Event-Based 64-Channel Binaural Silicon Cochlea with Q Enhancement Mechanisms

Shih-Chii Liu[1], Senior Member, IEEE, *André van Schaik*[2], Senior Member, IEEE,
Bradley A. Minch[3], Member, IEEE, *Tobi Delbruck*[1], Senior Member, IEEE
[1] University of Zurich and ETH Zurich, [2] University of Sydney, [3] Olin College of Engineering

***Abstract*— This paper describes an event-based binaural silicon cochlea aimed at spatial audition and auditory scene analysis. The chip has a matched pair of 64-stage cascaded analog second-order filter banks with 512 pulse-frequency modulated (PFM) address-event representation (AER) outputs. The spectral selectivity is sharpened through 2 different on-chip methods: an on-chip local Q DAC and an on-chip spatial sharpening through nearest neighbour lateral inhibition. The fabricated chip in a 4-metal 2-poly 0.35µm CMOS process consumes peak 25mW power for the digital circuits and 33mW for the analog core. Dynamic range to produce PFM output is 36dB (25mVpp to 1500mVpp at microphone preamp output). Event timing jitter is 2us for 250mVpp input. The peak output bandwidth is 10M events per second (eps) but typical speech scenarios show rates of 20keps.**

Keywords: AER, spike-based, neuromorphic, audition

I. Introduction

Embedded auditory devices with reduced post-processing compute power and sampling rate are desirable for localizing and classifying sound sources for source separation, beam forming, and classifying auditory scenes, e.g. quiet speech, public area, music, party. Conventional auditory signal processing is based on regular sampling of the auditory input signals at the necessary Nyquist frequency. Later processing must also run at this sample rate, at least until auditory 'frames' have been generated. Although parts of this digital processing have been highly optimized (e.g. 64 point FFT in hearing aids), high resolution ADCs running at Nyquist rates place a lower bound on power consumption and the sample rate limits the auditory timing resolution necessary for spatial audition, where a 1° change of angle changes inter-aural delay by at most 6us with microphones spaced by 10cm.

Event-based silicon cochleas such as the one proposed here model the basilar membrane biophysics as a large number of coupled filter stages, followed by rectification and asynchronous output quantization in the timing domain. A binaural event-based silicon cochlea which efficiently extracts spectrally selective auditory timing would be desirable for spatial audition tasks. The event-based cochlea output consists of a sparse stream of digital address-events, representing the addresses of active channels. These events preserve timing between the two ears and post-processing is cheaper because only this sparse stream of events needs to be processed.

Previous address-event representation (AER) silicon cochlea designs offer either only monaural operation [1-4], poor channel matching [1, 5, 6], do not integrate biasing circuits for temperature and process variation tolerance [1, 5], do not integrate microphone preamplifiers [1, 4, 5], or do not include any per-channel calibration capability [1, 5]. None of the prior work has open-sourced host software APIs and algorithms which open application scenarios [7].

The binaural cochlear system proposed here is the first integrated system that combines features of previous silicon cochlea designs that are robust to mismatch, along with novel features for easier programmability of the architecture and operating parameters. The chip includes local adjustment and enhancement of filter sharpness, on-chip digitally controlled biases [8], and prototype on-chip microphone preamplifiers [9]. A bus-powered USB board enables easy interfacing to standard PCs for control and processing.

The rest of this paper first describes the chip architecture, focusing on the improvement of filter sharpness. It then shows characterization measurements and concludes with a brief discussion.

II. Architecture

The binaural chip has two separate 64-stage cascaded filter banks allowing connection to two electret microphones. The architecture of one of the two cochleas on the chip is shown in Fig 1. Each cochlea consists of a 64-stage cascaded filter bank stage. The cascaded architecture [10, 11, 13] is preferred over a coupled bandpass architecture [1, 3, 6, 12] so we can achieve better matching and sharp high frequency roll-off.

