

### 30.3 A 360-Channel Speech Preprocessor that Emulates the Cochlear Amplifier

Bo Wen, Kwabena Boahen

University of Pennsylvania, Philadelphia, PA

Demand for speech recognition in portable devices, such as cell-phones and PDAs, remains unmet. These applications will undoubtedly benefit from a front-end that functions like the biological cochlea: hundreds of digital output channels, similar to the auditory nerve, for fine frequency discrimination and frequency-selective automatic gain control (AGC) to cope with large input dynamic range and combat noise. Mixed analog-digital cochlea-like front-ends (silicon cochleae) have promise in providing these desirable features in real-time at low power.

The first silicon cochlea, developed by Lyon and Mead [1], employed a cascade of second-order low-pass filters (LPFs) with exponentially decreasing resonant frequencies. However, the cascade structure accumulates noise and lacks fault-tolerance. Connecting filter banks in parallel, rather than in series, promised to eliminate these problems inherent in the cascade [2]. However, another problem emerges: destructive interference occurs when outputs are combined due to the large phase change at resonance [3].

To address the limitations of current silicon cochlea designs, inspiration was sought from the biological cochlea, which exhibits exquisite sensitivity and selectivity in detecting and analyzing sound. The cochlea achieves this performance through an active mechanism (often referred to as the cochlear amplifier) that is still not understood. However, its microanatomy provides clues, based on which we previously proposed active bidirectional coupling (ABC) as the amplification mechanism [4]. A silicon cochlea implementation is described here that uses ABC to overcome the limitations of existing architectures. In essence, our architecture is the first to employ negative damping (i.e., active behavior) instead of undamping (i.e., passive behavior).

A version of our novel cochlear architecture was fabricated with two 360×13 diffusive grids and 360 second-order sections, each driving 6 pulse-frequency modulators (PFMs) (Fig. 30.3.1). The second-order sections model the stiffness ( $S$ ), damping ( $\beta$ ), and mass ( $M$ ) of the basilar membrane (BM), the main vibrating organ in the cochlea. The BM interacts with the cochlear fluid, whose motion is governed by Laplace's equation, solved by the diffusive grid [5]. In our current-mode design, a current ( $I_{in}$ ) represents the fluid's velocity potential, whose spatial derivative is the fluid velocity, and another ( $I_{mem}$ ) the BM's velocity (see Fig. 30.3.1).  $I_{in}$  and  $I_{mem}$  are related (with physical analogs) by  $\partial^2 I_{in} / \partial t^2 = S I_{mem} + \beta \partial I_{mem} / \partial t + M \partial^2 I_{mem} / \partial t^2$ . This relation is realized using two interacting first-order LPFs, as described by:  $\tau_1 I_s s + I_s = I_o - I_{in}$ ,  $\tau_2 I_o s + I_o = I_{in} - b I_s$ , and  $I_{mem} = I_{in} + I_s - I_o$ . The physical analogs are thus obtained as  $S = (b+1)/\tau_1 \tau_2$ ,  $\beta = (\tau_1 + \tau_2)/\tau_1 \tau_2$ , and  $M = 1$ , with  $\tau_1$  and  $\tau_2$  increasing exponentially from section to section, to simulate the BM's nonuniform stiffness and damping. For a given choice of  $\tau_1$  and  $\tau_2$  (capacitance limited by silicon area), the gain factor  $b$  increases the quality factor  $Q = \sqrt{(b+1) / \{\sqrt{(\tau_1/\tau_2)} + \sqrt{(\tau_2/\tau_1)}\}}$ .

ABC is included in the cochlea design by exchanging currents between neighboring second-order sections (see Fig. 30.3.1). The second LPF thus becomes  $\tau_2 I_o s + I_o = I_{in} - b I_s - r_{fb}(b+1)T(I_{sb}) + r_{ff}(b+1)T(I_{sf})$ , where  $I_{sf}$  and  $I_{sb}$  are the outputs of the first LPF in the upstream and downstream neighbors, respectively;  $r_{ff}$  and  $r_{fb}$  are

the coupling strength in each direction; and  $T(I_s) = I_s \cdot \text{sat} / (I_s + I_{\text{sat}})$ , accomplished by a current-limiting transistor, which sets the saturation level  $I_{\text{sat}}$  (controlled by  $V_{\text{sat}}$ ).

