

Always-On 12-nW Acoustic Sensing and Object Recognition Microsystem for Unattended Ground Sensor Nodes

Seokhyeon Jeong, *Student Member, IEEE*, Yu Chen, Taekwang Jang, *Student Member, IEEE*, Julius Ming-Lin Tsai, David Blaauw, *Fellow, IEEE*, Hun-Seok Kim, *Member, IEEE*, and Dennis Sylvester, *Fellow, IEEE*

Abstract—This paper presents an ultra-low power acoustic sensing and object recognition microsystem for Internet of Things applications. The microsystem is targeted for unattended ground sensor nodes where long-term (decades) life time is desired without the need for battery replacement. The system incorporates an microelectromechanical systems microphone as a frontend sensor along with active circuitry to identify target objects. We introduce an algorithm-circuit cross optimization to realize a 12-nW stand-alone microsystem that integrates the analog frontend with the digital backend signal classifier. The frequency-domain analysis of target audio signals reveals that the system can operate with a relatively low bandwidth (<500 Hz) and SNR (>3 dB) which significantly relaxes power constraints on both analog frontend and digital backend circuits. To further relax the current requirement of the preceding amplifier, we propose an 8-bit SAR-analog-to-digital converter that is designed to have a highly reduced sampling capacitance (<50 fF). For the digital backend, we propose a feature extractor using the serialized tones-of-interest discrete Fourier transform, replacing a conventional high-power/area-consuming parallel feature extraction using the fast Fourier transform. This approach reduces area and thus leakage power which often dominates the overall power consumption. The proposed system successfully identifies a number of target objects including an electrical generator, a small car, and a truck with $>95\%$ reliability and consumes only 12 nW with continuous monitoring.

Index Terms—Amplifier, analog-to-digital converter (ADC), discrete Fourier transform (DFT), low-noise, microphone, subthreshold, support vector machine (SVM), ultra-low power (ULP).

I. INTRODUCTION

INTERNET of Things (IoT) devices are becoming increasingly intelligent and context aware. Such context awareness has been enabled by various types of sensors that are “always-on.” In the past, sensors on mobile platforms were activated periodically or passively by the user to constrain their battery power consumption. As a result, information loss

on surrounding activities was unavoidable, making them far from being “ambient intelligent.” On the other hand, recent mobile devices in the market are distinguished by their always-on functionality while adopting a full spectrum of sensors that ranges from accelerometers, gyroscopes, magnetometers to acoustic, image, and pressure sensors. Among these various sensory inputs used to realize context-aware intelligence, sound is an attractive sensory modality that is information rich but not as computationally demanding as other alternative modalities such as vision or radar. It has a relatively low bandwidth of <20 kHz and also serves as a natural user interface. The use of always listening technology has become popular in various applications such as voice activated intelligent personal assistants. New applications of always-on intelligent acoustic sensing includes agricultural monitoring to detect pests or precipitation, infrastructure health tracking to recognize acoustic symptoms of structural changes, and security/safety monitoring to identify intruders or dangerous conditions. This paper focuses on incorporating such always-on technology into small and extremely power constrained IoT devices.

The remainder of the paper is organized as follows. Section II provides an overview of the proposed system. Sections III through V present a detailed circuit description and analysis of system components: analog frontend, ADC, and digital backend. Section VI describes measurement results of the test chip. Finally, Section VII summarizes the key contributions of the work and concludes the paper.

A major challenge for the adoption of always-on, context-aware sensing in ultra-small unattended IoT devices is low-power consumption, as these devices require long-term operation without battery replacement. Ideally, unattended systems should be able to operate perpetually using harvested energy or should have decades of lifetime on small batteries. System power consumption less than 20 nW is required in order to run continuously with a 1-mm² solar cell under dim indoor light (100 lux), or to sustain ten years lifetime using a small coin cell battery (4.8 mm, 2 mAh). Current state-of-the-art acoustic sensing systems [1], [2] show power consumption in the microwatt range, which is more than two orders of magnitude higher than the proposed 20-nW target. A 3-nW ultra-low power (ULP) signal acquisition IC was introduced in [3] but it does not include backend signal classification.

Manuscript received April 15, 2017; revised June 20, 2017; accepted July 3, 2017. Date of publication September 4, 2017; date of current version December 26, 2017. This paper was approved by Guest Editor Pui-In Mak. This work was supported by the Defense Advanced Research Projects Agency (DARPA). (*Corresponding author: Seokhyeon Jeong.*)

S. Jeong, Y. Chen, T. Jang, D. Blaauw, H.-S. Kim, and D. Sylvester are with the University of Michigan, Ann Arbor, MI 48109 USA (e-mail: seojeong@umich.edu).

J. M.-L. Tsai is with InvenSense, San Jose, CA 95110 USA.

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.2017.2728787

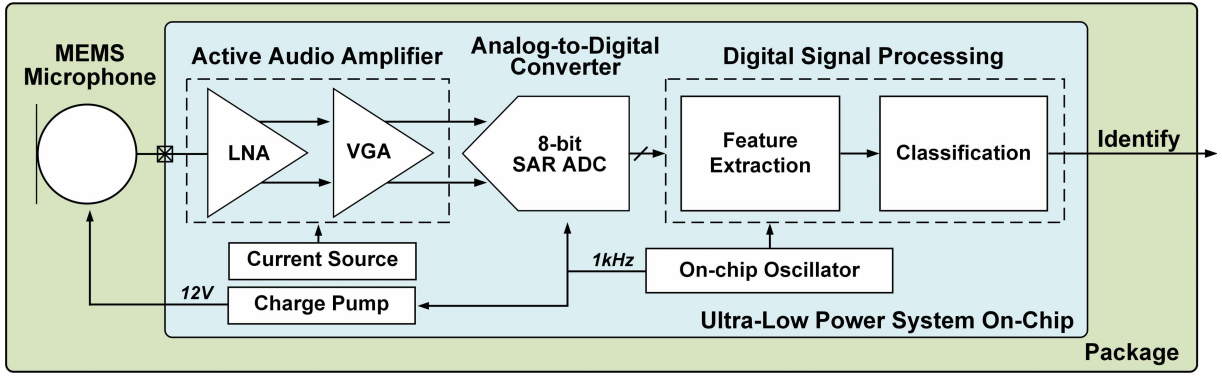


Fig. 1. Overall block diagram of the proposed system.

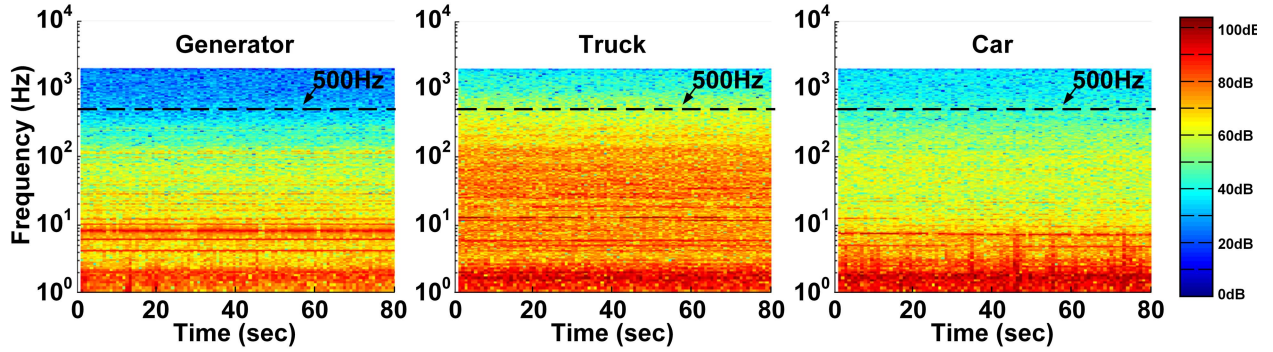


Fig. 2. Spectrograms of three different targets (generator, truck, and car).

