Transconductance Amplifier Structures With Very Small Transconductances: A Comparative Design Approach

Anand Veeravalli, *Student Member, IEEE*, Edgar Sánchez-Sinencio, *Fellow, IEEE*, and José Silva-Martínez, *Senior Member, IEEE*

***Abstract*—A family of CMOS operational transconductance amplifiers (OTAs) has been designed for very small G_m's (of the order of nanoamperes per volt) with transistors operating in moderate inversion. Several OTA design schemes such as conventional, using current division, floating-gate, and bulk-driven techniques are discussed. A detailed comparison has also been made among these schemes in terms of performance characteristics such as power consumption, active silicon area, and signal-to-noise ratio. The transconductance amplifiers have been fabricated in a 1.2-μm n-well CMOS process and operate at a power supply of 2.7 V. Chip test results are in good agreement with theoretical results.**

***Index Terms*—Bulk-driven transistors, current division, floating gates, OTA, small G_m.**

I. Introduction

In THE FIELD of medical electronics, active filters with very low cutoff frequencies (of the order of a few hertz) are needed due to the relatively slow electrical activity of the human body [1]. Another area of application of low-frequency circuits is ramp generation for analog-to-digital converter (ADC) testing [2] and in the field of neural networks [3]. Thus, there is a strong motivation for developing integrated solutions for circuits that are capable of operating at very low frequencies. For an operational transconductance amplifier-capacitor (OTA-C) filter implementation, such low frequencies imply large capacitors and very low transconductances [4], [5]. Thus, there are two entirely independent angles to the problem that need to be addressed. One is the design of OTAs with very low transconductances (typically of the order of a few nanoamperes per volt) and high linearity, while the other is the realization of very large capacitors (typically of the order of a few nanofarads) on chip. Keeping the foregoing in mind, different design techniques for obtaining low transconductances are analyzed here, and a comparative study has been made among these schemes in terms of performance characteristics such as power consumption, active silicon area, and signal to noise ratio (SNR). Special emphasis has been given to design in the moderate inversion region of operation of the MOS transistor due to the possibility of reaching a good tradeoff between power and area requirements.

II. OTA Topologies

Four different OTA topologies were designed in moderate inversion, using one equation all-region MOSFET model [6] for the same transconductance value of 10 nA/V, and the tradeoffs concerning design parameters such as power consumption, silicon area, and SNR were studied.

Manuscript received October 27, 2000; revised August 30, 2001.

A. Veeravalli is with Texas Instruments Incorporated, Dallas, TX 75243 USA.

E. Sánchez-Sinencio and J. Silva-Martínez are with the Department of Electrical Engineering, Texas A&M University, College Station, TX 77843 USA.

Publisher Item Identifier S 0018-9200(02)04937-5.

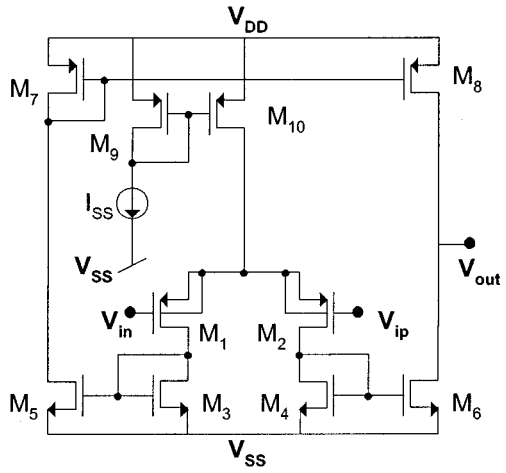

Fig. 1. Reference OTA.

A. Reference OTA (Design A)

The schematic is shown in Fig. 1. This OTA consists of a differential pair (M_1 and M_2) and three current mirrors. The overall transconductance of the amplifier G_m is the same as that of M_1, M_2 (with $M_3 = M_4 = M_5 = M_6$, $M_7 = M_8$). Depending on the value of the required transconductance, the current levels for this basic topology can be extremely small (of the order of several picoamperes for G_m's around several picoamperes per volt). This leads to W/L ratios of the order of 0.001 or less. Matching such geometries is a great challenge from a layout perspective. We have used an inversion level[1] $i_f = (I_d)/(I_s)$ of 10 for the drivers M_1 and M_2 in order to obtain the required transconductance ($\sim$ 10 nA/V), at the same time making sure that their lengths are not too large. The inversion levels [6] for the current mirrors were chosen to be 80 to allow them to operate closer to strong inversion for better matching. The same holds for the following designs as well.

