LOW-POWER SILICON NEURONS, AXONS, AND SYNAPSES

John Lazzaro and John Wawrzynek
Computer Science Division
UC Berkeley
Berkeley, CA, 94720

Power consumption is the dominant design issue for battery-powered electronic devices. Biologically-inspired sensory preprocessors may be an important component of portable computing devices that require real-world visual or auditory input. The silicon neural design style presented in (Mead, 1989; Andreou et al., 1991) naturally supports low-power VLSI design; in this design style, MOS transistors typically operate in the weak-inversion regime. The low-power performance of this design style is outstanding; for example, (Watts et al., 1991) reports on a 51-stage silicon cochlea, that computes all outputs in real time and consumes 11 μW.

However, several popular circuits in this design style operate transistors outside the weak-inversion regime. These circuits, that model the spiking behavior of the axon hillock and the pulse propagation of axons, dominate power consumption in many neural chips (Lazzaro and Mead, 1989ab; Lazzaro, 1991; Horiuchi et al, 1991).

This chapter describes modified versions of these axon circuits. The modified circuits have been designed and fabricated and are fully functional; these circuits show a measured improvement in power consumption over the original circuits. Power consumption decreases of a factor of 10 to 1000 have been measured, depending on pulse width and spiking frequency. The density of the modified circuits is comparable to the original circuits. This chapter also describes several low-power synaptic circuits.

AXONAL CIRCUITS

Figure 1 shows the spiking neuron circuit from (Mead, 1989), that uses a high-gain voltage amplifier with a sigmoidal nonlinearity as a gain element. The circuit converts the unidirectional current I_i into a sequence of fixed-width, fixed-height voltage pulses of V_o. During a pulse $V_o = V_{dd}$, and between pulses V_o is at ground potential.

To understand circuit operation, consider the circuit condition after an output pulse has completed. In this state, V_o is at ground potential, and V_c is lower than the switching threshold of the amplifier. The discharge path of the state capacitor C is closed, and the charging path of C is open.

The circuit remains in this state until the input current I_i increases V_c to the switching threshold of the amplifier. At this point, V_o switches to V_{dd}. The feedback capacitor C_f ensures the secure switching of the circuit. The new value of V_c is above the switching threshold of the amplifier, and depends on the relative values of C_f and C.

Once a pulse begins, the discharge path of the state capacitor C is open, and the charging path of C is closed. The control voltage V_p sets the discharge rate of C, and thus the width of voltage pulse of V_o. The circuit remains in this state until V_c decreases to the switching threshold of the amplifier. At this point, V_o switches to ground potential, and the pulse is complete. The voltage V_c is reset to a value below the switching threshold of the amplifier, that depends on the relative values of C_f and C.

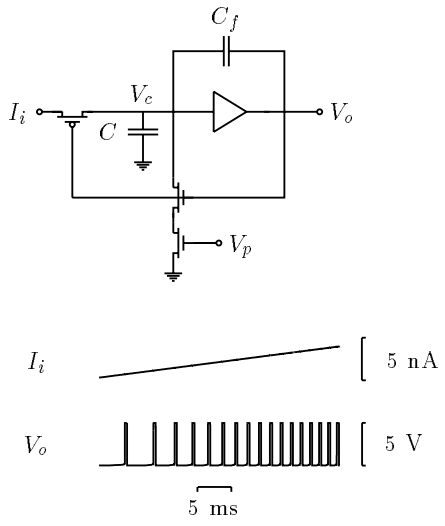

Figure 1. Spiking neuron circuit and function, with unidirectional current input I_i, voltage pulse output V_o, and pulse width control voltage V_p.

This circuit, as described in (Mead, 1989), uses two digital inverters in series as the high-gain amplifier; Figure 2(a) shows this amplifier implementation. When used in the spiking neuron circuit, the transistors in first inverter of this amplifier are biased outside the weak-inversion regime. In many designs, the static current consumption of these transistors dominates the current consumption of the chip.

Figure 2(b) also shows a low-power implementation of a high-gain amplifier suitable for use in the spiking neuron circuit. The control voltages K_1 and K_2 limit the static current consumption of the amplifier. The response time of the amplifier is not symmetric; the speed of crossing the amplifier threshold in a positive direction is not limited by K_1 and K_2, but the speed of crossing the threshold in a negative direction is directly dependent on K_1 and K_2. In many applications, this asymmetry allows the bias currents of the amplifier to be set in the weak-inversion regime. The diode-connected transistor acts to raise the switching threshold of the amplifier.

