Analog Integrated Circuits and Signal Processing, 13, 123–151 (1997)

A Low-Power Wide-Linear-Range Transconductance Amplifier

RAHUL SARPESHKAR,[1] RICHARD F. LYON[2], AND CARVER MEAD[3]

[1]*rahul@pcmp.caltech.edu,* [2]*lyon@apple.com,* [3]*candace@pcmp.caltech.edu*

[1]*Physics of Computation Laboratory, California Institute of Technology, Pasadena, CA, 91125,* [2]*Apple Computer, One Infinite Loop, Cupertino, CA 95014,*
[3]*Physics of Computation Laboratory, California Institute of Technology, Pasadena, CA, 91125*

Received April 29, 1996; Accepted September 16, 1996

Abstract. The linear range of approximately ± 75 mV of traditional subthreshold transconductance amplifiers is too small for certain applications—for example, for filters in electronic cochleas, where it is desirable to handle loud sounds without distortion and to have a large dynamic range. We describe a transconductance amplifier designed for low-power ($< 1\mu$W) subthreshold operation with a wide input linear range. We obtain wide linear range by widening the tanh, or decreasing the ratio of transconductance to bias current, by a combination of four techniques. First, the well terminals of the input differential-pair transistors are used as the amplifier inputs. Then, feedback techniques known as source degeneration (a common technique) and gate degeneration (a new technique) provide further improvements. Finally, a novel bump-linearization technique extends the linear range even further. We present signal-flow diagrams for speedy analysis of such circuit techniques. Our transconductance reduction is achieved in a compact 13-transistor circuit without degrading other characteristics such as dc-input operating range. In a standard 2μm process, we were able to obtain a linear range of ± 1.7V. Using our wide-linear-range amplifier and a capacitor, we construct a follower–integrator with an experimental dynamic range of 65 dB. We show that, if the amplifier's noise is predominantly thermal, then an increase in its linear range increases the follower–integrator's dynamic range. If the amplifier's noise is predominantly $1/f$, then an increase in its linear range has no effect on the follower–integrator's dynamic range. To preserve follower–integrator bandwidth, power consumption increases proportionately with an increase in the amplifier's linear range. We also present data for changes in the subthreshold exponential parameter with current level and with gate-to-bulk voltage that should be of interest to all low-power designers. We have described the use of our amplifier in a silicon cochlea [1, 2].

Key Words: transconductance, amplifier, wide-linear-range, degeneration, noise, dynamic range, cochlea, low-power

1. Introduction

In the past few years, engineers have improved the linearity of MOS transconductor circuits [3]–[12]. These advances have been primarily in the area of above-threshold, high-power, high-frequency, continuous-time filters. Although it is possible to implement auditory filters (20Hz–20khz) with these techniques, it is inefficient to do so. The transconductance and current levels in above-threshold operation are so high that large capacitances or transistors with very low W/L are required to create low-frequency poles, and area and power are wasted. In addition, it is difficult to span 3 orders of magnitude of transconductance with a square law, unless we use transistors with ungainly aspect ratios. However, it is easy to obtain a wide linear range above threshold.

In above-threshold operation, identities such as $(x-a)^2-(x-b)^2 = (b-a)(2x-a-b)$ are used to increase the wide linear range even further. In bipolar devices where the nonlinearity is exponential, rather than second-order, it is much more difficult to completely eliminate the nonlinearity. The standard solution has been to use the feedback technique of emitter degeneration, which achieves wide linear range by reducing transconductance, and is described by Gray [13]. A clever scheme for widening the linear range of a bipolar transconductor that cancels all nonlinearities up to fifth order, without reducing the transconductance, has been proposed by Wilson [14]. A method for getting perfect linearity in a bipolar transconductor by using a translinear circuit and a resistor has been demonstrated by Chung [15]. Both of the latter methods, however, require the use of resistors, and ultimately derive their linearity from the presence of a linear element in the

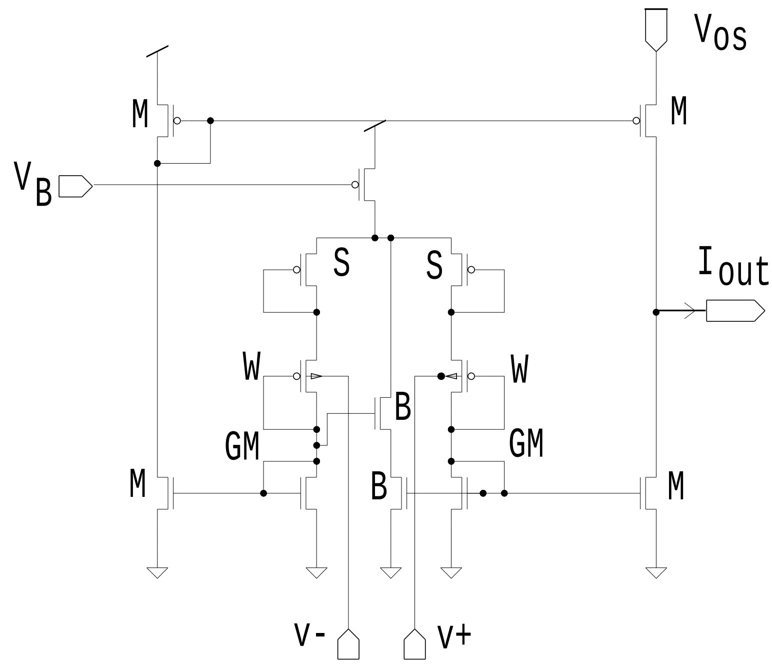

Fig. 1. The wide linear range transconductance amplifier. The figure shows the transconductance amplifier with v_+ and v_- well–input voltages and the I_{out} current output. The voltage V_B sets the bias current of the amplifier and the voltage V_{os} allows a fine adjustment of its offset if necessary. The S transistors reduce the transconductance of the differential-pair W transistors through source degeneration. The GM transistors reduce the transconductance of the W transistors through gate degeneration. They also serve to mirror the differential-pair currents to the output of the amplifier. The B transistors implement bump linearization. The M transistors form parts of the current mirrors in the circuit.

circuit. Resistors, however, cannot be tuned electronically, and require special process steps.

Various authors have used an MOS device as the resistive element in an emitter-degeneration scheme to make a BiCMOS transconductor—for example, the scheme proposed by Sanchez [16]. Another BiCMOS technique, reported by Weimin [17], uses an above-threshold differential pair to get wide linearity, and scales down the output currents via a bipolar Gilbert gain cell, to levels more appropriate for auditory frequencies. Above-threshold differential pairs, however, result in lower open-loop gain and higher voltage offset, and techniques such as cascode mirrors are required to improve these characteristics. Cascode mirrors, however, degrade dc output-voltage operating range and consume area. In addition, above-threshold operation results in higher power dissipation.

Subthreshold MOS technology, like bipolar technology, is based on exponential nonlinearities. Thus, it is natural to employ source-degeneration techniques. Methods for getting wider linear range that exploit the Early effect in conjunction with a source-degeneration method are described by Arreguit [19]. The Early voltage is, however, a parameter with high variance across transistors; thus, we cannot expect to get good transconductance matching in this method. Further, such schemes are highly offset prone, because any current mismatch manifests itself as a large voltage mismatch due to the exceptionally low transconductance.

The simple technique of using a diode as a source-degeneration element extends the linear range of a differential pair to about ± 150 mV, as described by Watts [18]. However, it is difficult to increase this linear range further by using two stacked diodes in series as the degeneration element—the wider linear range that is achieved is obtained at the expense of a large loss in dc-input operating range. If we operate within a 0 to 5V supply, the signal levels remain constrained to take on small values, because of the inadequate dc input operating range.

Three techniques for improving the linear range of subthreshold differential pairs have been described in [20]. The authors define the linear range to be the point where the transconductance drops by 1%. By that definition, the linear range of a conventional transconductance amplifier described by $I_{out} = I_B \tanh(x/V_L)$

is $V_L/5$. Their best technique achieved a value of $V_L = 584$ mV, and involved expensive common-mode biasing circuitry. In contrast, our technique yields a V_L of 1.7 V, and involves no additional biasing circuitry.

In [21], a 21-transistor subthreshold transconductance amplifier is described. From visual inspection of their data, the amplifier has a V_L of about 700 mV. They estimate from simple theoretical calculations that the effective number of shot-noise sources in their circuit is about 20. In contrast, our 13-transistor circuit has a V_L of 1.7 V, and the effective number of shot-noise sources in our circuit is around 5.3 (theory) or 7.5 (experiment).

We can solve the problem of getting wider linear range by interposing a capacitive divider between each input from the outside world and each input to the amplifier. Some form of slow adaptation is necessary to ensure that the dc value of each floating input to the amplifier is constrained. This approach, as used in an electronic cochlea, is described in [22]; it did not work well in practice because of its sensitivity to circuit parasitics. As we shall see later, capacitive-divider schemes bear some similarity to our scheme. We shall discuss capacitive-divider techniques in Section 5.4.

To get low transconductance, we begin by picking an input terminal that is gifted with low transconductance from birth: the well. We reduce the transconductance further by using source degeneration, and a new negative-feedback technique, which we call gate degeneration. Finally, we use a novel technique, which we call bump linearization, to extend the linear range even more; bump circuits have been described in [23]. The amplifier circuit that incorporates all four techniques is shown in Figure 1.

In Section 2, we present all the essential ideas and first-order effects that describe the operation of the amplifier. We describe second-order effects, such as the common-mode and gain characteristics, in Section 3. We discuss the operation of this amplifier as a follower–integrator filter in Section 4. We elaborate on noise and dynamic range in Section 5. In Section 6, we conclude by summarizing our contributions. Appendix A contains a quantitative treatment of common-mode effects on the amplifier's transconductance. Section A.1 describes the effects of changing transconductance; Section A.2 is on the effects of parasitic bipolar transistors present in our well-input amplifier. Normally, the amplifier operates in the 1V to 5V range, where these bipolar transistors are inactive.

2. First-Order Effects

We begin by expressing basic transistor relationships in a form that will be useful in our paper. We use standard IEEE convention for large-signal (i_{DS}), dc (I_{DS}), and small-signal (i_{ds}) variables.

2.1. Basic Transistor Relationships

The current in a subthreshold MOS well transistor in saturation is given by

$$i_{DS} = I_0 \exp\left(-\frac{\kappa v_{GS}}{U_T}\right) \exp\left(-\frac{(1-\kappa)\, v_{WS}}{U_T}\right), \quad (1)$$

where v_{GS} and v_{WS} are the gate-to-source and well-to-source voltage, respectively; κ is the subthreshold exponential coefficient; I_0 is the subthreshold current-scaling parameter; $U_T = kT/q$ is the thermal voltage; and $v_{DS} \gg 5U_T$.

Eq. (1) illustrates that the gate affects the current through a κ exponential term, whereas the well affects the current through a $1-\kappa$ exponential term. Thus, when the gate is effective in modulating the current, the well is ineffective, and vice versa. By differentiating Eq. (1), we can easily show that the gate, well, and source transconductances are

$$\begin{aligned} g_{gt} &= \frac{\partial i_{DS}}{\partial v_G} = \frac{i_{ds}}{v_g} = -\kappa \frac{I_{DS}}{U_T}, \\ g_{wl} &= \frac{\partial i_{DS}}{\partial v_W} = \frac{i_{ds}}{v_w} = -(1-\kappa)\frac{I_{DS}}{U_T}, \\ g_s &= \frac{\partial i_{DS}}{\partial v_S} = \frac{i_{ds}}{v_s} = \frac{I_{DS}}{U_T}, \end{aligned} \quad (2)$$

respectively. Thus if and only if $\kappa > 0.5$—which is almost always the case—then the well transconductance has a lower magnitude than the gate transconductance, and the well is preferable over the gate as a low-transconductance input.

It is convenient to work with dimensionless, small-signal variables: If i_d and v_d are arbitrary small-signal variables, and we define the dimensionless variables $i = i_d/I_D$, $v = v_d/U_T$, then a relation such as $i_d = g_d v_d = \kappa I_D v_d/U_T$ takes the simple form $i = \kappa v$. We notice then that κ plays the role of a dimensionless transconductance; that is, $\kappa = g_d/(I_D/U_T)$ is the dimensionless transconductance that we obtain by dividing the real transconductance g_d by I_D/U_T. We shall use the dimensionless variable forms to do most of our calculations, and then shall convert them back to the real forms. For convenience, we denote the dimensionless variable by the same name as that of the variable from which it is derived. Thus, Eq. (2) when converted to its dimensionless form, simply reads $g_{gt} = -\kappa$, $g_{wl} = -(1-\kappa)$, and $g_s = 1$.

Figure 2 shows a well transistor, its small-signal-equivalent circuit, and a signal-flow diagram that represents its small-signal relations. In this paper, we shall ignore the capacitances that would be represented in a complete small-signal model of the transistor.

2.2. *Transconductance Reduction Through Degeneration*

The technique of source degeneration is well known, and was first used in vacuum-tube design; there it was referred to as *cathode degeneration*, and was described by Landee [24]. Later, it was used in bipolar design, where it is referred to as *emitter degeneration* [13]. The idea behind source degeneration is to convert the current flowing through a transistor into a voltage through a resistor or diode, and then to feed this voltage back to the emitter or source of the transistor to decrease its current.

Gate degeneration has never been reported in the literature to our knowledge. This lacuna probably occurs because most designs use the gate as an input and thus never have it free to degenerate. The vacuum-tube literature, however, shows familiarity with a similar concept, called *screen degeneration*, as described in Landee [24]. The idea behind gate degeneration is to convert the current flowing through a transistor into a voltage through a diode, and then to feed this voltage back to the gate of the transistor to decrease its current.

Figure 3a shows a half-circuit for one differential arm of the amplifier of Figure 1, *if we neglect the B transistors for the time being*. The source-degeneration diode is the *p*FET connected to the source of the well-input transistor; the gate-degeneration diode is the *n*FET connected to the drain of the well-input transistor. The gate-degeneration diode is essentially free in our circuit, because it is part of the current mirror that feeds the differential-arm currents to the output. The voltage V_C represents the common-node voltage of the differential arms. In differential-mode ac analysis, the common-node voltage is grounded, as explained in the following paragraph.

In Figure 1, if $v_+ = v_-$, and the amplifier is perfectly matched, then a quiescent current of $\frac{I_B}{2}$ flows through each branch of the amplifier and i_{OUT} will be 0. If we now vary the differential voltage, $v_d = v_+ - v_-$, by a small amount, the current changes by $i_{out} = g v_d$, where g is the transconductance of the amplifier. We would like to compute g. If we apply v_d, such that v_+ changes by $+\frac{v_d}{2}$ and v_- changes by $-\frac{v_d}{2}$, then the common node of the two differential halves (the source of the S transistors) does not change in voltage. For the purposes of small-signal analysis, we can treat the common node as a virtual ground. Thus, if g_h is the transconductance of the half-circuit shown in Figure 3a, the output current is $g_h \frac{v_d}{2} - (-g_h \frac{v_d}{2}) = g_h v_d$. Hence, the transconductance of the half-circuit, biased to the current level of $I_B/2$, is the transconductance of the amplifier.

The circuit of Figure 3a yields the small-signal circuit of Figure 3b: The source-degeneration diode is represented by a dimensionless resistor of value $1/\kappa_p$, the gate-degeneration diode is represented by a dimensionless resistor of value $1/\kappa_n$, the gate-controlled current source of Figure 2 is represented by a dimensionless resistor of value $1/\kappa$ (the gate is tied to the drain), and the well-controlled current source of Figure 2 is represented by a dependent source, as shown.