B. OTA With Current Division and Source Degeneration (Design B—SD+CD)

This topology is described in [7] and [8]. This circuit is actually a combination of two schemes, i.e., current splitting and source degeneration. Fig. 2(a) illustrates the idea behind current splitting where the effective G_m is given by

[1] $I_s = \mu_n C_{ox}(\varphi_t^2/2)(W/L)$, φ_t is the thermal voltage.

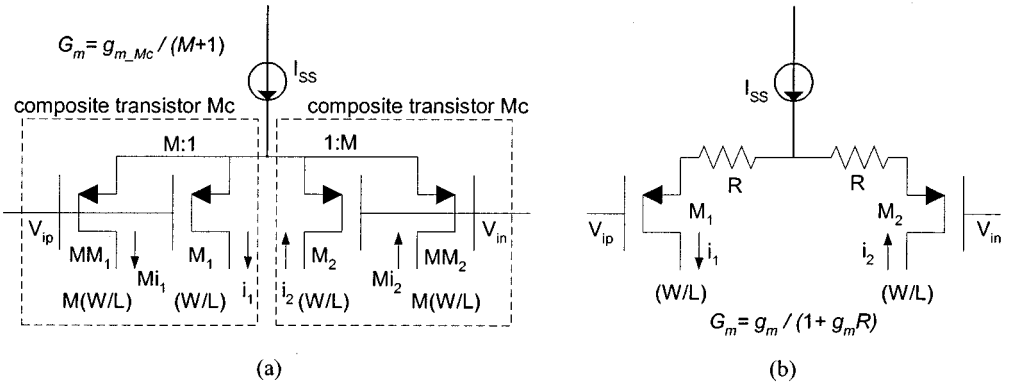

Fig. 2. Transconductance reduction techniques. (a) Current splitting. (b) Source degeneration.

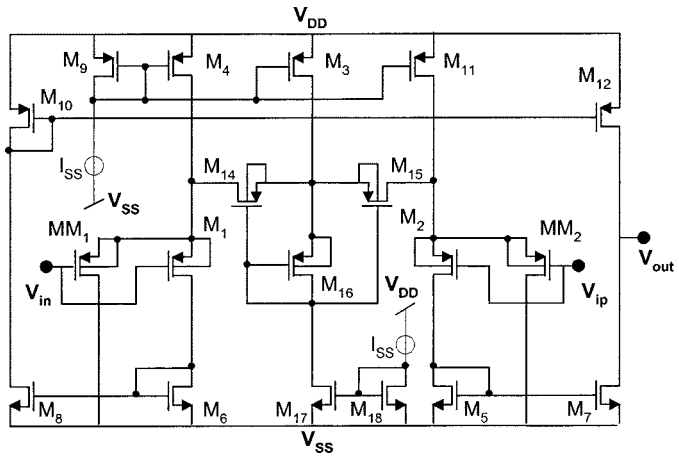

Fig. 3. OTA with current division and source degeneration.

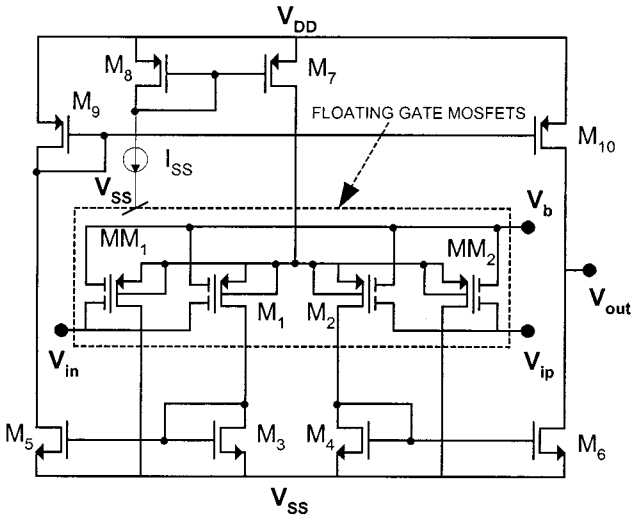

Fig. 4. Floating-gate OTA with current division.