An ECG monitoring system [4] includes backend signal classification but its power consumption exceeds 60 nW. To our knowledge, no sub-20-nW sensing microsystems have demonstrated complete operation with an integrated sensor, an analog frontend, and the digital backend for signal classification.

This paper reports an ULP acoustic sensing microsystem that meets the aforementioned 20-nW power constraint. The system continuously monitors its environment and identifies an event-of-interest while satisfying the stringent power budget without compromising event identification accuracy. This is accomplished through several features: 1) an micro-electromechanical systems (MEMS) microphone integrated in package with the rest of the electronics to provide a low-capacitance interface; 2) an SAR-analog-to-digital converter(ADC) exploiting a unique digital-to-analog converter (DAC) topology with extremely small (<50 fF) input capacitance to enable sufficient gain-bandwidth product for a frontend amplifier with nanowatt-level power consumption; 3) a serialized discrete Fourier transform (DFT) feature extraction performed only on discrete tones-of-interest (ToI) to avoid a high-power/area-consuming conventional parallel feature extraction using the fast Fourier transform (FFT); and 4) a power efficient classification engine using a programmable support vector machine (SVM).

II. SYSTEM OVERVIEW

Fig. 1 shows the overall system block diagram. The system includes a signal chain consisting of an MEMS microphone, an active audio amplifier followed by an ADC, and a digital signal processing (DSP) unit. Other peripherals include a

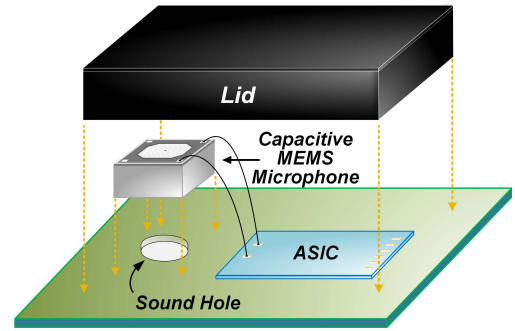


Fig. 3. System integration approach.

current source, charge pump, and clock source. Other than the MEMS microphone, the system is fully integrated on a single chip and does not need any external voltage/current sources.

The system aims to detect various machinery targets such as a generator, truck, and car. Spectrograms of these three different targets are shown in Fig. 2. Target sound in the frequency domain shows that features are mainly concentrated within a relatively narrow bandwidth of under 500 Hz. This allows the system to use a fairly slow clock of 1 kHz for all active components. Also, target features are concentrated in a few narrow sparse tones with high-power levels. This, along with the proposed DSP backend algorithm (presented in Section V), allows the system to operate with only 3-dB SNR (at the output of ADC). Overall, these two key factors enable the microsystem to continuously search for targets with only 12-nW total power consumption.

Fig. 3 shows the proposed system integration approach. The proposed system uses a custom printed circuit board (PCB)

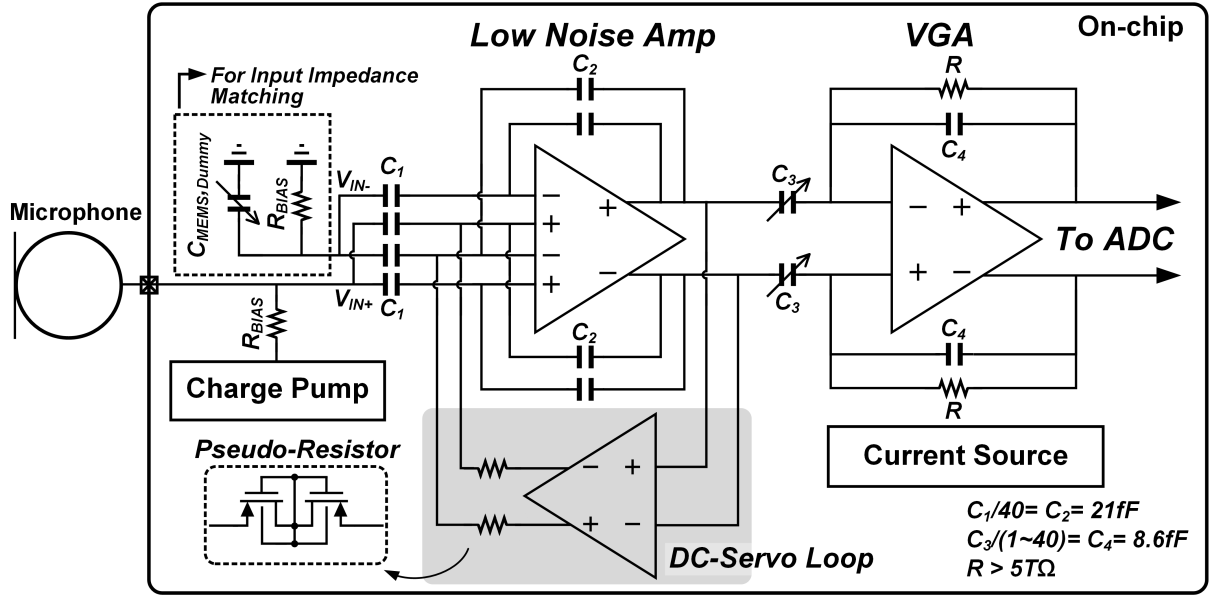


Fig. 4. Analog frontend architecture.

to place the passive MEMS microphone device close to the application specific integrated circuit (ASIC). This helps minimize the wirebonding length and reduces parasitic capacitance, which is important to minimize signal amplitude degradation. For packaging, a 3-D-printed lid is used to provide a back chamber for the passive MEMS microphone. The lid also covers the ASIC to avoid any light sensitivity. Avoiding light exposure is particularly important in the ULP circuit design as picoampere-biased components are susceptible to photo-generated currents.

III. ANALOG FRONTEND

Fig. 4 shows the detailed structure of the analog frontend. A capacitive (passive) MEMS microphone is used as a transducer to convert acoustic signal into an electrical signal. It is the most commonly used microphone type due to its low-power consumption and high sensitivity [5]. The following active audio amplifier consists of two stages: a fixed-gain low-noise amplifier and a variable gain amplifier. On-chip charge pump and current source provide the bias voltage and current for the MEMS microphone and amplifier, respectively. Following sections provide implementation details on each component in the analog frontend shown in Fig. 4.

A. MEMS Microphone

We first discuss the capacitive microphone package and its operating principle depicted in Fig. 5. The MEMS microphone consists of a perforated back plate (fixed) and a sensing membrane (movable). Once a bias voltage is applied, the charges between the backplate and membrane will be trapped through a high resistance node connection, creating a near-constant charge condition. The acoustic wave enters the sound port of the package and actuates the membrane motion, resulting in a voltage change between the membrane and the backplate. The voltage change is sensed by the following audio amplifier.

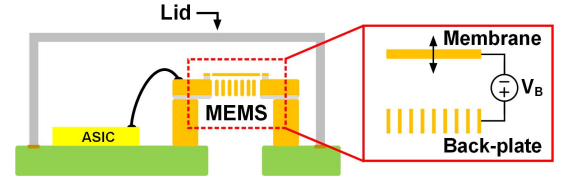


Fig. 5. Capacitive MEMS microphone package and its operating principle.