The left half of Figure 3c represents the signal-flow diagram for the well-input transistor, as derived in Figure 2. The right half of Figure 3c represents the blocks due to the source or gate degeneration diodes feeding back to the source or gate. Thus, we have two negative-feedback loops acting in parallel to reduce the transconductance. One loop feeds back the output current to the source via a $-1/\kappa_p$ block; the other loop feeds back the output current to the gate via a $1/\kappa_n$ block. Since the magnitude of the loop gains of the source-degeneration and gate-degeneration loops are $A_s = \frac{1}{\kappa_p}$ and $A_g = \frac{\kappa}{\kappa_n}$ respectively, the well transconductance is attenuated by $\frac{1}{1+A_s+A_g}$; that is to say, the transconductance is

$$g = \frac{1-\kappa}{1 + \frac{1}{\kappa_p} + \frac{\kappa}{\kappa_n}}. \tag{3}$$

We multiply the dimensionless transconductance thus computed by $\frac{I_B}{2U_T}$ to get the actual transconductance, since I_{DS} in each differential arm is $\frac{I_B}{2}$.

The W, S, and G transistors of each differential arm may be regarded as a single transistor with I–V char-

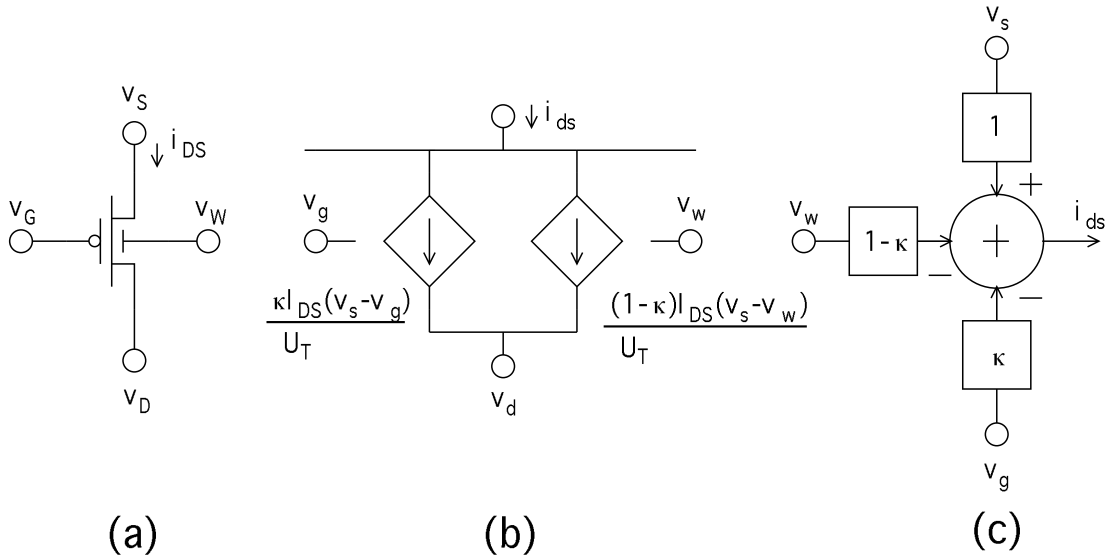

Fig. 2. Signal-flow diagram for a saturated well transistor. (a) A well transistor with marked voltages and currents. (b) The small-signal equivalent circuit for the well transistor. (c) The signal-flow diagram represents the dimensionless relationships between the small-signal variables of the transistor. In our paper it shall prove to be the most useful and insightful way of analyzing transconductance relationships, rather than the more conventional circuit representation shown alongside it. In the signal-flow diagram, it is understood that all currents are normalized by I_{DS}, all voltages are normalized by U_T, and all transconductances are normalized by I_{DS}/U_T. Thus, the small-signal relationship of the transistor takes the simple form $i_{ds} = v_s - \kappa v_g - (1-\kappa)v_w$.

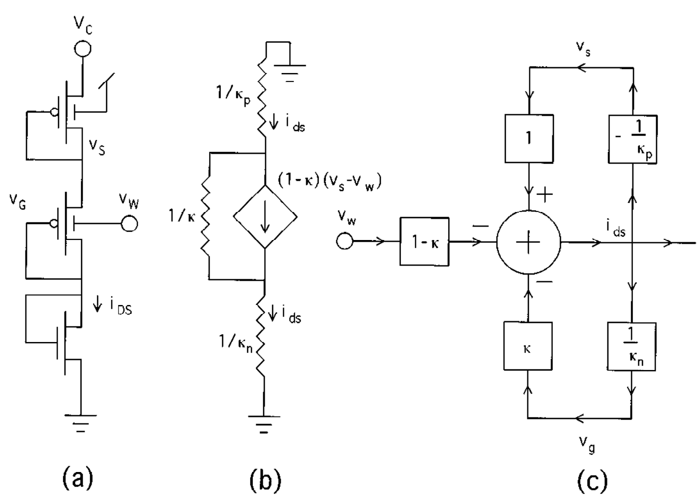

Fig. 3. Transconductance reduction through degeneration. (a) A half circuit for one differential arm of our amplifier, if we ignore the B transistors of Figure 1 for now. (b) A small-signal equivalent circuit for this configuration. (c) A signal-flow diagram for this configuration. The signal-flow diagram shows that the transconductance reduction is achieved through two parallel feedback loops. The top loop is due to source degeneration and the bottom loop is due to gate degeneration. A more detailed description of the feedback concepts assosciated with transconductance reduction may be found in Section 2.2.

acteristics given by

$$I \propto e^{-(v_S - g v_W)/U_T}. \tag{4}$$

By following the steps outlined by Mead [25], we can easily derive that

$$i_{OUT} = I_B \tanh\left(\frac{g(v_+ - v_-)}{2U_T}\right), \tag{5}$$

where g is given by Eq. (3).

2.3. Bump Linearization

Bump linearization is a technique for linearizing a tanh and extending the linear range of a subthreshold differential pair [26]. We shall first explain how it works for the simple differential pair; then, we shall extend this explanation to our amplifier in a straightforward fashion.

A bump differential pair has, in addition to its two outer arms, a central arm containing two series-connected transistors [23]. The current through the transistors in the central arm is a bump-shaped function of the differential voltage, so we call these transistors *bump transistors*. Thus, the differential output current from the outer two arms, I, is the usual tanh-like function of the differential voltage, V, except for a region near the origin, where the bump transistors steal current. By ratioing the W/L of the bump transistors to be w with respect to the transistors in the outer arms, we can control the properties of this I–V curve. A small w will essentially leave it unchanged. A large w will cause a flat zone near the origin, where there is current stealing. The larger the w, the larger the width and flatness of the zone. At intermediate values of w, the expansive properties of the curve due to the bump compete with the compressive properties of the curve due to the tanh, and a curve that is more linear than is a tanh is obtained. At $w = 2$, the curve is maximally linear.

For a simple subthreshold bump amplifier with bump scaling w, the differential output current can be shown to be

$$i_{out} = \frac{\sinh x}{\beta + \cosh x}, \tag{6}$$

where

$$\beta = 1 + \frac{w}{2}, \tag{7}$$

$$x = \frac{q\kappa(v_+ - v_-)}{kT}. \tag{8}$$

The bias current is assumed to be 1, without loss of generality. Simple calculus shows that the first and second derivatives of Eq. (6) are

$$\frac{di_{out}}{dx} = \frac{1 + \beta\cosh x}{(\beta + \cosh x)^2}, \tag{9}$$

$$\frac{d^2 i_{out}}{dx^2} = \frac{\sinh x\left(\beta^2 - \beta\cosh x - 2\right)}{(\beta + \cosh x)^3}. \tag{10}$$

If we require that the i_{out}-vs.-x curve have no points of inflection in it except at the origin, then it must always be convex or concave in the first or third quadrant; this requirement is necessary to ensure that there are no strange kinks in the I–V curve. In Eq. (10), we must then have

$$\left(\beta^2 - \beta\cosh x - 2\right) \le 0 \tag{11}$$

for all x. Note that in deriving Eq. (11), we have used the facts that the denominator term of Eq. (10) is always positive, and that $\sinh x$ changes sign at only the origin. The worst possible case for not meeting the constraint given by Eq. (11) occurs when $\cosh x$ is at its minimum value of 1 at $x = 0$. If we set $\cosh x = 1$ and solve the resulting quadratic for β, we obtain

$$\beta \le 2, \tag{12}$$

$$\Rightarrow w \le 2. \tag{13}$$

Thus, at $w = 2$, we are just assured of satisfying this constraint. At $w = 2$, the i_{out}-vs.-x curve is maximally linear: If we Taylor expand Eq. (6) at $\beta = w = 2$, we find that it has no cubic-distortion term. In comparison, the function $\tanh x/2$, which is to what Eq. (6) reduces when $w = 0$ or $\beta = 1$, has cubic distortion.

$$\tanh\frac{x}{2} = \frac{x}{2} - \frac{x^3}{24} + \frac{x^5}{240} - \frac{17x^7}{40320} + \cdots \tag{14}$$

$$\frac{\sinh x}{2 + \cosh x} = \frac{x}{3} - \frac{x^5}{540} + \frac{x^7}{4536} - \frac{x^9}{77760} + \cdots \tag{15}$$

At $x = 1$, the tanh function has cubic distortion of approximately 8%, as compared to the linearized tanh function which has no cubic distortion whatsoever, but has only fifth-harmonic distortion of less than 1%. Thus, we have linearized the tanh. We shall show later, in Section 5, that bump linearization is a particularly useful technique because it increases the linear range of an amplifier without increasing that amplifier's noise.

Our circuit topology in Figure 1 implements a wide-linear-range bump differential pair by ratioing the W/L of the the nFET B transistors to be w with respect to the W/L of the nFET GM transistors. However, the mathematical analysis of the circuit is identical to that of a simple bump differential pair if we replace the κ of Eq. (8) with the g of Eq. (3).

2.4. Experimental Data

In summary, from Eqs. (3) and (5), we get the transfer function of the amplifier *with the B transistors absent*:

$$I_{out} = I_B \tanh(V_d/V_L), \tag{16}$$
$$V_d = V_+ - V_-, \tag{17}$$
$$V_L = (2kT/q)\left(\frac{1 + 1/\kappa_p + \kappa/\kappa_n}{1-\kappa}\right). \tag{18}$$

From Eqs. (3), (5), (6), (7), and (8), the transfer function of the overall amplifier has the form

$$I_{out} = I_B \frac{\sinh(2x)}{1 + w/2 + \cosh(2x)}, \tag{19}$$

where $x = V_d/V_L$. The W/L of the B transistors is w times the W/L of the GM transistors.

We fabricated our transconductance amplifier in a standard 2μm CMOS n-well process, and obtained data from it. We have also used the amplifier in a working silicon cochlea [1], [2]. Figure 4a shows experimental data for our amplifier for values of $w = 0$, 1.86, and 66, fitted to Eq. (19). Note that, even at rather large w, the nonlinear characteristics are still gentle. This property ensures that even fairly large mismatches in w do not degrade the operation of the amplifier significantly. For $w \approx 2$, Figure 4b shows that the simple tanh fit of Eq. (16) is a good approximation to Eq. (19) if $V_L \to (3/2)V_L$—that is to say, from 1.16 V for $w = 0$ to 1.7 V for $w = 2$. Considering the leading-order terms of Eqs. (14) and (15), the factor of 3/2 seems natural. However, we shall show, from the follower–integrator data of Section 4, that the tanh approximation is inadequate for large differential inputs, because the curve is more linear than is that of a tanh.

We verify Eq. (18) experimentally in Section A.1. Since κ varies with the well-to-gate voltage, and consequently with the common-mode voltage, the verification is involved; we have relegated the details to the appendix.

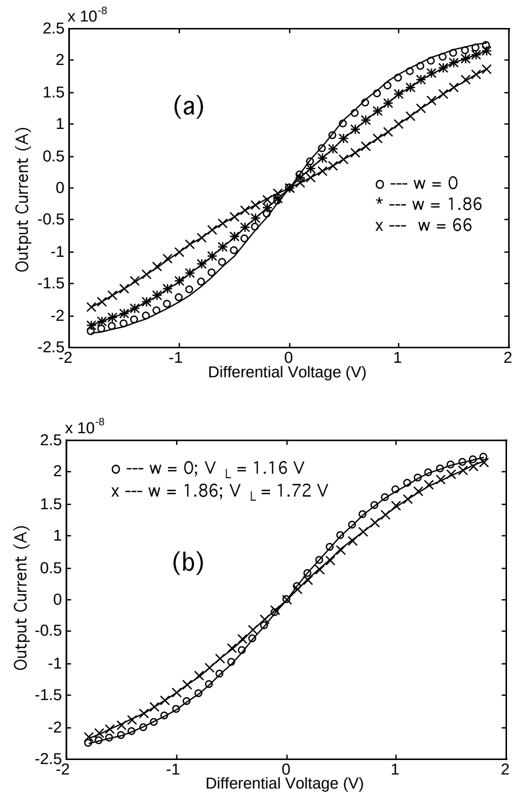

Fig. 4. I–V curves for the amplifier. (a) Experimental data for our amplifier for values of $w = 0$, 1.86, and 66, fitted to Eq. (19). (b) For $w \approx 2$, the simple tanh-fit of Eq. (16) is a good approximation to Eq. (19), if $V_L \to (3/2)V_L$.

3. Second-Order Effects

We shall now discuss several second-order effects in our amplifier. We begin by describing the common-mode characteristics.

3.1. Common-Mode Characteristics

Below an input voltage of 1V, the well-to-source junction becomes forward biased, and the parasitic bipolar transistors, which are part of every well transistor, shunt the current of the amplifer to ground. Thus, as Figure 5a shows, the output current of the amplifier falls. Figure 5b illustrates the same effects, but from the perspective of varying the differential-mode voltage, while fixing the common-mode voltage. In this figure, we can see that changes in κ increase the transconductance of the amplifier as the common-mode voltage is lowered. The increase is exhibited as a rise in the slope at the origin. At very low common-mode voltages, at a particular differential voltage, the input voltage for one arm or the other falls below 1V, and the bipolar transistors in that arm begin to shunt the current to ground. The lower the common-mode voltage, the smaller the differential voltage at which shunting begins. Thus, at low common-mode voltages the current starts to fall at small magnitudes of differential voltage.

We clarify the bipolar effect in Figure 6: Figure 6a shows a vertical crosssection of a well transistor in an n-well process; Figure 6b shows that the equivalent circuit of this well transistor contains two parasitic bipolar transistors. Typically, the well-to-source and well-to-drain junctions are always reverse biased, so that these bipolar transistors are turned off. For our amplifier, and in most cases, the source-to-drain voltage is sufficiently positive that the bipolar transistor at the source is the one that is turned on, if any is, whereas the one at the drain is hardly turned on. Thus, in Figure 6c, we have indicated the bipolar transistor at only the source. The bipolar effect is described in quantitative detail in Section A.2.

Typically, we operate the amplifier at a common-mode voltage of about 3V, where small dc offsets do not significantly affect its transconductance, and where the action of the bipolar transistors is negligible. When the bipolar transistors do turn on, there is no danger of latchup, because the current that is fed to the substrate is at most the tiny subthreshold bias current of the amplifier. Our input operating range of 1V to 5V is about the same as that of a simple subthreshold nFET differential amplifier. These amplifiers also show transconductance changes, due to κ changes, that are most abrupt at low common-mode voltages.

As the common-mode input voltage of our amplifier is decreased, the well-to-gate voltage falls, the depletion region beneath the channel of the transistor shrinks, κ decreases, the transconductance of the amplifier rises from Eq. (3), and, at a given differential voltage, there is more current. Thus, the data of Figure 5a show that the current rises as the common-mode voltage is decreased from 5V to 1V. Note also that the transconductance changes are greatest near 1V where the depletion region is thin, and are least near 5V, where the depletion region is thick. The κ changes are described in quantitative detail in Section A.1.

