

A 0.5 V 55 μ W 64 \times 2 Channel Binaural Silicon Cochlea for Event-Driven Stereo-Audio Sensing

Minhao Yang, *Member, IEEE*, Chen-Han Chien, *Student Member, IEEE*,
Tobi Delbrück, *Fellow, IEEE*, and Shih-Chii Liu, *Senior Member, IEEE*

Abstract—This paper presents a 64×2 channel stereo-audio sensing front end with parallel asynchronous event output inspired by the biological cochlea. Each binaural channel performs feature extraction by analog bandpass filtering, and the filtered signal is encoded into events via asynchronous delta modulation (ADM). The channel central frequencies f_0 are geometrically scaled across the human hearing range. Two design techniques are highlighted to achieve the high system power efficiency: source-follower-based bandpass filters (BPFs) and asynchronous delta modulation (ADM) with adaptive self-oscillating comparison. The chip was fabricated in $0.18 \mu\text{m}$ 1P6M CMOS, and occupies an area of $10.5 \times 4.8 \text{ mm}^2$. The core cochlea system operating under a 0.5 V power supply consumes $55 \mu\text{W}$ at an output rate of 100k event/s. The measured range of f_0 is from 8 Hz to 20 kHz, and the BPF quality factor Q can be tuned from 1 to almost 40. The 1σ mismatch of f_0 and Q between two ears is 3.3% and 15%, respectively, across all channels at $Q \approx 10$. Reconstruction of speech input from the event output of the chip is performed to validate the information integrity in event-domain representation, and vowel discrimination is demonstrated as a simple application using histograms of the output events. This type of silicon cochlea front end targets integration with embedded event-driven processors for low-power smart audio sensing with classification capabilities, such as voice activity detection and speaker identification.

Index Terms—Adaptive adaptive self-oscillating comparison comparison, asynchronous delta modulator (ADM), audio sensing, bandpass filter (BPF), central frequency, event-driven, quality factor, reconstruction, silicon cochlea, source follower, speech, spike, vowel discrimination.

I. INTRODUCTION

UNBIQUITOUS smart sensing is envisioned to facilitate our daily life and work. For example, we would like to control our mobile/wearable devices via acoustic interfaces, to operate appliances at home or office, which are wirelessly

Manuscript received April 6, 2016; revised June 30, 2016; accepted August 23, 2016. Date of publication September 22, 2016; date of current version October 29, 2016. This paper was approved by Guest Editor Wentai Liu. This work was supported in part by SNF NCCR Robotics, in part by SNF FRAS under Grant 153565, in part by EU SEEBETTER under Grant 270324, and in part by EU COCOHA under Grant 644732. The work of M. Yang was also supported by the Swiss National Science Foundation (SNF) Early Postdoc Mobility fellowship.

M. Yang was with the Institute of Neuroinformatics, University of Zürich and ETH Zürich, 8057 Zürich, Switzerland. He is now with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: yangmh.ic@gmail.com).

C.-H. Chien, T. Delbrück, and S.-C. Liu are with the Institute of Neuroinformatics, University of Zürich and ETH Zürich, 8057 Zürich, Switzerland (e-mail: chenhan@ini.uzh.ch; shih@ini.uzh.ch; tobi@ini.uzh.ch).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSSC.2016.2604285

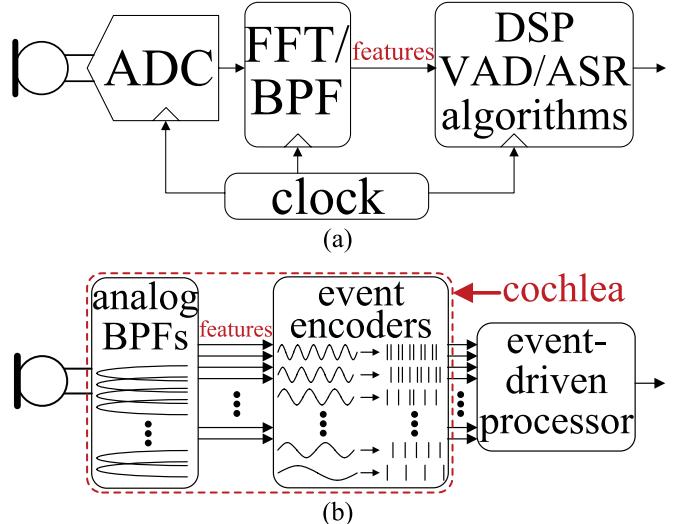


Fig. 1. (a) Conventional audio processing system. (b) Bioinspired audio processing system.

connected to scattered acoustic sensor nodes [1] by voice commands, or to deploy in field ambiently powered outdoor sensors to locate particular classes of wildlife or artificial sounds. Nowadays, these tasks may be easily done by sending the raw audio data to the “cloud” for analysis, and the end-user devices receive the analyzed results. This sensing paradigm, however, raises the concern of security and power constraint on raw data transmission, and therefore, embedded processing within sensor nodes is preferred. Low-power consumption is one of the most important design criterions for the longevity of a smart sensing system, particularly in the era of Internet of Things, and to this end, low-power chips based on the conventional audio processing chain, as shown in Fig. 1(a), have been built for voice activity detection (VAD) [2] and automatic speech recognition (ASR) [3]. In general, a high precision ADC converts sound input into digital representation, and the sound features (e.g., mel-frequency cepstrum coefficients) are then extracted via digital FFT/bandpass filter (BPF). The extracted features are processed by a dedicated DSP implementing VAD or ASR algorithms. All these signal acquisition and processing building blocks are clocked. One of the main drawbacks of such a clocked system is that the Nyquist-rate sampling results in high data redundancy when the varying signal bandwidth is mostly smaller than the Nyquist frequency and in turn leads to prolonged processing time and excessive

processing power, especially for sporadic signals, such as sound. The technique of compressive sampling can be used to reduce data redundancy by exploiting signal sparsity in some representation domain. Nonetheless, the compressed data often need time- and power-consuming recovery before feature extraction to ensure, e.g., good classification accuracy, except for few simple signals, such as heart rate, examined by photoplethysmogram [4].

Bioinspired design is a promising approach of tackling the power efficiency of smart sensing systems. Starting from Carver Mead's lab back in the late 1980s, the pursuit of silicon cochlea as the audio sensing front end has been ongoing for nearly three decades [5]–[15]. A complete silicon cochlea normally consists of two main functional parts, the analog BPFs with geometrically scaled central frequencies f_0 for feature extraction and the event encoders to convert analog signals into asynchronous event streams, as shown in Fig. 1(b). The parallel output events from the silicon cochlea channels can be sent directly to an event-driven processor for high-level cognitive audio processing. The benefits of this type of clockless system are threefold: 1) feature extraction in analog domain can save power, particularly when the required feature SNR is low to medium [16], [17]; 2) asynchronous event encoding helps reduce data redundancy, because the output event rate is proportionally dependent on the input activity and becomes zero if the input is quiescent [18], [19]; and 3) event-driven processors, such as continuous-time (CT) DSPs [19] and spiking neural networks [20], have adaptive power consumption that is correlated with the incoming event rate, and their processing latency can be reduced because of low data redundancy; low processing latency is particularly important for real-time applications involving human–machine interactions, where the machine's immediate reaction to human expressions, such as speech, is often necessary. Despite the aforementioned potential advantages, the silicon cochlea is yet to prove its competitiveness in power efficiency compared with the state-of-the-art conventional digital systems such as the one for cochlea implants [21] with a similar parallel BPF bank architecture.

This work [22] addresses the power efficiency enhancement of a silicon cochlea system mainly through two design techniques: the source-follower-based (SFB) BPF and the adaptive self-oscillating comparison in asynchronous event encoding. The first technique of constructing the SFB-BPF is based on the SFB-LPFs [23], [24]. Compared with other common filter types, such as active-RC and g_m -C, SFB-LPFs have shown the best power/pole/Hz metric thanks to the simple topology and intrinsically high linearity of source followers. An exemplary biquad is depicted in Fig. 2(a). The diode-connected nFETs and C_1 give the first pole, and the cross-coupled pFETs and C_2 give the second pole. The cross-coupled pFETs are essential for synthesizing complex poles. Several SFB-BPFs have been proposed [25]–[27], but they are not readily applicable in the cochlea design because of various deficiencies: reference [25] has large passband gain loss when multiple biquads are cascaded due to its low input impedance; reference [26] cannot achieve a quality factor $Q > 0.5$ due to absence of cross-coupled transistors;

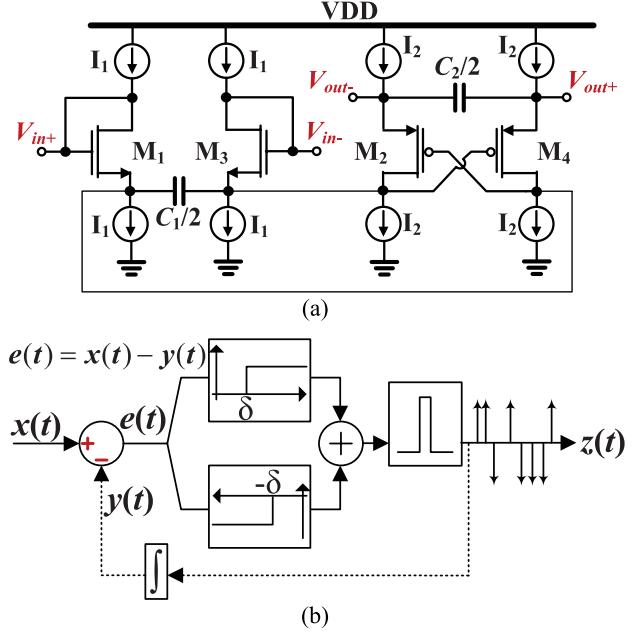


Fig. 2. (a) Exemplary SFB-LPF biquad [24]. (b) ADM model.

the transfer function of [27] is highly sensitive to the input common-mode level due to poorly defined bias currents of transistors. To circumvent the problems, a new SFB-BPF topology is introduced in this paper together with the design technique to optimize the tradeoff between variations of f_0 and Q . Our BPF design does not aim to rigorously model the biophysics or to replicate the biologically realistic transfer functions, but rather we see the BPF bank as a method of performing feature extraction. Nevertheless, the transfer function of our proposed fourth-order BPF approximately follows the form of the so-called one-zero gammatone filter equation in [12], which is derived from the well-known gammatone filter widely used in the auditory modeling community.

The second technique concerns the adoption of latched comparators in clockless data conversion, which requires asynchronous comparison of input signals with one or multiple thresholds. As an example, asynchronous delta modulation (ADM) needs two thresholds [18], [28], the upper threshold δ and the lower threshold $-\delta$ for generating ON and OFF events, respectively, as illustrated in Fig. 2(b). To the best of our knowledge, the circuit implementations of comparison in clockless data converters have always been CT-comparators [19], [29]–[34]. Latched comparators, such as the StrongARM latch [35], are more power efficient than CT-comparators under the condition of the same comparison speed. This efficiency gap is even enlarged at a low supply voltage where transistors in CT-comparators need to have large W/L ratios to stay in saturation, causing considerable parasitic capacitance, and thus power dissipation. The problem of adopting latched comparators in a clockless system is that the regular clock signal needed to reset the latch after each comparison is not available. A self-oscillating comparison scheme, similar to the ones in SAR ADCs [36], [37], is

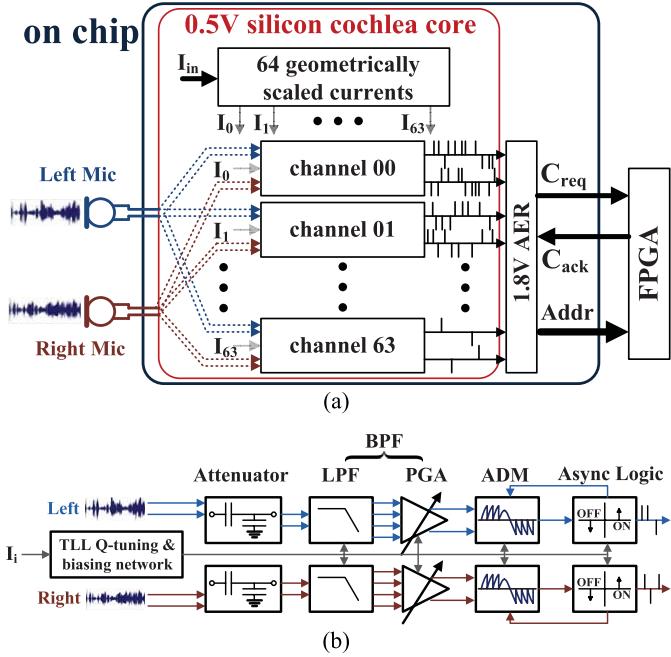


Fig. 3. (a) System architecture of the silicon cochlea. (b) Building blocks within one binaural channel.

used in this paper to create the pseudo-clock signal for reset. Adaptively changing the oscillation frequency according to the output event rate is conceived to further reduce the average power consumption.