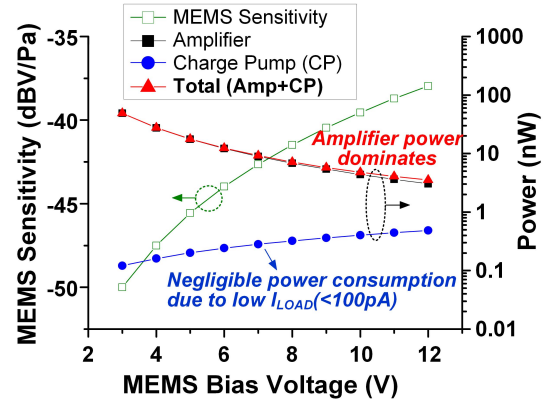


Fig. 6. Relationship between MEMS bias voltage and power consumption.

Fig. 6 shows the relationship between the MEMS microphone sensitivity and its bias voltage where the sensitivity is proportional to the MEMS bias voltage. As the bias voltage increases, a larger amount of charge is trapped between the membrane and backplate. As a result, a larger voltage variation is generated for an equivalent sound pressure level (SPL). However, the bias voltage cannot be increased indefinitely as there exists a critical bias voltage called the pull-in voltage, where the membrane will collapse and the device ceases to operate properly. Increased MEMS sensitivity relaxes the noise constraint on the following amplifier. As a result, the required current consumption of the amplifier is inversely proportional to the MEMS bias voltage. We assume a noise efficiency

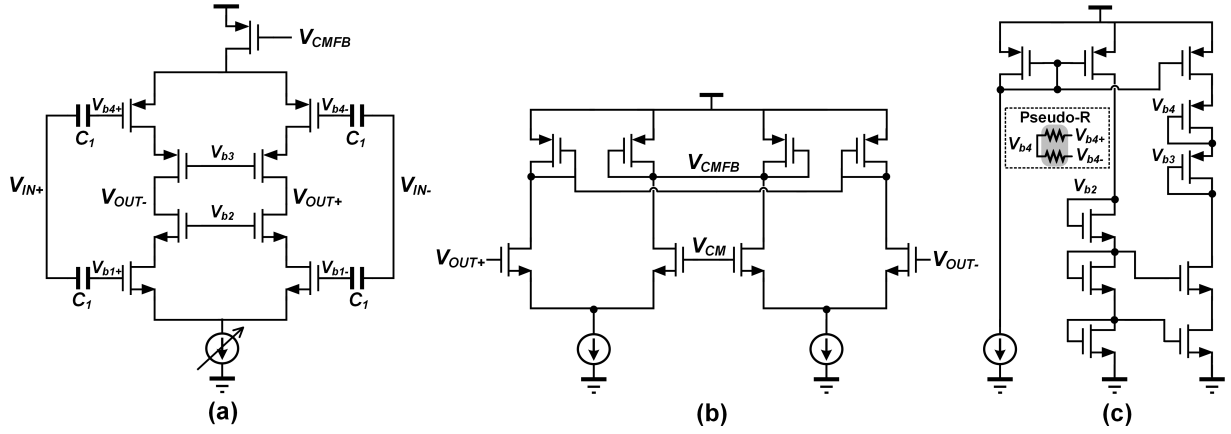


Fig. 7. (a) Detailed structure of the first-stage amplifier. (b) CMFB. (c) Bias generation.

factor (NEF) of 2- and 3-dB SNR at 40-dB Z-weighted (i.e., zero frequency weighting) [6], [7] SPL to estimate the amplifier power plotted in Fig. 6. The MEMS bias voltage is generated from a charge pump whose power consumption increases with the bias voltage. Power in Fig. 6 is estimated using a Dickson charge pump architecture. Since the capacitive MEMS microphone is a passive device, there is negligible load current on the charge pump output. Thus, the charge pump consumes only 100 s of pWs as it can operate at a low frequency. In the end, the total power consumption is dominated by the amplifier as the charge pump overhead is negligible. In this paper, we bias the MEMS device as high as possible, close to the pull-in voltage, to minimize the overall power consumption.

The MEMS microphone is dc biased with a charge pump through a resistor. In this scheme, the microphone capacitance (C_{MEMS}) and bias resistance (R_{BIAS}) sets the high-pass cut-off frequency ($=1/2\pi R_{BIAS}C_{MEMS}$). Therefore, R_{BIAS} should be large enough to set the corner frequency to be below the MEMS microphone input frequency range (20 Hz–20 kHz). In this paper, a C_{MEMS} of 5.4 pF requires >1.5 G Ω R_{BIAS} , which would consume a large area with a passive resistor implementation. Thus, subthreshold MOSFET-based pseudo-resistor has been used. Although, pseudo-resistors are not highly controllable, the corner has been pushed far enough to the low-frequency range (<20 Hz) to ensure proper operation across process corners. Since the following amplifier uses a differential structure, a dummy capacitor and a resistor are added on the other side to match the input impedance.

B. Amplifier

The system operates with a relatively low SNR (3 dB) and bandwidth (<500 Hz), which reduces the burden on the amplifier noise performance. Under this specification, the amplifier can be implemented to consume only a few nanowatt unlike conventional MEMS microphone readout circuitry [8], [9] that typically consumes at least a few microwatt. Even with a relaxed noise constraint, the first-stage amplifier design is still noise limited due to the low system power budget. The first-stage amplifier determines the overall noise

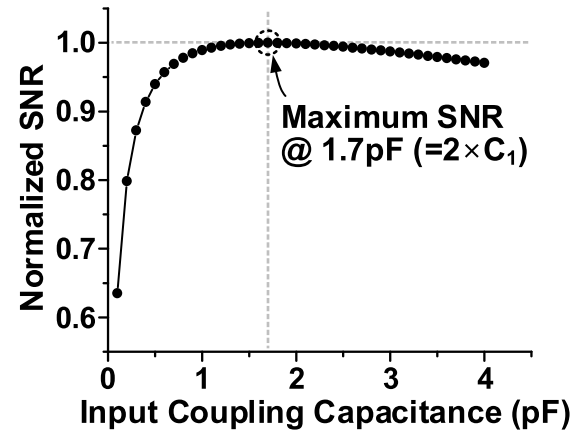


Fig. 8. Relationship between input coupling capacitor (C_1) and SNR at the amplifier output.

performance as the target noise specification on the amplifier renders the noise from the MEMS microphone to be negligible.

Fig. 7 shows the structure of the first-stage amplifier, the common-mode feedback (CMFB), and the bias generation. To achieve high noise efficiency, the amplifier uses a current-reuse topology where an inverter-based input stage is used to double the input transconductance (g_m). Also, input transistors are biased in a subthreshold regime to maximize g_m . In low-bandwidth applications, the flicker noise often dominates the overall noise level. The effects of the flicker noise can be minimized by having an input pair with large gate areas. Note that the dynamic offset-cancellation techniques such as auto-zeroing or chopping are not suitable for this work due to a limited power budget. The relaxed noise constraint allows the flicker noise to become negligible even with moderate size of input pairs ($W/L = 45 \mu\text{m}/500 \text{ nm}$). However, as increased input pair size also increases the noise gain of the amplifier, it cannot be increased indefinitely.