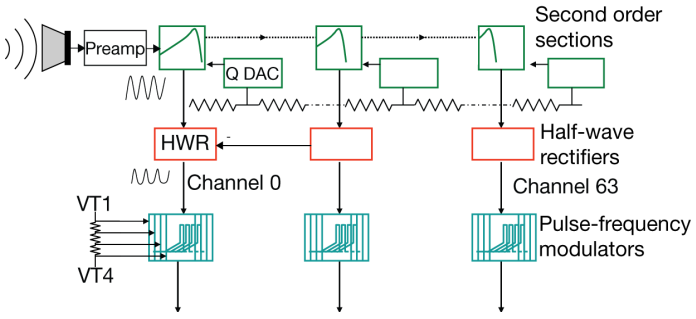

Fig 1 Architecture of 64 channel cochlea with 4 PFMs per channel.

The coupled architecture is particularly susceptible to destructive interference at mismatched stages [1, 3].

We choose a voltage-mode implementation on this chip [10-13] instead of a current-mode implementation [1, 3, 4, 6] because of better robustness to fabrication variances. The impact of the smaller linear input range is reduced by including global automatic gain control (AGC) on the front of the filters, using off-chip microphone preamplifiers. The voltage-mode filter bank implementation also reduces variability compared with log-domain current-mode implementations which are very susceptible to current copying mismatch [1, 3]. However, noise accumulation and time delay along the cascade favor a small number of sections per octave, making it harder to maintain high Q. But maintaining acceptably high Q is important for spectral selectivity and is why this chip includes Q adjustment and lateral suppression circuits, as will be explained.

Each filter stage (Fig 2) consists of a second-order-section (SOS) filter which is biased by a Complementary Lateral Bipolar Transistor (CLBT) ladder to improve matching [11]. A differential readout of each SOS output drives its own half-wave rectifier (HWR) circuit, and the HWR output drives 4 pulse-frequency modulators (PFMs). The PFM circuits implement an integrate-and-fire model with a threshold (VT). The four PFMs have individual global thresholds (VT1 to VT4), allowing volume encoding by selective activation of PFMs. Compared with regularly-sampled audio systems, the PFM outputs are transmitted asynchronously, reducing latency to the analog delay along the filter bank and increasing temporal resolution to microseconds.

In the cascaded architecture, using a defined number of sections per octave and at a particular input amplitude, the Q's of the filters can only be increased up to a limit without introducing nonlinearities due to amplifier saturation [13]. To sharpen the Q of the filter output, we have incorporated two methods which are based on on-chip circuits.

The first method consists of a local 5 bit DAC based on a current splitter (Fig 2). The DAC output current goes to a diffuser circuit. The resulting diffused current I_q is used to set the fraction αI_q current of the local I_Q (from the CLBT network) that biases the feedback amplifier in the SOS, allowing local Q adjustment.

The second method uses a nearest-neighbor lateral inhibition scheme. The input to the n'th HWR circuit is the difference of the outputs of the forward transconductance amplifiers A_1 and A_2:

$$V_{\text{out}}(s) - V_1(s) = \frac{\tau_n s}{1 + \tau_n s/Q + (\tau_n s)^2} V_{\text{in}}(s)$$

where $Q=1/(2-g_Q/g_\tau)$ and $\tau_n=C/g_\tau$. g_τ and g_Q are the transconductances of amplifiers $A_{1,2}$ and A_3 respectively. This differential readout adds a desirable zero to the transfer function without introducing undesirable gain proportional to frequency, as would occur with a temporal high pass filter [11]. The output of the half wave-rectified circuit at first approximation is given by

$$I_{\text{HWR}} = \max\left(0, g_m\left(V_1 - V_{\text{out}}\right) + I_{\text{off}}\right)$$

where g_m is controlled by V_{gain}. The next stage's I_{HWR} is subtracted from I_{HWR} to produce I_{IHC} which drives the PFMs:

$$I_{\text{IHC}} = \max\left(0, g_m\left(\tau_n s - \frac{\tau_{n+1} s}{1 + \tau_{n+1} s/Q + (\tau_{n+1} s)^2}\right) V_{\text{out}}(s)\right)$$

This subtraction of the downstream section (which is tuned to slightly lower best frequency) suppresses the 20dB/decade rising response of I_{HWR} to its peak, further sharpening the response.