The remainder of this paper is organized as follows. The cochlea system architecture is described in Section II; the circuit implementations of the building blocks are elaborated in Section III; the chip characterization results are shown in Section IV; the conclusion and discussions are given in Section V.

II. SYSTEM ARCHITECTURE

The system architecture of the silicon cochlea is illustrated in Fig. 3(a); 64 parallel binaural channels approximately covering a human-hearing frequency range from 20 Hz to 20 kHz convert input sound into parallel output event streams. The bias currents I_i ($i = 0-63$) for the 64 channels are derived from a single input current $I_{in} = 50$ nA, and are geometrically scaled with a designed ratio of 1.11 between neighboring channels. The asynchronous events are transmitted to an off-chip FPGA via the address event representation (AER) block with a tree-like fair arbiter [38], and the FPGA records the addresses and the timestamps of the events with a time resolution of 1 μ s.

Within one binaural channel as shown in Fig. 3(b), the bias current I_i is distributed to all the building blocks by a biasing network. The translinear loop (TLL) shared by the two identical signal chains receiving inputs from the left and right microphones, respectively, is used for the BPF Q tuning. The capacitive attenuator is used to attenuate input signals with large amplitude to control the distortion level of the SFB-BPF. The SFB-LPF and the programmable gain amplifier (PGA) constitute the BPF. The asynchronous delta modulation (ADM) encodes the filtered

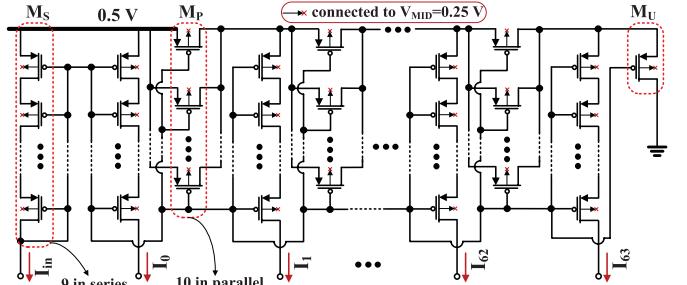


Fig. 4. Current divider circuit that generates the 64 geometrically scaled bias currents.

signal into asynchronous events. The asynchronous (Async) logic provides the feedback control signals for the ADM and communicates the generated events to the peripheral AER; 20 latches are used to store the configuration bits shared by the two signal chains. The configuration procedure is similar to the one used in a programmable bias current array [39].

III. CIRCUIT IMPLEMENTATION

A. Geometrically Scaled Bias Currents

The geometrically scaled currents I_0 – I_{63} define the f_0 scaling of the BPFs. Two prior methods of generating these currents in a silicon cochlea are: MOSFETs in subthreshold with linearly scaled gate voltages [6], [8] and CMOS-compatible lateral bipolar transistors [10], [13]. The first method requires two bias currents to define the upper and lower f_0 boundaries and two buffers to drive the resistive divider that generates the gate voltages. It never successfully demonstrated generation of monotonically scaled f_0 probably because of badly controlled variation of subthreshold currents. The second method produces much improved monotonic f_0 scaling, but the large V_{BE} of a BJT prevents the use of a 0.5 V supply voltage. The MOSFET-based current division technique [40] is adopted in this work. The complete circuit is shown in Fig. 4. Only a single input current I_{in} is needed. All the unit pFETs have the same W and L . The numbers of vertical pFETs in series (lumped as M_S) and horizontal pFETs in parallel (lumped as M_P) are designed to be $N_S = 9$ and $N_P = 10$, respectively, and the divider chain is terminated with a single unit pFET M_U . The choice of N_S and N_P is justified as follows. As indicated in [41], for a current scaling ratio of $r = 1.11$, the size ratio R_{SP} of M_S to M_P should be

$$R_{SP} = \frac{(r-1)^2}{r} \approx 0.011. \quad (1)$$

With the chosen N_S and N_P , we find that $1/N_S/N_P \approx 0.011$, the same as the calculated R_{SP} . The body terminals of all the pFETs are biased at $V_{MID} = 0.25$ V to lower the threshold voltage for a reduced W/L ratio. The theoretical ratio of $r = 1.11$ limits the f_0 variation of each channel to be within $\pm 5.5\%$ to avoid nonmonotonicity in f_0 scaling. Monte Carlo simulations were performed during design to ensure maximum 1% 1σ variation of I_0 – I_{63} .

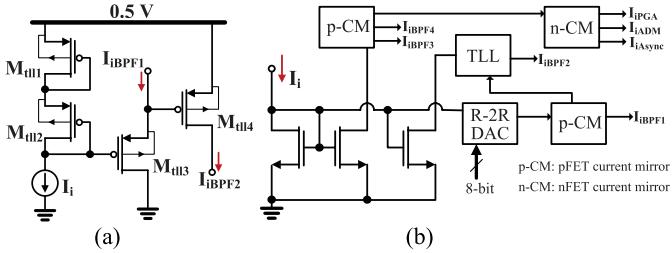


Fig. 5. (a) TLL circuit. (b) Bias current distribution network.

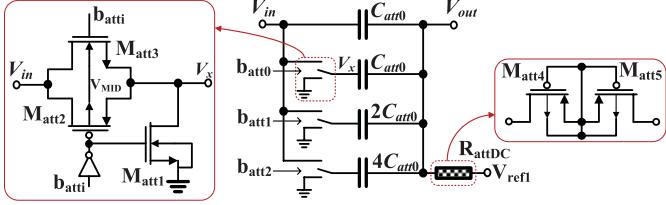


Fig. 6. 3 bit programmable capacitive attenuator.

B. Translinear Loop, Biasing Network, Attenuator

As will be clear in the Section III.C, the Q tuning of the SFB-BPF needs two bias currents whose ratio is tunable and whose product stays constant. A TLL circuit, as shown in Fig. 5(a), can fulfill this requirement. The pFETs $M_{\text{TLL}1}$ – $M_{\text{TLL}4}$ are in subthreshold, and in light of the translinear principle [42], the currents satisfy the relationship below

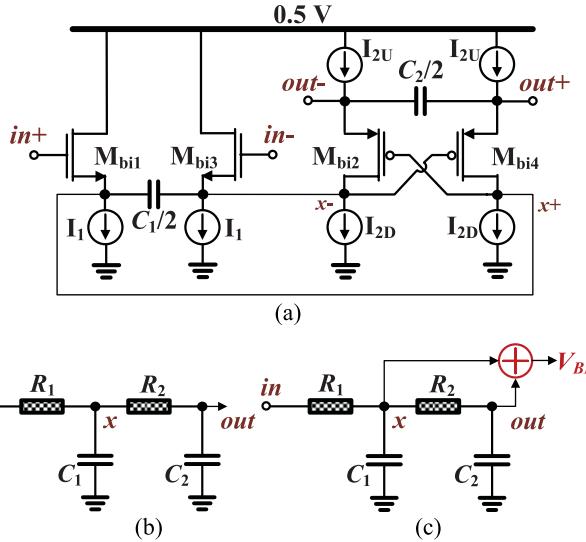
$$I_i^2 = I_{\text{iBPF}1} \times I_{\text{iBPF}2}. \quad (2)$$

$I_{\text{iBPF}1}$ is generated by a digitally programmable MOSFET-based R-2R DAC [39] and $I_{\text{iBPF}2}$ is generated by the TLL. They are the bias currents of the first SF-biquad of the complete fourth-order BPF in the i th channel. Fig. 5(b) shows the biasing network. $I_{\text{iBPF}3}$ and $I_{\text{iBPF}4}$ are the bias currents of the second SF-biquad, and their product also equals to the square of I_i . I_{iPGA} , I_{iADM} , and I_{iAsync} are the biases for the PGA, ADM, and Async Logic blocks, respectively.

A 3 bit programmable attenuator (Fig. 6) with maximum 18 dB attenuation is used to extend the system upper bound input signal range while keeping the distortion of the following BPF low; inclusion of this attenuator is further motivated by the fact that active microphones can produce large amplitude output voltages with close-by loud sound even without stand-alone preamplifiers, which may cause large-signal oscillation of BPFs at high Q [43]. The switch transistors $M_{\text{att}1}$ – $M_{\text{att}3}$ are sized with sufficiently low ON-resistance to make sure that the lowpass corner frequency of the attenuator is always much larger than 20 kHz and the associated third-order intermodulation (IM) always less than –60 dBc [44]. The output is directly connected to the gate input of the first SF-biquad in the BPF, and therefore, V_{refl} is needed to set the dc bias voltage through the pseudoresistor R_{attDC} composed of $M_{\text{att}4}$ and $M_{\text{att}5}$. The generation of V_{refl} will be described in the Section III.C.

C. Source-Follower-Based Bandpass Filter

The SF-biquad used in the BPF and its small signal RC equivalent circuit are depicted in Fig. 7(a) and (b), respectively.

Fig. 7. (a) SF-biquad used in the SFB-BPF design and (b) its small signal RC -equivalent circuit. (c) Second-order SFB-BPF small signal RC -equivalent circuit.

Assuming no mismatch, $I_{2U} = I_{2D} = I_2$. $R_1 = 1/g_{m1}$ and $R_2 = 1/g_{m2}$, where g_{m1} and g_{m2} are the transconductances of the nFETs $M_{\text{bi}1}/M_{\text{bi}3}$ and the pFETs $M_{\text{bi}2}/M_{\text{bi}4}$, respectively. Note that R_2 has a negative resistance due to the cross-coupling of the pFETs [24]. This biquad is a second-order LPF with its output at node out . To obtain a BPF transfer function, an additional zero is created by summing out and x (the intermediate node after the first RC stage), as shown in Fig. 7(c). The transfer function $H_{\text{BPF}}(s)$ and the parameters f_0 and Q of the second-order BPF are derived as

$$H_{\text{BPF}}(s) = \frac{s \cdot \frac{C_2}{g_{m2}}}{s^2 \cdot \frac{C_1 C_2}{g_{m1} g_{m2}} + s \cdot \left(\frac{C_1}{g_{m1}} + \frac{C_2}{g_{m2}} - \frac{C_2}{g_{m1}} \right) + 1} \\ f_0 = \frac{1}{2\pi} \sqrt{\frac{g_{m1} g_{m2}}{C_1 C_2}}, \quad Q = \frac{\sqrt{\frac{g_{m1} g_{m2}}{C_1 C_2}}}{\frac{g_{m1}}{C_1} + \frac{g_{m2}}{C_2} - \frac{g_{m2}}{C_1}}. \quad (3)$$

In a bank of BPFs, as mentioned in Section III-A, monotonic f_0 scaling imposes the constraints of maximum $\pm 5.5\%$ f_0 variation. However, as explained below, a small f_0 variation is in conflict with a small Q variation, and an optimized tradeoff is necessary. We first consider two special cases, the equal capacitance ($C_1 = C_2$) case and the equal transconductance ($g_{m1} = g_{m2}$) case, to gain insights into the sources that cause the variations of f_0 and Q .

Let $C_1 = C_2 = C_0$ in the equal capacitance case. f_0 and Q are simplified as

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{g_{m1} g_{m2}}{C_0}}, \quad Q = \sqrt{\frac{g_{m2}}{g_{m1}}}. \quad (4)$$

In subthreshold, the transconductance of an MOSFET is proportional to its bias current, and thus, $I_2 = I_1 = I_0$ at $Q = 1$. In practice, the mismatch between I_{2U} and I_{2D} makes the current in $M_{\text{bi}1}/M_{\text{bi}3}$ deviate from I_1 . If $I_{2D} = (1 + 1\%) \times I_{2U} = 1.01 I_0$ due to 1% mismatch, f_0 is 0.5% off compared with its value in the case of no mismatch.

At $Q = 10$, to keep the nominal f_0 unchanged, $I_2 = 10I_0$ and $I_1 = 0.1I_0$. The 1% mismatch makes the current in $M_{\text{bi}1}/M_{\text{bi}3}$ increases to $0.2I_0$, and the f_0 offset becomes 41%, much larger than the required 5.5% limit.