Amplifier current consumption is set to 3 nA to meet the SNR requirement at the lower bound of the target sensitivity SPL of 40 dB (Z-weighted; this represents a very quiet environment). Given the resulting gain-bandwidth product, a closed-loop gain of 32 dB can be achieved while meeting the target signal bandwidth (0.5 kHz). A telescopic structure is used to provide sufficient open-loop gain (>70 dB in simu-

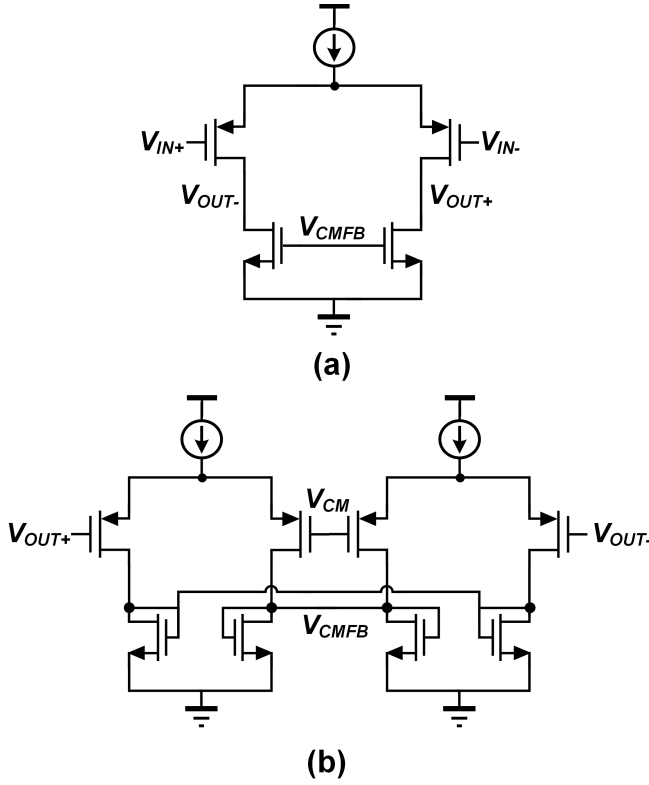


Fig. 9. (a) Detailed schematic of the second-stage amplifier and (b) its CMFB.

lation) to minimize gain nonlinearity errors [10]. Furthermore, pMOS and nMOS input transistors are separately biased, increasing headroom and thereby output swing to preserve linearity over the entire dynamic range (40–94 dB SPL). A dc servo loop is added to minimize dc offset to prevent output saturation. This loop also sets the high-pass corner, which is designed to be less than 20 Hz to avoid signal degradation. The dc servo-loops set the dc bias voltage of the nMOS input pair and the bias generation circuit [Fig. 7(c)] generates the dc bias voltages of pMOS input pair and the core amplifier (V_{b2} and V_{b3}).

The signal and noise gain at the output of the first-stage amplifier can be expressed as follows:

$$\text{Signal Gain} = \frac{C_{\text{MEMS}}C_1}{C_2(C_{\text{MEMS}} + C_1 + C_{\text{pkg}})} \quad (1)$$

$$\text{Noise Gain} = \frac{C_1(C_{\text{MEMS}} + C_{\text{pkg}}) + C_2 + C_{\text{in}}}{C_2} \quad (2)$$

where C_{pkg} is a parasitic capacitance between MEMS and ASIC interface and C_{in} is a parasitic capacitance at the amplifier input. Since, MEMS and parasitic capacitances are known, the size of the series capacitor C_1 is chosen for a fixed closed-loop gain so that the amplifier operates at the optimal SNR point (Fig. 8).

The following second-stage amplifier is bandwidth limited unlike the noise-limited first stage. Hence, its current consumption is directly impacted by the load capacitance. This motivates the extremely low input loading capacitance (<50 fF) of the proposed ADC, allowing the

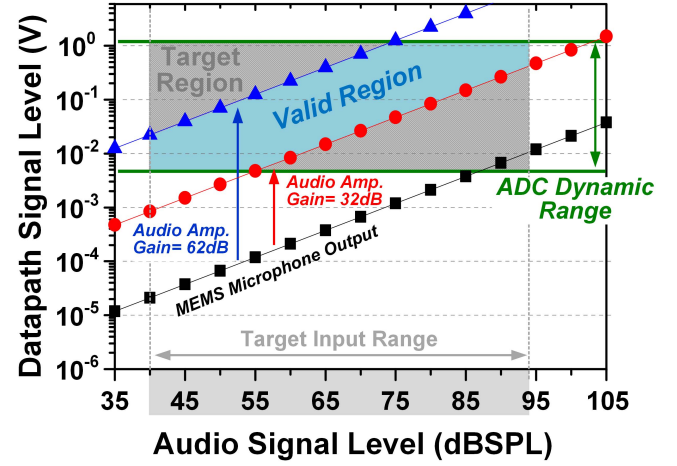


Fig. 10. Datapath signal level planning.

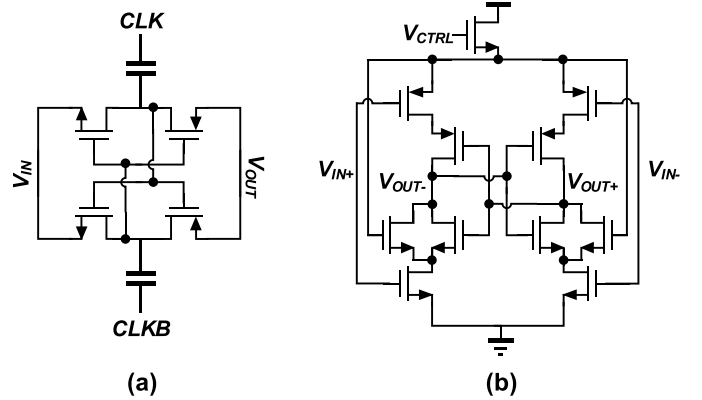


Fig. 11. (a) Schematic of voltage doubler [11]. (b) Delay cell used in the oscillator [13].

second-stage bias current to be reduced to 100 s of pAs. As the amplifier design is not noise limited, a simple single-stage op amp (Fig. 9) is used instead of the current-reuse topology to provide better output swing. The second-stage amplifier provides a closed-loop gain up to 32 dB by adjusting the C_3 , which is sufficient to cover the target input signal range (40–94 dB SPL). Fig. 10 summarizes the overall datapath signal level planning from the first-stage amplifier input to the second-stage amplifier output (i.e., ADC input).

C. Peripheral Circuits

The charge pump provides a bias voltage across the MEMS sensor, setting the sensitivity of MEMS device. The charge pump consists of cascaded voltage doublers (see Fig. 11(a), [11]) with 12 stages in total. The first two stages can be bypassed, allowing the output voltage to vary from 11.19 to 13.35 V based on the configuration (10× to 12×, simulation results). As there is negligible conduction loss, small transistors are used to minimize voltage drop due to charge sharing. A decoupling capacitor is added at the charge pump output to reduce output ripple. This capacitor is implemented with a metal–oxide–metal (MOM) capacitor to avoid breakdown. The operating frequency is variable from 7.81 Hz to 1 kHz in a binary fashion.

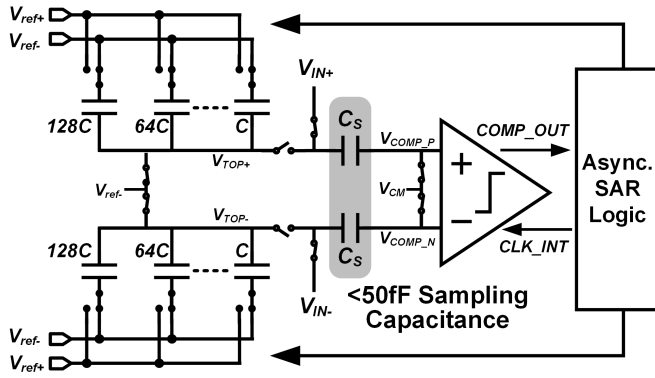


Fig. 12. Proposed ADC architecture with low sampling capacitance.

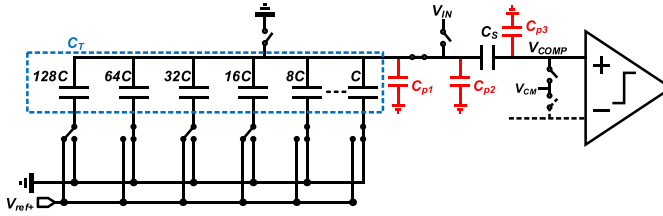


Fig. 13. Parasitic capacitors affecting ADC performance.