$G_m = (g_{m\_M_C})/(M+1)$, where M_C is the composite transistor (before splitting) as shown in Fig. 2(a). The small-signal currents in transistors M_1 and MM_1 are split by the factor M of the ratio in their sizes and only the currents i_1 and i_2 are used. Thus, the effective transconductance is reduced by the factor $M+1$ compared to that before current splitting [8]. Fig. 2(b) shows the principle behind source degeneration where the effective G_m is given by

$$G_m = \frac{g_{m\_M1,2}}{1 + g_{m\_M1,2} R}$$

which gives an effective transconductance reduction by the factor $1 + G_m R$.

The overall schematic of the OTA obtained by a combination of both the above-mentioned schemes is shown in Fig. 3. This structure has a source degeneration linearization and an additional transconductance reduction by implementing current division through MM_1 and MM_2. Small-signal analysis gives the overall G_m as

$$G_m = \left(\frac{g_{m\_M1,2}}{1 + \frac{(M+1) g_{m\_M1,2}}{g_{0\_M14}}} \right) \quad M = \frac{g_{m\_MM1}}{g_{m\_M1}} \tag{1}$$

$$g_{0\_M14} = n\mu C_{0x} \frac{W_{M14}}{L_{M14}} \left(V_{SG\_M14} - |V_{TP}| \right)$$

$$= n\mu C_{0x} \frac{W_{M14}}{L_{M14}} \sqrt{\frac{2 I_{SS}}{n\mu C_{0x}} \frac{L_{16}}{W_{16}}} \tag{2}$$

where g_m and $g_o (= 1/R)$ are, respectively, the transconductance and output conductance of the MOS transistor. G_m can be changed by changing g_{o_M14}, which is controlled by the bias current I_{SS}. The transistors M_{14} and M_{15} are biased in the triode region and thus act as source-degeneration resistors. The purpose of M_3, M_{16}, M_{17}, and M_{18} is to control the V_{SG} of M_{14} and M_{15}, and thus, their resistance. MM_1 and MM_2 divert a significant portion of the bias current to the rail, thus reducing G_m by the factor $1 + M$.

As discussed earlier, to realize extremely small G_m's, we need very small currents, which are not easy to generate and are not well controlled. Also, transistors with very long lengths are required and these are difficult to match from a layout perspective. For these reasons, we use the current division scheme, which enables us to increase the current levels while maintaining very low transconductance levels. From a layout perspective, transistor M_1 is used as the unit and MM_1 is built up using M fingers of M_1 for better matching.

C. Floating-Gate OTA (Design C—FG+CD)

This schematic is shown in Fig. 4. In this scheme, the input transistors are floating-gate MOS transistors [9], [10] with two inputs each (input and bias). Since floating-gate techniques have a natural attenuation due to the voltage division at the input

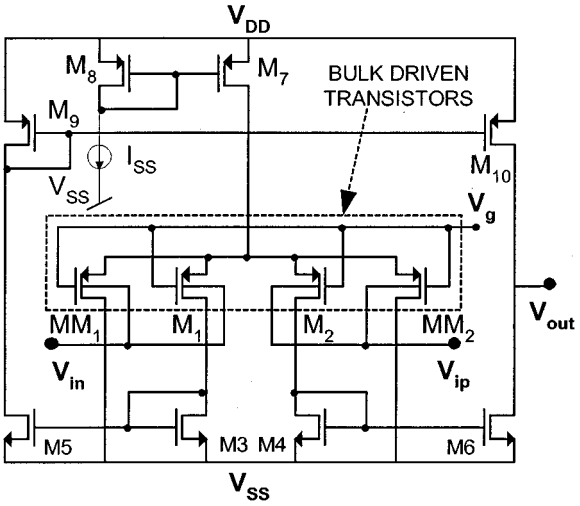

Fig. 5. Bulk-driven OTA with current division.

