# A 6×5×4mm³ General Purpose Audio Sensor Node with a 4.7µW Audio Processing IC

Minchang Cho[1], Sechang Oh[1], Seokhyeon Jeong[1], Yiqun Zhang[1], Inhee Lee[1], Yejoong Kim[1], Li-Xuan Chuo[1], Dongkwun Kim[1],
Qing Dong[1], Yen-Po Chen[1], Martin Lim[2], Mike Daneman[2], David Blaauw[1], Dennis Sylvester[1], Hun-Seok Kim[1]

[1]University of Michigan, Ann Arbor, MI, USA, [2]Invensense, San Jose, CA, USA

## Abstract

We present a complete, fully functional energy-autonomous audio sensor node with 6×5×4mm³ form factor. The system uses a new audio processing IC integrated with a MEMS microphone, general purpose 32-bit processor, 8Mb Flash, RF transceiver with custom 3D antenna, PV cells for energy harvesting and battery. The 4.7µW audio processing IC performs audio acquisition with 4-32× compression. The complete stand-alone system achieves 38mins of speech recording and energy-autonomous operation in room light.

## Introduction

Realizing a millimeter-scale audio processing platform enables a number of new IoT applications such as distributed audio recording, event logging, and security monitoring. While several efforts [1-2] have sought to miniaturize audio sensors, their centimeter-scale volume and >20mW power severely limit use as an unobtrusive, self-powered sensing node. This work is the first to demonstrate a fully functional and self-contained audio sensor node in millimeter-scale size including recording, storage and transmission of the recording over a 20m wireless link all when operating on a 16µAh thin-film battery. The complete audio sensor is enabled by stacked-die integration of a new audio processing IC with other system components. The audio IC consumes only 4.7µW for signal acquisition and compression.

## Audio Processing IC

Fig. 1 shows the architecture of the proposed audio processing IC that integrates AFE, ADC and compression engine. The AFE and ADC operate at 1.4V and 0.9 V respectively, directly provided by on-chip LDOs from the battery (3.6-4.2V) to decouple them from the noisy digital supply. The compression engine operates at 0.6V with standard $V_{th}$ transistors. These logic blocks are power gated in sleep mode. The 1.2V bus controller and configuration register file are always on and thus use high $V_{th}$ devices, reducing leakage.

The AFE shown in Fig. 2 consists of LNA, variable gain amplifier (VGA) and charge pump. The charge pump biases the MEMS transducer at 10V. The LNA gain (29dB) is set by $C_M/C_{F1}$, and the gain and bandwidth of VGA are tuned by $C_{I2}$ and $C_{L2}$, respectively. $R_{F1, 2}$ sets input common mode voltage and removes offset. To maximize noise efficiency, $OTA_1$ and $OTA_2$ use inverter-based cascode amplifier and their input-pair transistors operate in subthreshold region. Fig. 2 shows measured input referred noise (IRN) spectrum of AFE with the MEMS transducer. The LNA and VGA consume 1.54µW and 1.11µW, respectively, to achieve 20.1µV$_{rms}$ IRN with 4kHz of bandwidth. The 8bit synchronous SAR ADC (Fig. 2) operates on two separate clocks: an 8kHz clock (CLK_S) for sample and hold obtained from a 32kHz crystal and an inaccurate 150kHz clock (CLK_F) for internal ADC control from a power-efficient RO at 0.6V.

Compression of the audio stream is critical to minimize Flash storage size, access power and RF transmission power. In the proposed compression algorithm (Fig. 3 (a)), incoming samples are first converted to the frequency domain using polyphase subband filtering. The power of each subband is accumulated during 1 frame (6 samples) and then N subbands with the highest power are selected. All subbands whose power is lower than a programmable threshold are eliminated. To reduce power, we apply a mathematically-equivalent but computationally efficient polyphase quadrature filtering (PQF) [3] that uses inverse FFT for its inverse discrete cosine transform (Fig. 3 (d)). Complexity is reduced by 94% in total from this algorithm optimization as shown in Fig. 3 (c). A comparison between the proposed and other off-the-shelf algorithms is shown in Fig. 3 (b). Maintaining similar sound quality, the proposed algorithm has 1000× lower complexity than CELP and 3.9× better compression than ADPCM.

The compression engine architecture is shown in Fig. 4. The proposed PQF operates on 512 consecutive samples. Data streaming is handled in a block basis using a 32-entry FIFO. Until the next block shift, registers are clock gated, resulting in 32× power reduction. We observe that 25% of PQF coefficients are zero, allowing us to inactivate unused samples. Compression is performed in a frame basis with clock gating to avoid unnecessary data switching on shared data bus, buffers and computation logic (59% power reduction; simulated). The proposed sorting unit uses a tree structure (Fig. 5) where all PEs compare and forward their inputs in the 1st cycle to obtain the top result. Then, in each subsequent cycle the winning PE zeros its value and only its path is updated to produce the next highest values. Compared with a conventional parallel sorter, such as bitonic, this implementation shows 42% less dynamic energy for sorting top 16 out of 32. After pruning, subband power values are log-domain quantized with leading-one detector, implemented with round approximation.