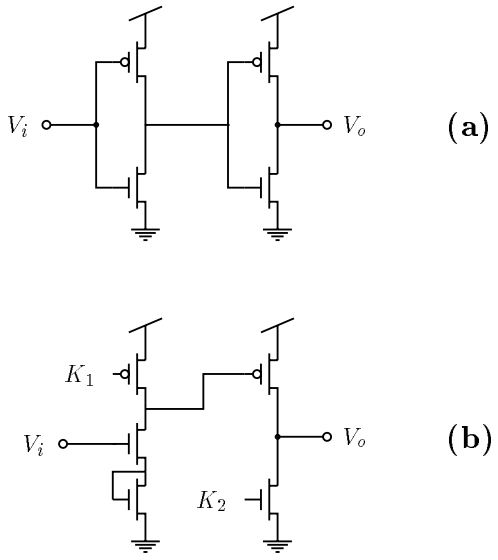

Figure 2. Original gain stage (a) and low-power gain stage (b). K_1 and K_2 are control voltages, set to ensure all transistors operate in the weak-inversion regime.

AXONAL EXPERIMENTAL DATA

Figure 3 shows the static current consumption of the two amplifiers shown in Figure 2, as a function of the input voltage V_i. This figure shows data from a test chip fabricated in the 2μm double polysilicon n-well Orbit process as supplied by MOSIS. The current meter used in this measurement is not able to measure currents below 5 nA. As expected, the low-power amplifier consumes negligible current below its switching threshold, and a constant current above its switching threshold.

Figure 4 shows the current consumption of the spiking neuron circuit of Figure 1, implemented with the original amplifier and the low-power amplifier. As expected, the current consumption of the low-power circuit is a linear function of spiking frequency, and increases with the size of the pulse width. All data was taken with the values of K_1 and K_2 necessary for correct circuit function with 10μs pulses; the current consumption for longer pulse-width operation can be reduced by adjusting K_1 and K_2.

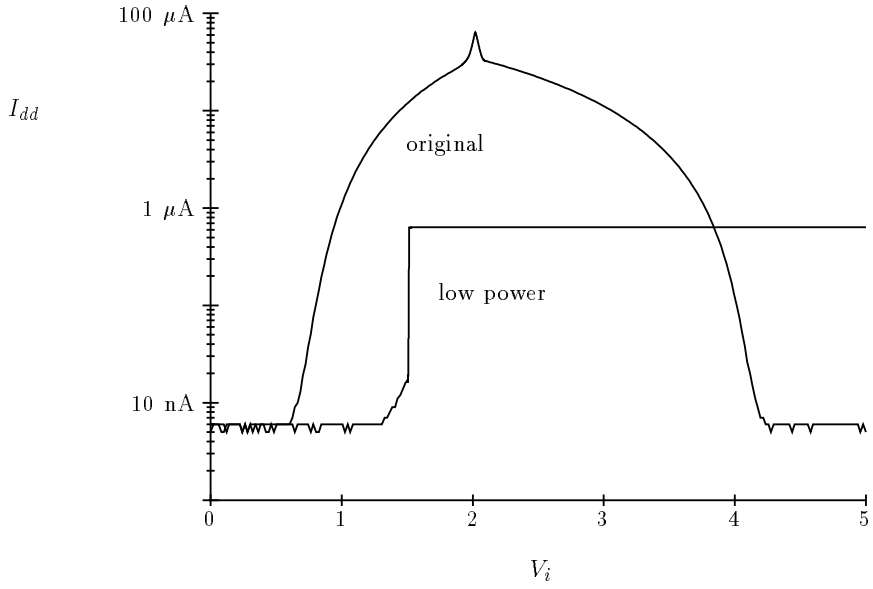

Figure 3. Power supply current I_{dd} for original gain stage and low-power gain stage, as a function of V_i. Note log scale for current.