Let $g_{m1} = g_{m2} = g_{m0}$ in the equal transconductance case. f_0 and Q are simplified as

$$f_0 = \frac{1}{2\pi} \frac{g_{m0}}{\sqrt{C_1 C_2}}, \quad Q = \sqrt{\frac{C_2}{C_1}}. \quad (5)$$

The Q tuning can be achieved by changing the capacitance ratio. The problem here is the sensitivity of Q on the variation of g_{m2}/g_{m1} . Define the transconductance ratio as $n = g_{m2}/g_{m1}$ and the capacitance ratio as $k = C_2/C_1$, and the sensitivity of Q to n is derived as

$$S_Q^n = \frac{\partial Q}{\partial n} / \frac{Q}{n} = \frac{1}{2} \frac{n - k + nk}{n + k - nk}. \quad (6)$$

At $Q = 10$, $n = 1$, $k = 100$, and $S_Q^n = 99.5$, which means that a 1% variation in n results in 99.5% variation in Q , a value that is too large for a BPF bank to have relatively uniform Q values. Therefore, the tradeoff between f_0 and Q variation leads to the choice of $k = 26/15$ used in the fabricated design. With this k value, at $Q = 10$, the f_0 offset due to 1% mismatch between I_{2U} and I_{2D} is reduced to 1%, and S_Q^n to about 8.

At low Q , the main sources of output noise contribution in the proposed BPF biquad are the pFETs $M_{\text{bi}2}/M_{\text{bi}4}$ and the current sources I_{2U} in Fig. 7(a), which is beneficial especially to low-frequency channels, because pFETs normally have much lower flicker noise than nFETs in a $0.18 \mu\text{m}$ process; at high Q , the most dominant in-band noise contributions are from the current sources I_{2U} and I_{2D} [45]. The capacitance values are determined by the aimed input-referred noise (IRN), about $25 \mu\text{V}_{\text{rms}}$ at $Q = 1$, and therefore, $C_1 = 7.5 \text{ pF}$ and $C_2 = 13 \text{ pF}$.

The complete fourth-order biomimetically asymmetrical SFB-BPF is illustrated in Fig. 8, together with the circuits of V_{ref} generator and PGA. The “2nd-order SF” block (SF block in short) is the SF-biquad circuit in Fig. 7(a) without capacitors. The four poles come from the two cascaded second-order SFB-LPFs, and the one zero comes from the summation of the differential $\text{out}^+/\text{out}^-$ and x^+/x^- of the second SF block using the PGA. The two SF blocks are ac-coupled through C_{AC} , so that the internal dc voltages of each block can be separately set to ensure maximum signal swing. The input dc levels of the first and second SF blocks are set to $V_{\text{ref}1}$ via R_{attDC} (in Fig. 6) and $V_{\text{ref}2}$ via R_{BPFDC} , respectively. The pseudoresistor R_{BPFDC} has the same transistor implementation as R_{attDC} . $V_{\text{ref}1}$ and $V_{\text{ref}2}$ are generated by the V_{ref} generator, where the nFET M_{ref} has the same size as the $M_{\text{bi}1}/M_{\text{bi}3}$ in Fig. 7(a), so that the dc voltages of the x^+/x^- nodes of both SF blocks track V_{dcref} . This is important to guarantee the robustness of the SFB-BPF across process variations under a low supply voltage. The input capacitors $C_{\text{pga}0}$ of the PGA are also the loading capacitors of the second SF block. For the two cascaded second-order SFB-LPFs to have the same cutoff frequency, $C_{\text{pga}0}$ needs to satisfy

$$C_{\text{pga}0}^2 = C_1 \times C_2 \quad (7)$$

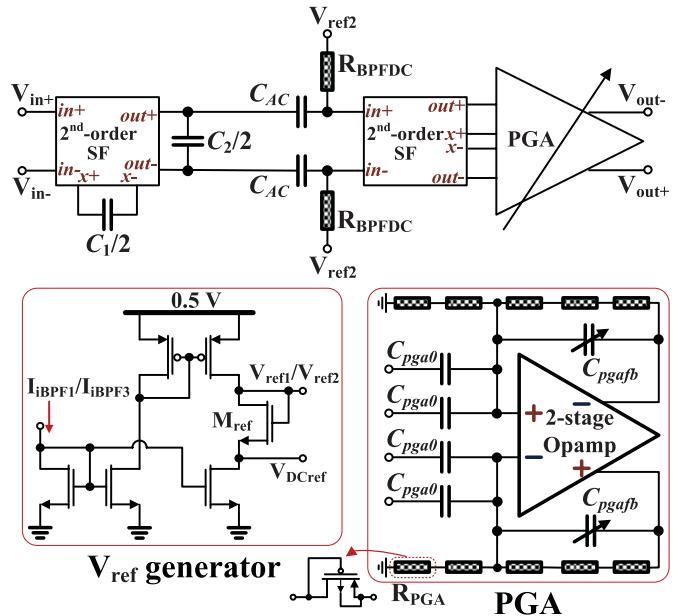


Fig. 8. Complete fourth-order biomimetically asymmetrical SFB-BPF, together with the circuits of V_{ref} generator and PGA.

which makes $C_{\text{pga}0} = 10 \text{ pF}$. The dc output of the PGA is V_{MID} for maximum swing. Its input dc is set to 100 mV by resistive dividers to relax the design headroom of the pFET input stage of the two-stage Opamp. The unit pseudoresistor in the resistive dividers is implemented by a diode-connected pFET [46]. The PGA gain is programmable from 18 to 40 dB with 2 b by changing the feedback capacitance C_{pgafb} . Sufficient gain from PGA is crucial to mitigate the negative impact of uncalibrated comparison threshold variation in ADM on the quality of event encoding.

The schematic of the two-stage fully differential Opamp is shown in Fig. 9. A complementary input stage using both nFETs and pFETs is ideal for a high noise efficiency factor [46], but a simple pFET input stage is used here, because the Opamp has to accommodate a wide range of current scaling with a 0.5 V supply. Pseudocascode compensation with C_{a1}/C_{a2} and C_{a4}/C_{a5} is adopted to improve phase margin compared with traditional Miller compensation and to avoid zero-nulling resistors [47], [48]. Feedforward capacitors C_{a3}/C_{a6} are used to mitigate the potential insufficient gain margin due to gain peaking beyond the unity-gain frequency [49]. All $C_{a1}-C_{a6}$ are made programmable in accordance with the feedback capacitor C_{pgafb} in Fig. 8 to keep a constant PGA IRN about $12 \mu\text{V}_{\text{rms}}$, which is determined by the closed-loop gain and compensation capacitance [50], [51]. The common-mode feedback (CMFB) of the first-stage is achieved by two diode-connected pFETs M_{a1}/M_{a2} . The CMFB of the second-stage is achieved by an error amplifier to set the output dc level to 250 mV. The unity-gain bandwidth of the two-stage CMFB loop is set to be only several times larger than the PGA bandwidth to save power. The rationale is that any large input common-mode signal with frequencies higher than the BPF f_0 is already filtered by the fourth-order SFB-LPF. Compared with the conventional three-stage CMFB

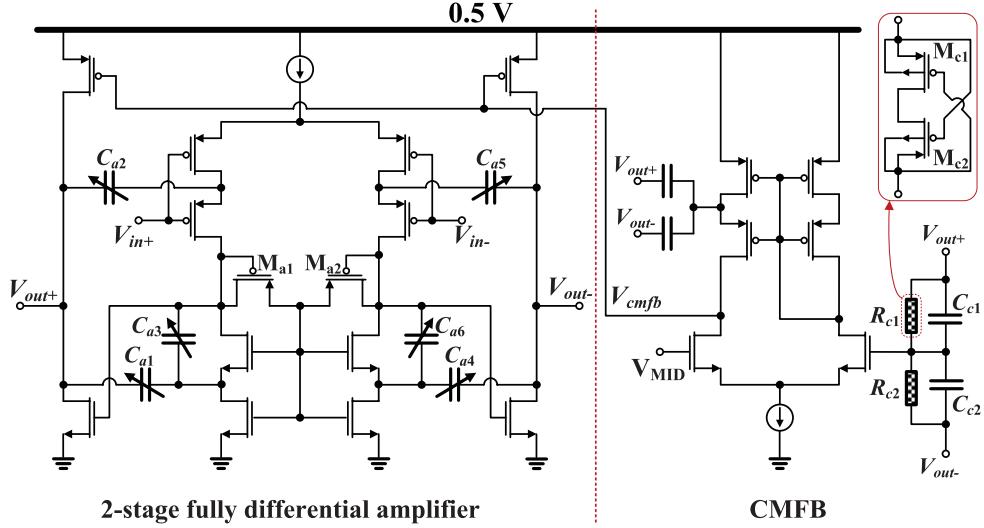


Fig. 9. Schematic of the two-stage fully differential Opamp with its CMFB.

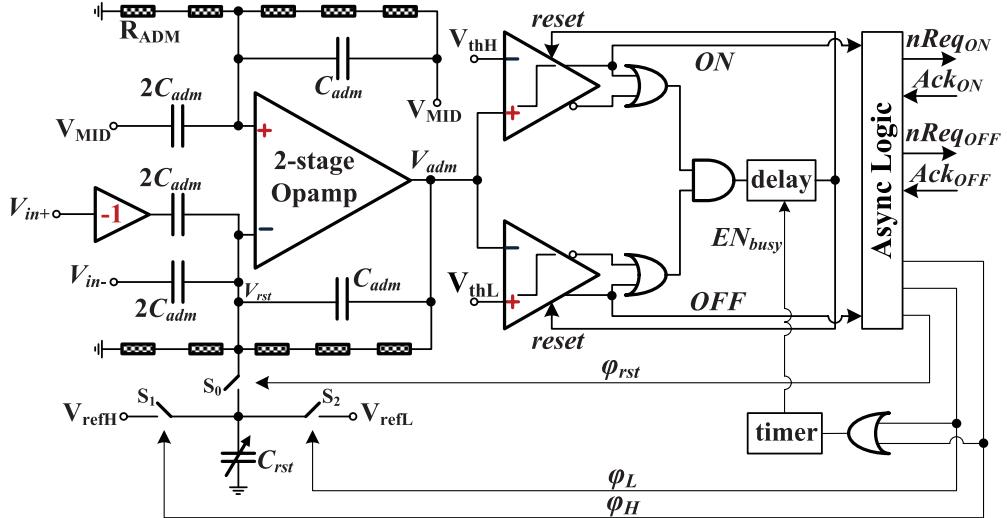


Fig. 10. Block diagram of the ADM.

loop [52], the power consumed by the CMFB circuitry is reduced from $>20\%$ to $<1\%$ of the whole Opamp. The common-mode output is sensed through R_{c1}/R_{c2} and C_{c1}/C_{c2} . Normal resistors, such as nonsilicided poly, are not suitable for compact realization of R_{c1}/R_{c2} because of the required large resistance considering the large output impedance of the Opamp. Most pseudoresistors like diode-connected pFETs are not suitable either because of their asymmetrical $I-V$ characteristics, which causes the sensed common-mode voltage and, in turn, the Opamp output dc to drift away from V_{MID} . The proposed symmetrical pseudoresistor adapted from the one in [53] is composed of M_{c1} and M_{c2} .

D. Asynchronous Delta Modulator With Adaptive Self-Oscillating Comparison

The circuit implementation of the ADM, as illustrated in Fig. 10, is adapted from the one in a prior event-driven vision sensor [51]. The differential input V_{in+}/V_{in-}

of the ADM from the PGA output is combined to form a single-ended input to the ADM amplifier by using the inverting amplifier “ -1 ” block on V_{in+} . The amplifier output V_{adm} is compared with an upper threshold V_{thH} and a lower threshold V_{thL} via two latched comparators. Whenever V_{adm} crosses above V_{thH} or goes below V_{thL} , an ON or OFF event is generated. The ON or OFF signal is then sent to the asynchronous logic to trigger a four-phase handshake with the peripheral AER for event transmission and also to generate the feedback signals φ_{rst} , φ_H , and φ_L to control the S_0-S_2 switches in the switched-capacitor (SC) network for event integration and δ subtraction [see Fig. 2(b)].