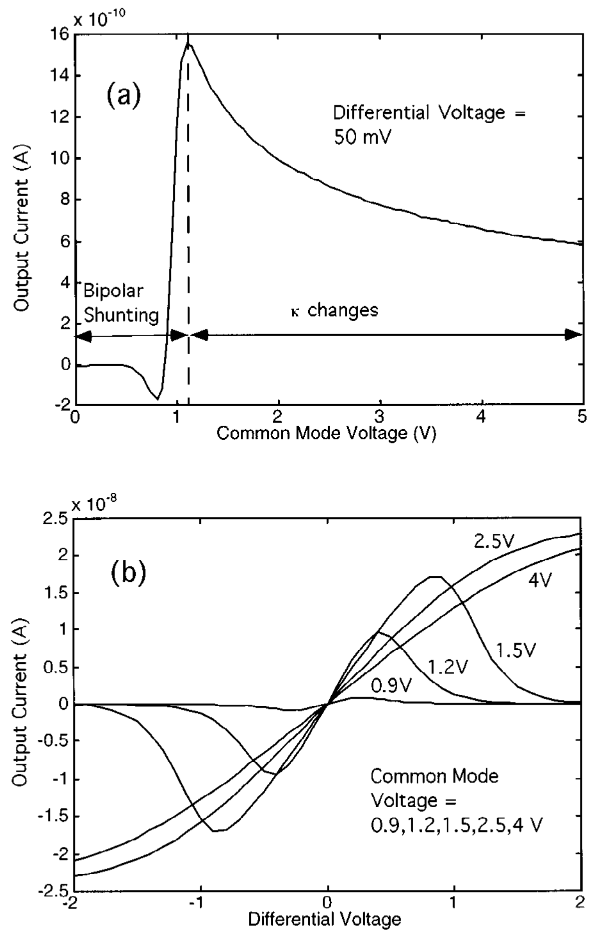

Fig. 5. Differential and common-mode curves. (a) The data shows that the current rises as the common-mode voltage is decreased from 5V to 1V, because changes in κ increase the transconductance of the amplifier. Below 1V, the well-to-source junction becomes forward biased, and the parasitic bipolar transistors, which are part of every well transistor, shunt the current of the amplifer to ground. (b) The same effects as in (a) but from the perspective of varying the differential-mode voltage, while fixing the common-mode voltage. Further details may be found in Section 3.1.

3.2. Bias-Current Characteristics

Figure 7a shows the bias-current characteristics of our amplifier as a function of the bias voltage V_B. The amplifier is capable of operating with little loss in common-mode range from bias currents of pA to μA. Figure 7b illustrates that the linear range begins to increase at above-threshold current levels, because of an increase in the I/g_m ratio of a transistor at these lev-

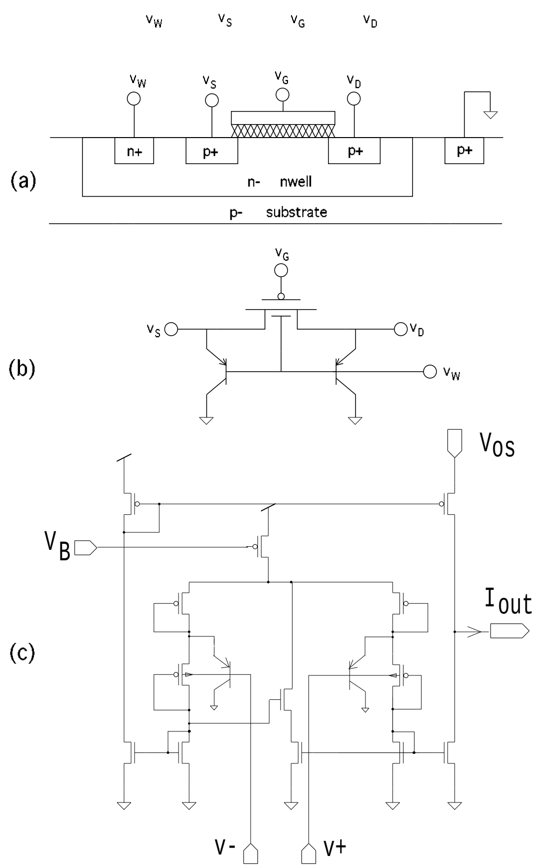

Fig. 6. The effect of the parasitic bipolar transistor. (a) A vertical crosssection of a well transistor in an n-well process. (b) The equivalent circuit of this well transistor contains two parasitic bipolar transistors. (c) The bipolar transistor at the source is responsible for shunting the amplifier's current to ground at low common-mode voltages, while the bipolar transistor at the drain plays a negligible part.

els. That is to say, the characteristic scaling voltage in Eq. (18) begins to increase above $2kT/q$.

Note that the changes in κ observed with bias current and common-mode voltage are not unique to our amplifier: They occur in every subthreshold MOS transistor. By measuring the slope of the I–V curve in a simple nFET differential pair, we obtained changes in κ with common-mode voltage and with bias current. Figure 8 shows these data. Note that κ changes are most abrupt at low common-mode voltages—that is to say, at low gate-to-bulk voltages; κ also decreases with increasing bias current.

3.3. Gain Characteristics

The voltage gain of the amplifier is determined by the ratio of its transconductance g_m to its output conductance g_o. When $w = 0$, the transconductance is I_B/V_L. The output conductance is $I_B/2V_0$, where V_0 is the effective Early voltage of the output M transistors in Figure 1. The effective Early voltage of the output M transistors is the parallel combination $V_0^n V_0^p/(V_0^n + V_0^p)$, where V_0^n and V_0^p are the Early voltages of the n and p output transistors, respectively. When $w = 2$, g_m is given by $I_B/(3/2V_L)$, since the linear range increases by a factor of 3/2. The value of g_o is given by $I_B/3V_0$, as the output current in each arm falls from $I_B/2$ to $I_B/3$. Thus, effectively, the $w = 2$ case corresponds to the $w = 0$ case, with I_B being replaced by $2I_B/3$. So the gain is unchanged, because g_m and g_o change proportionately. Independent of whether $w = 0$ or $w = 2$, the gain is $2V_0/V_L$. Figure 9a shows that the conclusions of the previous paragraph are borne out by experimental data. As the common-mode voltage is lowered, the gain increases because of the increasing transconductance, or decreasing V_L. The $w = 0$ and $w = 2$ amplifiers have almost identical gain. Figure 9b illustrates that, as a function of bias current, the gain initially rises, because the Early voltage increases with current. As the bias current starts to go above threshold, however, V_L drops faster than V_0 increases, and the gain starts to fall. From Figure 7b and Figure 9b, we see that the location of the gain peak is at a bias current (10 nA) where the linear range starts to change significantly, as we would expect. Figure 9b shows, however, that the $w = 0$ and $w = 2$ cases do not have identical gains at all bias currents, although the gains are similar. We attribute these small differences to differences in g_m and g_o at bias-current levels of $I_B/2$ versus $I_B/3$.

For the amplifier for which data were taken, the output transistors had a channel length of 16μm in a 2μm n-well Orbit analog process. If higher gain is desired, these channel lengths should be increased, or the positive and negative output currents can be cascoded via pFET and nFET transistors respectively.

3.4. Offset Characteristics

A current offset of i_k can be compensated for by a voltage offset of $v_l = i_k/g_m$, where g_m is the transconductance of the $i_K - v_L$ relation (i_K and v_L are arbitrary variables). Therefore, if the fractional cur-

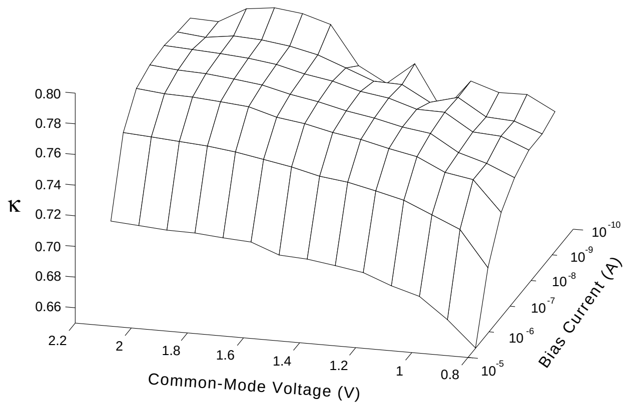

Fig. 8. The changes in κ in a simple *n*FET differential pair. The changes in κ are most abrupt at low common-mode voltages, i.e., at low gate-to-bulk voltages; κ also decreases with increasing bias current.

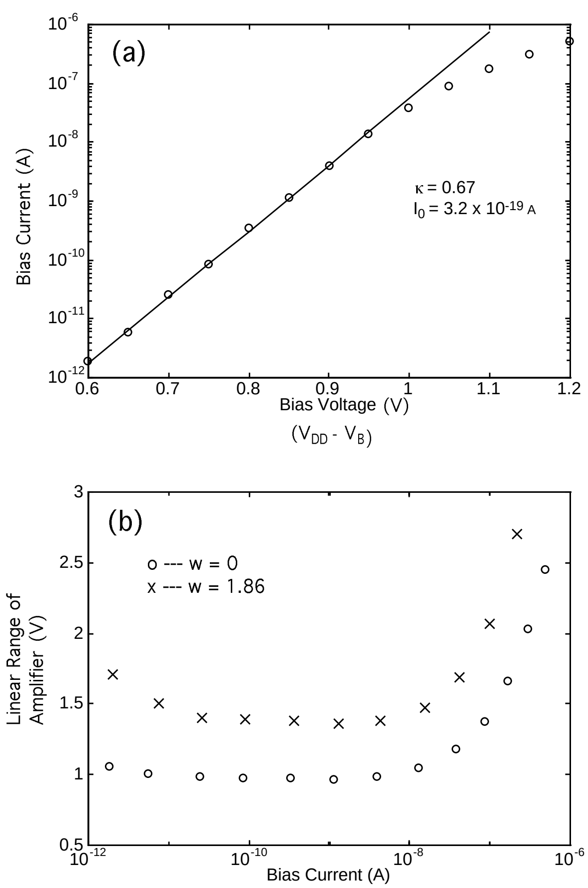

Fig. 7. Linear range vs. bias current. (a) The bias current I_B as a function of the bias voltage V_B. (b) The linear range of the amplifier V_L as a function of the bias current I_B.

rent offset caused by a threshold or geometry mismatch is $i_k/I_K = \delta$, then the voltage offset v_l is $i_k/g_m = (I_K/g_m)(i_k/I_K) = (I_K/g_m)\delta$. The I_K/g_m ratio is thus the scale factor that converts percentage mismatches to voltage offsets. A transconductance amplifier has an I_B/g_m ratio of of $2U_T/g$. Thus, the same percentage mismatch results in a larger voltage offset for an amplifier with a lower g. This general and well-known result causes the lowering of transconductance to be associated with an increase in voltage offset. However, we show next that, although this increase in voltage offset is fundamental and unavoidable, certain transistor mismatches matter more than do others. By designing the amplifier carefully, such that more area is expended where mismatches are more crucial, we can minimize the offset. In our amplifier, mismatches may arise in the two *n*FET GM–M mirrors, in the *p*FET M mirror, or between the W, S, and GM transistors in the two arms. In a first-order analysis, the B transistors do not affect the voltage offset, except that g is replaced by $(2/3)g$. We shall therefore start our analysis with the case of $w = 0$, and then extend our analysis to the $w = 2$ case.

Suppose the net mismatch due to all mirror mismatches is such that $i_+/i_- = \frac{I_B}{2}(1+\frac{\delta_M}{2})/\frac{I_B}{2}(1-\frac{\delta_M}{2}) \approx 1+\delta_M$. The mismatch in currents may be referred back to the input as a voltage offset of $+V_M/2$ on one input,

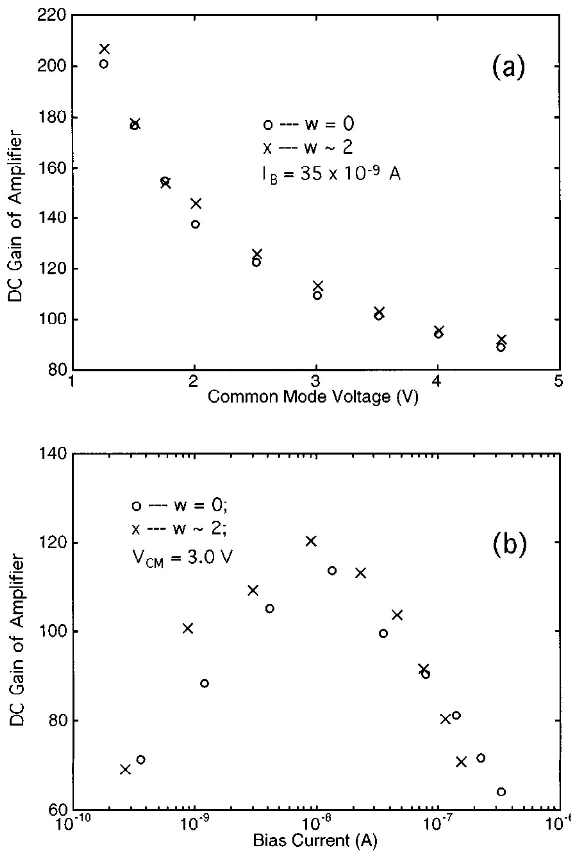

Fig. 9. Gain characteristics. (a) The gain of our amplifier as a function of the common-mode voltage. At low common-mode voltages, the gain increases because of the rising transconductance of the amplifier. (b) The gain of our amplifier as a function of the bias current. Initially, the gain rises because of the increasing Early Voltage of the amplifier, and then falls because of the increasing linear range. A more detailed discussion may be found in Section 3.3.

and a voltage offset of $-V_M/2$ on the other input:

$$\frac{V_M}{2} = \left(\frac{I_K}{g_m}\right)\frac{\delta_M}{2} = \left(\frac{\frac{I_B}{2}}{\frac{gI_B}{2U_T}}\right)\frac{\delta_M}{2}$$
$$\Rightarrow \quad V_M = \delta_M\frac{U_T}{g}. \tag{20}$$

Since $\frac{U_T}{1-\kappa}$ is the I_K/g_m ratio of the W transistors, a mismatch of δ_W between those transistors introduces a voltage offset of $\frac{U_T}{1-\kappa}\delta_W$ on the well inputs. Because $\frac{U_T}{\kappa_n}$ is the I_K/g_m ratio of the S transistors, a mismatch of δ_S between those transistors results in a voltage offset of $\frac{U_T}{\kappa_p}\delta_S$ on their gates. This voltage offset is fed back to the sources of the W transistors and is amplified by $1/(1-\kappa)$ when referred back to the those transistors' well-inputs. Similarly, a mismatch of δ_G between the G transistors causes a voltage offset of $\frac{U_T}{\kappa_n}$ on the gates of the GM transistors that is amplified by $\kappa/(1-\kappa)$ when referred back to the well inputs. The total offset, V_{of}, is just the sum of the voltage offsets introduced by all the mismatches, and is given by

$$V_{of} = U_T\left(\frac{1}{g}\delta_M + \frac{1}{1-\kappa}\delta_W + \frac{1}{\kappa_p}\frac{1}{1-\kappa}\delta_S\right) + U_T\left(\frac{1}{\kappa_n}\frac{\kappa}{1-\kappa}\delta_G\right). \tag{21}$$

If we use the expression from Eq. (3) to substitute for g, then

$$V_{of} = \frac{U_T}{1-\kappa}\left(\left(1+\frac{1}{\kappa_p}+\frac{\kappa}{\kappa_n}\right)\delta_M + \delta_W\right) + \frac{U_T}{1-\kappa}\left(\frac{1}{\kappa_p}\delta_S + \frac{\kappa}{\kappa_n}\delta_G\right). \tag{22}$$

We notice that the greatest contributor to the offset is the mirror mismatch δ_M. Thus, it is important that all mirror transistors in the amplifier be big. The matching of the S transistors is more important than is the matching of the W transistors since $1/\kappa_p > 1$. The matching of the GM transistors is more important than is that of the W transistors if $\kappa/\kappa_n > 1$, which is usually the case; for κ/κ_n ratios exceeding $1/\kappa_p$, it is also more important than the matching of the S transistors. The amplifier is thus laid out with big pFET M transistors, moderate GM transistors, moderate nFET M transistors (to match the GM transistors), moderate S transistors and small W transistors. It is interesting that the transistors that matter the least for matching are the input transistors.