capacitors, they are a natural choice for obtaining small G_m's. To further reduce the G_m, current division has also been incorporated. The overall G_m in terms of the model parameters, assuming that the parasitic capacitances between the floating gate and the source, drain, and bulk terminals are negligible compared to C_A and C_B, is approximately given by

$$G_m \cong \left(\frac{C_A}{C_A+C_B}\right) g_{m\_\mathrm{FG1}} = \left(\frac{C_A}{C_A+C_B}\right)\left(\frac{I_{SS}}{\varphi_t n\left(1+\sqrt{1+i_{f\_M1}}\right)}\right) \cdot \left(\frac{1}{M+1}\right) \tag{3}$$

where C_A is the capacitance coupling at input A to the floating gate, C_B is the capacitance coupling at input B to the floating gate, and g_{m_FG1} is the transconductance of the floating-gate transistor M_1. For proper input voltage scaling, C_A and C_B should be significantly larger than the total parasitic capacitance seen at the floating gate. A good compromise would be to make C_A and C_B around 5–10 times this parasitic capacitance. In our design, $C_A = C_B \sim 8C_P$.

D. Bulk-Driven OTA (Design D—BD+CD)

In this topology, shown in Fig. 5, the inputs of the OTA are driven through the bulks of the input transistors rather than the gates [11], [12]. Bulk-driven transconductance g_{mb} is typically around 0.2–0.4 times G_m, but it is very process dependent. Current division has also been included to further reduce G_m levels. Analysis yields the overall OTA transconductance in terms of the model parameters as

$$G_m = \left(\frac{\gamma_0}{2\sqrt{2\varphi_{FB}+|V_{\mathrm{BS}}|}}\right) g_{m\_M1} = \left(\frac{\gamma_0}{2\sqrt{2\varphi_{\mathrm{FB}}+|V_{\mathrm{BS}}|}}\right)\left(\frac{I_{SS}}{\varphi_t n\left(1+\sqrt{1+i_{f\_M1}}\right)}\right) \cdot \left(\frac{1}{M+1}\right) \tag{4}$$

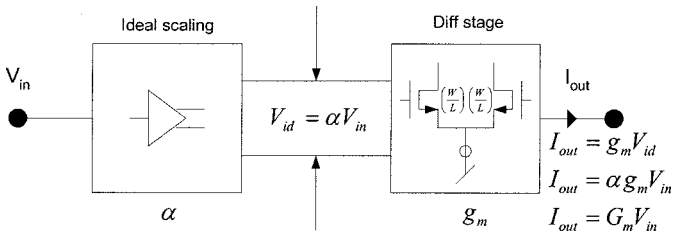

Fig. 6. Linearization model.

where γ_0 is the body effect parameter (typically 0.7 $V^{1/2}$), φ_{FB} is the bulk Fermi potential (typically 0.35 V), and g_{m_M1} is the gate transconductance.

It is worth mentioning here that the bulk-driven transistors need to be isolated in separate wells. Another drawback is the finite input impedance of the OTA.

E. Approximate Expressions for the Signal-to-Noise Ratio

The different designs presented in the previous sections can all be modeled as shown in Fig. 6. Essentially, all four designs have a certain transconductance reduction factor $\alpha(<1)$ and a differential transconductance stage G_m such that the overall $G_m = \alpha G_m$ is the same for all four designs. We now obtain approximate analytical expressions for the input signal that can be applied for a given harmonic distortion component HD_3, the input referred thermal and flicker noise voltages, and finally, the SNR for the model of Fig. 6. From these expressions, we strive to obtain an insight into the various design tradeoffs that exist. Assuming that the attenuator α is linear, the HD_3 as a function of the peak input signal αV_{in} is given by

$$\mathrm{HD}_3 = \frac{1}{32}\left(\frac{\alpha V_{\mathrm{in}}}{V_{\mathrm{DSAT}}}\right)^2 \tag{5}$$

where αV_{in} and V_{DSAT} are the peak value of the incoming signal and the saturation voltage, respectively [13]. After some algebraic manipulations, (5) expresses the rms input signal $V_{\mathrm{in\_rms}}$ as