The inverting amplifier shown in Fig. 11(a) has a simple common-source amplifier core that consists of M_{inv1} and M_{inv2} to minimize its power overhead. Its bias current is mirrored from the diode-connected M_{inv3} , and the pseudoresistor R_{invDC} with the same transistor implementation as R_{attDC} in Fig. 6 is to prevent the input signal from being highpass filtered. The

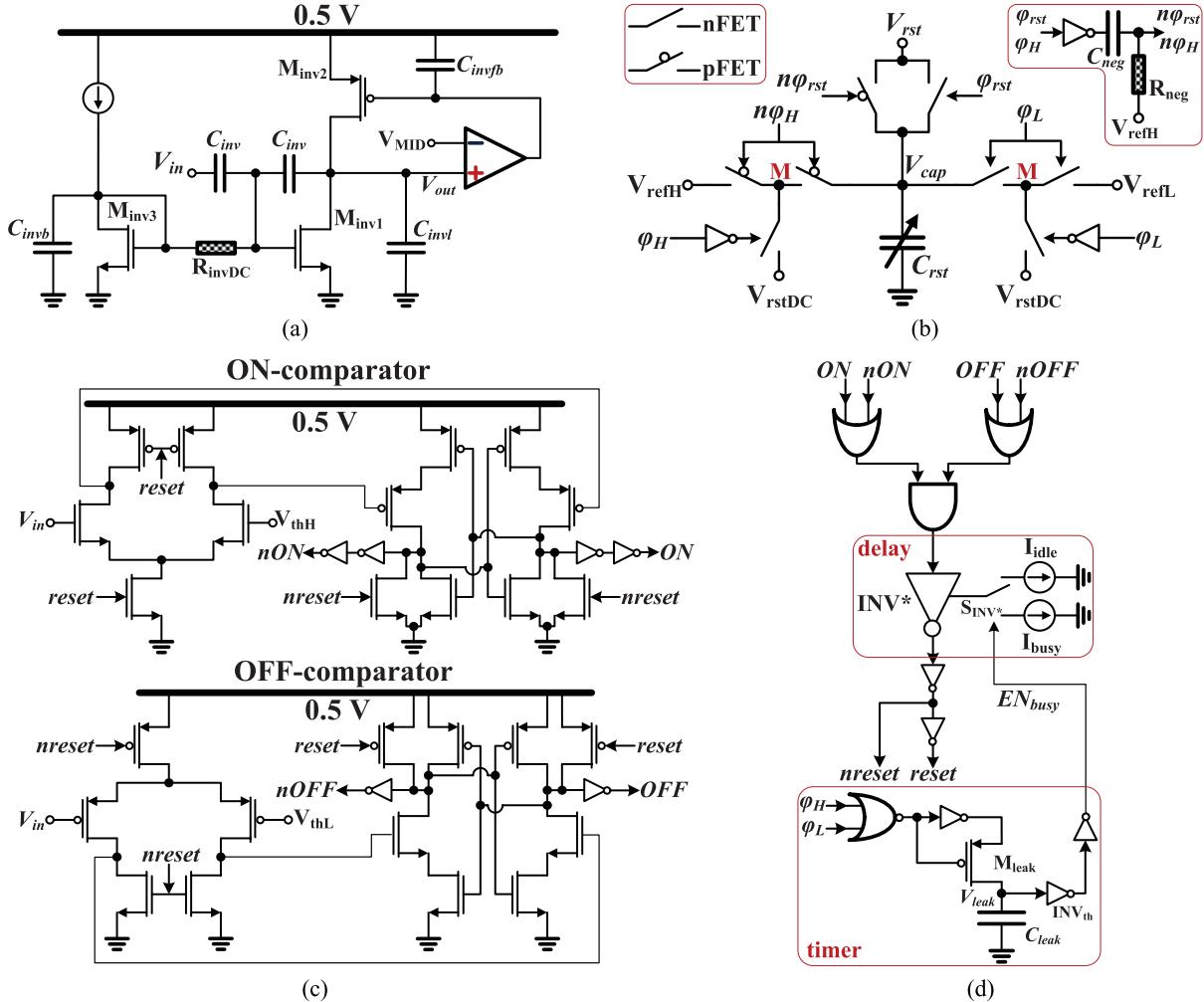


Fig. 11. (a) Inverting amplifier circuit, i.e., the “ -1 ” block in Fig. 10. (b) SC network. (c) ON- and OFF-comparators [55]. (d) Delay and timer circuits.

output dc is set to 250 mV by an error amplifier, which is biased such that the highpass corner is at least $10\times$ lower than the BPF f_0 for a minimized phase shift of the input; the same reason for the inverting amplifier to have a lowpass corner $\times 10$ higher than f_0 . The two-stage Opamp with a single-ended output in the ADM has the same topology as that in [51], and is biased for fast settling during δ subtraction.

The SC network is illustrated in Fig. 11(b). Each switch S_0-S_2 in Fig. 10 is implemented by multiple transistors. The V_{rst} dc is at 100 mV complying with virtual ground, and for symmetrical δ subtraction, $V_{refL} = 100 - \delta$ mV and $V_{refH} = 100 + \delta$ mV. The maximum δ is 100 because V_{refL} cannot go below ground. C_{rst} is made programmable, so that the actual subtracted voltage can be set between δ and 2.5 δ with a 0.5 δ step. Note that the **M** nodes in switches S_1 and S_2 are set to $V_{rstDC} = 100$ mV when V_{cap} is disconnected from V_{refH}/V_{refL} . This is to prevent V_{adm} from slow drift when S_0-S_2 are all disconnected due to the subthreshold leakage of the S_0 transistors by minimizing their V_{ds} [39], [54]. V_{gs} of the ON-state pFET switches driven by $n\varphi_{rst}$ and $n\varphi_H$ is only 100 and 200 mV, respectively, if the active-low state of $n\varphi_{rst}$ and $n\varphi_H$ is at ground and $V_{refH} = 200$ mV.

To maximize the V_{gs} of the ON-state pFETs, so that they can have sufficiently low ON-resistance, $n\varphi_{rst}$ and $n\varphi_H$ are generated by the circuit in the top-right insert. R_{neg} has the same transistor implementation as R_{attDC} in Fig. 6. The active-low state of $n\varphi_{rst}$ and $n\varphi_H$ can now reach to -300 mV, giving 400 mV V_{gs} for the pFET in S_0 and 500 mV in S_1 .

CT-comparators in the prior ADM [51] are replaced by latched comparators in Fig. 10 because of the superior power efficiency and speed of the latter. The “reset” signal needed to initiate each comparison by resetting the latch is generated through self-oscillation by adding two OR gates, one AND gate, and one delay block. The two comparators [55] are illustrated in Fig. 11(c). The ON- and OFF-comparators have an nFET and a pFET input stage, respectively. Typical threshold voltages are set to $V_{thH} = 400$ mV and $V_{thL} = 100$ mV. The delay block controls the oscillation frequency to be geometrically scaled across the 64 channels and is implemented as a simple starved inverter INV^* shown in Fig. 11(d). To further save power, the oscillation frequency within each channel is switchable between a low value and a high value. When no event occurs, EN_{busy} is low and the switch S_{inv^*} connects INV^* to a small starved current I_{idle} for slow oscillation. Whenever an ON or

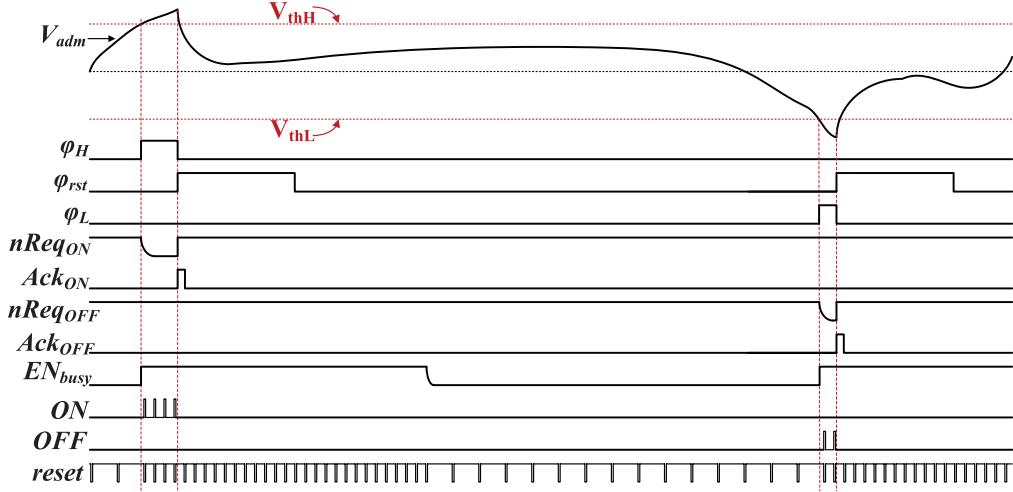


Fig. 12. Exemplary timing diagram of the ADM.

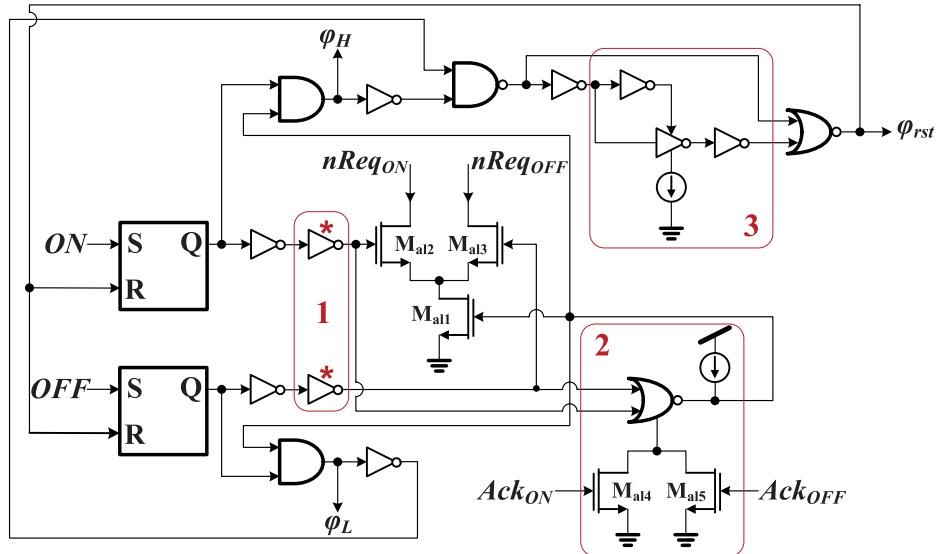


Fig. 13. Asynchronous logic schematic.

OFF event occurs, ϕ_H or ϕ_L becomes high which sets EN_{busy} high, and INV^* is connected to a large starving current I_{busy} for fast oscillation to reduce the delay and delay dispersion in event generation, which is important for high encoding quality [33]. At the moment when ϕ_H or ϕ_L goes back low, M_{leak} starts to discharge C_{leak} through its leakage current. EN_{busy} flips to low once the threshold of the inverter INV_{th} is crossed, or stays high if another event occurs before the threshold crossing of INV_{th} . An exemplary timing diagram containing the key signals in the ADM is shown in Fig. 12.

E. Asynchronous Logic

The schematic of the asynchronous logic is shown in Fig. 13. Because the output of a latched comparator is only valid at the end of a comparison phase and becomes invalid during reset, the two SR latches are indispensable for storing a generated event and to avoid repeated multiple

transmissions of one event. Take an ON-event transmission as an example. An ON event immediately sets ϕ_H high to charge the C_{rst} to V_{refH} in ADM. The inverter in the red box "1" has a pFET with long L , and hence, the request $nReqON$ becomes active low after a certain delay, which determines the pulselength of ϕ_H . An active high $AckON$ is then sent back from the peripheral AER and brings $nReqON$ back high as long as either M_{al1} or M_{al2} stays off after $AckON$ goes low; otherwise, $nReqON$ is pulled low again, causing a false event transmission. The circuit in the red box "2" is to guarantee the overlapped OFF-state of M_{al1} and M_{al2} after $AckON$ goes low. The weak pull-up current source charges up the gate of M_{al1} slowly in order to give sufficient time margin for M_{al2} to be turned off by active high ϕ_{rst} , which sets the output of the SR latch low. Once the gate of M_{al2} goes low following the output of the SR latch, the NOR gate in the red box "2" turns on M_{al1} immediately, ready for the next event transmission. The circuit in the red box "3" is to generate the ϕ_{rst} pulse. The

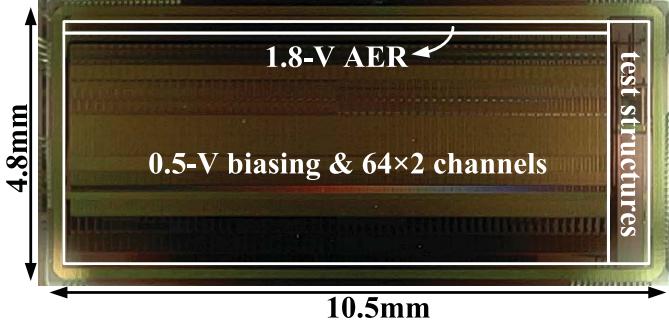


Fig. 14. Chip microphotograph.

current-starved inverter controls the φ_{rst} pulsewidth to ensure sufficient settling time of the ADM amplifier. The current is proportionally derived from the channel bias I_i , and thus, the pulsewidth is geometrically scaled across channels. Note that although the four handshake signals all have a 1.8 V high-state voltage level in contrast to the 0.5 V supply voltage for the asynchronous logic circuits, no level shifter is needed.