If the B transistors are present, a similar analysis shows that the matching of the M transistors becomes even more important, because the g in Eq. (21) is replaced by $(2/3)g$. In Eq. (22), we then have $\delta_M \rightarrow (3/2)\delta_M$.

The sizes of the transistors that we used for the amplifier were 12/12 for the W transistors, 12/12 for the bias transistor, 29/12 for the S transistors, 14/16 for the GM and nFET M transistors, and 26/16 for the pFET M transistors. The total dimensions of the amplifier were 85μm $\times$ 190μm in a 2μm process. Our random offset is about 5 to 10 mV.

Due to Early voltage effects, the cascaded gains of the GM–M mirror and the pFET M mirror in the

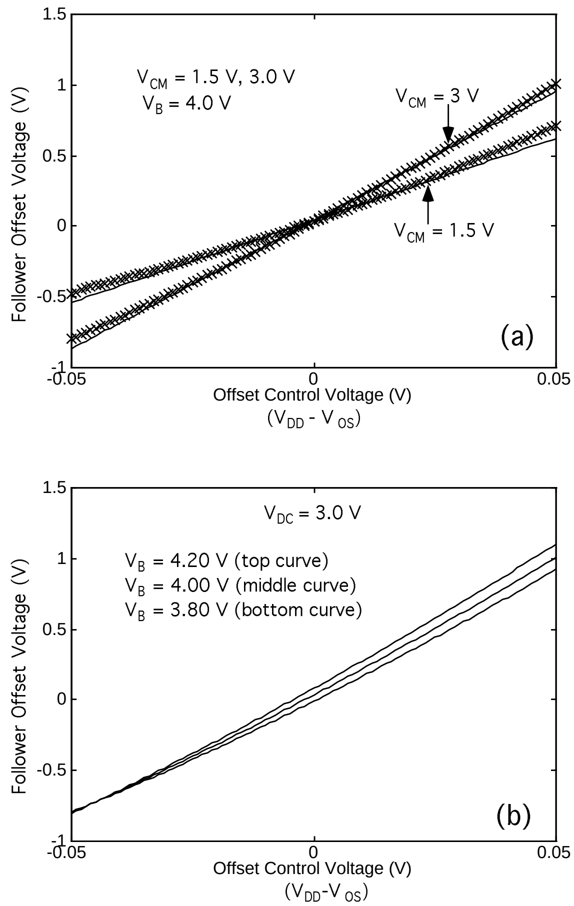

Fig. 10. Offset characteristics. (a) The offset characteristics of the amplifier at two different common-mode voltages. (b) The same characteristics at three different bias currents. More details may be found in Section 3.4.

positive-output-current path, exceed those of the GM–M mirror in the negative-output-current path. Hence, with $V_{OS} = V_{DD}$, the amplifier typically has a systematic positive voltage offset that is on the order of 10 to 20 mV. The voltage V_{OS} is operated approximately 0.5 mV below V_{DD}, to cancel this offset. We have also built automatic offset-adaptation circuits to control the offset of our amplifiers; we shall not elaborate on those schemes here.

The data of Figure 10a show the offset voltage of a follower built with this amplifier as a function of the V_{OS} voltage. The two current curves correspond to two different common-mode voltages. The slope of the input–output relation is the ratio of the source transconductance at the V_{OS}-input g_s to the transconductance of the overall amplifier g. At high common-mode voltages, the offset is more sensitive to the V_{OS} voltage, because g decreases with common-mode voltage. Figure 10b shows that the offset voltage is less sensitive to the V_{OS} voltage at high bias-current levels, because g_s decreases faster than does g with bias current.

4. Follower–Integrator Characteristics

One of the most common uses of transconductance amplifiers is in building filters. Thus, it is important to study the properties of a follower–integrator. The follower–integrator is the simplest filter that can be built out of a transconductance amplifier; it is a first-order, lowpass filter. Figure 11a shows the basic configuration. The voltage V_T is identical to V_B in Figure 1, and, with the capacitance C determines the corner frequency (CF) of the filter. Figure 11b shows the dc characteristics of such a filter built with our amplifier. Except for very low input voltages, where the parasitic bipolar transistors shunt the amplifier's current to ground, the output is a faithful replica of the input. Figure 12 reveals the effects of the parasitic bipolar transistors in detail. At input voltages that are within the normal range of operation of the amplifer (1V to 5V), such as those shown in the top portion of the figure, the output of the follower is a simple lowpass-filtered version of the input. In the bottom portion of the figure, we see that, as long as the input voltage is below about 0.7 V, the amplifier's bias current is shunted to ground, and the output voltage barely changes. The constancy of the output voltage occurs because there is no current at the output of the amplifier to charge or discharge the capacitor such that it barely changes. When the input voltage is outside this range, the output voltage tries to follow the input voltage by normal follower action. We show these data to illustrate the consistency of our earlier results with the behavior of the follower–integrator; we do *not* recommend operation of the filter in this regime. We shall now describe more useful linear and nonlinear characteristics of the follower–integrator in detail.

4.1. Linear and Nonlinear Characteristics

Figure 13a and Figure 13b show data for the gain and phase characteristics of a follower–integrator. The phase curve is much more sensitive to the presence of parasitics than is the gain curve, which remains ideal for a much wider range of frequencies. The higher sensitivity of the phase curve to parasitics is reasonable, because the effect of a parasitic pole or zero on

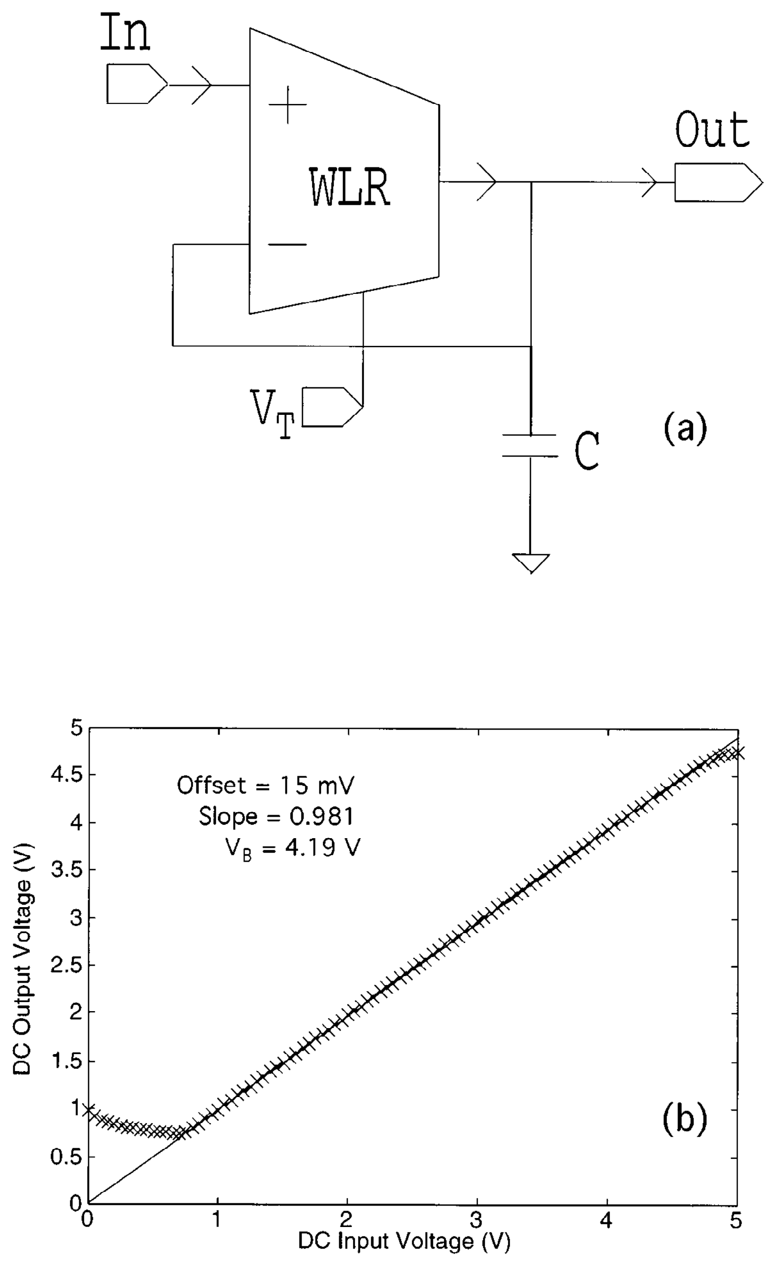

Fig. 11. Follower DC characteristics. (a) The basic circuit for a follower–integrator. (b) The output is a faithful replica of the input except at very low input voltages.

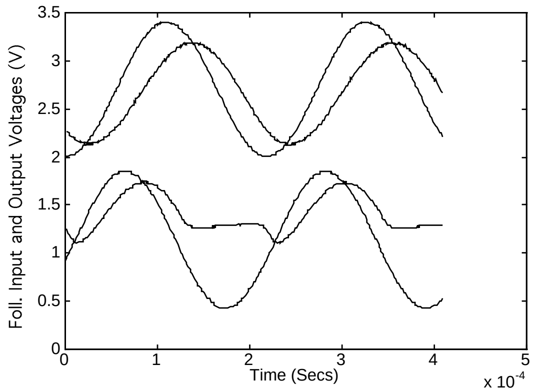

Fig. 12. Bipolar effects in the follower–integrator. The top portion of the figure shows normal follower–integrator operation, i.e., the filtering of a sinewave input. The output waveform is attenuated and phase-shifted with respect to the input waveform. The bottom portion of the figure shows operation of the filter at low input voltages. An explanation for the strange output curve is given in Section 4, and is due to the effects of the parasitic bipolar transistors in our amplifier.

the gain is appreciable only at or beyond the pole or zero frequency location. The effect of the same pole or zero on the phase is significant a full decade before the frequency location of the pole or zero.

We have studied the origin of the parasitic capacitances in detail, but we shall not delve into this subject here. The parasitics are caused by the large sizes of transistors that we use to reduce our $1/f$ noise and offset; the dominant parasitics are the well-to-drain capacitance of the W transistor, which causes a right-half-plane zero, and the gate-to-bulk capacitance of the GM transistors, which causes a left-half-plane pole. Rather than make detailed models of the parasitics, we model their effect by simply having a different corner frequency (CF) for the gain and phase curves. As Figure 13a and Figure 13b show, the phase CF is slightly lower than the gain CF, because of the excess phase due to the parasitics.

It is possible to reduce the influence of parasitics in our amplifier by having the *n*FET GM–M mirrors be attenuating. Then, the differential-arm parasitics are at higher frequencies, compared with the CF of the filter. Another alternative is to use a larger output capacitor. However, for cochlear-filtering applications, for which this amplifier is mainly intended, there is a steep amplitude rolloff beyond the CF of the lowpass filter, and fine phase effects are not important at these frequencies. In addition, the second-order filter, of which this filter is a part, is only an approximation to the more complicated filtering that occurs in a real cochlea. Thus, we have not expended energy optimizing the parasitics to get an ideal first-order filter.

Figure 13c shows the first, second, and third harmonic rms output amplitudes when the input rms amplitude is 1V (peak-to-peak of 2.8V). The data were collected with an EGG3502 lockin amplifier. Because of the wide linear range of our amplifiers, there is little distortion even at these large amplitudes. At frequencies that are within a decade of the CF, the second harmonic is less than 4×10^{-2} the amplitude of the first harmonic. The third-harmonic distortion is negligible at these frequencies. The total harmonic distortion is less than 4%.

Figure 14 plots the relative amplitude distortion—that is to say, the ratio of the magnitude of the second and third harmonic amplitude to that of the first. Note that the X-axis is plotted in multiples of the CF, and that we are uninterested in effects more than a decade beyond the 1 point. The chief features of interest are that the addition of the bump transistors reduces third-harmonic distortion as we expect from Eq. (14) and Eq. (15). However, it also increases the second-harmonic distortion at frequencies just below the CF, for reasons unknown to us. Usually, second-harmonic distortion occurs along with dc shifts. Figure 15 shows these shifts for an input signal with an rms input amplitude of 1V. These shifts exert only a mild influence in cochlear designs, because input-signal amplitudes far beyond the CF are small due to successive filtering from prior stages. The dc shifts are not significant for small inputs. The dc shifts are observed in OTA–C follower–integrators as well; they are typically positive for amplifiers with *n*FET differential pairs and negative for amplifiers with *p*FET differential pairs.

Figure 16a shows the shifts in gain and phase CF for the $w = 0$ and $w = 2$ cases as a function of the input rms amplitude. Note that the gain and phase CF shifts are similar. The addition of the B transistors decreases the amount of CF shift due to the linearization of the tanh. Figure 16b shows the shifts in CF as a function of the input dc voltage to the follower–integrator. As we expect from prior data on the amplifier (Figure 5a), the CF shifts track the shifts in transconductance with dc input voltage. The tracking is seen for gain and phase CF curves, and for the $w = 0$ and $w = 2$ cases.

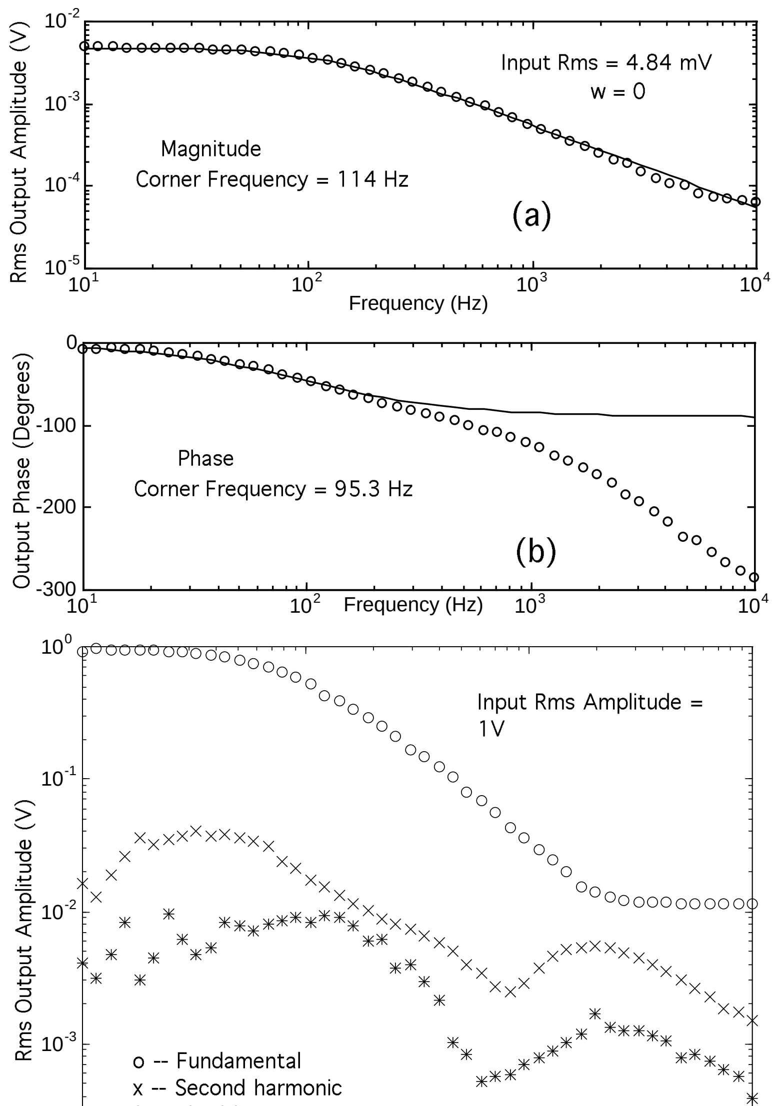

Fig. 13. Follower–integrator frequency response at a dc-input voltage of 3V. (a) and (b) show the gain and phase characteristics of the follower–integrator along with fits to lowpass filter transfer functions. The phase characteristics are more sensitive to parasitics and we model their effect by fitting the phase curves to a different corner frequency than the gain curves. (c) The first, second, and third harmonic rms output amplitudes at an input rms amplitude of 1V (2.8 V peak-to-peak). Because of the wide linear range of the amplifier, the total harmonic distortion is low even at this amplitude.