$$V_{\mathrm{in\_rms}} = \frac{G_m L}{\sqrt{2}K_{\mathrm{P}} W\alpha}\sqrt{32\mathrm{HD}_3} \tag{6}$$

where K_{P} is the transconductance parameter, and W is the width and L is the length of the transistors of the differential stage. The linear range V_{in} can be increased by decreasing α, W, or by increasing G_m, L. If the noise is dominated by the OTA differential stage, the input-referred rms thermal noise (7) integrated from frequency f_1 to f_2 is given by

$$V_{\mathrm{nt\_rms}} = \frac{1}{\alpha}\sqrt{\frac{8kT}{3G_m}}\sqrt{f_2-f_1}. \tag{7}$$

Thermal noise can be reduced by increasing α. The input-referred rms flicker noise integrated from frequency f_1 to f_2 is given by

$$V_{\mathrm{nf\_rms}} = \frac{1}{\alpha}\sqrt{\frac{K_{\mathrm{F}}}{WLC_{\mathrm{ox}}^2}}\sqrt{\ln\left(\frac{f_2}{f_1}\right)} \tag{8}$$

where C_{ox} is the oxide capacitance per unit channel area and K_{F} is the flicker-noise coefficient. Flicker noise can be reduced

TABLE I
NUMERICAL VALUES FOR KEY PARAMETERS

DESIGN	α	G_m (nA/V)	V_{in} (mVpp)	V_{nt_rms} (μV)	V_{nf_rms} (μV)	SNR (dB)
Reference	1.00	10.0	168	2.4	14.1	72.1
SD+CD	0.60	16.6	247	3.0	21.2	71.7
FG+CD	0.45	22.2	344	4.1	31.5	71.8
BD+CD	0.23	43.5	947	4.9	82.9	72.1

TABLE II
SUMMARY OF SIMULATION RESULTS

PARAMETER \ DESIGN	REFERENCE	SD+CD	FG+CD	BD+CD
G_{M0} (nS)	11.6	11.55	11.51	11.24
Linearity@1% HD_3, 1 Hz (mVpp)	162	240	330	900
Input referred noise (μVrms)	12.04	17.29	26.03	70.3
Signal to Noise ratio (SNR) (dB)	73.5	73.9	73.0	73.1
Max Common Mode input (V)	0.18	0.2	1.5	1.6
Bias Iss (nA)	2	100	200	500
Supply $V_{DD} = \|V_{SS}\|$ (V)	1.35	1.35	1.35	1.35
Power (μW)	0.0162	1.35	1.62	4.05
Total area (mm^2)	1.44	0.21	4.65	0.22
Inversion level of driver transistors i_f	10	15	17	31
Normalized power	1	83.3	100	250
Normalized area	6.71	1.00	21.61	1.01

by increasing the gate area or increasing α. From (6)–(8), we observe a direct tradeoff between linearity and noise with respect to α. The smaller the α, the higher the linearity, and at the same time, the higher the noise. The total rms input noise voltage $V_{\mathrm{n\_rms}}$ is given by $\sqrt{V_{\mathrm{nt\_rms}}^2 + V_{\mathrm{nf\_rms}}^2}$, therefore the SNR becomes

$$\begin{aligned}\mathrm{SNR} &= 20\log\left(\frac{V_{\mathrm{in\_rms}}}{V_{\mathrm{n\_rms}}}\right)\\ &= 20\log\left(\frac{\dfrac{G_m L}{\alpha K_{\mathrm{P}} W}\sqrt{32\mathrm{HD}_3}}{\sqrt{2}\sqrt{\dfrac{8kT\alpha}{3G_m}(f_2-f_1)+\dfrac{K_{\mathrm{F}}}{WLC_{\mathrm{ox}}^2}\ln\left(\dfrac{f_2}{f_1}\right)}}\right).\end{aligned} \tag{9}$$

An approximate estimate for the SNR considering only flicker noise is given in (10).