IV. EXPERIMENTAL RESULTS

The silicon cochlea chip named CochLP was fabricated in 0.18 μm 1P6M CMOS with deep N-well and MIM capacitors occupying an area of $4.8 \times 10.5 \text{ mm}^2$, as labeled in the microphotograph in Fig. 14. The 0.5 V core, including the geometrically scaled current generator and the 64×2 channels, consumes about $55 \mu\text{W}$ at a 100k event/s output rate, and the analog part of the core consumes more than 96% of the total power. Because channel power consumption scales with f_0 , most of this power is consumed by the highest frequency channels. The 1.8 V AER interface reused from previous designs [51] consumes about $300 \mu\text{A}$ at 100k event/s, i.e., $540 \mu\text{W}$. The dominant power of the AER interface in this prototype chip can be either significantly reduced by redesigning the AER circuits under a 0.5 V supply with transistors working in the near-threshold region [56], [57] or completely avoided in an embedded system where the event streams are directly sent in parallel to the back-end event-driven processor on the same chip, as illustrated in Fig. 1(b). Each channel has pFET source-follower buffers run off a 1.8 V supply for measuring the BPF characteristics. These buffers are biased to have low noise and are enabled in one channel at a time through a 128 b long scanner. The firmware logic, USB interface, and host-side codes in jaER¹ for sending the recorded event addresses and timestamps to PC and for visualization of the output events are based on existing designs. The Section IV.A, IV.B, IV.C, IV.D below present the detailed chip characterization results and comparison with prior works.

A. Bandpass Filter Bank

An SR780 network signal analyzer is used for all frequency-related measurements. The measured transfer functions of the right ear BPF array at $Q \approx 1$ and $Q \approx 10$ are shown

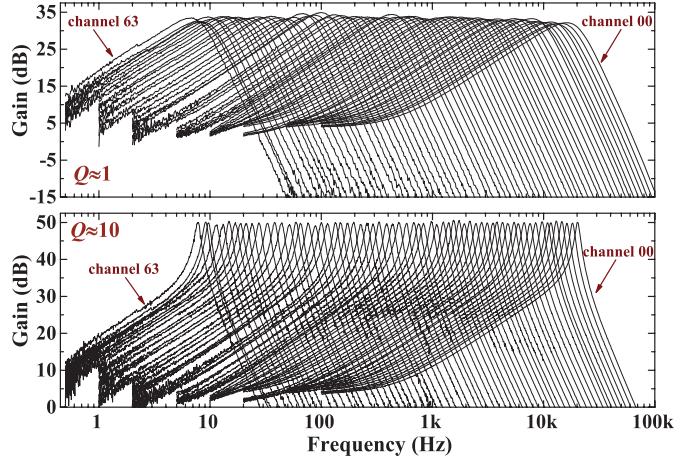


Fig. 15. BPF transfer functions of the right ear at $Q \approx 1$ and $Q \approx 10$.

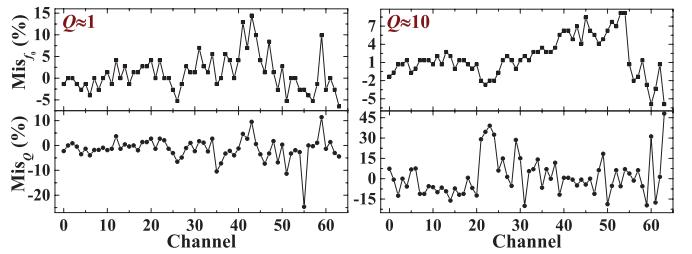


Fig. 16. Binaural mismatch of f_0 and Q .

in Fig. 15. The high Q setting has an approximately 17 dB more peak gain than the low Q setting. In both the cases, f_0 monotonically scales from about 8 Hz to 20 kHz, giving an average scaling ratio of 1.13. The measured mismatches of f_0 and Q between the two ears are plotted in Fig. 16. The 1σ variation of f_0 and Q mismatch across the array is 4.2% and 4.8% at $Q \approx 1$, and 3.3% and 15% at $Q \approx 10$. The larger Q mismatch at a high Q setting is consistent with the sensitivity analysis using (6). Q tuning is demonstrated in four channels, as shown in Fig. 17. The lowest Q starts from about 1 and the highest Q goes up to almost 40.

The noise power spectral density (PSD) at the output of the PGA is measured for all four PGA gain settings, and Fig. 18 shows the exemplary plots from channel 00 and 54. The noise floor increases proportionally as the PGA gain increases, and has obvious gain peaking at f_0 at $Q \approx 10$. With the same PGA gain and Q , the two channels have about the same integrated output referred noise (ORN), e.g., with an 18 dB PGA gain, at $Q \approx 1$, the ORN of channel 00 and 54 is 0.30 and 0.32 mV_{rms}, respectively, and at $Q \approx 10$, the ORN is 0.76 and 0.79 mV_{rms}, respectively. Fig. 19 shows examples of the measured distortion, including IM and harmonic distortion (HD), with a PGA gain of 18 dB. In the HD plots, the third-order harmonic dominates instead of the second-order one thanks to the differential signal path. Note that it is actually the difference between the signal and the “nonharmonic” peak in the HD plot of channel 00 that accounts for the 1% distortion. The maximum SNR at 1% distortion is calculated to be 55 dB at $Q \approx 1$ and 42 dB at $Q \approx 10$ for channel 00,

¹<https://sourceforge.net/p/jaer/wiki/Home/>

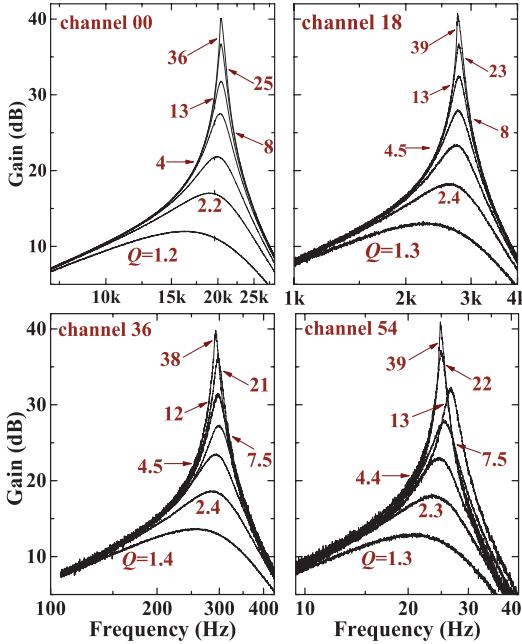


Fig. 17. Q tuning examples from the BPFs of four different channels, channels 00, 18, 36, and 54.

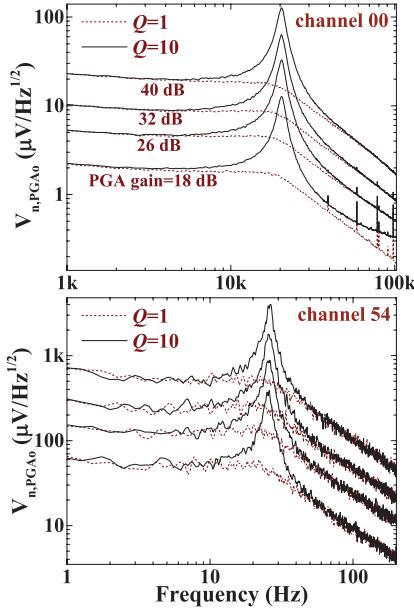


Fig. 18. Noise PSD of channels 00 and 54 at different PGA gain settings.

and 55 and 40 dB for channel 54. The maximum SNR at larger PGA gains is reduced due to an elevated distortion with a larger output swing. The system input signal range with the 1% HD specification can be extended by another 18 dB because of the attenuator.

Fig. 20 shows the example of the measured CMRR and PSRR from channel 00 at the output of the ADM amplifier, i.e., V_{adm} in Fig. 10, when the self-oscillation is disabled. At f_0 , the CMRR and PSRR are 51 and 48 dB, respectively. The significant PSRR degradation below f_0 is deemed to be aggravated by the inverting amplifier in Fig. 11(a), because at low frequencies, I_{ds} of M_{inv2} is modulated by VDD, whereas at high frequencies, C_{invfb} helps stabilize the V_{gs} of M_{inv2} .

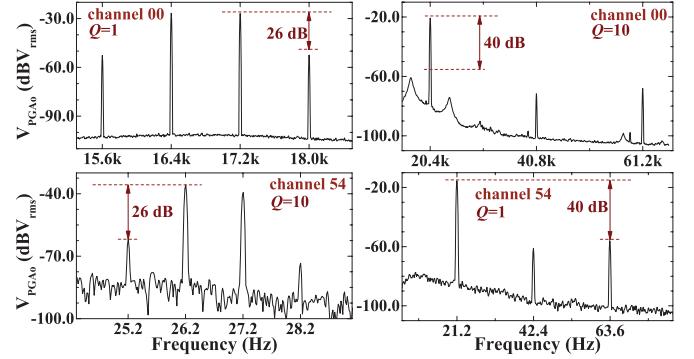


Fig. 19. Distortion of channels 00 and 54 including IM and HD.

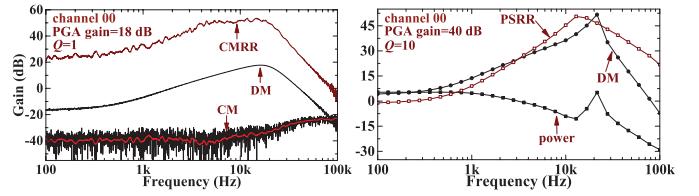


Fig. 20. CMRR and PSRR of channel 00 at the output of the ADM amplifier.

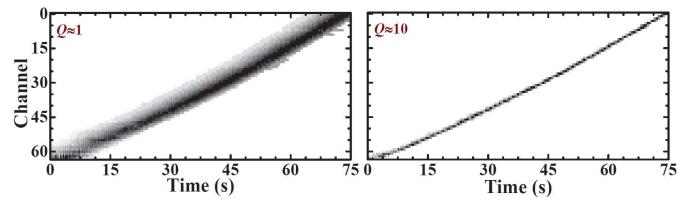


Fig. 21. Event cochleagram in response to a chirp input sweeping from 6 Hz to 21 kHz.

B. Event Output

With BPF transfer functions set as in Fig. 15, the event cochleograms of the chip in response to an exponential chirp input sweeping from 6 Hz to 21 kHz is shown in Fig. 21. The input amplitude is scaled to account for the peak gain difference between the two different Q settings. To plot the grayscale cochleogram, the events are binned every 0.1 s and the event number (EN) in each time bin is normalized by the maximum number in each channel. At every time point, the number of responding channels, i.e., channels that produce events, is obviously less in the high Q case than in the low Q case because of the higher frequency selectivity.

A two-tone test is performed with the BPFs set at $Q \approx 10$. The input signal is the combination of a 2.8 kHz sinusoid and a 1.4 kHz sinusoid each with a 5 mV_{pk} amplitude. The combined output event stream over all channels is shown in Fig. 22(a) within a 4 ms time window, and the event stream from channel 24 alone is in Fig. 22(b). The filtered analog input to the ADM of channel 18 (channel 24) is reconstructed from the event stream of only channel 18 (channel 24) in MATLAB by applying a BPF with corner frequencies at 300 Hz and $(f_{0i} + f_{0j})/2$ ($j = i + 1$, $i = 18, 24$) to the integrated events [equivalent to $y(t)$ in Fig. 2(b)]. The reconstructed waveforms are shown in Fig. 22(c) and (d). The highpass filtering is to avoid slow dc drift of the

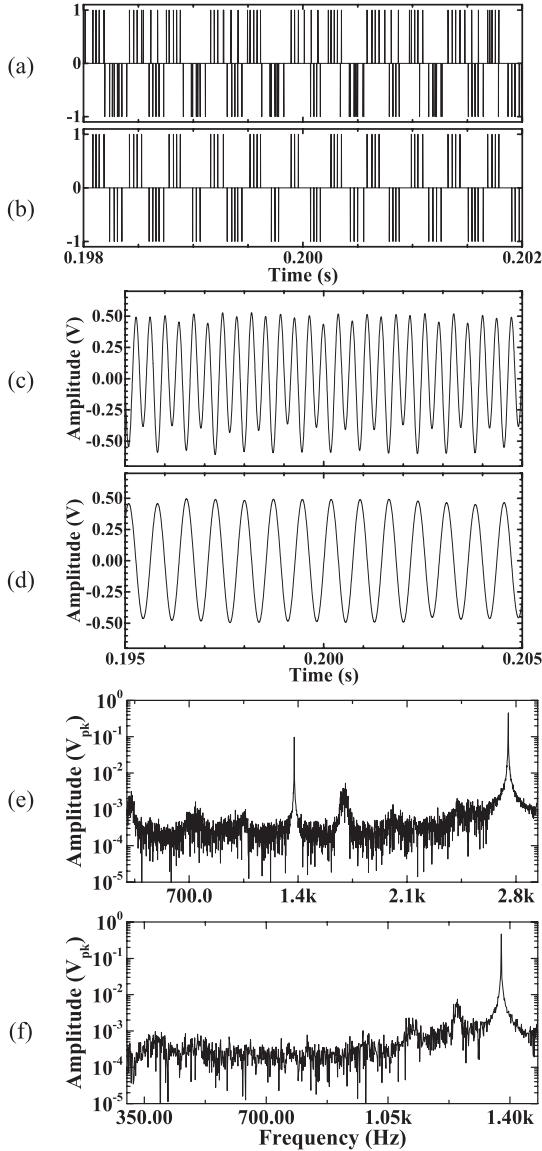


Fig. 22. (a) Total output event stream of the cochlea chip in response to a combinational tone of a 1.4 and a 2.8 kHz sinusoid. (b) Output event stream of channel 24. (c) and (d) Reconstructed analog signals from the output event streams of channels 18 and 24. (e) and (f) Amplitude spectra of (c) and (d).

reconstructed signal due to unbalanced ON and OFF δ subtraction caused by mismatch, and the lowpass filtering is to filter out the high-frequency harmonics because of the coarse quantization by the ADM, i.e., less than ten events per one sinusoidal cycle. The spectra of the reconstructed signals are shown in Fig. 22(e) and (f). The reason for the -25 Hz shift of the 1.4 kHz tone and the -50 Hz shift of the 2.8 kHz tone is not yet clear. The 1.4 kHz peak is about 15 dB lower than the 2.8 kHz peak in Fig. 22(e), consistent with the transfer function of channel 18.