5. Noise and Dynamic Range

The largest signal that a filter can process without excessive distortion is determined by the linear range of that filter. The smallest signal that a filter can process is one whose rms amplitude is just above the filter's input-referred noise. The dynamic range of a filter is given by the ratio of the power of the largest signal to the power of the smallest signal. We have already discussed linear range and distortion in the follower–integrator. Now, we focus on the other half of the story: noise in the follower–integrator. After we compute and measure the noise, we can tell whether there has been an improvement in the dynamic range of the follower–integrator. Since there are no free lunches in nature, we also want to discover what price we have paid in our design, in terms of the increase in power and area. We shall discuss noise and dynamic range in our design in Sections 5.1–5.3. We shall discuss the dynamic range of capacitive-divider schemes in Section 5.4, since those schemes bear similarity to our technique of using the well as an input.

5.1. Theoretical Computations of Noise in the Amplifier and Follower–Integrator

We now compute the noise in a follower–integrator built with our amplifier. We tie the v_+ and v_- to a common constant voltage source, and the output to another constant voltage source. Then, we replace each of the 13 transistors in Figure 1 with its small-signal equivalent circuit. We introduce a noise current source between each transistor's drain and source terminals in the small-signal equivalent circuit. For each noise source, we compute the ac transfer function between its current and the differential output current. We sum the noise contributions from each of the 13 sources incoherently; that is to say, the square of the net output current noise is the sum of the squares of the current noise from each source. The input-referred voltage noise per unit bandwidth is the output current noise per unit bandwidth divided by the transconductance of the amplifier. The total voltage noise is the voltage noise per unit bandwidth integrated over the bandwidth of the follower–integrator. The bandwidth of the follower–integrator is determined by the transconductance of the amplifier and the capacitance. If the amplifier is used in a system, where the bandwidth is determined by some other part of the system, then this bandwidth determines the interval of integration. The parasitic

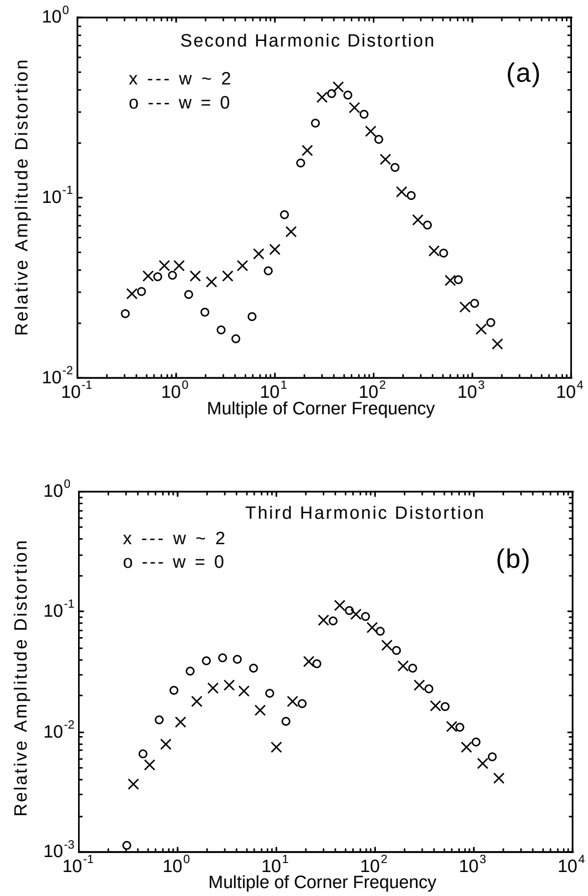

Fig. 14. Follower–integrator distortion characteristics. (a) The ratio of magnitude of the second harmonic to that of the first harmonic vs. normalized frequency, i.e., frequency normalized by the CF. (b) The same as (a) except that the third harmonic is plotted instead. Note that the addition of the *B* transistors decreases third harmonic distortion but increases second harmonic distortion.

capacitances in the amplifier set an upper bound on the maximum possible bandwidth.

Although there are 13 transistors in Figure 1, we do *not* get 13 transistors' worth of current noise at the output. Our calculations will show that we get about 5.3 transistors' worth of current noise. This figure is only slightly higher than the 4 transistors' worth of noise obtained from a 5-transistor ordinary transconductance amplifier (OTA). The reduction in noise occurs for three reasons. First, for each noise source, there is a total or partial cancellation of its noise current at the output, due to opposing contributions from the two circuit arms. As an example, the noise current from the bias transistor in Figure 1 (with gate voltage V_B) makes no contribution to the output noise current,

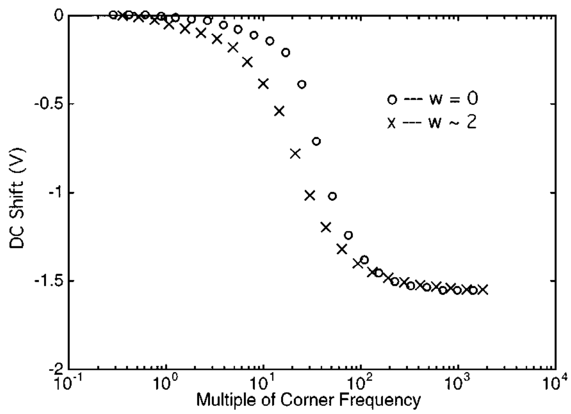

Fig. 15. DC shifts in the follower–integrator. The shifts are shown at an input rms amplitude of 1V. These shifts exert only a mild influence in cochlear designs. See Section 4.1 for details.

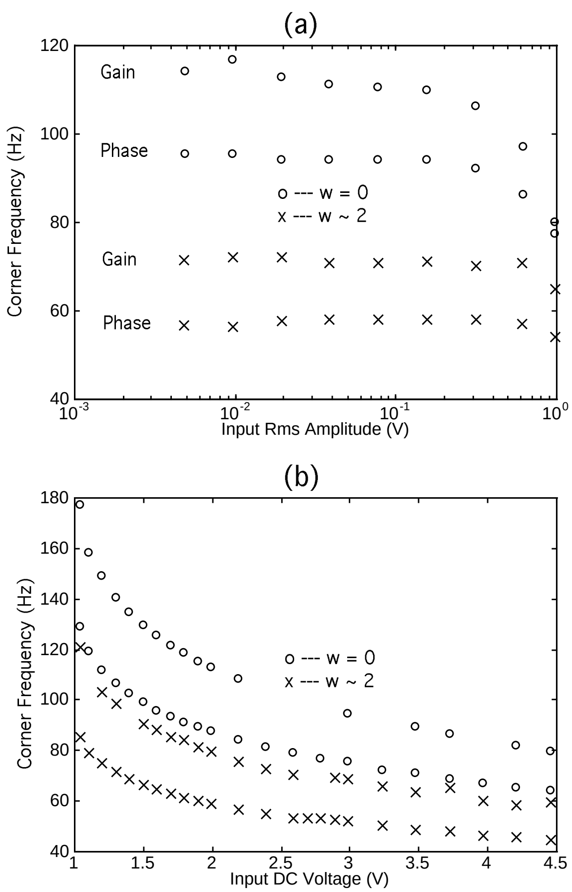

Fig. 16. Corner-frequency shifts in the follower–integrator. (a) CF-shifts vs. input rms amplitude and (b) CF-shifts vs. input dc voltage. The lower curve is the phase CF for the $w = 0$ and the $w = 2$ cases.

because it branches into equal portions in the two differential arms, which cancel each other at the output. Similarly, other noise sources, such as those from the B transistors, contribute no noise. The sources from the GM transistors have a partial noise cancellation. Second, some of the noise current from each source is prevented from contributing to the output by local shunting circuit impedances. As an example, the noise currents from the W transistors contribute only 0.16 transistors' worth of current noise to the output because most of the noise current gets shunted by the W transistors themselves. Third, when we compute the incoherent noise average across many sources, a given source's contribution to the net noise is proportional to the square of its fractional current gain; that is to say, to the square of the output noise current divided by source current. Therefore, weak contributions are weakened further.

If we define $\alpha_B, \alpha_S, \alpha_W, \alpha_G$, and α_M to be the current gain between the output current of the amplifier and the input drain-to-source noise current of a B,S,W,G, or M transistor, respectively, we can show that

$$\alpha_B = 0, \tag{23}$$

$$\alpha_S = \frac{\kappa_n}{\kappa_n + \kappa_n\kappa_p + \kappa\kappa_p}, \tag{24}$$

$$\alpha_W = \frac{\kappa_n\kappa_p}{\kappa_n + \kappa_n\kappa_p + \kappa\kappa_p}, \tag{25}$$

$$\alpha_G = \frac{\kappa_n + \kappa_n\kappa_p}{\kappa_n + \kappa_n\kappa_p + \kappa\kappa_p}, \tag{26}$$

$$\alpha_M = 1. \tag{27}$$

Since each transistor has the same dc current flowing through it (i.e. $I_B/2$ in the $w = 0$ case, or $I_B/3$ in the $w = 2$ case), the magnitude of the noise current source across each transistor is the same. Thus, the output-current noise of the amplifier is the noise current due to one transistor times an effective number of transistors N, where, N is given by

$$N = 2\alpha_S{}^2 + 2\alpha_W{}^2 + 2\alpha_G{}^2 + 4\alpha_M{}^2. \tag{28}$$

For our amplifier with $\kappa \approx 0.85$, $\kappa_n \approx 0.7$, and $\kappa_p \approx 0.75$, the numbers work out such that $N = 5.3$. The dominant contribution to N comes from the four M transistors which contribute a whole transistor each. The two G transistors contribute 0.865 transistors. The two S transistors contribute 0.28 transistors. The two W transistors contribute 0.16 transistors. The B transistors and the bias transistor contribute no noise. The most noise-free linearization techniques, in decreasing order of merit, are bump linearization, the use of the well as an input, source degeneration, and gate degeneration. Bump linearization is the only technique of the four that adds no noise whatsoever. Note that, depending on the circuit configuration, the relative noise efficiencies of the use of the well as an input, source degeneration, and gate degeneration may vary. For example, in a well-input amplifier with source degeneration but no gate degeneration, $\alpha_W = \alpha_S = \kappa_n/(\kappa + \kappa_n)$. In the latter case, the use of the well as an input and gate degeneration each contribute 0.41 transistors' worth of noise.

The magnitudes of the individual noise current sources depend on the dc current flowing through the transistor, and are well described by a white-noise term for low subthreshold currents [27]. At high subthreshold currents, there is also a $1/f$-noise term. Our experimental data in Section 5.3 reveal the amount of the $1/f$ contribution; we shall model this term empirically, because no satisfactory theory for $1/f$ noise currently exists.

In the first paragraph of this section, we explained the procedure for calculating the noise. We shall now perform the calculations. As usual, we begin by analyzing the case for $w = 0$ and then extend our analysis to the $w = 2$ case. The output-current noise of the amplifier $\overline{i_o^2}$ is given by

$$\overline{i_o^2} = \int_{f_l}^{\infty} N\left(2q\left(\frac{I_B}{2}\right) + \frac{K(I_B/2)^2}{f}\right) df, \tag{29}$$

where the first and second terms in the integral correspond to white noise and $1/f$ noise, respectively; I_B is the bias current, and K is the 1/f noise coefficient. We also assume that there is low-frequency adaptation in the amplifier, so that frequencies below f_l are not passed through. This assumption is necessary if we are to prevent the $1/f$ noise from growing without bound at low frequencies. In our amplifier, we have an offset-adaptation circuit that keeps f_l around 1 Hz. Also, typically the K for pFETs is smaller than is that for nFETs, and scales inversely with the channel area of the transistor. However, we assume a transistor-invariant K, for simplicity.

The corner frequency of the follower–integrator f_c is

$$f_c = \frac{I_B}{2\pi C V_L}, \tag{30}$$

so from Eq. (29), the input-referred voltage noise $\overline{v_i^2}$ is

$$\begin{aligned}\overline{v_i^2} &= \int_{f_l}^{\infty} \frac{N\left(qI_B + \frac{KI_B^2}{4f}\right)}{(I_B/V_L)^2}\left(\frac{1}{1+(f/f_c)^2}\right) df \\ &= \left(\frac{NqV_L{}^2}{I_B}\right)\frac{\pi f_c}{2} + \frac{NKV_L^2}{8f}\ln\left(1+(f_c/f_l)^2\right) \\ &= \frac{NqV_L}{4C} + \frac{NKV_L{}^2}{8}\ln\left(1+\left(\frac{I_B}{2\pi f_l C V_L}\right)^2\right).\end{aligned} \tag{31}$$

In evaluating Eq. (31), we computed two integrals:

$$\int_0^{\infty} \frac{dx}{1+x^2} = \frac{\pi}{2}, \tag{32}$$

$$\int_{f_l}^{\infty} \frac{dx}{x(1+x^2)} = \frac{1}{2}\ln\left(1+\frac{1}{f_l^2}\right). \tag{33}$$

Note that the $1/f$ noise rises with bias current because of the increasing bandwidth. The magnitude of the $1/f$ noise depends on the bandwidth; that is to say, it depends on I_B/CV_L. The white noise is current invariant, because the noise per unit bandwidth and the bandwidth scale as $1/I_B$ and I_B, respectively. The white noise increases linearly with V_L, whereas the $1/f$ noise increases quadratically with V_L.

5.2. Theoretical Computations of Dynamic Range

The dynamic range D_R is defined to be the ratio of rms input power in a signal with amplitude V_L to the input-referred noise power; this definition is implicitly based on our willingness to accept the distortion present at this amplitude, which from Figure 13 is seen to be reasonable. If we want to be more conservative with respect to distortion, then we simple scale V_L by the necessary fractional amount. So, we have from Eq. (31) that

$$\begin{aligned}D_R &= \frac{V_L^2/2}{\overline{v_i^2}} \\ &= \frac{V_L^2/2}{\frac{NqV_L}{4C} + \frac{NKV_L{}^2}{8}\ln\left(1+\left(\frac{I_B}{2\pi f_l C V_L}\right)^2\right)} \\ &= \frac{1}{\frac{Nq}{2CV_L} + \frac{NK}{8}\ln\left(1+\left(\frac{I_B}{2\pi f_l C V_L}\right)^2\right)}.\end{aligned} \tag{34}$$

In Eq. (34), the white-noise term is typically much larger than the $1/f$ term, because of the small value of K in our large, mostly pFET transistors, because frequencies below 1Hz are filtered out by offset adaptation, and because white noise is still relatively high at the low subthreshold current levels. Even at high bias currents, where the $1/f$ noise appears to be important on a log–log plot of noise amplitude versus frequency, the net area under the plot is dominated by white noise on a linear–linear plot. Thus, from Eq. (34), the dynamic range is nearly

$$D_R = \frac{2CV_L}{Nq}. \tag{35}$$

We see from Eq. (31) that the white-noise power increases like V_L. The maximum signal power increases like V_L^2. *Thus, if white noise dominates, wide linear range implies more dynamic range, as Eq. (35) predicts*. The noise per unit bandwidth scales like V_L^2, but the bandwidth scales like $1/V_L$ such that the overall noise scales like V_L. If bandwidth is to be preserved, then power proportional to V_L has to be expended, according to Eq. (30). After the bandwidth has been restored to its original value, however, the noise still remains at a value proportional to V_L. The noise remains at the same value because thermal noise is current invariant. The current invariance arises because the noise per unit bandwidth and the bandwidth scale like $1/I_B$ and I_B, respectively. The extra dynamic range has been earned at the price of an increase in power.