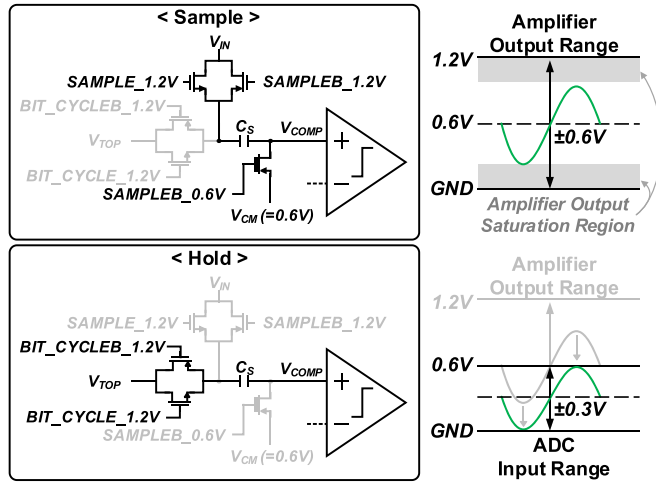


Fig. 14. Detailed sample and hold operation for matching signal range.

The current source is based on a self-biased Nagata current reference [12]. Subthreshold biased MOSFETs are used to generate 100 s of pA of current. Also, a pseudo-resistor is used instead of a passive resistor to save area. Its resistance is controlled by adjusting the MOSFET gate voltage. Voltage control uses a 128-stage diode stack from 1.2-V supply voltage realizing 10-mV steps.

The voltage-controlled oscillator (VCO) is used as the clock source. An nMOS header adjusts the virtual supply of the oscillator while the unit cell is implemented with a thyristor-based oscillator (see Fig. 11(b), [13]) to avoid short-circuit current. The oscillator consists of ten stages where different phases of the clock are combined to generate a duty-cycled clock. The duty ratio can be adjusted from 60% to 90% of the entire clock period with 10% step.

A single commercial I/O pad in the given technology consumes >1 nA which is intolerable. Therefore, custom pads have been designed which only includes small diodes

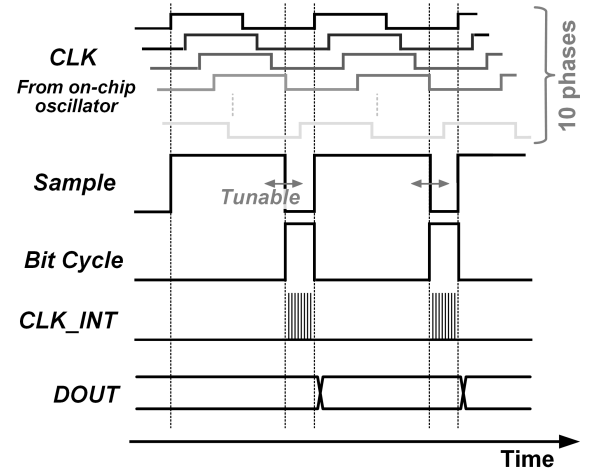


Fig. 15. Timing diagram of ADC operation.

($20 \mu\text{m} \times 20 \mu\text{m}$) reducing current consumption to 1 pA (at 25°C in simulation). Furthermore, pads for interconnecting the MEMS and the amplifier are composed of metal layers only to have ultra-low current level. Note that pW-level I/O pads are nearly as robust to electrostatic discharge events as conventional commercial I/O pads [14].

IV. ANALOG-TO-DIGITAL CONVERTER

In conventional low resolution (<10 b) SAR-ADCs, the unit capacitance of the binary DAC array is usually limited by mismatch rather than kT/C noise. In our target process technology, 4-fF unit capacitance is required to ensure worst case differential non-linearity (DNL) (3σ) is less than 1 LSB (10 k Monte Carlo runs). This translates into 1 pF of the load capacitance for the preceding variable gain amplifier, which would require more than 10 nA of current consumption to achieve the required performance (32-dB gain, settling within 1 LSB at 1-kHz sampling rate). To overcome this issue, we separate the sampling capacitor (C_s) from the capacitive DAC (Fig. 12). This is done by inserting the sampling capacitor in series between the DAC and the comparator. In this way, the size of the sampling capacitor is reduced to less than 50 fF. As a result, the load capacitance seen by the preceding amplifier is dramatically reduced while the unit capacitance of the DAC stays the same. Since the preceding variable gain amplifier (VGA) is bandwidth limited, this enables significant power savings.

During the sampling phase, the DAC is disconnected from C_s and purges all its charge while the input is sampled on C_s . At the beginning of the bit-cycling phase, the DAC top plate (V_{TOP}) is connected to C_s , and the rest of the conversion process is identical to a conventional approach due to charge conservation on node V_{COMP} . While the proposed scheme reduces the size of the sampling capacitance, it raises two issues. First, kT/C noise is determined by the small C_s rather than the DAC capacitance as in conventional scheme. However, resulting noise is still negligible ($\sim 300 \mu\text{V}_{\text{rms}}$) compared to quantization noise ($>1 \text{ mV}_{\text{rms}}$). The second issue, which is more significant, is that the ADC becomes

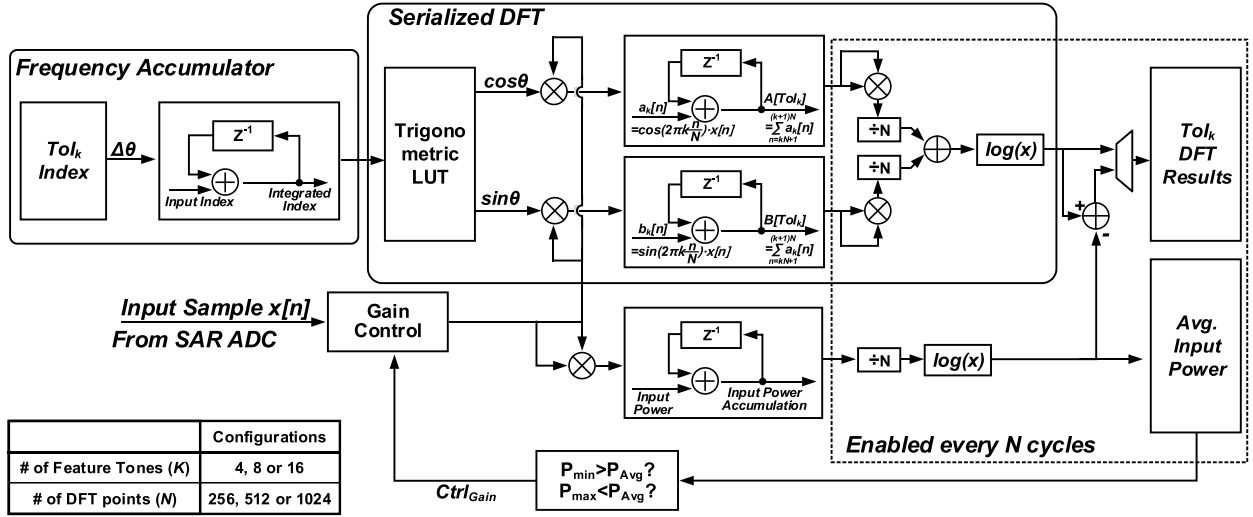


Fig. 16. Feature extraction using serialized ToI DFT.