$$\begin{aligned}\mathrm{SNR} &= 20\log\left(\frac{V_{\mathrm{in\_rms}}}{V_{\mathrm{n\_rms}}}\right)\\ &= 20\log\left[\left(\frac{\dfrac{G_m}{K_{\mathrm{P}}}\sqrt{32\mathrm{HD}_3}}{\sqrt{\dfrac{2K_{\mathrm{F}}}{C_{\mathrm{ox}}^2}\ln\left(\dfrac{f_2}{f_1}\right)}}\right)\left(\frac{L^{1.5}W^{0.5}}{\alpha}\right)\right].\end{aligned} \tag{10}$$

Notice that this equation is valid if α is a noiseless attenuator, otherwise, its noise must be added. From (10), we observe that the SNR is a function of the device dimensions, G_m, α, and HD_3. Now, if we fix G_m, HD_3, and $(L^{1.5}W^{0.5}/\alpha)$ for all topologies we can obtain the same SNR. Table I summarizes the approximate numerical values for the different parameters ($\alpha, G_m, V_{\mathrm{in\_pp}}$ at $1\%\mathrm{HD}_3, V_{\mathrm{nt\_rms}}, V_{\mathrm{nf\_rms}}$, and SNR) calculated using the above equations for the different OTA topologies.

The peak input V_{in} has been computed for $1\%\mathrm{HD}_3$ and the noise has been integrated between $f_1 = 10$ mHz and $f_2 = 10$ Hz. From Table I, it is clear that flicker noise is the dominant component of the total noise. The transconductance reduction factor α, G_m, and the device sizes are different for each design but are related in such a way as to yield the same SNR for the different designs.

III. SIMULATION RESULTS

All the above circuits were designed and simulated using the AMI 1.2-μm n-well CMOS technology with BSIM3 models available through MOSIS. The results are summarized in Table II. The current division factor M was set at 49. As we can see in Table II, we gain a lot in terms of linearity as we move from design reference to BD+CD (bulk), but pay in terms of power consumption and total noise. The area of designs SD+CD and BD+CD are more or less the same but less than the reference. It is interesting to note here that the floating-gate design (design FG+CD) consumes a huge amount of area because of the large input capacitors. In our design, the input capacitors were about ten times the parasitic

TABLE III
EXPERIMENTAL RESULTS FOR THE DIFFERENT OTA DESIGNS

PARAMETER	REFERENCE	SD+CD	FG+CD	BD+CD
G_M (nA/V)	9.4	9.3	9.2	9.4
HD_3 (%)	0.9@162mVpp	1.0@242mVpp	1.1@330mVpp	0.9@900mVpp
Input noise (μVrms)	18.1	26.1	39.1	104.7
SNR @1% HD_3 (dB)	69.9	70.3	69.5	69.6
I_{BIAS} (nA)	2.6	120	232	560

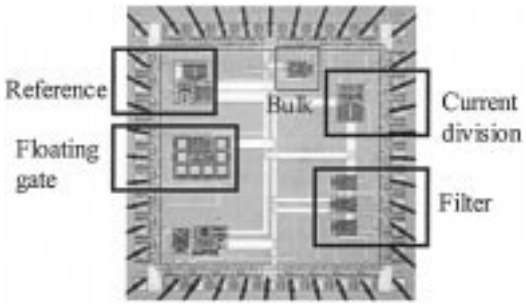

Fig. 7. Chip microphotograph.

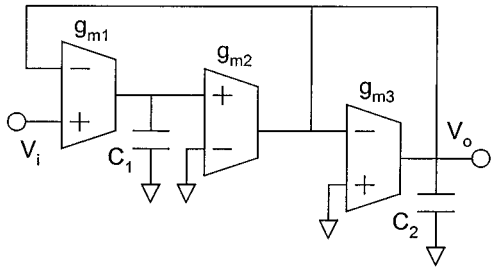

Fig. 8. Low-pass filter.