C. Speech Reconstruction and Vowel Discrimination

Sample speeches from different databases are also used as the input to the chip, and reconstruction from the output events is demonstrated to validate the information integrity after the feature extraction of the BPF bank and the sparse

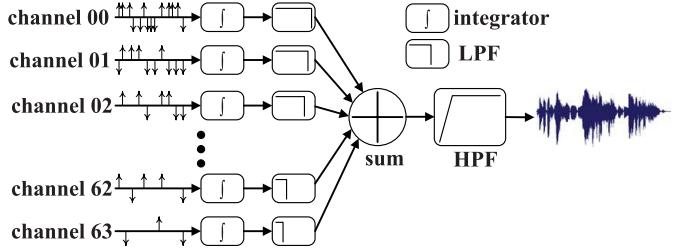


Fig. 23. Method of speech reconstruction from parallel output events.

event encoding. The linear reconstruction method is illustrated in Fig. 23. The 64 parallel event streams are separately integrated and then filtered by LPFs with corner frequencies at the mean of the f_0 of channel i and $i + 1$ ($i = 0 \sim 63$). The filtered signals are summed together and the summed signal is further filtered by an HPF with a 300 Hz corner frequency to obtain the final reconstruction. Fig. 24 shows the reconstruction results on the original digits ‘18174’ spoken by a male speaker from the TIDIGITS database. In both low Q and high Q cases, the reconstructed waveforms [Fig. 24(d) and (f)] capture the envelope of the input sound [Fig. 24(b)], and their spectrograms [Fig. 24(c) and (e)] look similar to the spectrogram of the original waveform [Fig. 24(a)]. The audio files of the reconstructed speeches from the aforementioned male speaker and another female speaker saying a sentence (from the TIMIT database) are available in the Multimedia Supplementary Materials online. The words are well recognizable in all the reconstructed audio, although the distortion generally sounds larger in the high Q case, possibly in part due to higher BPF nonlinearity. Nevertheless, as is evident in comparing the event cochleograms in Fig. 24(g) and (h), a high Q BPF bank helps extract more distinctive acoustic features, which may facilitate processing in the following event-driven processor.

The benefit of a high Q setting of the BPF array is further demonstrated in a vowel discrimination task. The sounds “heed” with the vowel /i/ and “had” with /æ/ are fed as the inputs to the chip. Fig. 25(a) and (b) shows their spectrograms with the two most energy-significant formants labeled: F1, F2, and F4 as the first, second, and forth formant, respectively. Fig. 25(c)–(f) shows the simple normalized histograms of the output EN across the 64 channels. The EN peaks correspond to the formant frequencies and are more distinguishable at $Q \approx 10$ than those at $Q \approx 1$. In light of the peak locations in the histograms at $Q \approx 10$, it is easy to separate the two words with different vowels.

D. Comparison With Prior Works

The comparison of CochLP with prior event-driven silicon cochleae is given in Table I. The normalized power consumption P_{norm} to a 20 kHz f_0 channel is calculated by

$$P_{norm} = \frac{P_{total}(1 - r_f)}{1 - r_f^i} \cdot \frac{20k}{f_H}, r_f = \left(\frac{f_L}{f_H}\right)^{\frac{1}{i-1}} \quad (8)$$

where P_{total} is the total power consumption, f_H and f_L are the highest and lowest channel central frequencies, and

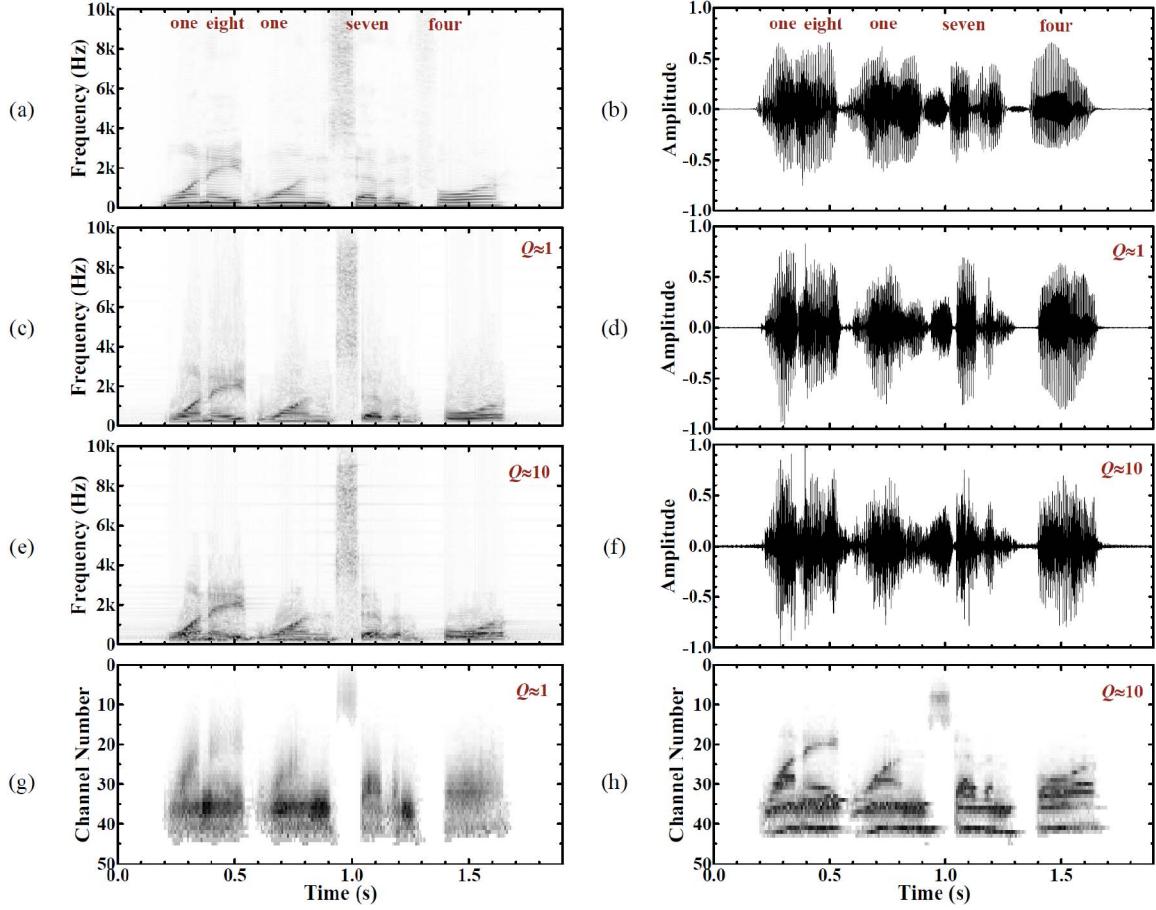


Fig. 24. Spectrograms and waveforms of (a) and (b) original input sound, a male speaker saying digits “18174”, (c) and (d) reconstructed sound from output events at BPF $Q \approx 1$, and (e) and (f) at BPF $Q \approx 10$. (g) and (h) Event cochleograms at $Q \approx 1$ and 10.

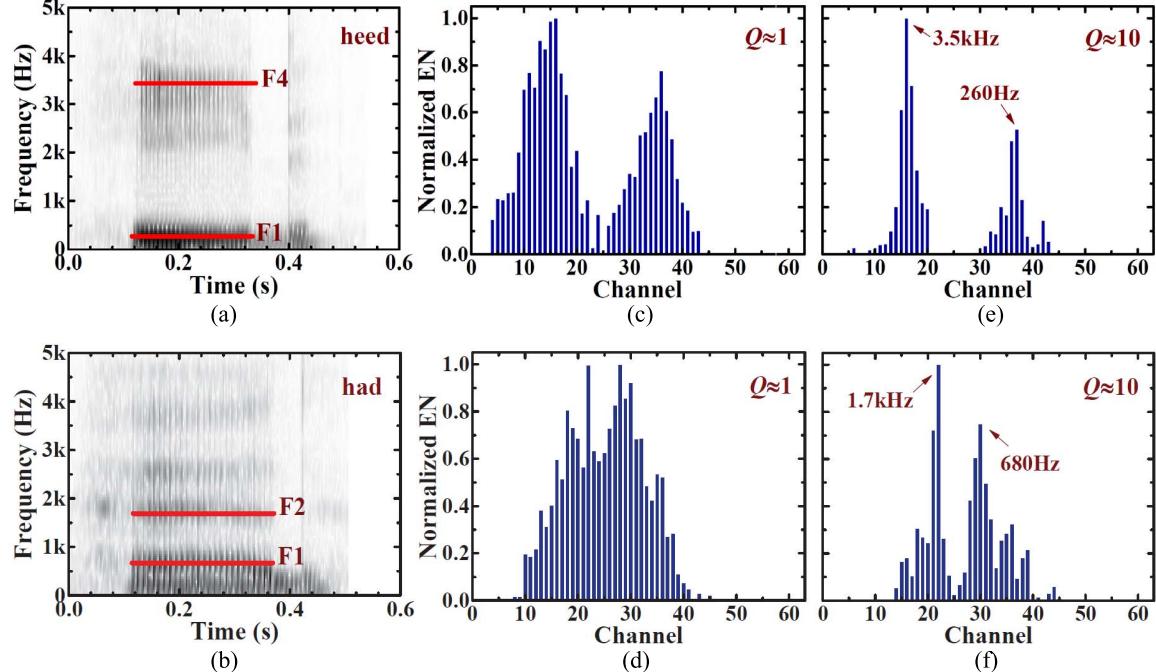


Fig. 25. (a) and (b) Spectrograms of words “heed” and “had” with most-energy-significant vowel formants labeled. EN histograms of the two words (c) and (d) at BPF $Q \approx 1$ and (e) and (f) at BPF $Q \approx 10$.

i is the channel number. CochLP has the lowest P_{norm} , about 18 \times lower than the best prior art designed by

Sarpeshkar *et al.* [7]. Even if the power supply in [7] could be lowered to 0.5 V, the power efficiency of CochLP would still

TABLE I
COMPARISON WITH PRIOR EVENT-DRIVEN SILICON COCHLEAE

	This work 2016	Sarpeshkar 2005 [7]	Fragniere 2005 [8]	Wen 2006 [9]	Liu 2014 [13]
Architecture	Parallel	Parallel	Passive coupling	Active coupling	Cascade
Technology (μm)	0.18 CMOS	1.5 BiCMOS	0.5 CMOS	0.25 CMOS	0.35 CMOS
Power supply (V)	0.5	2.8	3.3	2.5	3.3
Power (μW)	55	60	1700	35900	14000
Channel number	64 \times 2	16	100	360	64 \times 2
Frequency range (Hz)	8-20k	100-5k	200-20k	210-14k	50-50k
Q tuning range	1.3-39 ^a	<10	0.25-12	1.16 \pm 0.92 ^b	1.5 \pm 0.4
Normalized power (μW)	3.2	56	78	605	291
Area per channel (mm^2)	0.26	5.5	0.17	0.030	0.11
Input range (dB)	73 ^c @ $Q \approx 1$	75 ^d	50	52	52 ^e
Event encoding	ADM	Zero-crossing (log ADC)	Threshold-crossing	PFM	PFM
Event readout	AER	Scanning	Scanning	AER	AER
Binaural	Yes	No	No	No	Yes
Building blocks included for power	Cochlea core including BPF and ADM	BPF, envelope detector and bias	Filter bank and DAC for Q tuning	Filter bank with active coupling	Cochlea core, preamplifier and AER

^a from channel 18; ^b Q_{10} , 10 dB below peak gain; ^c including 18 dB of the attenuator;

^d with AGC, 55 dB without AGC;

^e with preamplifier gain control, 36 dB without gain control.

be 3 \times better. In addition, CochLP has a competitive dynamic range and a wide Q tuning range. At the highest $Q \approx 39$, the so-called Q_{10} , a metric often used in physiology and calculated by using a bandwidth measured 10 dB below the peak gain, is about 11, much larger than any previous designs by using active [9] or passive [8] coupling between neighboring channels and multistage ultrasteep roll-off filtering [15]. CochLP also shows unprecedented matching and monotonicity after the event encoding with a large number of channels, and for the first time shows the well-recognizable speech directly reconstructed from the massively parallel asynchronous output events without learning the mapping from events to input speech as in [58].