sensitive to parasitic capacitors on the top plate. Specifically in this scheme, parasitic capacitors C_{p1} through C_{p3} create a gain error, common-mode voltage shift, and comparator gain loss (Fig. 13, which assumes single-ended view for simplicity). Taking these parasitic capacitors into account, the comparator input (V_{comp}) during conversion can be expressed as follows:

$$V_{COMP} = k \left[(1 + n)V_{CM} - V_{in} + \frac{C_1}{C_T} m V_{REF} \right] \quad (3)$$

where C_1 is sum of all capacitance connected to the reference voltage (V_{REF}) during conversion and C_T is the total capacitance of the DAC array. In (3), n is the common-mode shifting factor, k is comparator gain loss, and m is gain error, which are expressed as follows, respectively:

$$n = \frac{C_{p2} + \frac{C_{p3}}{C_s}(C_T + C_{p1} + C_{p2} + C_{p3}/C_s)}{C_T + C_{p1} + C_{p3}/C_s - C_s} \quad (4)$$

$$k = \frac{C_s}{C_s + C_{p3}} \frac{C_T + C_{p1} + C_{p3}/C_s - C_s}{C_T + C_{p1} + C_{p2} + C_{p3}/C_s} \quad (5)$$

$$m = \frac{C_T}{C_T + C_{p1} + C_{p3}/C_s - C_s}. \quad (6)$$

It is evident that the C_{p3} to C_s ratio has the most significant impact and therefore should be kept low. To minimize this ratio, Metal–insulator–Metal capacitors are used for C_s where the top plate faces the comparator input to minimize C_{p3} . To further reduce the effect of parasitic capacitors, the DAC unit capacitor structure uses the bottom plate to enclose the top plate. Consequently, the gain error reduces the ADC dynamic range by 8% but this is acceptable as the overall SNR is limited by the amplifier noise. Also, the resulting 12% comparator gain loss is not an issue because comparator power in an 8-bit configuration is insignificant and the common-mode voltage shift is negligible (<1%).

The amplifier operates at 1.2 V while the ADC operates at 0.6 V to reduce power. Therefore, the common-mode voltages of these two blocks do not match. To solve this

issue, we sample the amplifier output against its common-mode voltage 0.6 V to shift the common mode-voltage down to 0.3 V during the hold phase (Fig. 14). With this approach, we can match the dynamic range of the amplifier output with the ADC input. Fig. 15 provides a timing diagram of the ADC. The ADC operates asynchronously to avoid a high-speed clock, reducing power. The SAR control logic is implemented with I/O devices to minimize leakage. The sampling time is maximized by adjusting the duty cycle of an on-chip oscillator to further save power on the preceding amplifier.

V. DIGITAL SIGNAL PROCESSING BACKEND

The DSP backend consists of three main blocks. First, the ADC output is directly fed into the feature extraction block that generates target features. Then, the following classifier generates classification results based on these features. Finally, post processing is performed to generate an object identification result.

We first discuss feature extraction that is based on the frequency-domain analysis. A conventional FFT generates the entire spectrum in parallel but consumes high-power and large area. Existing ULP FFT designs (see [15]) still consume >400 nW, far exceeding our system power budget. To eliminate the need for full FFT computation, we use the stationary and sparse signal property that is evident in our target machinery objects' spectrogram as shown in Fig. 2. We propose a feature extraction scheme based on a partial DFT that is performed only on discrete ToIs instead of the entire spectrum. Furthermore, DFT is performed on a tone-by-tone basis using serialized computation, where the system only monitors one frequency band (tone) at a time and then switches to the next tone. The DFT of a particular ToI index k is obtained by the following equation:

$$X[ToI_k] = \sum_{n=kN+1}^{(k+1)N} e^{-j \frac{2\pi n ToI_k}{N}} x[n] \quad (7)$$

where $x[n]$ is ADC output, N is the DFT size, and the set of integer ToI indices is defined as $ToI_1, ToI_2, ToI_3, \dots, ToI_K$.

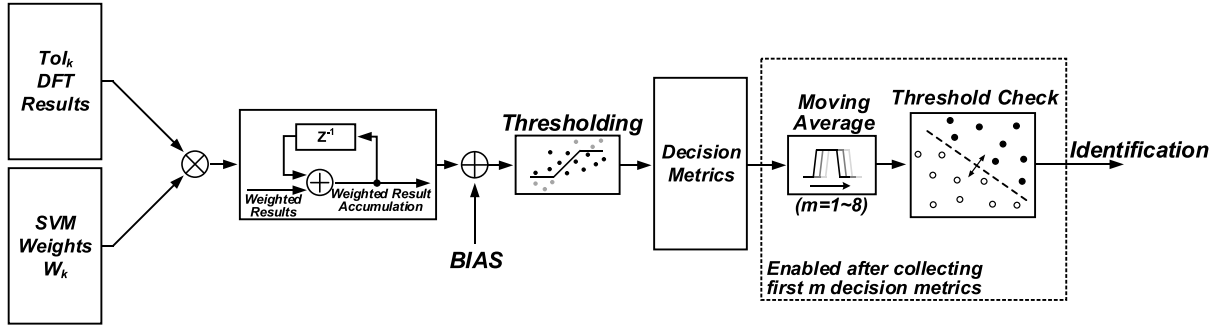


Fig. 17. Classifier using linear SVM.

These ToI indices consist of feature tones of target objects and some background noise tones to be used to obtain reference noise level. The ToI DFT output $X[\text{ToI}_k]$ for the k th ToI is computed using samples arriving at $kN + 1, kN + 2, \dots, (k + 1)N$ time indices. That is, each ADC sample $x[n]$ is used only once for a particular ToI index, and not for other indices. As the DSP runs at a low clock frequency of 1 kHz, its power consumption is dominated by static rather than dynamic power. By listening to one tone at a time, logic area is reduced significantly which consequently reduces the leakage power. It is important to note that serial computation does not degrade the system performance because the system is designed to recognize machinery target object sounds that are inherently stationary.

Fig. 16 shows the overall serialized ToI DFT architecture, which greatly simplifies the computation to be performed in each clock cycle. By simply changing the phase accumulation rate ($\Delta\theta$), the DFT weight sequence on any particular ToI can be obtained. Trigonometric functions ($\sin \theta$ and $\cos \theta$) for DFT are implemented using a compact look-up table storing only one quadrant ($0 \leq \theta \leq \pi/4$) of cosine values. The other quadrants along with the sine values are generated by manipulating the address and negating stored values using the symmetry of the trigonometric function. This technique reduces power by 30% (simulation) compared with a coordinate rotation digital computer [16]. Generated sine and cosine values are multiplied with input samples and accumulated. After 512 cycles (DFT size, programmable), the final frequency domain power for a ToI is computed by squaring and adding two accumulated values (real and imaginary). The ToI power value is converted to log scale and then stored into a local memory. The design also includes digital gain control by monitoring the average input power. The digital gain control scales the 8-bit input into 4-bit representation to reduce the complexity of the DSP system and thus to save power consumption. Due to the relatively low SNR requirement, quantization noise from such reduction does not degrade the overall performance of the system. The design supports three configurations with different number of feature tones and DFT sizes, allowing tradeoffs in latency and feature bandwidth for feature extraction.

While the ToI DFT computation is always-on, the classifier logic can be clock gated until DFT sequentially generates all desired ToI results. Due to the low ToI DFT throughput

(e.g., a period of 4.096 s for 8 ToIs, 512-point DFT, 512 cycles per ToI, and 1-kHz clock), a moderately sophisticated classification scheme can be used with a negligible power overhead. Fig. 17 shows the overall classifier architecture, which is based on the SVM. Computationally, demanding SVM training is performed off-line, resulting in a $1 \times K$ weight vector and a constant offset (BIAS) per object to be classified. These training results are stored in a programmable memory. Computations in SVM classification are also performed serially by accumulating multiplication results to reduce power consumption. Before obtaining the SVM output, a fixed threshold is applied to suppress any significant outliers. Then, SVM outputs undergo a post processing step consisting of a moving average window to perform low pass filtering. The size of the moving average window (m) sets the initial latency of the system. For example, averaging five SVM outputs ($m = 5$) results in 20 s of initial latency (5×4.096 s for 8 ToIs). The averaged result is compared against a constant (programmable) threshold to make a final decision. Once the initial latency is elapsed, the final decision is updated whenever a new SVM output is available (every 4.096 s for the same example). Note that the relatively long initial latency is a secondary concern when the system is always-on and the detection target is stationary.