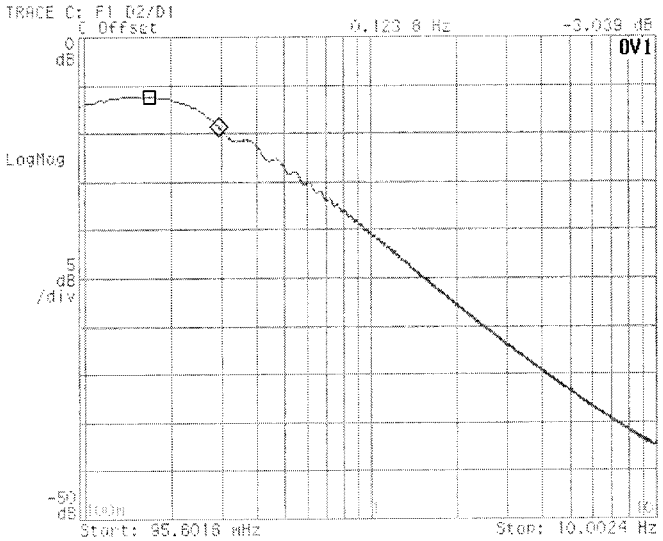

Fig. 9. Low-pass filter magnitude response.

capacitance. From the standpoint of very small power levels in the range of nanowatts, the reference topology becomes preferable. However, performance is poor in terms of linearity and silicon area. On the other hand, if power levels of the order of microwatts are tolerable, then designs SD+CD, FG+CD, and BD+CD are all better than design reference in terms of the above-mentioned performance parameters. Among these designs, while design SD+CD has the least area of the three, design BD+CD is very good in terms of linearity but worst in terms of noise.

IV. EXPERIMENTAL MEASUREMENTS

The above-described OTAs have been fabricated in a 1.2-μm CMOS process available through MOSIS. The chip microphotograph is shown in Fig. 7. The total die area is 1.9 mm $\times$ 1.9 mm. The test die consists of the four different transconductance amplifiers, a second-order low-pass filter, and some other sample circuits.

A. Operational Transconductance Amplifiers

Measurement results for the different OTAs are tabulated in Table III. We observe reasonably good agreement between theoretical results with those measured. The SNR is about the same for each design, much like the predictions based on the simulation results, though the measured noise is higher than the simulated values. Moreover, due to process variations, the bias currents had to be adjusted. The reference design is particularly affected by these variations because of the extremely small nominal bias current. The supply voltages used for all topologies were $+/-1.35$ V.

Second-Order Low-Pass Filter: The chip also contains a second-order low-pass filter built using the bulk-driven OTA so as to test it in a sample application. The topology of the filter [4] is shown in Fig. 8.

The low-pass filter was tested for functionality and the measured magnitude response is shown in Fig. 9. The output spectrum for a 150-$\mathrm{mV_{pp}}$ input at 0.1 Hz is shown in Fig. 10. The transconductance $g_{m1}(g_{m1} = g_{m2} \sim 0.7g_{m3})$ was set at 10 nS and the capacitors ($C_1 = C_2$) were external to the chip (10 nF). The measured -3-dB cutoff frequency was around 0.17 Hz which is close to the theoretical value of 0.16 Hz. The rolloff of the filter is about -25 dB/dec instead of the -40 dB/dec. This may be attributed to board parasitics, transistor output impedance, and finite input impedance of the bulk-driven OTAs. The measured $\mathrm{HD_3}$ is about -45 dB (SPICE result is about -48 dB) for $v_{\mathrm{in}} = 150\,\mathrm{mV_{pp}}$. Measured results for the filter are summarized in Table IV.

We would like to mention here that the power dissipation of 8.2 μW is including the bias network which is approximately the same as that of the filter itself.

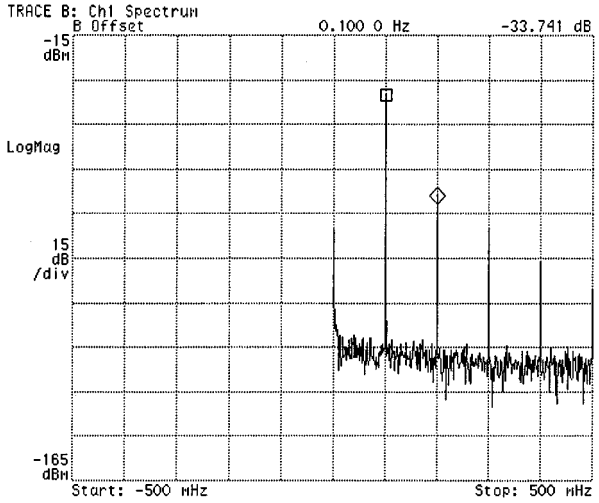

Fig. 10. Low-pass filter output spectrum.