III. Implementation

The AEREAR2 chip was fabricated in 0.35um 4M 2P CMOS (Fig 3). A prototype bus-powered USB board based on [15] with integrated microphones (Fig 4) interfaces to

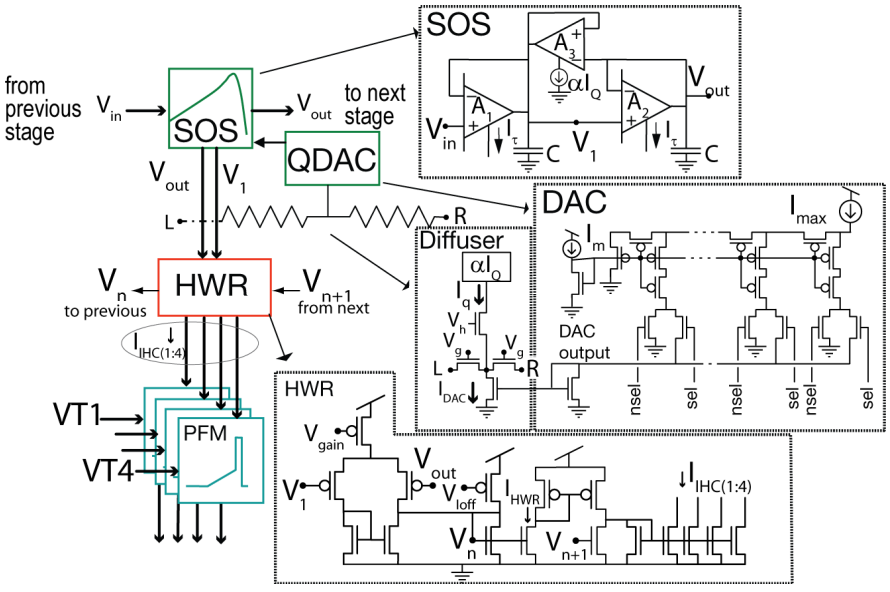

Fig 2 Single stage circuit details including Q adjustment by local DAC and HWR circuit. $\alpha=I_q/(I_q+I_T)$, where I_T is another globally adjustable bias current. Cascodes in mirrors and differential pairs omitted.

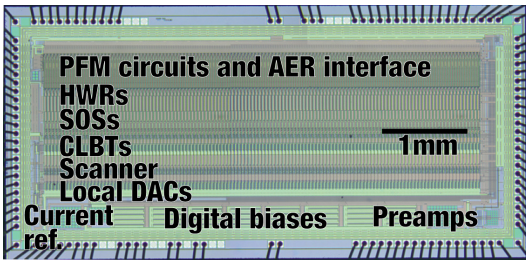

Fig 3 Die photo of AEREAR2.

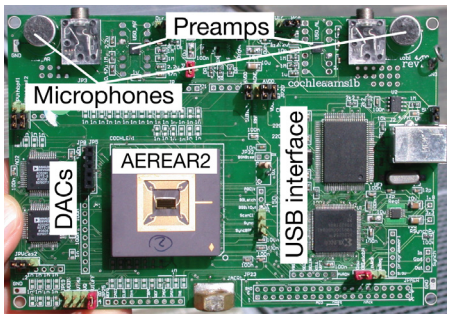

Fig 4 Prototype USB board.

jAER, an open-source software project for processing AER output [7]. The time-stamped events are sent to a PC where they are further processed digitally for applications. A mistake in the design of the on-chip preamps made them unusable. Off-chip MAX9814 microphone preamplifiers with AGC were used for natural sounds and input was applied from a PC sound card after the preamplifiers for analysis of channel responses.