THE AXONAL DELAY CIRCUIT

Figure 5 shows one section of the axonal delay line circuit described in (Mead, 1989). In engineering terms, the circuit is a cascade of non-retriggerable monostables, each with a voltage output V_k, that is either at V_{dd} or at ground potential. To understand circuit operation, consider the condition $V_k = V_{dd}$, $V_{k+1} = 0$, $F_{k+1} = 0$. In this case, the voltage V_c as shown in Figure 5 is below the switching threshold of the amplifier. The discharge path of the capacitor C is closed, and C is charged by a current set by the voltage $V_k - V_p \equiv V_{dd} - V_p$. When the voltage V_c increases to the switching threshold of the amplifier, V_{k+1} changes to V_{dd}, and the axon stage associated with V_{k+2} begins a similar charging cycle. Once this cycle has completed, the voltage F_{k+1} switches to V_{dd} and the capacitor C associated with V_{k+1} quickly discharges, causing the voltage V_{k+1} to switch back to ground. In this way, a constant width voltage pulse propagates down the structure.

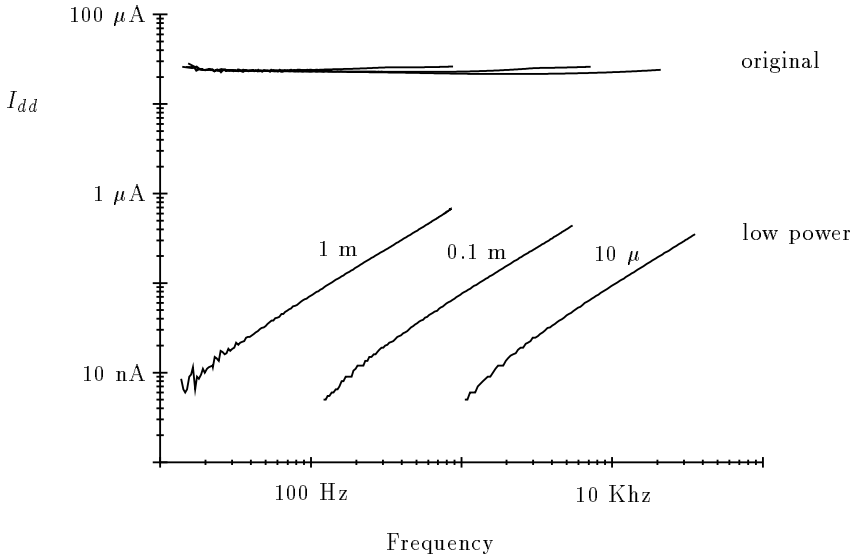

Figure 4. Power supply current I_{dd} for original spiking neuron circuit and low-power spiking neuron circuit, as a function of spiking frequency. Labels next to graphs indicate pulse width of spikes, in units of seconds. Note log scale for frequency and current.

As with the spiking neuron circuit, the amplifier in each neuron circuit, if implemented with two digital inverters in series, consumes appreciable static current. Replacing the original amplifier with the low-power amplifier reduces the current consumption of an axonal delay circuit. Figure 6 shows the current consumption of an 18-stage axonal delay circuit, implemented with the original amplifier and the low-power amplifier.

For pulse widths of 100 μs and 1 ms, current consumption of the low-power axon circuit is a linear function of spiking frequency. For a pulse width of 10 μs, the state capacitors C are not fully discharged with a single pulse, and some additional static current consumption occurs. As with the spiking neuron circuit, the current consumption of the low-power axon circuit also depends on pulse width length. All data was taken with the values of K_1 and K_2 necessary for correct circuit function with 10μs pulses; the current consumption for longer pulse width operation can be reduced by adjusting K_1 and K_2.

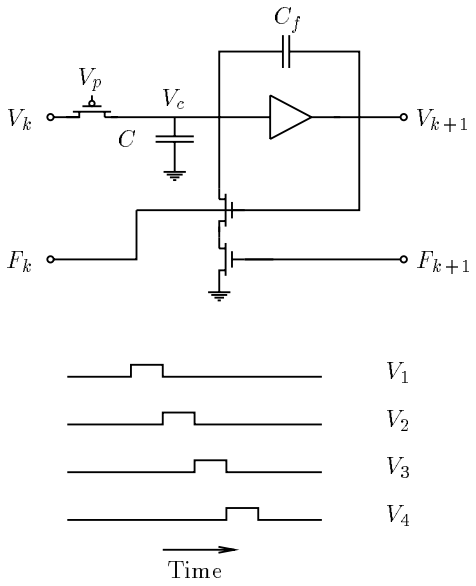

Figure 5. Axon circuit and function. Labels include ports V_k and F_k to previous stage, ports V_{k+1} and F_{k+1} to next stage, pulse width control voltage V_p, state voltage V_c, state capacitor C and feedback capacitor C_f.