The power efficiency of CochLP is still competitive when compared with the state-of-the-art digital solution for cochlea implants, where SAR ADC and digital BPFs are used [21]. The SAR ADC with a 52 dB dynamic range consumes 0.8 μW when linearly scaled to 40 kS/s and the P_{norm} of the digital BPFs is about 2.2 μW . The area per channel is estimated to be about 0.2 mm^2 . However, it is not feasible to scale the f_0 of digital BPFs with a small fractional scaling ratio, not to mention the omitted clock power and the output data redundancy. If an FFT is used instead for fine spectrum separation, the power consumption increases significantly even in an advanced 32 nm technology [2]. A cochlea-like front end without event encoding was adopted in a 6 μW VAD [59] with a P_{norm} of about 13 μW as well as in other low-power sound processing systems [60], [61], further proving the efficiency and efficacy of analog feature extraction.

V. CONCLUSION AND DISCUSSION

A 64 \times 2 channel binaural silicon cochlea with parallel event output and 55 μW core power consumption at a 0.5 V supply

is presented. Two design techniques to improve the system power efficiency, including SFB-BPF and event encoding with adaptive self-oscillating comparison, are detailed. Speech reconstruction and vowel discrimination using the chip output events are demonstrated. By exploiting the timing difference in the binaural output events, this chip can also be used for sound source localization [13].

The defect in the current PGA design is that its closed-loop bandwidth should be at least twice the channel central frequency f_0 to avoid peak gain degradation because of the summing operation. But we did not realize this until the end of tape-out and the PGA bandwidth was only set to about 25% larger than f_0 . Consequently, up to 6 dB peak gain loss was found. The method of creating the one zero to obtain a bandpass transfer function from the SFB-LPF does not have to be the active summation by using the power-hungry PGA. Passive summation via capacitors with a source-follower buffer, as illustrated in Fig. 26(a), can result in much higher power efficiency, though the PGA gain is necessary in CochLP to combat the uncalibrated comparator input offset as mentioned in Section III-C. An alternative way of creating one zero without any summation is to use the super-source-follower (SSF) topology shown in Fig. 26(b). If the output is at the x node, the circuit forms an LPF as analyzed in [62], and if the output is at the out node, a BPF transfer function can be obtained with the characteristic parameters given below

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{g_{m1}g_{m2}}{C_1C_2}}, Q = \sqrt{\frac{g_{m1}C_1}{g_{m2}C_2}}, A_{f0} = \frac{g_{m1}C_1}{g_{m2}C_2} \quad (9)$$

where g_{m1} and g_{m2} are the transconductances of M_{ssf1} and M_{ssf2} , respectively, and A_{f0} is the peak gain. The main drawback of the SSF-based BPF is that the peak gain is the square of the BPF Q , making its dynamic range very

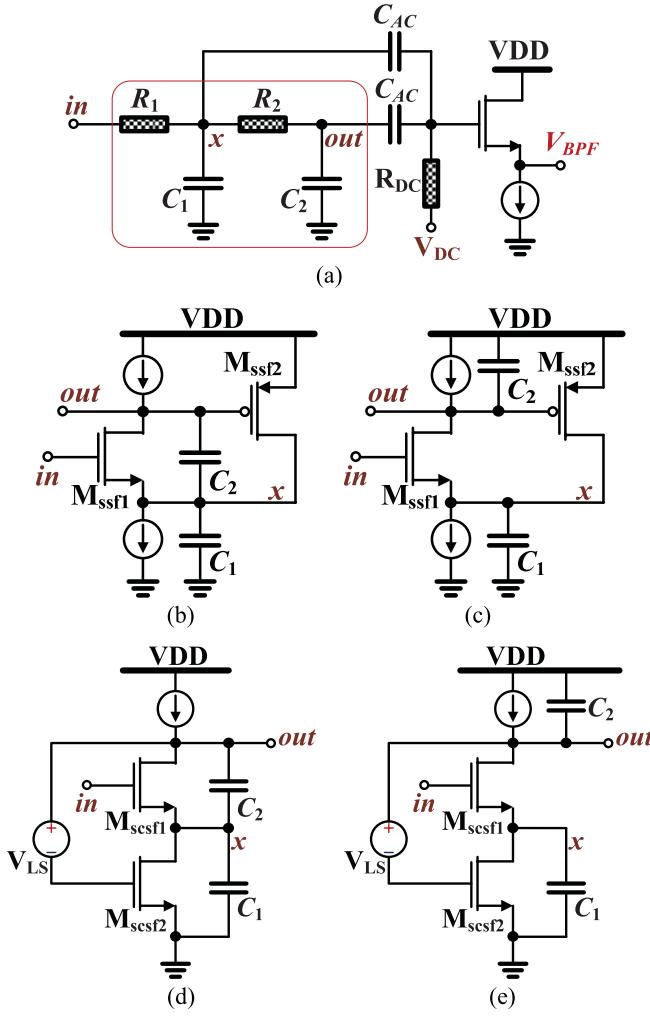


Fig. 26. (a) Passive summation by two capacitors C_{AC} . (b) and (c) Two SSF-based BPFs. (d) and (e) Two self-coupled SFB-BPFs, where V_{LS} is a level-shifting dc voltage source.

limited at a high Q setting. A mutation of the SSF-based BPF similar to the topology in [63] is shown in Fig. 26(c) with the characteristic parameters given below

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{g_{m1}g_{m2}}{C_1C_2}}, Q = \sqrt{\frac{g_{m2}C_1}{g_{m1}C_2}}, A_{f0} = \frac{C_1}{C_2}. \quad (10)$$

The quadratic dependence of A_{f0} on Q is avoided, but a large ratio of g_{m2}/g_{m1} and/or C_1/C_2 is still needed to obtain a high Q , which brings challenges in low-voltage and area-constraint design. Another two BPF topologies shown in Fig. 26(d) and (e) adapted from the recently published self-coupled SFB-LPF [64] are in principle the same as the SSF-based BPF, and consequently Fig. 26(d) and (e) has the same characteristic parameters as in (9) and (10), respectively, except that the transconductance values of M_{scsf1} and M_{scsf2} are forced to be the same in subthreshold due to current reuse.

The quality of the reconstructed speech from the chip output events cannot yet compete with conventional digital representation enabled by high-precision clocked ADCs. However, we share the same viewpoint as in [17] that such a cochlea-like front end, which uses low-power analog signal processing for

feature extraction, does not aim for conventional multimedia applications where the quality of reconstructed signals may be of paramount importance, but rather it targets applications in ubiquitous smart sensing where power consumption needs to be minimized. Although [17] argues that the front-end nonidealities, such as f_0 and Q variation, can be absorbed by the machine-learning back-end, resulting in a complete system that may still maintain good classification/recognition accuracy, the variation reduction technique presented in this paper is still valuable in our opinion. It is an open question as quantitatively to what extent the back-end processing can tolerate the front-end nonidealities with acceptable degradation of performance in terms of classification/recognition accuracy, and whether such chip-by-chip machine-learning calibration methods are affordable in mass production.

ACKNOWLEDGMENT

The authors would like to acknowledge the support from R. Berner and C. Brändli with AER, V. Villanueva with PCB, L. Longinotti with firmware, and TowerJazz with chip fabrication. They would also like to thank the anonymous reviewers for their helpful comments.

REFERENCES

- [1] A. Bertrand, "Signal processing algorithms for wireless acoustic sensor networks," Ph.D. dissertation, Dept. Elect. Eng., Katholieke Univ. Leuven, Leuven, Belgium, 2011.
- [2] A. Raychowdhury, C. Tokunaga, W. Beltman, M. Deisher, J. W. Tschanz, and V. De, "A 2.3 nJ/frame voice activity detector-based audio front-end for context-aware system-on-chip applications in 32-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 8, pp. 1963–1969, Aug. 2013.
- [3] M. Price, J. Glass, and A. P. Chandrakasan, "A 6 mW, 5,000-word real-time speech recognizer using WFST models," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 102–112, Jan. 2015.
- [4] P. V. Rajesh et al., "A 172 μ W compressive sampling photoplethysmographic readout with embedded direct heart-rate and variability extraction from compressively sampled data," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, Jan./Feb. 2016, pp. 386–387.
- [5] R. F. Lyon and C. Mead, "An analog electronic cochlea," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 36, no. 7, pp. 1119–1134, Jul. 1988.
- [6] L. Watts, D. A. Kerns, R. F. Lyon, and C. A. Mead, "Improved implementation of the silicon cochlea," *IEEE J. Solid-State Circuits*, vol. 27, no. 5, pp. 692–700, May 1992.
- [7] R. Sarapeshkar, M. W. Baker, C. D. Salthouse, J.-J. Sit, L. Turicchia, and S. M. Zhak, "An analog bionic ear processor with zero-crossing detection," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, Feb. 2005, pp. 78–79.
- [8] E. Fragniere, "A 100-channel analog CMOS auditory filter bank for speech recognition," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, Feb. 2005, pp. 140–141.
- [9] B. Wen and K. Boahen, "A 360-channel speech preprocessor that emulates the cochlear amplifier," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, Feb. 2006, pp. 556–557.
- [10] V. Chan, S.-C. Liu, and A. van Schaik, "AER EAR: A matched silicon cochlea pair with address event representation interface," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 1, pp. 48–59, Jan. 2007.
- [11] T. J. Hamilton, C. Jin, A. van Schaik, and J. Tapson, "An active 2-D silicon cochlea," *IEEE Trans. Biomed. Circuits Syst.*, vol. 2, no. 1, pp. 30–43, Mar. 2008.
- [12] A. G. Katsiamis, E. M. Drakakis, and R. F. Lyon, "A biomimetic, 4.5 μ W, 120-dB, log-domain cochlea channel with AGC," *IEEE J. Solid-State Circuits*, vol. 44, no. 3, pp. 1006–1022, Mar. 2009.
- [13] S.-C. Liu, A. van Schaik, B. A. Minch, and T. Delbrück, "Asynchronous binural spatial audition sensor with 2 \times 64 \times 4 channel output," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 4, pp. 453–464, Aug. 2014.
- [14] G. Yang, R. F. Lyon, and E. M. Drakakis, "A 6 μ W per channel analog biomimetic cochlear implant processor filterbank architecture with across channels AGC," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 1, pp. 72–86, Feb. 2015.