IV. Characterization

In contrast to previous work, the results in this paper are measured exclusively from the final PFM outputs of the chip. The PFM output addresses are transmitted asynchronously off-chip using the AER protocol [16]. The USB interface time-stamps the events to 1us resolution.

Fig 5 shows event rasters in response to speech at a distance of 1m from the PCB at normal speech volume (65dB LAF SPL, measured with Bruel & Kjaer 2250). Each dot is one event and ears are different colors. The mean event rate is 17keps.

Fig 6 shows the raw PFM outputs of the 64 channels of the two cochleas in response to a frequency-swept chirp. All channels respond to only a limited frequency range of the chirp.

Fig 7 shows measurements of frequency responses of 3

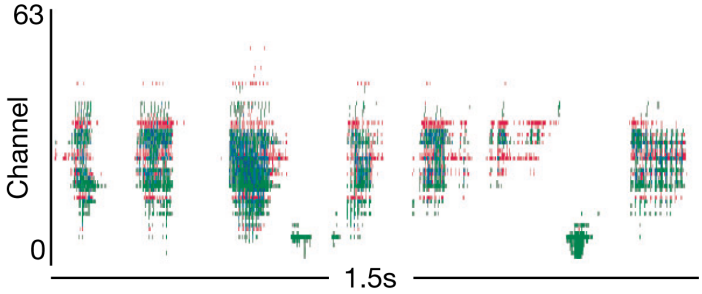

Fig 5 Response to speech "The quick red fox jumped over the lazy dog".

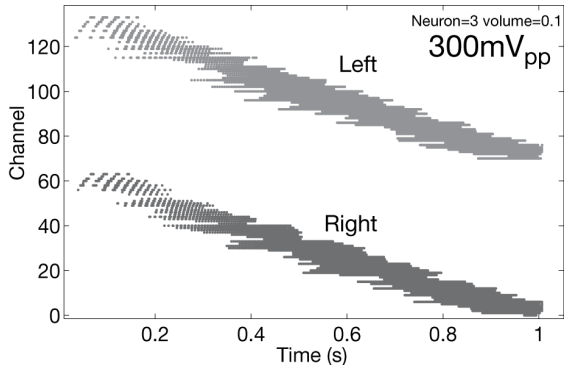

Fig 6 Event rasters recorded from the 64 channels of both ears. Frequency is logarithmically swept from 30Hz to 10kHz with input amplitude 300mV$_{pp}$.

selected channels for 7 different sound volumes. Channel responses broaden with volume as more PFMs go above threshold, but the upper cut-off frequencies and roll-off slopes do not change.

From these responses, we can extract the best characteristic frequency (BCF) and quality factor (Q) of each filter (Fig 8). The BCFs are logarithmically distributed over the 64 channels, and for the 39 of 64 channels where both ears respond to the 150mVpp input, the BCFs are matched to ±16% between the two cochleas. The Qs are computed as (BCF/width at 0.7 of BCF) and are matched between ears to ±27% at an input level of 450mV$_{pp}$; Fig 7 shows that Q strongly depends on sound volume. As of yet, no local QDAC calibration has been used to reduce variation.

As described in Section II and shown in Fig 9, the Q's of

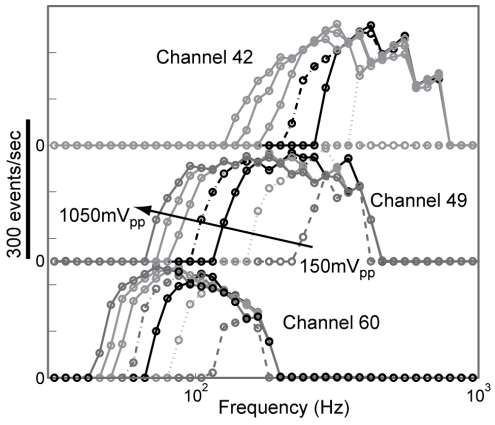

Fig 7 PFM frequency and amplitude responses.