The original axon circuit has a single free parameter, V_p, that sets both the speed of pulse propagation and the width of each pulse. The low-power axon circuit provides two additional parameters, K_1 and K_2. These parameters permit the pulse width and propagation speed to be set independently, allowing overlapping pulses in successive taps.

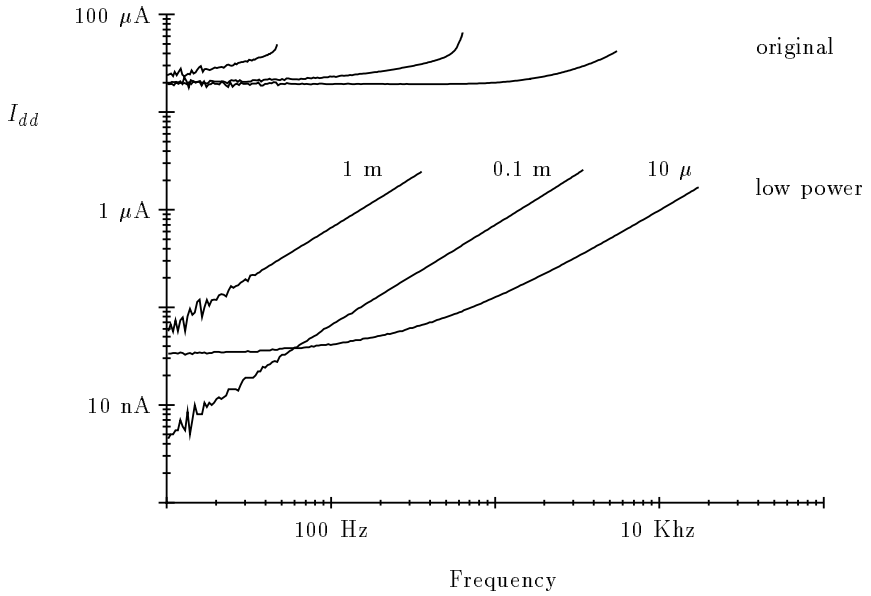

Figure 6. Power supply current I_{dd} for original axon circuit and low-power axon circuit, as a function of spiking frequency (18 sections). Labels next to graphs indicate pulse width of spikes, in seconds. Note log scales for current and frequency.

SYNAPTIC CIRCUITS

This section describes several different types of low-power synaptic circuits. Referring to Figure 1, synaptic circuits convert the pulse output representation of the signal V_o into a unidirectional weak-inversion current signal suitable for connection in I_i. Different synaptic circuits perform different types of signal processing during the conversion. All of the circuits in this section have been fabricated as components in functional systems, except where indicated.

Figure 7 shows simple synaptic circuits. The circuit in Figure 7(a) (Mead, 1989) is a very simple synapse, converting a downward voltage pulse into the unidirectional current output I_o. The magnitude of the current pulse is set by

the control voltage V_w, and the width of the current pulse reflects the width of the input voltage pulse.

In some situations, the width of the input voltage pulse is very small and not under voltage control; in particular, pulse outputs from spiking neuron circuits used in off-chip communication have this property. (Lazzaro et al., 1993). The synapse circuit shown in Figure 7(b) is designed to produce an output current pulse I_o that is wider than the input voltage pulse. The control voltage V_τ sets the width of I_o, and the control voltage V_w sets the magnitude.

The circuits in Figure 7 allows voltage pulses to be multiplied by a weight set by V_w, and summed together using Kirchoff's current law. Circuits shown in Figure 8 are useful for performing signal processing on this aggregate current signal. These circuits transform a unidirectional input current I_i into a unidirectional output current I_o.

The circuit in Figure 8(a) includes two control voltages, V_1 and V_2, which can be used together to scale I_i by factors greater than or less than unity. If all transistors are operating in the weak-inversion regime, the equation

$$I_o = I_i e^{(\kappa/2)(V_1 - V_2)/V_o}$$

describes the operation of the circuit, where $V_o = kT/q$ and κ is a fabrication constant as described in (Mead, 1989).