## Complete Integrated System

The complete system (Fig. 6) consists of 6 heterogeneous stacked ICs: 1) The proposed audio processor acquires and compresses audio signal. 2) 8Mb of custom embedded Flash [4] stores compressed audio at 120pJ/bit. 3) An RF transceiver co-designed with a 3D antenna [5] communicates with a gateway up to 20m away. 4) The energy harvester charges the battery using stacked photovoltaic (PV) cells [6] and also protects it from reverse current. 5) The power management unit [7] converts battery voltage to 1.2V and 0.6V to provide multiple voltage domains to the ICs. 6) An ARM Cortex M0 processor coordinates system operation and enables additional signal processing such as event detection. The stacked ICs communicate via ultra-low power Mbus [8]. The system integration strategy is carefully devised to achieve minimal volume. On the bottom side of a custom 6×5 mm² PCB substrate, we stack 2 rechargeable thin-film Li batteries together with ICs. We place a MEMS transducer directly adjacent to the stacked ICs to minimize the system volume and improve SNR by limiting the parasitic capacitance between the transducer and Audio IC. A 3D-printed custom lid covers all electronics, including a 32kHz crystal and 3 caps for the RF transceiver to generate an acoustic back chamber. By combining the cavity for the sound chamber with the location of all electronics, system volume is aggressively reduced and also protects the electronics from light. At the same time, air volume is also increased compared to a commercial package which improves the microphone's sensitivity and low frequency response. The top side contains a sound hole for air passage, and a 3D magnetic dipole antenna. The magnetic dipole does not require physical separation from the electronics, further enabling compact integration. PV cells are mounted on top of the antenna and covered with clear epoxy to provide protection while allowing light to reach the PV cells.

## Measurements

The proposed audio processing IC is fabricated in 180nm CMOS. Measured A-weighted input referred noise of AFE is 13.2µVrms (Fig. 2), showing 61dBA of SNR at 94dBSPL (1kHz) input sound. As shown in Fig. 7, the proposed IC compresses audio signals with a variable rate, >15x for speech, enabling 38mins of recording with 8Mb custom Flash. Measured compression ratio and quality tradeoff is controlled by N (Fig. 8) and power threshold setting. The audio processing IC consumes 4.73µW including 1.44µW from the compression engine (Fig. 9). Fully functional operation, including audio acquisition, compression, storage and RF transmission, of a mm-scale unit identical to that pictured in Fig. 6 was demonstrated when operating stand-alone powered only by its internal battery and energy harvesting. Measured power profile of the stand-alone operation is shown in Fig. 11. With harvesting from 2.6×3mm² PV cell (1klux), 10.5hr of charge recovery time is needed after 38mins of recording. The measured RF transmission power is 79µW, and the system sleep power is 7.2nW. Measured system parameters are summarized in Fig. 12.
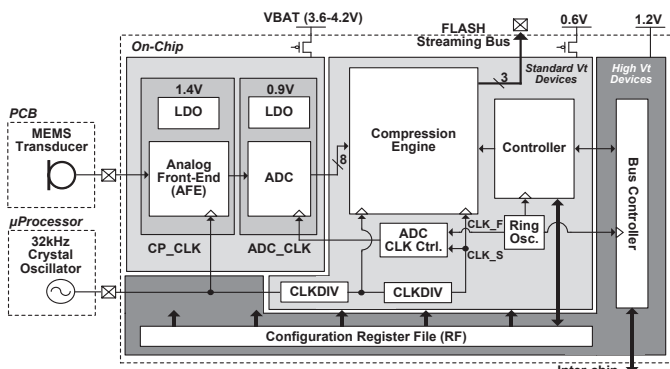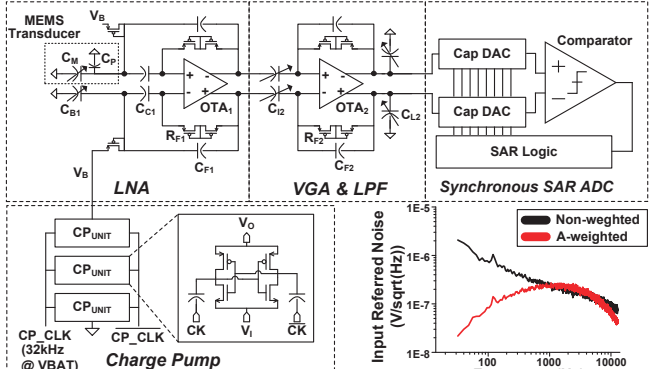
**Fig. 1. Overall architecture of audio processing IC.**