We see from Eq. (31) that the $1/f$-noise power increases like V_L^2. The maximum signal power increases like V_L^2. *Thus, if $1/f$ noise dominates, wide linear range does not affect the dynamic range*. The noise per unit bandwidth scales like V_L^2, and the bandwidth scales like $1/V_L$. If bandwidth is to be preserved, then power proportional to V_L has to be expended, according to Eq. (30). After the bandwidth has been restored to its original value, however, the noise increases like V_L^2 because $1/f$ noise is not current invariant, but rather is bandwidth invariant. The bandwidth invariance arises because the noise per unit bandwidth depends on only V_L^2 and is current invariant. So, the total noise over a given bandwidth depends on only that bandwidth. The advantage of linearity in such a situation is merely that the signal and noise levels have been scaled to be suited to inputs that themselves have a higher noise floor and large amplitudes. There is also an area cost associated with extra dynamic range: To reduce $1/f$ noise and offset, we must use transistors with large areas.

If $w = 2$, the analysis for $\overline{v_i{}^2}$ and D_R proceeds similarly. For the currents through the transistors, $I_B/2 \to I_B/3$. For the bandwidth and linear range, $V_L \to (3/2)V_L$. If a $w = 0$ and a $w = 2$ follower–integrator have the same bandwidth, then we can show that they have identical input-referred noise spectra. For the bandwidths for the $w = 2$ and $w = 0$ cases to be equal, however, the bias current for the $w = 2$ case has to be increased by a factor of (3/2) over that for the $w = 0$ case. Theoretically, the dynamic range of the $w = 2$ case is $(3/2)^2$ times the dynamic range of the $w = 0$ case because of the increase in linear range. In practice, however, if the linear range is already large in the $w = 0$ case, the increased linearity in the $w = 2$ case does not reduce distortion further. The distortion is typically limited by some other mechanism. For example, in our amplifier the dominant distortion is due to κ shifts caused by the nonlinear well-to-channel capacitance. Nevertheless, the increased linearity does reduce third-harmonic distortion and CF shifts with input amplitude, as shown in Figure 14b and Figure 16b.

5.3. Experimental Noise Curves

Figure 17a shows noise spectra at low bias-current levels for a follower–integrator with a capacitance C = 1.02 pF. The data were taken with a HP3582A spectrum analyzer. From Eq. (29), we would expect white or thermal noise to be dominant at low bias currents. We observe from the figure that, even at the lowest frequencies, no $1/f$ noise is visible. We were able to fit the data with lowpass-filter transfer functions as shown. Note that, for the low bias currents of Figure 17a, where $1/f$ noise is hard to discern, we did not use any $1/f$ terms in our fit. The terms in the integral of Eq. (31) predict that the noise spectra reveal relatively more $1/f$ noise at high bias current levels because the white-noise term decreases and the $1/f$ term remains roughly constant. The data of Figure 17b illustrates that this prediction is approximately borne out. However, we were able to empirically fit the voltage noise per unit bandwidth $\overline{v_{if}^2}$ more accurately by a term of the form

$$\overline{v_{if}^2} = \left(\frac{K_B}{f^n} + A\right)\left(\frac{1}{1 + (f/f_c)^2}\right). \tag{36}$$

Figure 18a shows a typical fit to the data in more detail. From the first line of Eq. (31), we would expect $K_B = V_L^2 K/4$, n = 1, and $A = {V_L}^2 Nq/I_B$. Since K_B and n are empirical, they do not yield substantial theoretical information, although they are useful practically. In the following paragraph, we show how to extract the value of N of Eq. (35) from the value of A.

From the value of f_c obtained from the fit to the data, from our knowledge of C, from measurements of I_B, and from Eq. (30) we obtain V_L. Given V_L and the fit parameter A, we obtain N, the effective number of transistors contributing shot noise. Figure 18b shows a plot of N versus the bias current I_B. We observe that N is roughly 7.5 in subthreshold, and decreases as the bias current goes above threshold and space-charge smoothing sets in. The value of N in subthreshold is within a factor of 1.4 of our theoretical prediction of 5.3. Above threshold, the space-charge smoothing, that is to say, the modulation of the mobile charge concentration by the mobile charges themselves, reduces the noise to a value below what we would expect from shot noise.

Figure 19 shows a plot of how the K_B and n of Eq. (36) vary with bias current I_B. Since $K_B = K V_L^2/4$, part of the increase in K_B arises from the increase in V_L and part of it arises from the increase in K. It is also interesting that, as the bias current increases, the $1/f$ noise power systematically rises from about 0.67 to about 0.95.

The noise measurements of Figure 17 through 19 were taken for an amplifier with $w = 0$. We also experimentally confirmed that the noise in a $w = 2$ amplifier was identical to that in a $w = 0$ amplifier of the same bandwidth.

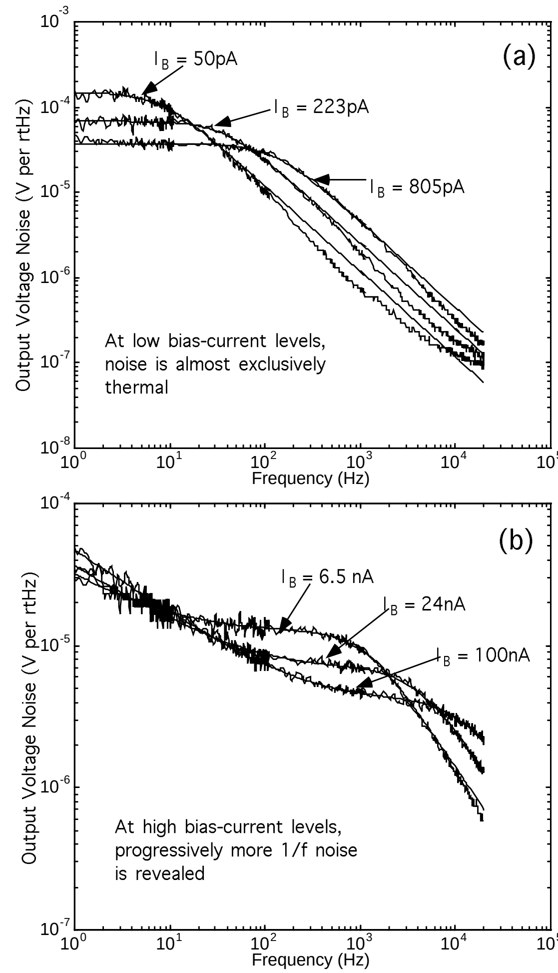

Fig. 17. Noise spectra at various current levels. (a) At low bias currents the noise is almost solely white or thermal. The bold lines are lowpass-filter fits to the data. (b) At high bias currents, there is relatively more $1/f$ noise. Nonetheless, the dominant contribution to the noise, which is the area under this curve in a linear–linear plot, remains thermal. The bold lines are fits to the sum of a $1/f$ term and a white-noise term.

5.4. Capacitive-Divider Techniques

Our use of the well as an input implicitly involves a capacitive-divider technique: The gate, surface potential, and well form three terminals of a capacitive divider. We chose the well as an input because coupling ratio of the well to the surface potential, $1 - \kappa$, is smaller than the coupling ratio of the gate to the surface potential, κ. The advantage of this implicit capacitive-divider scheme is that the divider is inherently part of the transistor; so, we exploit a parasitic capacitance rather than avoiding one. Also, no additional floating-gate adaptation circuits or control voltages are needed. The disadvantage of the technique is that the capacitive-divider ratio is fixed by the physical parameters of the process, and is slightly nonlinear. If the divider ratio is not as small as desired, we must use other circuit techniques like source degeneration, gate degeneration or bump linearization to obtain wider linear range. Luckily, in our circuit, the additional transistors used to widen the linear range do not increase the noise greatly, but they do cost more area. It is logical to ask whether we can do better in area consumption with circuits that have explicit capacitive dividers.

We shall discuss two simple schemes where capacitive dividers are explicitly used around OTA's. We assume that any floating-gate inputs of the amplifiers

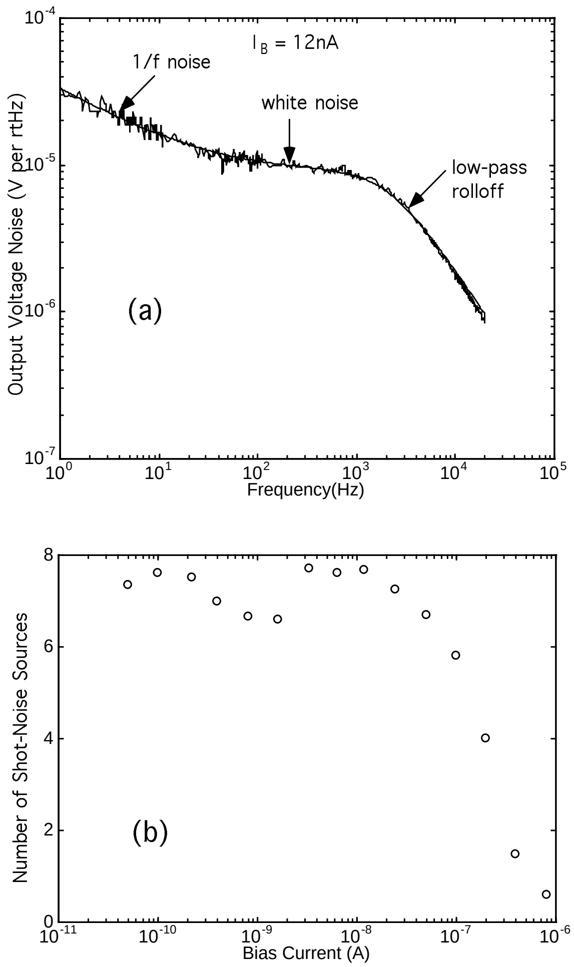

Fig. 18. Basic noise characteristics. (a) A typical noise spectrum. (b) The effective number of transistors contributing shot noise in our circuit as a function of the bias current. As the bias current goes above threshold, the effective number of transistors contributing shot noise decreases because of space-charge smoothing.

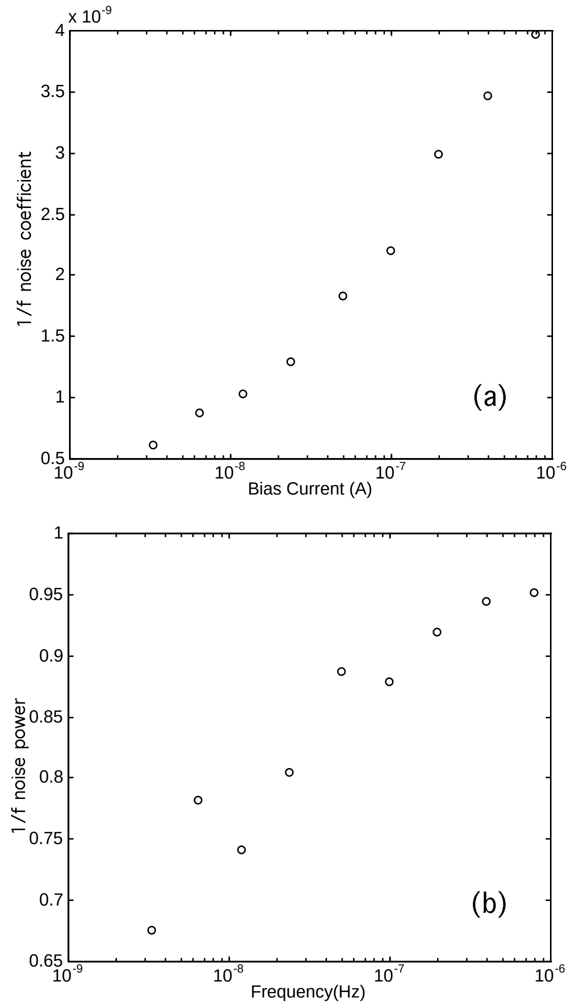

Fig. 19. Characteristics of $1/f$ noise. (a) The $1/f$ noise coefficient K_B, used in Eq. (36), as a function of the bias current I_B. (b) The $1/f$ noise power n as a function of the bias current I_B.

are held at a proper dc value by low-frequency adaptation circuits. We further assume that the adaptation circuits do not affect noise in the amplifier's transistors or in the circuit. In practice, this assumption may not be true of certain adaptive schemes.

Figure 20a shows a simple scheme. The voltage V_T determines the bias current in the OTA. In practice, parasitic capacitances between the output and input terminals of the OTA can hurt the design significantly. If V_L is the linear range of the OTA, and N is the effective number of noise-contributing transistors in the OTA, then it can be shown that the dynamic range D_R is

$$D_R = \frac{2\left(C_{out} + \frac{C_1 C_2}{C_1+C_2}\right)\left(\frac{C_1+C_2}{C_1}\right)V_L}{Nq}. \quad (37)$$

The analysis leading to the previous equation is similar to that preceding Eq. (35). We assume that thermal noise dominates. From Eq. (37), we see that C_{out} needs to be moderately large. If not, any improvement in dynamic range over that of an OTA–C follower–integrator arises only at the expense of an extremely large value of C_2. For example, if C_{out} were 0, we would need C_2 to be approximately 15 pF to get a dynamic range improvement of 15 over that of an OTA–C follower–integrator with 1pF.

Similarly, for the inverting configuration of Figure 20b, we get

$$D_R = \frac{2\left(C_{out} + \frac{C_2(C_{in}+C_1)}{C_2+C_{in}+C_1}\right)\left(\frac{C_2+C_{in}+C_1}{C_2}\right)V_L}{Nq}. \qquad (38)$$

Once again, we observe that C_{out} must be moderately large. If not, any improvement in dynamic range arises only at the cost of an extremely large value of $(C_{in} + C_1)$. This configuration also introduces an RHP zero. If the effects of this zero are to occur at frequencies well past the CF of the follower–integrator, then

$$C_{out}\left(\frac{C_2 + C_{in} + C_1}{C_2}\right) + C_{in} + C_1 \gg C_2. \qquad (39)$$

Parasitic capacitances can also hurt this design significantly, especially if the explicit capacitors in the circuit are small.

Actually, the circuit of Figure 20b does not even need an OTA, as the reference input of the OTA is not really used. The OTA can be replaced by a two-transistor amplifier, but, in that case, V_L also should be replaced by $V_L/2$. Thus, from Eq. (35), as in a normal OTA, N is still effectively 4.

Theoretically, by making capacitive-divider ratios appropriately small, and by spending power, the dynamic range may be increased to values beyond that attained in our design. A floating-gate adaptation scheme combined with a two-transistor version of Figure 20b is being explored [30,31].