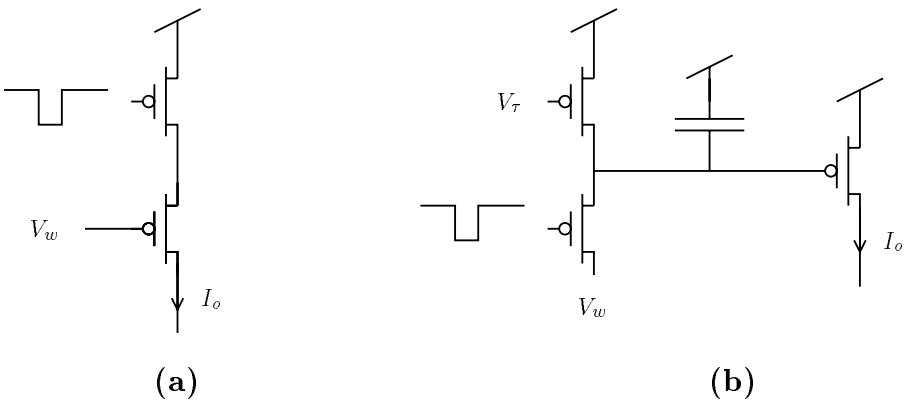

Figure 7. Synapses circuits, producing output currents I_o.

The circuit in Figure 8(b) is similar to the circuit in Figure 8(a), but uses the currents I_1 and I_2 to scale the input current. If all transistors are operating in the weak-inversion regime, the equation

$$I_o = I_i \sqrt{I_1 / I_2}$$

describes the operation of the circuit.

The scaling circuits of Figure 8 can be used as components for circuits that implement adaptation or facilitation. Figure 9(a) shows an adaptation circuit, that has been used in an auditory nerve model (Lazzaro, 1992). The primary input of this circuit is I_i, a current usually obtained by summing the response of many simple synapses. The output of the circuit, I_o, is a scaled version of I_i. If the auxiliary input of this circuit (marked with a pulse) is inactive for an extended period of time, the circuit performs unity scaling. If the auxiliary input is active, however, the output I_o is a reduced version of I_i. The temporal characteristics of the adaptation are set by the control voltages V_u and V_d; depending on the values of these parameters, the circuit can produce time constants as long as several seconds.

Figure 9(b) shows a facilitation circuit that operates on a similar principal to the adaptation circuit of Figure 9(a). In this circuit, auxiliary input activity scales I_o to be larger than I_i; a lack of auxiliary input activity performs unity scaling. An addition control voltage, V_l, limits the maximum scaling of the circuit. This circuit simulates correctly but has not be used in a fabricated system.

The circuits in Figure 9 are both controlled by a single auxiliary input. This input can be replaced by a logical combination of several inputs, to produce a circuit suitable for learning algorithms.

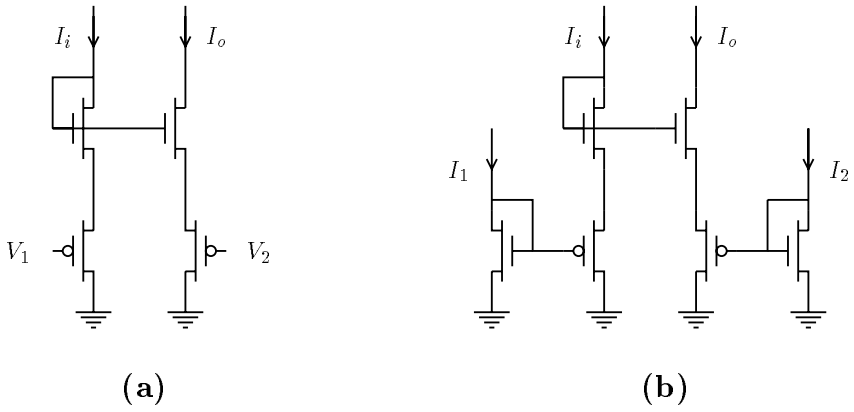

Figure 8. Scaling circuits for currents: $I_o = I_i\ \alpha(I_1, I_2, V_1, V_2)$.