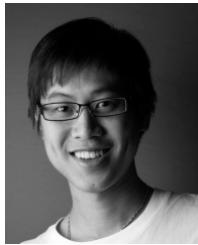
- [15] S. Wang, T. J. Koickal, A. Hamilton, R. Cheung, and L. S. Smith, "A bio-realistic analog CMOS cochlea filter with high tunability and ultra-steep roll-off," *IEEE Trans. Biomed. Circuits Syst.*, vol. 9, no. 3, pp. 297–311, Jun. 2015.
- [16] R. Sarpeshkar, "Analog versus digital: Extrapolating from electronics to neurobiology," *Neural Comput.*, vol. 10, no. 7, pp. 1601–1638, Oct. 1998.
- [17] M. Verhelst and A. Bahai, "Where analog meets digital: Analog-to-information conversion and beyond," *IEEE Solid-State Circuits Mag.*, vol. 7, no. 3, pp. 67–80, Sep. 2015.
- [18] H. Inose, T. Aoki, and K. Watanabe, "Asynchronous delta-modulation system," *Electron. Lett.*, vol. 2, no. 3, pp. 95–96, Mar. 1966.
- [19] B. Schell and Y. Tsividis, "A continuous-time ADC/DSP/DAC system with no clock and with activity-dependent power dissipation," *IEEE J. Solid-State Circuits*, vol. 43, no. 11, pp. 2472–2481, Nov. 2008.
- [20] P. A. Merolla *et al.*, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668–673, Aug. 2014.
- [21] M. Yip, R. Jin, H. H. Nakajima, K. M. Stankovic, and A. P. Chandrakasan, "A fully-implantable cochlear implant SoC with piezoelectric middle-ear sensor and arbitrary waveform neural stimulation," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 214–229, Jan. 2015.
- [22] M. Yang, C.-H. Chien, T. Delbrück, and S.-C. Liu, "A 0.5V 55 μ W 64×2-channel binaural silicon cochlea for event-driven stereo-audio sensing," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, Jan./Feb. 2016, pp. 388–389.
- [23] S. D'Amico, M. Conta, and A. Baschirotto, "A 4.1-mW 10-MHz fourth-order source-follower-based continuous-time filter with 79-dB DR," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2713–2719, Dec. 2006.
- [24] S. D'Amico, M. De Matteis, and A. Baschirotto, "A 6th-order 100 μ A 280 MHz source-follower-based single-loop continuous-time filter," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, Feb. 2008, pp. 72–73.
- [25] M. Yang, J. Liu, Y. Xiao, and H. Liao, "14.4 nW fourth-order bandpass filter for biomedical applications," *Electron. Lett.*, vol. 46, no. 14, pp. 973–974, Jul. 2010.
- [26] C. Sawigun, W. Ngamkham, and W. A. Serdijn, "A 0.5-V, 2-nW, 55-dB DR, fourth-order bandpass filter using single branch biquads: An efficient design for FoM enhancement," *Microelectron. J.*, vol. 45, no. 4, pp. 367–374, Apr. 2014.
- [27] Y. Chen, P.-I. Mak, L. Zhang, and Y. Wang, "0.07 mm², 2 mW, 75 MHz-IF, fourth-order BPF using source-follower-based resonator in 90 nm CMOS," *Electron. Lett.*, vol. 48, no. 10, pp. 552–554, May 2012.
- [28] R. Steele, *Delta Modulation Systems*. Mountain View, CA, USA: Pentech, 1975.
- [29] L. C. Gouveia, T. J. Koickal, and A. Hamilton, "An asynchronous spike event coding scheme for programmable analog arrays," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 4, pp. 791–799, Apr. 2011.
- [30] M. Trakimas and S. R. Sonkusale, "An adaptive resolution asynchronous ADC architecture for data compression in energy constrained sensing applications," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 5, pp. 921–934, May 2011.
- [31] Y. Li, D. Zhao, and W. A. Serdijn, "A sub-microwatt asynchronous level-crossing ADC for biomedical applications," *IEEE Trans. Biomed. Circuits Syst.*, vol. 7, no. 2, pp. 149–157, Apr. 2013.
- [32] W. Tang *et al.*, "Continuous time level crossing sampling ADC for bio-potential recording systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 6, pp. 1407–1418, Jun. 2013.
- [33] C. Weltin-Wu and Y. Tsividis, "An event-driven clockless level-crossing ADC with signal-dependent adaptive resolution," *IEEE J. Solid-State Circuits*, vol. 48, no. 9, pp. 2180–2190, Sep. 2013.
- [34] S. Patil, A. Ratiu, D. Morche, and Y. Tsividis, "A 3–10 fJ/conv-step 0.0032 mm² error-shaping alias-free asynchronous ADC," in *Proc. IEEE Symp. VLSI Circuits (VLSI Circuits)*, Jun. 2015, pp. 160–161.
- [35] B. Razavi, "The StrongARM latch [a circuit for all seasons]," *IEEE Solid-State Circuits Mag.*, vol. 7, no. 2, pp. 12–17, Jun. 2015.
- [36] S.-W. M. Chen and R. W. Brodersen, "A 6-bit 600-MS/s 5.3-mW asynchronous ADC in 0.13- μ m CMOS," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, pp. 2669–2680, Dec. 2006.
- [37] P. Harpe, E. Cantatore, and A. van Roermund, "A 10b/12b 40 kS/S SAR ADC with data-driven noise reduction achieving up to 10.1b ENOB at 2.2 fJ/conversion-step," *IEEE J. Solid-State Circuits*, vol. 48, no. 12, pp. 3011–3018, Dec. 2013.
- [38] K. A. Boahen, "A burst-mode word-serial address-event link—I: Transmitter design," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 7, pp. 1269–1280, Jul. 2004.
- [39] M. Yang, S.-C. Liu, C. Li, and T. Delbrück, "Addressable current reference array with 170 dB dynamic range," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2012, pp. 3110–3113.
- [40] K. Bult and G. J. G. M. Geelen, "An inherently linear and compact MOST-only current division technique," *IEEE J. Solid-State Circuits*, vol. 27, no. 12, pp. 1730–1735, Dec. 1992.
- [41] B. Linares-Barranco and T. Serrano-Gotarredona, "On the design and characterization of femtoampere current-mode circuits," *IEEE J. Solid-State Circuits*, vol. 38, no. 8, pp. 1353–1363, Aug. 2003.
- [42] B. Gilbert, "Translinear circuits: A proposed classification," *Electron. Lett.*, vol. 11, no. 1, pp. 14–16, Jan. 1975.
- [43] C. Mead, *Analog VLSI and Neural Systems*. Reading, MA, USA: Addison-Wesley, 1989.
- [44] W. Cheng, M. S. O. Alink, A. J. Annema, G. J. M. Wienk, and B. Nauta, "A wideband IM3 cancellation technique for CMOS II- and T-attenuators," *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 358–368, Feb. 2013.
- [45] M. Yang, "Silicon retina and cochlea with asynchronous delta modulator for spike encoding," Ph.D. dissertation, Dept. Phys., ETH Zürich, Zürich, Switzerland, 2015.
- [46] D. Han, Y. Zheng, R. Rajkumar, G. Dawe, and M. Je, "A 0.45V 100-channel neural-recording IC with sub- μ W/channel consumption in 0.18 μ m CMOS," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, Feb. 2013, pp. 290–291.
- [47] M. Taherzadeh-Sani and A. A. Hamoui, "A 1-V process-insensitive current-scalable two-stage opamp with enhanced DC gain and settling behavior in 65-nm digital CMOS," *IEEE J. Solid-State Circuits*, vol. 46, no. 3, pp. 660–668, Mar. 2011.
- [48] M. Yang, S.-C. Liu, and T. Delbrück, "Subthreshold DC-gain enhancement by exploiting small size effects of MOSFETs," *Electron. Lett.*, vol. 50, no. 11, pp. 835–837, May 2014.
- [49] A. Yoshizawa and Y. Tsividis, "A channel-select filter with agile blocker detection and adaptive power dissipation," *IEEE J. Solid-State Circuits*, vol. 42, no. 5, pp. 1090–1099, May 2007.
- [50] R. R. Harrison, "The design of integrated circuits to observe brain activity," *Proc. IEEE*, vol. 96, no. 7, pp. 1203–1216, Jul. 2008.
- [51] M. Yang, S.-C. Liu, and T. Delbrück, "A dynamic vision sensor with 1% temporal contrast sensitivity and in-pixel asynchronous delta modulator for event encoding," *IEEE J. Solid-State Circuits*, vol. 50, no. 9, pp. 2149–2160, Sep. 2015.
- [52] M. De Matteis, S. D'Amico, and A. Baschirotto, "A 0.55 V 60 dB-DR fourth-order analog baseband filter," *IEEE J. Solid-State Circuits*, vol. 44, no. 9, pp. 2525–2534, Sep. 2009.
- [53] X. Zou, X. Xu, L. Yao, and Y. Lian, "A 1-V 450-nW fully integrated programmable biomedical sensor interface chip," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1067–1077, Apr. 2009.
- [54] K. Ishida, K. Kanda, A. Tamtrakarn, H. Kawaguchi, and T. Sakurai, "Managing subthreshold leakage in charge-based analog circuits with low-V_{TH} transistors by analog T-switch (AT-switch) and super cut-off CMOS (SCCMOS)," *IEEE J. Solid-State Circuits*, vol. 41, no. 4, pp. 859–867, Apr. 2006.
- [55] M. van Elzakker, E. van Tuijl, P. Geraedts, D. Schinkel, E. A. M. Klumperink, and B. Nauta, "A 10-bit charge-redistribution ADC consuming 1.9 μ W at 1 MS/s," *IEEE J. Solid-State Circuits*, vol. 45, no. 5, pp. 1007–1015, May 2010.
- [56] S. Hanson *et al.*, "A low-voltage processor for sensing applications with picowatt standby mode," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1145–1155, Apr. 2009.
- [57] R. G. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, and T. Mudge, "Near-threshold computing: Reclaiming Moore's law through energy efficient integrated circuits," *Proc. IEEE*, vol. 98, no. 2, pp. 253–266, Feb. 2010.
- [58] A. T. Zai, S. Bhargava, N. Mesgarani, and S.-C. Liu, "Reconstruction of audio waveforms from spike trains of artificial cochlea models," *Frontiers Neurosci.*, vol. 9, 347, pp. 1–13, Oct. 2015.
- [59] K. M. H. Badami, S. Lauwereins, W. Meert, and M. Verhelst, "A 90 nm CMOS, 6 μ W power-proportional acoustic sensing frontend for voice activity detection," *IEEE J. Solid-State Circuits*, vol. 51, no. 1, pp. 291–302, Jan. 2016.
- [60] B. Rumberg, D. W. Graham, V. Kulathumani, and R. Fernandez, "Hibernets: Energy-efficient sensor networks using analog signal processing," *IEEE Trans. Emerg. Sel. Topics Circuits Syst.*, vol. 1, no. 3, pp. 321–334, Sep. 2011.

- [61] S. Ramakrishnan, A. Basu, L. K. Chiu, J. Hasler, D. Anderson, and S. Brink, "Speech processing on a reconfigurable analog platform," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 22, no. 2, pp. 430–433, Feb. 2014.
- [62] M. De Matteis, A. Pezzotta, S. D'Amico, and A. Baschirotto, "A 33 MHz 70 dB-SNR super-source-follower-based low-pass analog filter," *IEEE J. Solid-State Circuits*, vol. 50, no. 7, pp. 1516–1524, Jul. 2015.
- [63] Z. Gao, J. Ma, M. Yu, and Y. Ye, "A fully integrated CMOS active bandpass filter for multiband RF front-ends," *IEEE Trans. Circuits Syst. II, Express Briefs*, vol. 55, no. 8, pp. 718–722, Aug. 2008.
- [64] Y. Xu, S. Leuenberger, P. K. Venkatachala, and U. Moon, "A 0.6 mW 31 MHz 4th-order low-pass filter with +29dBm IIP3 using self-coupled source follower based biquads in 0.18 μ m CMOS," in *Proc. IEEE Symp. VLSI Circuits*, Jun. 2016, pp. 132–133.



Minhao Yang (S'11–M'16) received the Ph.D. degree in physics from ETH Zürich in 2015.

He is currently a Post-Doctoral Fellow with Columbia University funded by SNF Early Postdoc Mobility Fellowship. His current research interests include spike coding and processing, low-power spiking sensors with embedded processing, and silicon retina and cochlea.



Chen-Han Chien (S'13) received the B.S. and M.S. degrees in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 2006 and 2008, respectively. He is currently pursuing the Ph.D. degree with the Institute of Neuroinformatics, University of Zürich and ETH Zürich, Zürich, Switzerland.

His current research interests include probabilistic neural computation, neuromorphic VLSI design, low-power event-based visual and auditory sensors, and spike coding and processing.



Tobi Delbrück (M'99–SM'06–F'13) received the B.Sc. degree in physics and applied mathematics from the University of California, San Diego, CA, USA, and the Ph.D. degree from the California Institute of Technology, in 1986 and 1993, respectively.

He has been a Professor of Physics with the Institute of Neuroinformatics, University of Zürich and ETH Zürich, Switzerland, since 1998. His group focuses on neuromorphic sensory processing. He worked on electronic imaging at Arithmos, Synaptics, National Semiconductor, and Foveon.

Dr. Delbrück has co-organized the Telluride Neuromorphic Cognition Engineering summer workshop and the live demonstration sessions at International Symposium on Circuits and Systems. He is also co-founder of iniLabs and Insightness. was the Chair of the IEEE CAS Sensory Systems Technical Committee, is currently the Secretary of the IEEE Swiss CAS/ED Society, and an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS. He has received nine IEEE awards.



Shih-Chii Liu (M'02–SM'07) received the Bachelor's degree in electrical engineering and the Ph.D. degree in the computation and neural systems program from the California Institute of Technology in 1997.

She worked at various companies, including Gould American Microsystems, LSI Logic, and Rockwell International Research Labs. She is currently a Group Leader with the Institute of Neuroinformatics, University of Zürich and ETH Zürich, Switzerland. Her current research interests include neuromorphic visual and auditory sensors, cortical processing circuits, and event-driven circuits, networks, and algorithms.

Dr. Liu was the Chair of the IEEE CAS Sensory Systems and Neural Systems and Applications Technical Committees. She is currently the Chair of the IEEE Swiss CAS/ED Society and an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS and *Neural Networks Journal*.