6. Conclusions

We conclude by summarizing our key findings:

1. If the amplifier's noise is predominantly thermal, then an increase in its linear range increases the follower–integrator's dynamic range. If the amplifier's noise is predominantly $1/f$, then an increase in its linear range has no effect on the follower–integrator's dynamic range. To preserve follower–integrator bandwidth, power consumption increases proportionately with an increase in the amplifier's linear range according to Eq. (30).
2. In subthreshold, the noise is predominantly due to thermal noise, even at high bias currents, where some $1/f$ noise is present. The theory described in [27] accurately modeled our thermal noise. Empirical expressions in the paper modelled our $1/f$ noise.
3. In subthreshold circuits where thermal noise dominates, a simple formula for the dynamic range of a follower–integrator is $D_R = 2CV_L/Nq$. The capacitance of the follower–integrator is C, the linear range of the amplifier is V_L, the charge on the electron is q, and the effective number of noise-contributing transistors in the amplifier is N. A more complicated formula that includes $1/f$ noise is given by Eq. (34).
4. Experimentally, we obtained a dynamic range of 65.2 dB in a follower–integrator with a capacitance of 1pF. A signal with an input rms amplitude of 1V yielded 4% total harmonic distortion. The total measured noise of the follower–integrator was 0.55 mV. A simple OTA–C follower–integrator has a theoretical linear range of 75 mV, and a theoretical noise floor of 110 μV. Thus, we obtained a dynamic range improvement of at least 8.5 dB over the OTA–C follower–integrator. In practice, due to offchip noise floors on the order of 0.5–1 mV, the improvement can be as much as 20 dB.
5. Bump linearization is our most efficient linearization technique, because it increases the linear range of our amplifier without increasing its noise.
6. Gate degeneration is a useful transconductance-reduction technique. It can be generalized to the notion of the current increase from one input degenerating another input. The technique could be useful in multiple-gate-input circuits [32].
7. When the well is used as an input, the gate must be operated at as low a voltage as possible in order to obtain maximum dc-input operating range.
8. Capacitive-divider techniques that widen the linear range bear similarity to our technique of using the well as an input. If appropriate attention is paid to capacitor sizing, parasitic capacitances, and floating-gate adaptation in these techniques, then they may yield dynamic range improvements similar to ours.
9. Changes in κ, the subthreshold exponential parameter, are due to changes in dc current and to changes in well-to-gate voltage. These two effects may be studied separately through the techniques described in the paper.

Acknowledgments

We thank Lena Peterson for providing the data of Figure 8. We thank Paul Hasler and Bradley Minch for

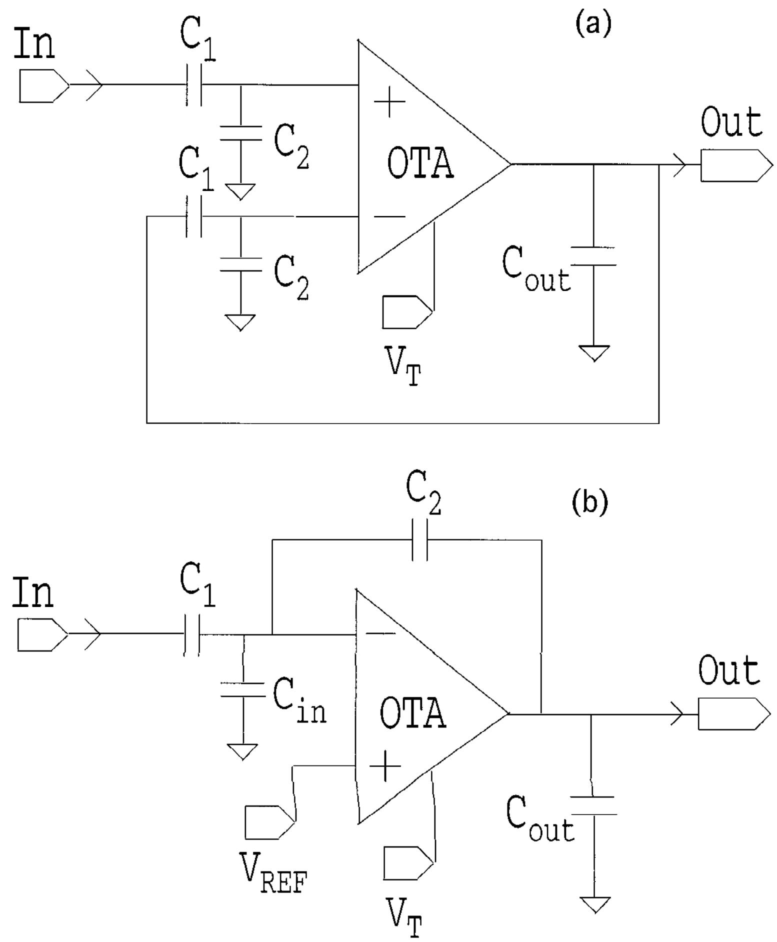

Fig. 20. Capacitive-divider schemes for widening the linear range. (a) and (b) show two different schemes. Section 5.4 contains further details.

useful discussions. This work was supported by the Beckman Hearing Center, and by the Center for Neuromorphic Systems Engineering as part of the National Science Foundation Engineering Research Center Program.

Appendix A

This appendix contains a quantitative discussion of the common-mode effects in our amplifier. The data in the appendix were taken for a $w = 0$ amplifier built in a p-well process, as opposed to the data in the rest of the paper, which were taken in an n-well process. The κ for the p-well process is lower, and consequently the linear range is near 0.6 V, rather than 1V. We also use the grounded-substrate convention [28]. This convention enables us to present the data as though they were from an n-well process, as in the rest of the paper. The grounded-substrate convention implements the following transformation from n-well space to p-well space: $V \rightarrow -V$, n-channel→ p-channel, and p-channel→n-channel. Note that the transformation is applied to all voltages implicitly defined in terms of the gate, source, drain, or well voltages in addition. For example, the flatband voltage is defined in terms of $v_G - v_W$, and changes sign as we move from n-well space to p-well space. Thus, if the flatband voltage is quoted as –0.75 V for an n-channel transistor, it's taken as –0.75 V for a native transistor in n-well space and as +0.75 V for a well transistor in p-well space.

A.1. The Effects of Changes in κ

In our amplifier, the gates of the W transistors are near ground. As we lower the voltages of the well inputs, the well-to-gate voltage decreases; consequently κ decreases; the transconductance, which is proportional to $1-\kappa$, increases. We now analyze this effect more quantitatively.

The parameter κ is a function of the gate-to-well voltage. We can show that

$$\kappa = 1 - \frac{\frac{\gamma}{2}}{\sqrt{\frac{\gamma^2}{4} - (v_G - v_W - V_{FB})}}, \qquad \text{(A1)}$$

where γ is the body-effect parameter and V_{FB} is the flatband voltage.

A well-input amplifier that has no source degeneration or gate degeneration has a transconductance of magnitude g_w, given by $g_w = (1-\kappa)$. By computing the transconductance at the origin for various common-mode voltages V_C, we measured g_w as a function of V_C at a bias current corresponding to $V_{DD} - V_B = 0.77$ V. From Eq. (40), if we plot $1/(1-\kappa)^2$ versus v_W, i.e., $1/{g_w}^2$ versus v_C, we get

$$\frac{1}{{g_w}^2} = \left(\frac{1}{\frac{\gamma^2}{4}}\right) v_C + \left(1 + \frac{V_{FB} - V_G}{\frac{\gamma^2}{4}}\right), \qquad \text{(A2)}$$

which is a straight line. Thus, we can compute γ and V_{FB} from the slope and y-intercept of this line, if we know V_G. For our experiment, we grounded the gate to allow maximum dc input operating range, so V_G was 0. From the data shown in Figure A.1a, we computed $\gamma = 1.06V^{\frac{1}{2}}$ and $V_{FB} = 0.68$ V. In comparison, the SPICE parameters from the MOSIS sheets quoted $\gamma = 1.05V^{\frac{1}{2}}$ and $V_{FB} = 0.75$V. The actual flatband voltages are negative; since the data were taken in a p-well process, we use the positive values as explained in the first paragraph of this appendix.

A well-input amplifier with gate degeneration has a transconductance of magnitude g_g, given by

$$g_g = \frac{1-\kappa}{1 + (\kappa/\kappa_n)}. \qquad \text{(A3)}$$

For such an amplifier, we can determine the functional variation of κ with V_C from Eq. (40), using the previously determined values of V_{FB} and γ, and with V_G being the amount of diode drop on a G transistor. By using measured well and native transistor parameters we estimate $V_G = 0.69$ V given that $V_{DD} - V_B = 0.77$ V, and also that $\kappa_n = 0.714$. By using these parametric values in Eq. (A3) and Eq. (A1), we predicted the dependence of g_g with v_C. The middle curve of Figure A.1b shows that changes of g_g with V_C were in good agreement with the theory of Eq. (A3) and Eq. (A1). The uppermost curve of Figure A.1b is that of g_w versus V_C and is also plotted for reference; it is simply a different way of plotting Figure A.1a.

A well-input amplifier with source and gate degeneration has a transconductance g given by Eq. (3). By using the functional variation of κ versus V_C, the values of κ_n, the value of V_G estimated in the previous paragraph, and $\kappa_p = 0.753$, we were able to predict the variation of g with V_C, as shown by the lowest curve of Figure A.1b. The data begin to deviate from theory at the lower input voltages, probably because of the change in κ_p with increasing well-to-gate voltage.

A.2. The Effects of the Parasitic Bipolar Transistor

To understand exactly when the parasitic bipolar transistor present in every MOS well transistor becomes significant, we find it instructive to analyze the interaction between the bipolar and MOS modes of operation for a well transistor: The subthreshold saturation current of an MOS transistor situated in a well, which is assumed to be an n-well without loss of generality, is given by

$$i_M = I_M e^{\left(\frac{v_S - \psi}{U_T}\right)}, \qquad \text{(A4)}$$

where ψ is the surface potential, and I_M is a constant pre-exponential factor. The constant I_M does have a weak dependence on ψ, described in [29], that we neglect for simplicity. If I_{T0} is the threshold current of the transistor, and if, at this point, the surface potential is below the source potential by an amount $2\phi_F$, then

$$I_M = I_{T0} e^{-2\phi_F/U_T}. \qquad \text{(A5)}$$

The constant I_{T0} is typically near $\mu C_{ox}(W/2L)(U_T/\kappa)^2$.

If all voltages are referenced to the well (i.e., $v_W = 0$), and we define $\bar{\psi} = -\psi$ and $\overline{v_F} = -(v_G - V_{FB})$, then we can show that

$$\bar{\psi} = \left(\sqrt{\overline{v_F} + \frac{\gamma^2}{4}} - \frac{\gamma}{2}\right)^2. \qquad \text{(A6)}$$

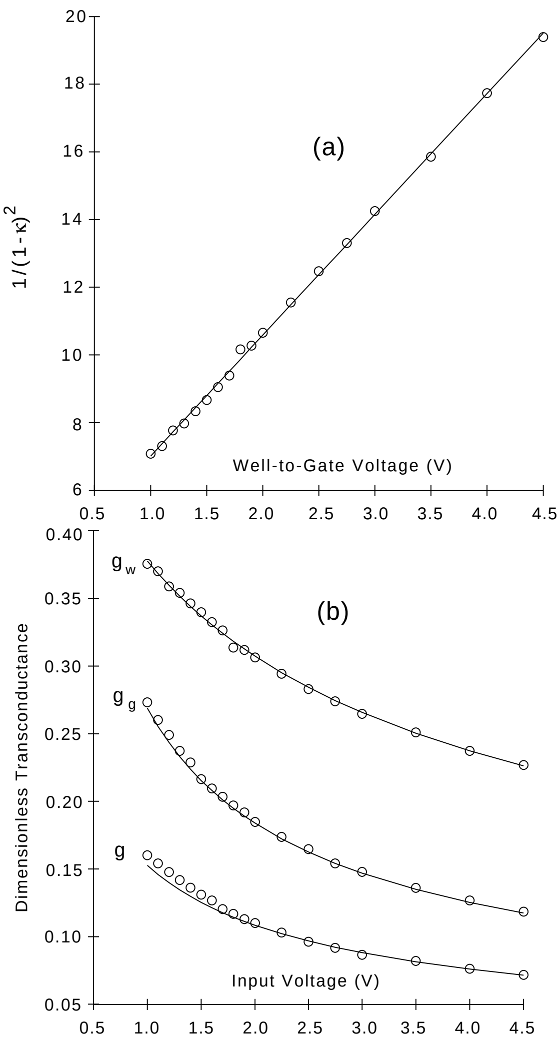

Figure A.1. The effects of changes in κ. (a) The changes in κ with well-to-gate voltage may be used to extract the body-effect parameter γ and the flatband voltage V_{FB}. The slope of the graph yields information about γ, and the intercept then yields information about V_{FB}. See Section A.1 for details. (b) Data for the change in transconductance of well-input amplifiers with no degeneration (g_w), with gate degeneration (g_g), and with gate and source degeneration (g). The solid lines are fits to theory.

We introduce the definitions of $\bar{\psi}$ and $\overline{v_F}$ because it is more convenient to work with $-\psi$ and $-(v_G - V_{FB})$ when dealing with transistors in the well.

The source current of a well transistor is split into an MOS component, called i_M, which reaches the drain of the transistor and a bipolar component, called i_B, which is shunted away to ground. The bipolar current is given by

$$i_B = I_B e^{\left(\frac{v_S - v_W}{U_T}\right)}, \tag{A7}$$

where I_B is the saturation current of the bipolar. The MOS current is given by Eq. (40).

The question that we now ask is this: When does the MOS current exceed the bipolar current ($i_M \geq i_B$)? The answer to this question provides insight into how a well transistor must be operated if it is to have as wide a range of MOS operation as possible. We notice that the MOS and bipolar transistors have the same dependence on the source voltage, v_S. Thus, in subthreshold, the answer is independent of the source voltage. The MOS pre-exponential factor, I_M, is usually 1000 to 10000 times smaller than the bipolar pre-exponential factor I_B. Thus, if the MOS transistor is to have any hope of competing with the bipolar transistor, its surface potential must be below that of the well by the amount that compensates for its pre-exponential handicap. Hence, the gate-to-well voltage must be below the flatband voltage by an amount needed to generate an adequate depth of depletion region. We now compute exactly how much this amount must be.

If $i_M \geq i_B$, then, from Eqs. (A7) and (A4), we must have

$$\begin{aligned} v_W - \psi \;&\geq\; U_T \ln\left(\frac{I_B}{I_M}\right) \\ &\geq\; V_{BM}, \end{aligned} \tag{A8}$$

where V_{BM}, defined as in the previous equation, is a voltage that yields a measure of by how much the bipolar's pre-exponential constant exceeds the MOS's. Thus, if we reference all voltages to the well, Eq. (A8) yields

$$\begin{aligned} -\psi \;&\geq\; V_{BM}, \\ \bar{\psi} \;&\geq\; V_{BM}, \\ \left(\sqrt{\overline{v_F} + \frac{\gamma^2}{4}} - \frac{\gamma}{2}\right)^2 \;&\geq\; V_{BM}. \end{aligned} \tag{A9}$$

After we perform simple manipulations on the previous equation, we finally obtain the criterion for the MOS transistor to dominate over the bipolar transistor:

$$(v_G - v_W) \;\leq\; V_{FB} - \sqrt{V_{BM}}\left(\sqrt{V_{BM}} + \gamma\right), \tag{A10}$$

$$\leq\; V_{CT} \tag{A11}$$

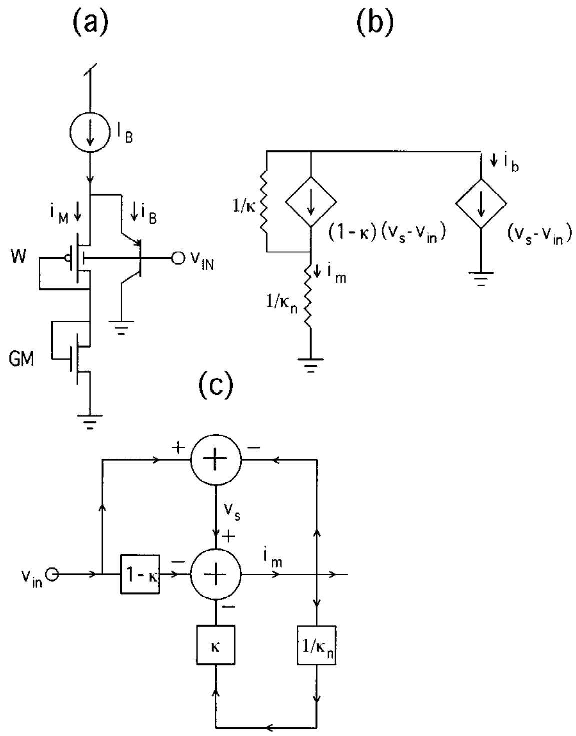

Figure A.2. The parasitic bipolar half circuit. (a) A simplified half circuit for our amplifier that is important for understanding the bipolar–mos interaction in it. (b) The small-signal equivalent circuit for (a). (c) A signal-flow diagram for the small-signal equivalent circuit.

where we define V_{CT} to be the RHS of (A10). If we require $i_M \geq 100 i_B$, for robust operation, we must increase V_{BM} by $U_T \ln 100$ before using Eq. (39). We now have a recipe for operating well transistors in subthreshold: As long as we ensure that the gate is sufficiently below the well, we are in no danger of turning on the bipolar transistor. *Thus, if the range of MOS operation is to be as wide as possible, the gate voltage must be set to as low a value as circuit requirements will allow*. The well is then free to operate from low values near the gate voltage to V_{DD}. Therefore, when the well is being used as an input, the gate should be at ground or near ground. In our amplifier, it is one diode drop above ground.