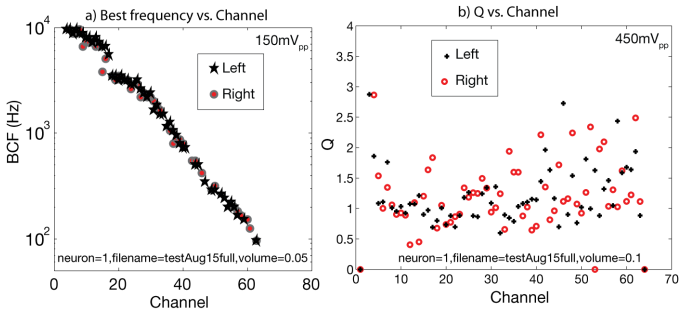

Fig 8 Extracted binaural BCFs and Qs.

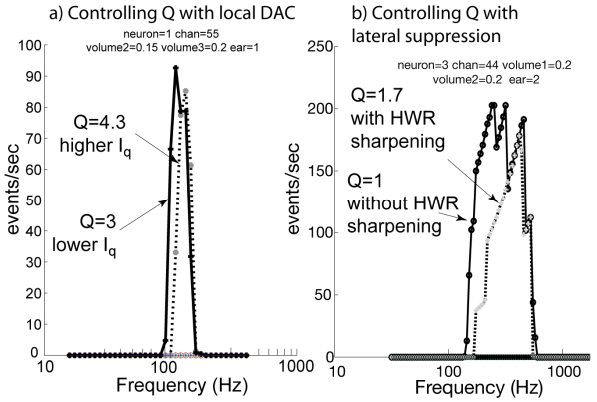

Fig 9 Two ways of increasing the Q of a filter. a) Sharpening of frequency response of one channel as a function of 2 locally programmed I_q values. Input amplitude adjusted to obtain same peak response rate. b) Sharpening of response due to HWR lateral suppression. Mean increase over all channels of Q is factor of 1.18.

the PFM outputs can be sharpened by using the local QDAC (Fig 9a) and by turning on the lateral inhibition from the nearest neighbor downstream (Fig 9b). The response of the filters could be further sharpened by subtracting the PFM responses of neighboring channels, since the filter responses roll off sharply.

V. CONCLUSION

TABLE I. summarizes design and characterization specifications. Compared with prior work, this cochlea achieves acceptable on-chip dynamic range of 36dB when combined with an off-chip preamplifier with global AGC. It proposes two novel on-chip mechanisms for sharpening spectral selectivity in cascaded filter bank designs, where they are particularly valuable. It also achieves usable matching between corresponding channels from the left and right ears, which is important for spatial audition. The integration of on-chip per-channel Q adjustment and highly usable USB implementation will be useful in future work on application scenarios.

Silicon cochlea designs have been explored over the past two decades. This design represents the most integrated version so far with multiple functions and easy use. The asynchronous output representation of this cochlea presents a novel front-end for various auditory tasks and is being investigated for spatial audition and speech recognition tasks.

ACKNOWLEDGMENTS

The project was funded by ONR NICOP grant N00014-05-1-0916, ETH Zurich, and the University of Zurich. We thank R. Berner for assistance in the CPLD firmware on the cochlea system, and H. Finger for the work in the frequency response measurement and localization code.

REFERENCES

[1] B. Wen and K. Boahen, "A 360-channel speech preprocessor that emulates the cochlear amplifier," in *ISSCC Dig. of Tech. Papers*, 2006, pp. 556-557.