Overall DSP power consumption is dominated by the always-on feature extraction block and its corresponding leakage power. We observed that leakage power consumption would be 21 nW while dynamic power consumption is only ~ 2 nW (simulation result with PEX) when standard cells for nominal devices are used for DSP implementation. To resolve this leakage dominated condition, custom library cells-based on I/O devices are implemented and used throughout the entire DSP design. As a result, leakage power is reduced by $160\times$ to 132 pW and total power is reduced by 93%.

VI. MEASUREMENTS

The test chip was implemented in 180-nm CMOS with an active area of 0.75 mm^2 . Fig. 18 shows the die photograph, and Fig. 19 shows the photographs of the assembled microsystem. The PCB is 6.5 by 3 cm and includes a sound hole on its backside for the MEMS device. Inside the lid, the chip is placed as close as possible to the MEMS device to minimize signal degradation.

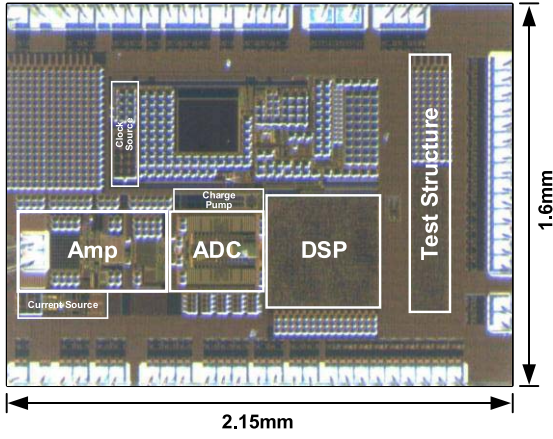


Fig. 18. Die photograph.

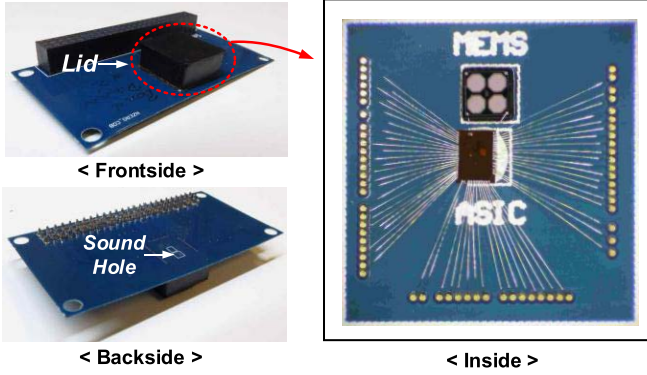


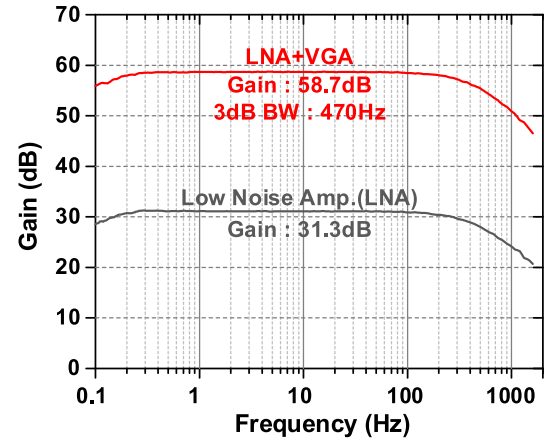
Fig. 19. Photographs of the integrated system.

A. Block Measurements

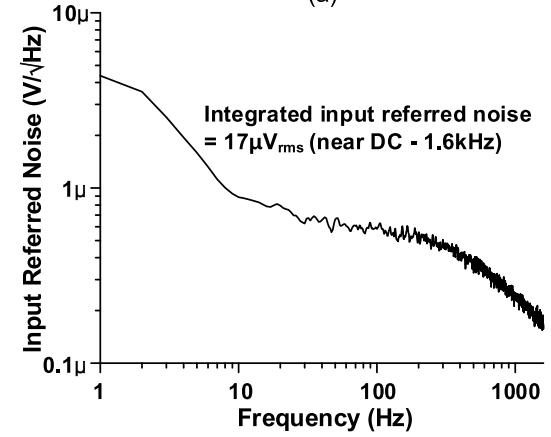
Fig. 20(a) shows the measured amplifier transfer function. Midband gain ranges from 32 to 59 dB with a 3-dB bandwidth of 470 Hz. Estimated low-frequency corner is ~ 85 mHz. The first-stage low-noise amplifier shows $800 \text{ nV}/\sqrt{\text{Hz}}$ noise floor at 100 Hz with a $1/f$ corner at ~ 20 Hz [Fig. 20(b)]. Integrated noise under this curve from near dc to 1.6 kHz gives $17 \mu\text{V}_{\text{rms}}$ of the input referred noise level. It can be seen that $1/f$ is not a dominant noise source as discussed in Section III-B. The first-stage amplifier consumes 3.4 nA resulting in a 1.8 NEF. Overall amplifier including the second stage consumes 4.6 nA achieving 2 NEF. Fig. 21 shows the measured amplifier output spectrum with 1.5% total harmonic distortion for 94 dB(z) SPL at 20 Hz. Measured power supply rejection ratio is >62 dB in the passband.

The ADC operates with a 0.6-V supply at a sampling rate of 1 kS/s. Measured DNL and integral non-linearity (INL) are shown in Fig. 22(a). The peak DNL and INL are $+0.34/-0.65$ LSB and $+0.67/-0.63$ LSB, respectively, across five different chips. Fig. 22(b) shows the measured power spectrum with input signal at near-Nyquist rate. Measured spurious-free dynamic range is 60.04 dB and signal-to-noise and distortion ratio (SNDR) is 48.26 dB, which corresponds to 7.7 b effective number of bits (ENOB). This SNDR is maintained throughout the full bandwidth. The ADC consumes 2.7 nW exhibiting a 13 fJ/convstep figure-of-merit.

The charge pump consumes 240–840 pW while operating from 31 Hz to 1 kHz (Fig. 23). In this operating range,



(a)



(b)

Fig. 20. Measured amplifier. (a) Frequency response. (b) Noise.

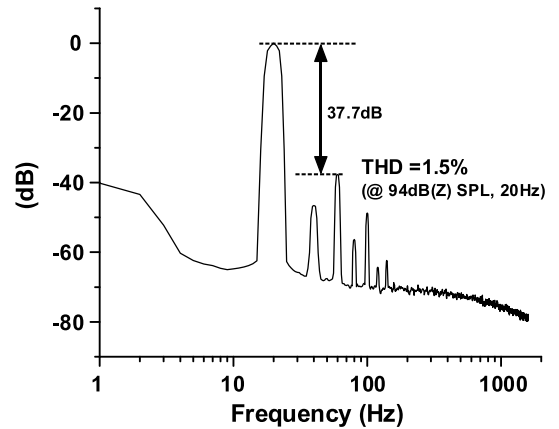


Fig. 21. Measured amplifier output spectrum with 94 dB(z) SPL at 20 Hz.

the output voltage varies from 9 to 13 V. In this paper, we use an operating point where the charge pump runs at 250 Hz to output 12.5 V.

