Nov. 2004.

[15] B. Y. Kamath, R. G. Meyer, and P. R. Gray, "Relationship between frequency response and settling time of operational amplifiers," *IEEE J. Solid-State Circuits*, vol. SSC-9, no. 6, pp. 347–352, Dec. 1974.

[16] G. Giustolisi and G. Palumbo, "An approach to test the open-loop parameters of feedback amplifiers," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 1, pp. 70–75, Jan. 2002.