A log-domain Class AB second-order section that constrains the common-mode dynamically, thus avoiding instability associated with negative feedback correction, was synthesized (Fig. 30.3.2). Following the approach in [6], a first-order LPF is described by  $\tau(I_{out}^+ - I_{out}^-)s + (I_{out}^+ - I_{out}^-) = I_{in}^+ - I_{in}^-$  and  $\tau I_{out}^+ I_{out}^- + I_{out}^+ I_{out}^- = I_q^2$ , where  $I_q$  (controlled by  $V_q$ ) sets the quiescent current. Hence, the positive path's nodal equation is  $CdV_{out}^+/dt = I_\tau(I_{in}^+ - I_{in}^-) - (I_{out}^+ - I_q^2/I_{out}^-)/(I_{out}^+ + I_{out}^-)$ , where  $I_\tau$  (controlled by  $V_\tau$ ) sets the time constant ( $\tau = C u_T / \kappa I_\tau$ ;  $\kappa$  is the subthreshold slope coefficient and  $u_T$  is the thermal voltage). The negative path simply swaps the superscripts + and -. These two paths interact in push-pull. To obtain sinusoidal current as the input to the BM circuit, we set the differential voltages applied to the diffusive grids to be the logarithm of a half-wave rectified sinusoid (with appropriate DC offset).  $I_{mem}$  is encoded by the PFMs, whose digital outputs are transmitted through an address event interface [7].

Tuned to frequencies from 210Hz to 14kHz, the chip exhibits relatively large amplification (Fig. 30.3.3) and compression at high input intensities (Fig. 30.3.4). Four octave-spaced pure tones elicit maximum responses at monotonically increasing channels. Increasing the input current level from 0 to 48dB yields 24dB compression at the peak; frequency tuning becomes broader as well, with  $Q_{10} = 1.84$  (i.e.,  $f_{\text{peak}}$  over  $\Delta f_{10\text{dB}}$ , the width 10dB below the peak) at 0dB and 1.14 at 48dB; phase accumulation remains the same. Thus, ABC sharpens frequency tuning and increases dynamic range. Further, the chip processes natural sounds in real time (Fig. 30.3.5).

A VLSI implementation of a 2D nonlinear cochlear model is presented that utilizes a novel active mechanism, ABC, which amplifies the traveling wave to a degree that decreases with input intensity, thus realizing frequency-selective AGC (Fig. 30.3.6 summarizes the specifications; Fig. 30.3.7 shows the die micrograph). Rather than detecting the wave's amplitude and implementing a control loop, our biomorphic architecture simply employs nonlinear interactions between adjacent neighbors, emulating the cochlear amplifier. In addition to AGC, this cochlear amplifier nonlinearity is thought to suppress weaker tones in favor of stronger ones, thereby enhancing formant perception in noisy environments. These biological features of our silicon cochlea are desirable in speech recognition systems that seek to match biological performance.