**Fig. 2. Analog front-end (AFE), ADC and the measured input referred noise spectrum of AFE (bottom right).**



$$s_t[i] = \sum_{k=0}^{63} y[k] \times \cos\left[\frac{(2i+1)(k-16)\pi}{64}\right] \rightarrow \text{PQF}$$

$$= \sum_{k=0}^{31} y'[k] \times \cos\left[\frac{(2i+1)k\pi}{64}\right] \rightarrow \text{IDCT conversion}$$

$$y'[k] = \begin{cases} y[16] & (k=0) \\ y[k+16] + y[16-k] & (k=1,2,\cdots,16) \\ y[k+16] - y[80-k] & (k=17,18,\cdots,31) \end{cases}$$

$$\text{let } X[k] = y'[k] \times e^{\frac{jk\pi}{2N}}, \quad k = 0,1,2,\cdots,31$$

$$x[m] = Re[FFT^{-1}(X[k])], \quad m = 0,1,2,\cdots,31$$

$$s_t[i] = \begin{cases} s_t[2i] = x[i], & i = 0,1,2,\cdots,15 \\ s_t[2i+1] = x[31-i], & i = 0,1,2,\cdots,15 \end{cases}$$
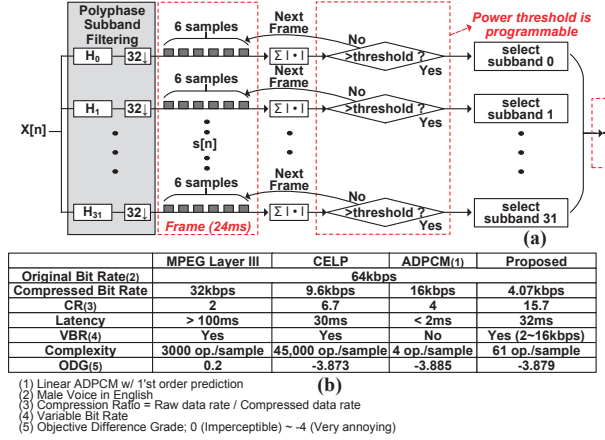
*Polyphase subband filtered sequence*

**Fig. 3. (a) Proposed compression algorithm and (b) performances with comparison; (c) The complexity reduction and (d) principles of optimization.**
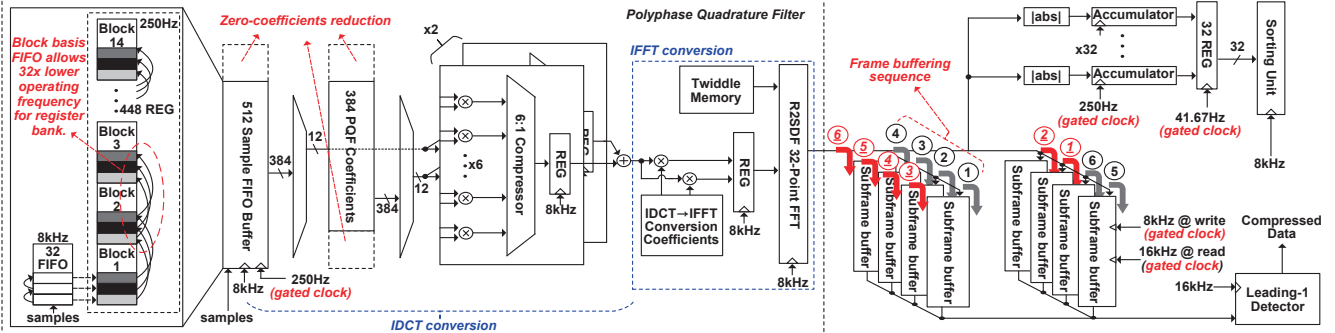
| | MPEG Layer III | CELP | ADPCM(1) | Proposed |
|---|---|---|---|---|
| Original Bit Rate(2) | 64kbps | | | |
| Compressed Bit Rate | 32kbps | 9.6kbps | 16kbps | 4.07kbps |
| CR(3) | 2 | 6.7 | 4 | 15.7 |
| Latency | > 100ms | 30ms | < 2ms | 32ms |
| VBR(4) | Yes | Yes | No | Yes (2~16kbps) |
| Complexity | 3000 op./sample | 45,000 op./sample | 4 op./sample | 61 op./sample |
| ODG(5) | 0.2 | -3.873 | -3.885 | -3.879 |