[2] R. Sarpeshkar, *et al.*, "An Analog Bionic Ear Processor with Zero-Crossing Detection," in *ISSCC Dig. of Tech. Papers*, 2005, pp. 78-79.

[3] E. Fragniere, "A 100-channel analog CMOS auditory filter bank for speech recognition," in *ISSCC Dig. of Tech. Papers*, 2005, pp. 140-589.

[4] J. Georgiou and C. Toumazou, "A 126-uW cochlear chip for a totally implantable system," *IEEE Journal of Solid-State Circuits,* vol. 40, pp. 430-443, 2005.

[5] V. Chan, *et al.*, "AER EAR: A matched silicon cochlea pair with address event representation interface," *IEEE Trans. on Circuits and Systems I: Regular Papers,* vol. 54, pp. 48-59, 2007.

[6] T. Hamilton, *et al.*, "An active 2-D silicon cochlea," *IEEE Trans. Biomedical Circuits and Systems,* vol. 2, p. 30, 2008.

[7] T. Delbruck. (2007, *jAER open source project*. Available: http://jaer.wiki.sourceforge.net

[8] T. Delbruck and P. Lichtsteiner, "Fully programmable bias current generator with 24 bit resolution per bias," in *2006 IEEE International Symposium on Circuits and Systems (ISCAS 2006)*, Athens, 2006, pp. 2849-2852.

[9] M. W. Baker and R. Sarpeshkar, "A low-power high-PSRR current-mode microphone preamplifier," *IEEE Journal of Solid-State Circuits,* vol. 38, pp. 1671-1678, Oct 2003.

[10] R. F. Lyon and C. Mead, "An analog electronic cochlea," *IEEE Transactions on Acoustics Speech and Signal Processing,* vol. 36, pp. 1119-1134, 1988.

[11] A. van Schaik, *et al.*, "Improved silicon cochlea using compatible lateral bipolar transistors," in *Advances in Neural Information Processing Systems 8*, 1996, pp. 671-677.

[12] L. Watts, *et al.*, "Improved implementation of the silicon cochlea," *IEEE J. of Solid State Circuits,* vol. 27, pp. 692-700, 1992.

[13] R. Sarpeshkar and R. F. Lyon, "A low-power wide-dynamic-range analog VLSI cochlea," *Analog Integrated Circuits and Signal Processing,* vol. 16, pp. 245-274, Aug 1998.

[14] S. Shamma, "Speech processing in the auditory system I: The representation of speech sounds in the responses of the auditory nerve," *J. Acoust. Soc. Am.,* vol. 78, 1985.

[15] R. Berner, *et al.*, "A 5 Meps $100 USB2.0 Address-Event Monitor-Sequencer Interface," in *ISCAS 2007*, 2007, pp. 2451-2454.

[16] K. A. Boahen, "A burst-mode word-serial address-event link-I transmitter design," *IEEE Transactions on Circuits and Systems I-Regular Papers,* vol. 51, pp. 1269-1280, Jul 2004.

Design	
Technology	0.35um 2P 4M CMOS
Chip size	3.5mm x 6mm (21mm^2)
Channel count	64x2
PFM outputs	512
Supply voltage	3.3V
Characterization (at final PFM output)	
Power consumption	18.4mW to 26mW (DVdd) 33mW (AVdd)
Dynamic range to produce PFM output	36dB (25mVpp to 1500mVpp) at microphone preamp output
Frequency range	50Hz to 50kHz (adjustable)
PFM Best characteristic frequency (BCF) matching	±16% between ears at 150mVpp
PFM Q and Q matching (BCF/width at 0.7 of BCF)	1.5 ± 0.4 (±27%) at 450mVpp
Event timing jitter, 1kHz ,	±2us at 250mVpp
PFM peak bandwidth	10M events/sec
Total PFM typical speech rate	20k events/sec

TABLE I. SUMMARY OF CHIP CHARACTERISTICS.