#### References:

- [1] R. F. Lyon and C. A. Mead, "An Analog Electronic Cochlea," *IEEE Trans. Acoust. Speech and Signal Proc.*, vol. 36, pp. 1119-1134, 1988.
- [2] L. Watts, "Cochlear Mechanics: Analysis and Analog VLSI," Ph.D. thesis, Pasadena, CA, Caltech, 1993.
- [3] E. Fragniere, "A 100-Channel Analog CMOS Auditory Filter Bank for Speech Recognition," *ISSCC Dig. Tech. Papers*, pp. 140-141, Feb., 2005.
- [4] B. Wen and K. Boahen, "A Linear Cochlear Model with Active Bi-directional Coupling," *IEEE EMBS*, pp. 2013-2016, 2003.
- [5] A. G. Andreou and K. A. Boahen, "Translinear Circuits in Subthreshold MOS," *Journal of Analog Integrated Circuits and Signal Processing*, vol. 9, pp. 141-166, 1996.
- [6] K. Zaghoul and K. A. Boahen, "An On-Off Log-Domain Circuit that Recreates Adaptive Filtering in the Retina," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 1, pp. 99-107, Jan., 2005.
- [7] K. Boahen, "A Burst-Mode Word-Serial Address-Event Channel-I: Transmitter Design," *IEEE Transactions on Circuits and Systems I*, vol. 51, no. 7, pp. 1269-1280, July, 2004.

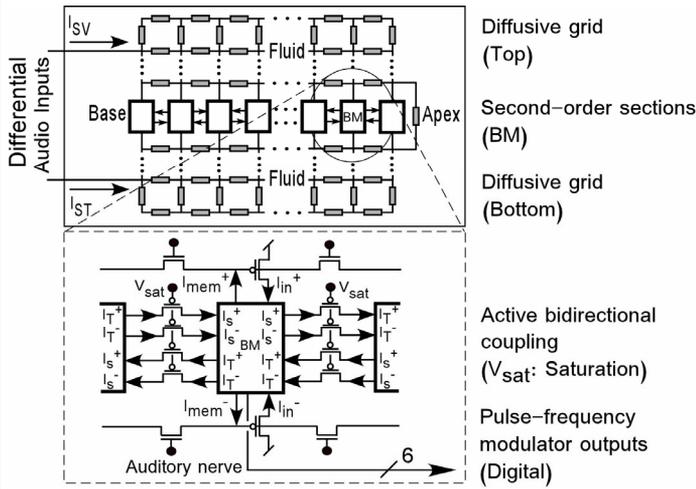


Figure 30.3.1: Cochlear chip architecture.

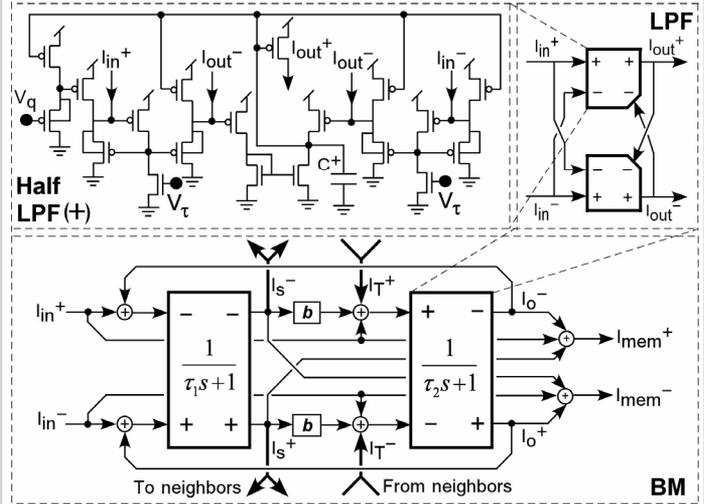


Figure 30.3.2: Basilar membrane (BM) circuit.

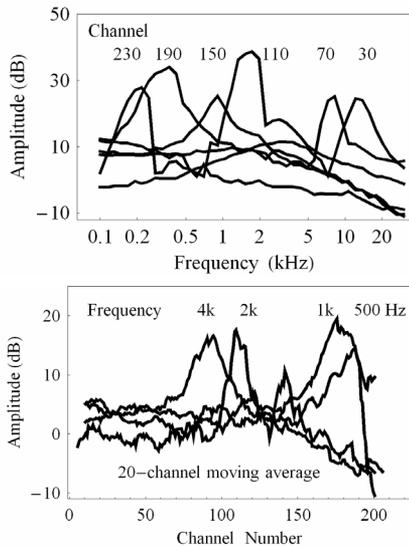


Figure 30.3.3: Top: Frequency responses. Bottom: Longitudinal responses.