TABLE IV
EXPERIMENTAL RESULTS FOR THE FILTER

PARAMETER	VALUE
Filter order	2
-3dB Bandwidth (Hz)	0.3
HD_3 @ V_{in}=150mVpp (dB)	-45
Total input noise (μVrms)	15.6
SNR (dB)@ HD_3 = -45dB	70.5
Power consumption (μW)	8.18
Power supply (V)	$\pm$1.35
Total filter area (mm^2)	0.06

V. CONCLUSION

This paper has presented different design techniques for obtaining very small transconductances, such as current division, source degeneration, floating-gate techniques, and bulk-driven techniques. In particular, the natural attenuating properties of the floating-gate and bulk-driven transistors have been advantageously utilized for realizing small transconductance values. Moreover, for obtaining a given transconductance value, the various tradeoffs involving key circuit parameters such as linearity, noise, and power consumption have been discussed and a detailed comparison has been made among the various designs. The designed OTAs have been fabricated in a 1.2-μm CMOS process and simulated and measured results are in good agreement. By choosing an appropriate level of inversion for the transistors based on (1), (3), and (4), it is possible to obtain an optimum balance between contradicting design considerations such as power consumption, silicon area, and noise.

REFERENCES

[1] L. C. Stotts, "Introduction to implantable biomedical IC design," *IEEE Circuits Devices Mag.*, pp. 12–18, Jan. 1989.
[2] M. R. Dewitt, G. F. Gross, and R. Ramachandran, "Built-in-self-test for analog to digital converters," U.S. Patent 5 132 685, Aug. 9, 1991.
[3] P. Kinget and M. Steyaert, "Full analog CMOS integration of very large time constants for synaptic transfer in neural networks," *Analog Integr. Circuits Signal Process.*, vol. 2, pp. 281–295, 1992.
[4] R. L. Geiger and E. Sánchez-Sinencio, "Active filter design using operational transconductance amplifiers—a tutorial," *IEEE Circuits Devices Mag.*, no. 1, pp. 20–32, 1985.
[5] W. H. G. Deguelle, "Limitations on the integration of analog filters below 10 Hz," in *Proc. IEEE ESSCIRC'88*, 1988, pp. 131–134.
[6] A. I. A. Cunha, O. C. Gouveia-Filho, M. C. Schneider, and C. Galup-Montoro, "A current-based model for the MOS transistor," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'97)*, vol. 3, 1997, pp. 1608–1611.
[7] J. Silva-Martínez and S. Solís-Bustos, "Design considerations for high-performance very-low-frequency filters," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'99)*, vol. 2, 1999, pp. 648–651.
[8] P. Garde, "Transconductance cancellation for operational amplifiers," *IEEE J. Solid-State Circuits*, vol. SC-12, pp. 310–311, June 1977.
[9] C. G. Yu and R. L. Geiger, "Very low voltage operational amplifier using floating-gate MOSFETs," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'93)*, vol. 2, 1993, pp. 1152–1155.
[10] L. Yin, S. H. K. Embabi, and E. Sánchez-Sinencio, "A floating-gate MOSFET D/A converter," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'97)*, vol. 1, 1997, pp. 409–412.
[11] R. Fried and C. C. Enz, "Bulk-driven MOS transconductor with extended linear range," *Electron. Lett.*, vol. 32, pp. 638–640, 1996.
[12] A. Guzinski, M. Bialko, and J. C. Matheau, "Body-driven differential amplifier for application in continuous-time active-C filter," in *Proc. IEEE Eur. Conf. Circuit Theory and Design (ECCTD'87)*, 1987, pp. 315–319.
[13] E. Sánchez-Sinencio and J. Silva-Martínez, "CMOS transconductance amplifiers, architectures and active filters—A tutorial," *Proc. IEE Circuits Devices Syst.*, vol. 147, no. 1, pp. 3–12, Feb. 2000.