(1) Linear ADPCM w/ 1'st order prediction
(2) Male Voice in English
(3) Compression Ratio = Raw data rate / Compressed data rate
(4) Variable Bit Rate
(5) Objective Difference Grade; 0 (Imperceptible) ~ -4 (Very annoying)

**(b)**



**Fig. 4. Proposed compression engine architecture with power reduction techniques.**



Complexity:
(sort N/2 out of N)
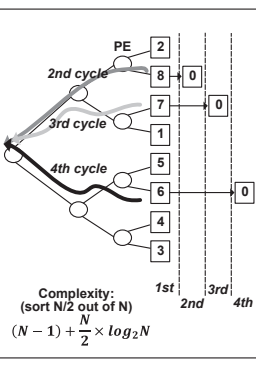
$$(N-1) + \frac{N}{2} \times log_2 N$$

**Fig. 5. Proposed sorting unit.**



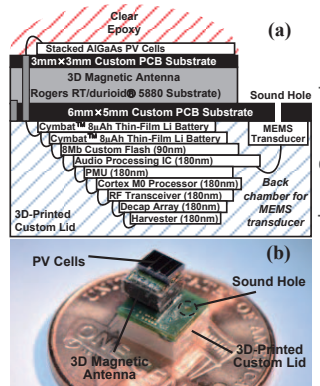**Fig. 6. The 6×5×4mm³ audio sensor node: (a) cross sectional diagram, (b) top-facing view, and (c) bottom-facing view.**
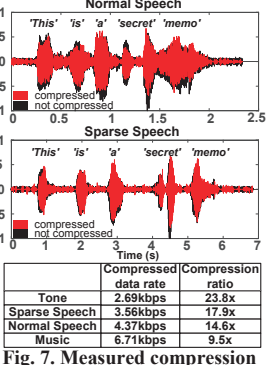


| | Compressed data rate | Compression ratio |
|---|---|---|
| Tone | 2.69kbps | 23.8x |
| Sparse Speech | 3.56kbps | 17.9x |
| Normal Speech | 4.37kbps | 14.6x |
| Music | 6.71kbps | 9.5x |

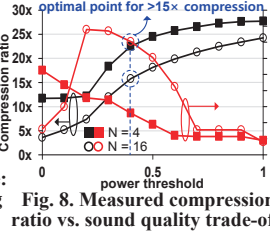**Fig. 7. Measured compression engine performance (averaged).**



**Fig. 8. Measured compression ratio vs. sound quality trade-off.**
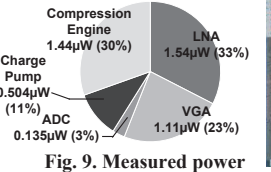


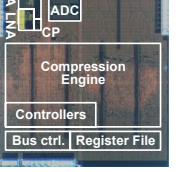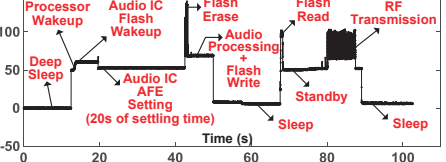**Fig. 9. Measured power breakdown of Audio IC.**



**Fig. 10. Die photo.**



**Fig. 11. Measured power profile of audio sensor node.**

| Technology | 180nm | Dimensions | 6×5×4 mm³ |
|---|---|---|---|
| Die Area | 2.5mm × 2.85mm | Storage | 8Mb Flash / 24Kb SRAM |
| Supply Voltage | 0.6V / 1.2V / 3.6-4.2V | | |
| AFE Gain | 30 ~ 44.7dB | Processor | ARM Cortex M0 |
| AFE Bandwidth | 4kHz | Battery | 16µAh Thin-Film |
| Input Referred Noise | 13.2µVrms (A-weighted) | Energy Harvesting | Solar, 4.07µW @ 1klux |
| SNR@94dBSPL | 60.6dBA | RF TX Energy | 3.67nJ/bit |
| Clock Frequency | Ctrl. 150kHz | Power Consumption | Sleep 7.2nW |
| | Engine 8/16kHz | | Standby 50µW |
| CR | 4-32×(speech:>15×) | | Audio processing +Flash write 68µW |
| Power | 4.73µW | | |

**Fig. 12. Performance summary of Audio IC (left) and the complete sensor system (right).**

**References**

[1] C. Pham *et al.*, *LCNW*, 2014.
[2] G. Zhao *et al.*, *WoWMoM*, 2011.
[3] J. H. Rothweiler, *ICASSP*, 1983.
[4] Q. Dong *et al.*, *ISSCC*, 2017.
[5] L.-X. Chuo *et al.*, *ISSCC*, 2017.
[6] I. LEE *et al.*, *ISSCC*, 2016.
[7] W. Jung *et al.*, *ISSCC*, 2016.
[8] Y.-S. Kuo *et al.*, *CICC*, 2014.