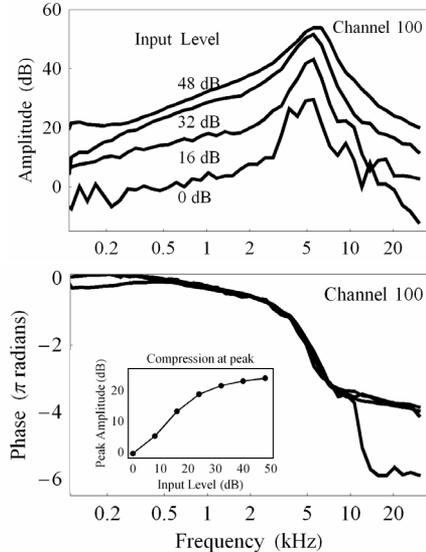


Figure 30.3.4: Automatic gain control.

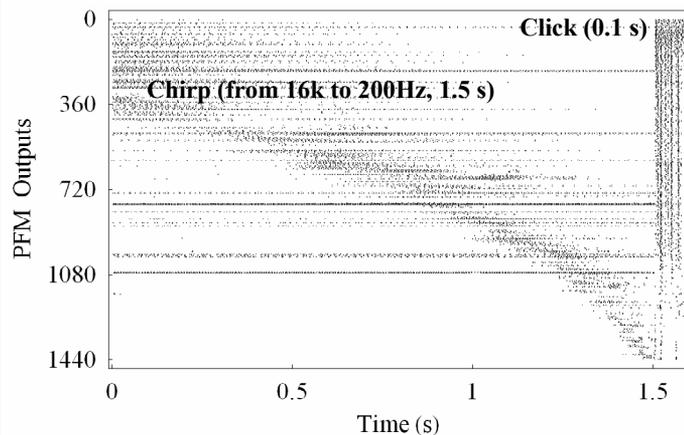


Figure 30.3.5: Cochleagram.

Specification	Values
Technology	0.25μm 1P5M CMOS
Chip die size	3.76 × 2.91mm <sup>2</sup>
Transistor count	115,000
Channel count	360
PFM outputs	2160
Supply voltages	Analog: 2.4V; Digital: 2.5V
Frequency range	210 ~ 14kHz
Dynamic range	17 ~ 42dB
Phase at peak	0.3 ~ 1.4cycles
$Q_{10}$	0.90 ~ 2.73
Cutoff slope	20 ~ 70dB/oct
AGC compression at peak	24dB
Power consumption	52mW

Figure 30.3.6: Chip specifications.

Continued on Page 672

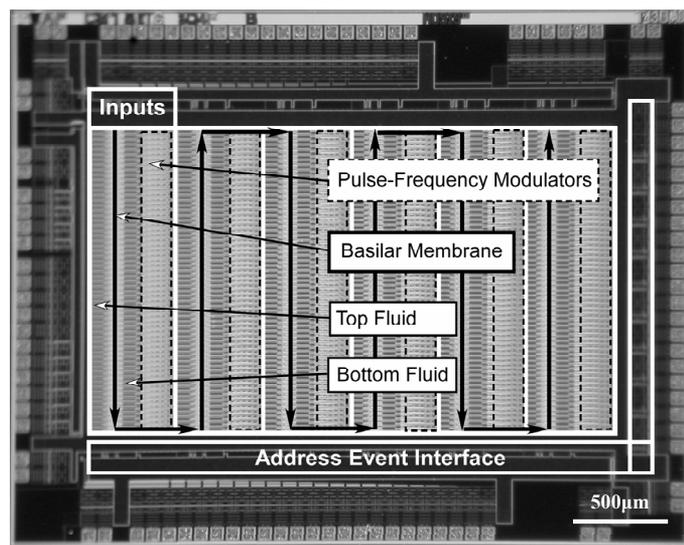


Figure 30.3.7: Cochlear chip micrograph.