To understand the effects of the parasitic bipolar transistor in our amplifier, we need only to understand the half-circuit shown in Figure A.2a, because the bipolar transistor(s) in the arm(s) of our amplifier is (are) activated if the voltage of either input gets too low. It is thus simpler to analyze the effects of the bipolar transistor in a single arm of the amplifier, and to extend the analysis to the case where the bipolar transistors in both arms are activated. Hence, we tie one input to V_{DD} to turn off completely one arm of the amplifier, and we concentrate on the turned-on arm.

The circuit of Figure A.2a is one half of the differential pair of the amplifier with the S transistor omitted, and the bipolar transistor drawn explicitly. The S transistor is omitted because it affects the drain voltage of the bias transistor (the bias-current transistor not shown), but has no effect on the currents i_B and i_M, which are our primary concern. Figure A.2b and Figure A.2c show the small-signal model and corresponding signal-flow diagram that describe the half-circuit of Figure A.2a.

When the well-input voltage v_{IN} is in the 1V to 5V range, the gate voltage, v_G, which is one diode drop above ground, is sufficiently below the well voltage that all the bias current is carried by the MOS transistor. As the well voltage begins to drop, it starts to get close to the gate voltage and the bias current starts to be diverted to the bipolar transistor. The data of Figure A.3a show that a differential-pair–like competition between the MOS and bipolar transistors yields a sigmoid curve in output current versus input voltage. The data curves of Figure A.3 were normalized by their saturation current values, and were fit by sigmoid equations of the form

$$i_N = \frac{1}{1 + e^{-0.326(v_{IN} - V_H)/U_T}}, \tag{A12}$$

with differing values of V_H for the three curves.

The input voltage at which the bipolar and MOS currents become equal, V_H, depends on the bias current. Higher bias currents drop larger voltages across the GM transistor, increase the gate voltage of the W transistor, and consequently cause the well-input voltage to approach the gate voltage at higher values. The data of Figure A.3b show our measurements of this effect, and also show that

$$V_H = 0.81\,(V_{DD} - V_B) + 0.019. \tag{A13}$$

Given the theory that we have developed for the bipolar–mos interaction (Eq. (39)), and the effects of changes in κ (Eq. (40)), we can predict what Eqs. (A12) and (A13) should be by analyzing the circuit of Figure A.2a. Using measured values for the constants $V_{FB}, \gamma, \kappa_n, \kappa_b$ (the κ of the bias-current transistor), and V_{BM} (determined by measurements of the bipolar and

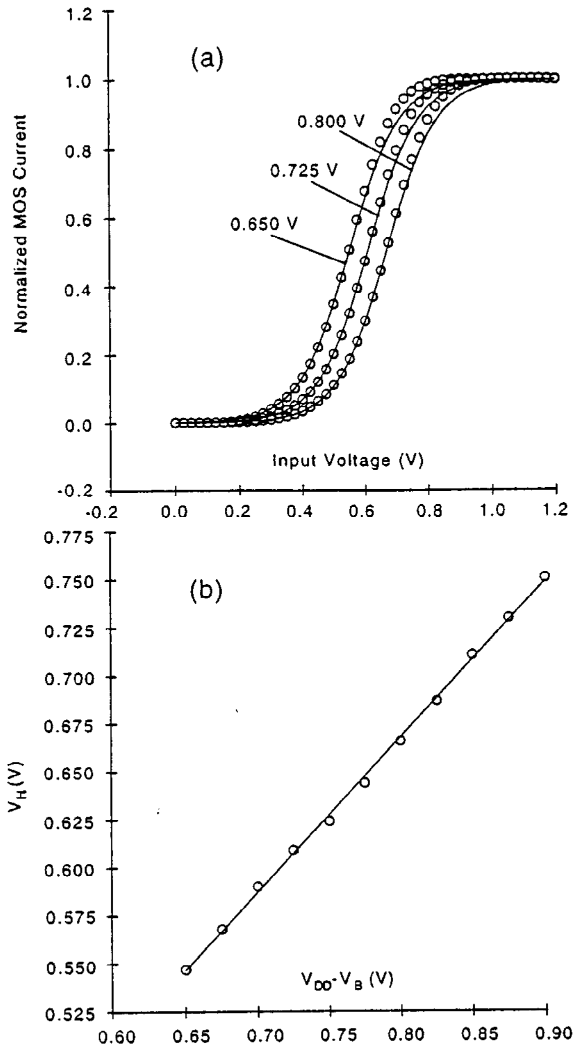

Figure A.3. The bipolar–MOS characteristics. (a) The sigmoid-like compeitition between the MOS and bipolar transistors as the input voltage is varied. The three parametric voltages refer to the value of the bias voltage V_B of Figure 1. (b) The threshold point of this sigmoid varies with V_B.

MOS pre-exponential constants), we were able to obtain reasonable agreement between theory and experiment. From the signal-flow diagram of Figure A.2c, we can show that the constant 0.326 of Eq. (40) is nearly

$$\alpha = 2\frac{(\kappa/2)\kappa_n}{\kappa/2 + \kappa_n}. \tag{A14}$$

Note that we evaluated κ with the surface potential at a value that was V_{BM} below the well voltage. We found that it was 0.467. Using Eq. (39) and elementary reasoning, we can show that Eq. (40) is derived from

$$V_H = \frac{\kappa_b}{\kappa_n}(V_{DD} - V_B) + \frac{U_T}{\kappa_n}\ln\left(\frac{I_0^B}{2I_0^N}\right) - V_{CT}, \tag{A15}$$

where I_0^B and I_0^N are the subthreshold scaling parameters for the bias-current transistor and for the G transistor respectively. We found that $\kappa_b = 0.5615$, $\kappa_n = 0.714$, $I_0^B = 1.12 \times 10^{-17}$ A, $I_0^N = 1.40 \times 10^{-16}$ A, $I_B = 2.12 \times 10^{-16}$ A, $I_{T0} = 1.8 \times 10^{-7}$ A, $2\phi_F = 0.749$ V, $V_{BM} = 0.2165$V, $\gamma = 1.05V^{\frac{1}{2}}$, $V_{FB} = 0.75$ V. The most interesting finding is that V_{BM} = 0.26 V, which implies that the MOS transistor's pre-exponential constant I_M is about 4000 times as weak as the bipolar transistor's pre-exponential constant I_B.

References

1. R. Sarpeshkar, R. F. Lyon, and C. A. Mead, "An analog VLSI cochlea with new transconductance amplifiers and nonlinear gain control," in *Proceedings of the IEEE Symposium on Circuits and Systems*, Vol. 3, Atlanta, May 1996, pp. 292–295.
2. R. Sarpeshkar, R. F. Lyon, and C. A. Mead, "Nonvolatile correction of Q-offsets and instabilities in cochlear filters," in *Proceedings of the IEEE Symposium on Circuits and Systems*, Vol. 3, Atlanta, May 1996, pp. 329–332.
3. H. Khorramabadi and P. R. Gray, "High frequency CMOS continuous time filters." *I.E.E.E. J. Solid State Circuits* SC-19(6), pp. 939–948, 1984.
4. A. Nedungadi and T. R. Viswanathan, "Design of linear CMOS transconductance elements." *I.E.E.E. Trans.* CAS-31(10), pp. 891–894, 1984.
5. Y. Tsividis, Z. Czarnul, S. C. Fang, "MOS transconductors and integrators with high linearity." *Electronics Letters* 22(5), pp. 245–246, 1986.
6. Z. Wang, and W. Guggenbuhl, "A voltage-controllable linear MOS transconductor using bias offset technique." *I.E.E.E. J. Solid State Circuits* SC-22(3), pp. 357–365, 1987.
7. F. Krummenacher and N. Joehl, "A 4 MHz CMOS continuous time filter with on-chip automatic tuning." *I.E.E.E. J. Solid-State Circuits* SC-23, pp. 750–758, 1988.
8. B. Nauta, and E. Seevinck, "Linear CMOS transconductance element for VHF filters." *Electronics Letters* 25, pp. 448–450, 1989.
9. S. T. Dupuie and M. Ismail, "High Frequency CMOS Transconductors," in *Analogue IC Design: The Current Mode Approach* (C. Toumazou, F. J. Lidgey and D. G. Haigh, eds.), Peter Peregrinus Ltd. on behalf of IEEE, 1990.
10. P. Wu, and R. Schaumann, "Tunable operational transconductance amplifier with extremely high linearity over very large input range." *Electronics Letters* 27(14), pp. 1254–1255, 1991.
11. M. C. H. Cheng, and C. Toumazou, "Linear composite MOSFETs (COMFETs)." *Electronics Letters* 27(20), pp. 1802–1804, 1991.

12. S. Szczepanski, J. Jakusz, and A. Czarniak, "Differential pair transconductor linearisation via electronically controlled current-mode cells." *Electronics Letters* 28(12), pp. 1093–1095, 1992.
13. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. John Wiley and Sons: New York, pp. 183–186, 1987.
14. G. Wilson, "Linearised bipolar transconductor." *Electronics Letters* 28(4), pp. 390–391, 1992.
15. W. Chung, K. Kim, and H. Cha, "A linear operational transconductance amplifier for instrumentation applications." *IEEE Trans. on Instrumentation and Measurement* 41(3), pp. 441–443, 1992.
16. J. Ramirez-Angulo and E. Sánchez-Sinecio, "Programmable BiCMOS transconductor for capacitor-transconductor filters." *Electronics Letters* 28(13), pp. 1185–1187, 1992.
17. W. Liu, "An analog cochlear model: Signal representation and VLSI realization," PhD. Thesis, John Hopkins University, Baltimore, Maryland, 1992.
18. L. Watts, D. A. Kerns, R. F. Lyon, and C. A. Mead, "Improved implementation of the silicon cochlea." *IEEE Journal of Solid-State Circuits* 27(5), pp. 692–700, May 1992.
19. X. Arreguit, "Compatible lateral bipolar transistors in CMOS technology: Model and applications." Ecole Polytechnique Federale de Lausanne DSc thesis, These no. 817, 1989.
20. P. Furth and A. G. Andreou, "Linearised differential transconductors in subthreshold CMOS." *Electronics Letters* 31(7), pp. 545–547, 30th March 1995.
21. I. E. Opris and G. T. A. Kovacs, "Large-signal subthreshold CMOS transconductance amplifier." *Electronics Letters* 31(9), pp. 718–720, 27th April 1995.
22. R. F. Lyon, "Analog implementations of auditory models." *DARPA Workshop on Speech and Natural Language*. Morgan Kaufmann Publishers: San Mateo CA, 1991.
23. T. Delbrück, "Bump circuits for computing similarity and dissimilarity of analog voltages," CNS Memo 26, California Institute of Technology, May 24, 1993.
24. R. W. Landee, D. C. Davis, and A. P. Albrecht, *Electronic Designers Handbook*. McGraw-Hill Book Co.: N.Y., 1957, pp. 3–18, pp. 3–34 to 3–38.
25. C. A. Mead, *Analog VLSI and Neural Systems*. Addison-Wesley Publishing Co.: Reading, MA, 1989, pp. 33–36, 67–82.
26. R. F. Lyon, T. Delbrück, and C. A. Mead, "Circuits for wide input range analog rectification and correlation," U.S. Patent 5,319,268, June 7th. 1994.
27. R. Sarpeshkar, T. Delbrück, and C. Mead, "White noise in MOS transistors and resistors." *IEEE Circuits and Devices* 9(6), pp. 23–29, November 1993.
28. R. F. Lyon and C. A. Mead, "The CMOS grounded-substrate convention," *Caltech Computation and Neural Systems Memo* #27, 1993.
29. Y. Tsividis, *Operation and Modeling of the MOS Transistor*. McGraw-Hill Book Company: New York, 1987, pp. 136–140.
30. P. Hasler, B. A. Minch, C. Diorio, and C. A. Mead, "An autozeroing amplifier using *p*FET hot-electron injection," in *Proceedings of the IEEE Symposium on Circuits and Systems*, Vol. 3, Atlanta, May 1996, pp. 325–328.
31. P. Hasler, Personal communication.
32. B. A. Minch, C. Diorio, P. Hasler, and C. A. Mead, "Translinear circuits using subthreshold floating-gate MOS transistors." *Analog Integrated Circuits and Signal Processing* 9(2), pp. 1–13, 1996.

Rahul Sarpeshkar double majored in Electrical Engineering and Physics at MIT between the years of 1986 and 1990. He interned at the T.J. Watson Research Center, IBM as an undergraduate. Since 1990 he has been a PhD student in Carver Mead's analog VLSI lab at the California Institute of Technology. There, he has done work on noise in devices, circuits, and systems, on adaptive silicon cochleas, and in visual motion detection. His interests lie in analog circuit design for sensory systems, and in the fundamental limits of computing due to physics. He has authored seventeen publications and holds one patent. His bachelor's thesis won an award at MIT. He is a member of Tau Beta Pi and Sigma Pi Sigma.

Richard F. Lyon received the B.S. degree in engineering and applied science from California Institute of Technology in 1974 and the M.S. degree in electrical engineering from Stanford University in 1975. He has worked on a variety of projects involving communication and information theory, digital system design, analog and digital signal processing, VLSI design and methodologies, and sensory perception at Caltech, Bell Labs, Jet Propulsion Laboratory, Stanford Telecommunications Inc., Xerox PARC, and Schlumberger Palo Alto Research. In his current position as Distinguished Scientist at Apple Computer, he does research in handwriting recognition, neural networks, and models of hearing. Dick is also in his 13th year on the Computation and Neural Systems faculty at Caltech, where he works on sensory modeling research and analog VLSI techniques for signal processing.

Carver A. Mead, Gordon and Betty Moore Professor of Engineering and Applied Science, has taught at the California Institute of Technology for more than 30 years. He has contributed in the fields of solid-state electronics and the management of complexity in the design of very large scale integrated circuits, and has been active in the development of innovative design methodologies for VLSI. He has written, with Lynn Conway, the standard text for VLSI design, *Introduction to VLSI Systems*. His recent work is concerned with modeling neuronal structures, such as the the cochlea and retina using analog VLSI systems. His newest book on this topic, *Analog VLSI and Neural Systems*, was published in 1989 by Addison-Wesley. Professor Mead is a member of the National Academy of Sciences, the National Academy of Engineering, the American Academy of Arts and Sciences, a foreign member of the Royal Swedish Academy of Engineering Sciences, a Fellow of the American Physical Society, a Fellow of the IEEE, and a Life Fellow of the Franklin Institute. He is also the recipient of a number of awards, including the Centennial Medal of the IEEE.