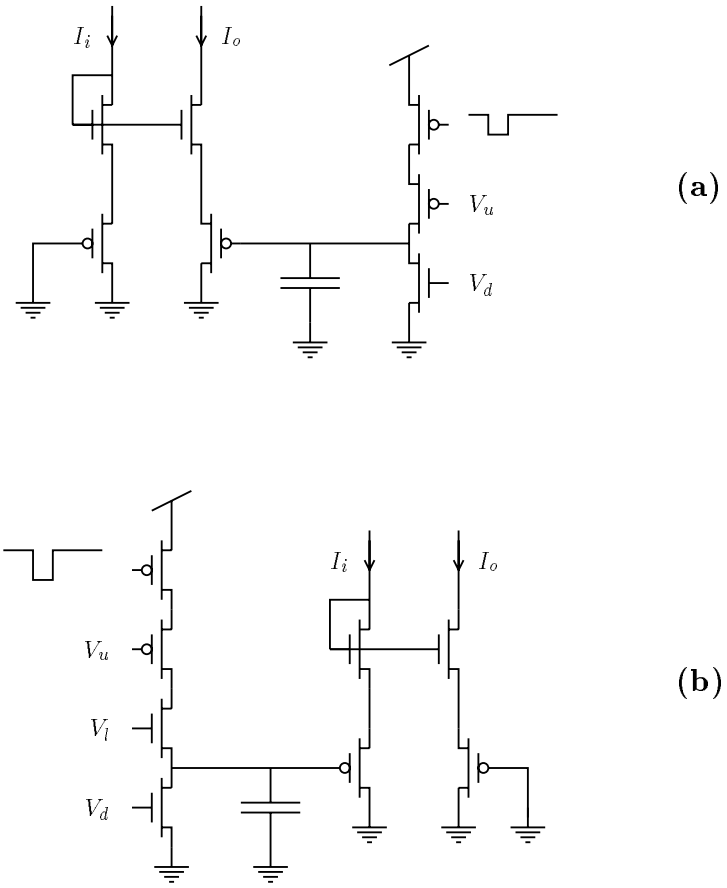

Figure 9. Short-term adaptation and facilitation circuits.

CONCLUSIONS

This chapter reviews low-power circuit technology for spiking neurons, axons, and synaptic circuits. Using the circuits from this chapter, the reader should be able to recast many of the circuits from other chapters of this book into low-power designs.

Acknowledgements

Spiking neuron and axon research was performed at CU Boulder, and funded by the National Science Foundation; all other research performed at UC Berkeley, and was funded by the NSF (PYI award MIPS-895-8568), AT&T, and the ONR (URI-N00014-92-J-1672). Thanks to Carver Mead for his encouragement of this research.

References

Andreou, A.G., Boahen K. A. , and Pouliquen, P. O. (1991). Current-mode subthreshold MOS circuits for analog VLSI neural systems, *IEEE Transactions on Neural Networks,* **2**:2, p. 205.

Horiuchi, T., Lazzaro, J. P., Moore, A., and Koch, C. (1991). A correlation-based motion detection chip, in Tourestzky, D. (ed), *Advances in Neural Information Processing Systems 3,* San Mateo, CA: Morgan Kaufmann Publishers.

Lazzaro, J. P. and Mead, C.A. (1989). Silicon models of auditory localization, *Neural Computation* **1**: 41–70.

Lazzaro, J. P. and Mead, C. (1989). Silicon models of pitch perception, *Proc. Natl. Acad. Sci. USA,* **86**, pp. 9597–9601.

Lazzaro, J. P. (1991). A silicon model of an auditory neural representation of spectral shape, *IEEE Journal Solid State Circuits,* **26**: 772–777.

Lazzaro, J. P. (1992). Temporal adaptation in a silicon auditory nerve, in Moody, J., Hanson, S., and Tourestzky, D. (eds) *Advances in Neural Information Processing Systems 4,* San Mateo, CA: Morgan Kaufmann Publishers.

Lazzaro, J. P., Wawrzynek, J., Mahowald, M., Sivilotti, M., Gillespie, D., (1993). Silicon auditory processors as computer peripherals, *IEEE Journal Neural Networks,* (in press).

Mead, C. A. (1989). *Analog VLSI and Neural Systems.* Reading, MA: Addison-Wesley.

Watts, L., Kerns, D. A., Lyon, R. F., and Mead, C. A. (1992). Improved implementation of the silicon cochlea, *IEEE Journal Solid State Circuits,* **27:** 5, 692-700.