Fig. 24 shows the measured system power breakdown. The complete system consumes 12 nW. The overall power is relatively evenly distributed with the most power-dominant component (frontend amplifier) consuming less than half of the total power. Table I summarizes the performance of the test chip and compares with prior works.

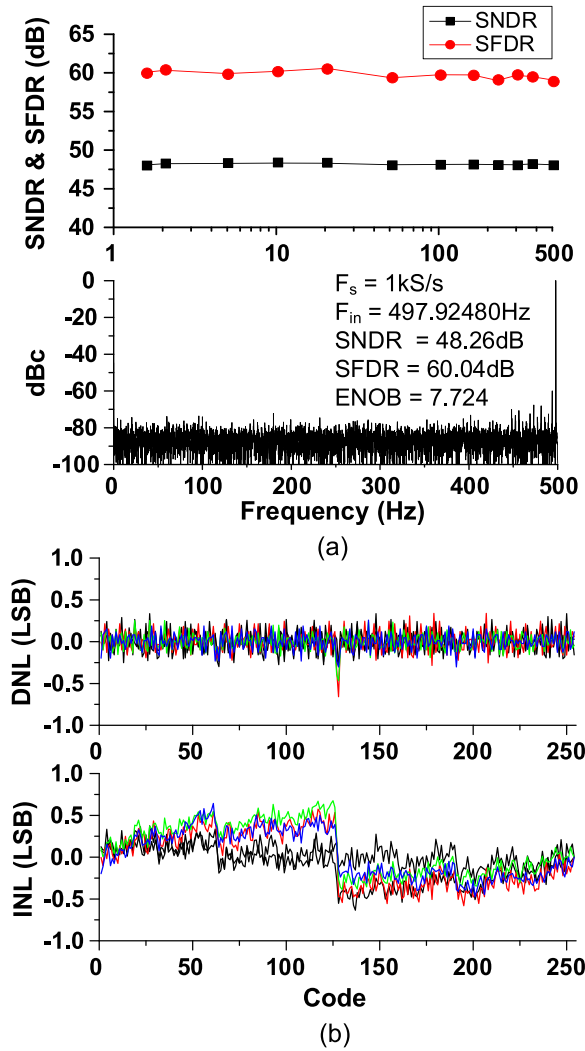


Fig. 22. Measured (a) FFT of ADC output data and (b) linearity.

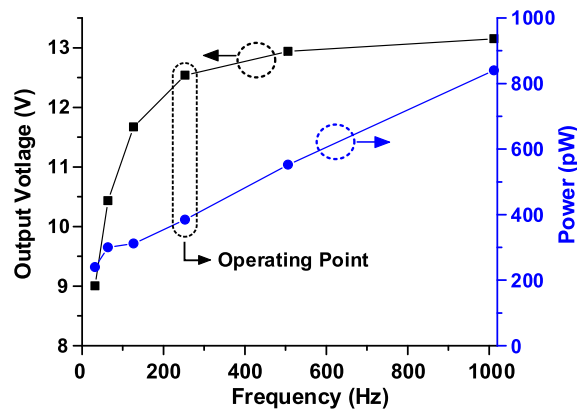


Fig. 23. Charge pump measurements.

B. System Measurements

Fig. 25 shows the testing environment for system level measurements. A desktop computer plays the target sounds, using an audio amplifier to drive a passive speaker. A reference microphone measures the actual SPL near the proposed system. The sound database is provided by a third party research lab for the three different target objects (a generator,

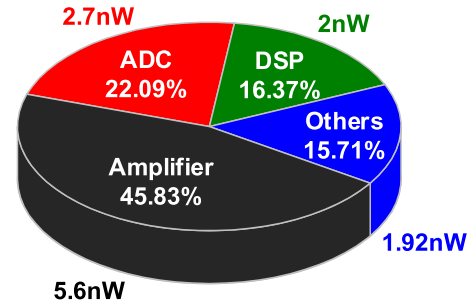


Fig. 24. System power breakdown.

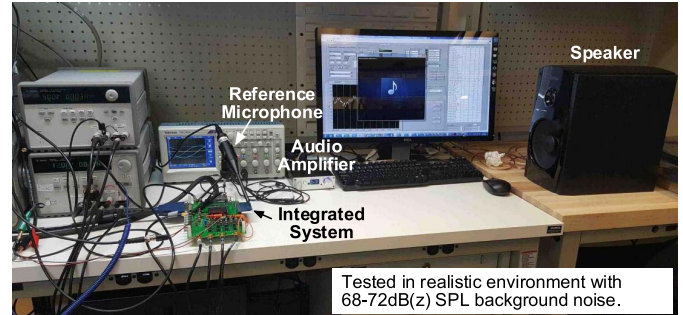


Fig. 25. Testing environment for system level measurements.

car, and truck), which are recorded in an anechoic chamber. Performance of the proposed system is tested in a realistic environment with 68–72 dB (Z) SPL background noise.

Fig. 26 shows the measured output spectrum of the amplifier, where feature tones appear only when each target is ON. Fig. 27 shows the classifier outputs for various test cases. Object identification is triggered when the filtered classification metric exceeds the threshold of 0. When the generator is running, the decision metric is well above 0 while it is below 0 when the generator is OFF. Notice the decision metrics of the other two targets, car and truck, are consistently below 0 when the generator sound is present. Similar results are observed when the truck is the target. The car is the most difficult to distinguish as seen by smaller separation between the decision metrics. There exists no threshold that can achieve 100% detection accuracy without a false alarm. False alarm is an important metric given that this type of system would serve as a wakeup to a more power-hungry and higher performance audio detection system. In such case, the average power consumed would depend on the false alarm (i.e., false positive rate) and the power of the more complex system. Assuming that upon detection a system that is $1000\times$ more power hungry is employed for a span of 4 s (detection rate of the wakeup system), a false positive rate of 0.1% would be an ideal target to balance the power between the wakeup and main systems. Relationship between the generator classification accuracy and SNR is analyzed with receiver operating characteristic curves shown in Fig. 28. As expected, generator detection accuracy monotonically improves as SNR increases. Distinguishing a generator from a truck is more challenging than generator versus car distinction. Fig. 28 shows classifying a generator eventually reaches a perfect classification point (0, 1) with ≥ 6 dB SNR. To render realistic environmental noise, the

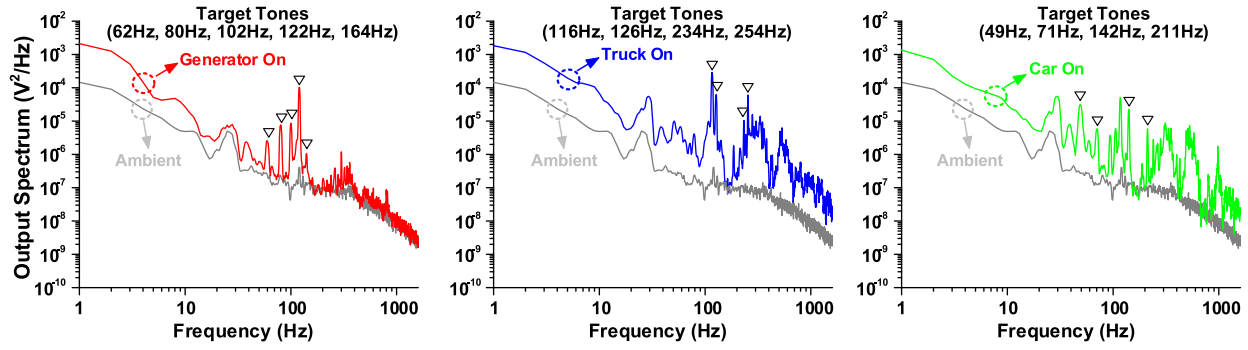


Fig. 26. Amplifier output spectrum with different target.

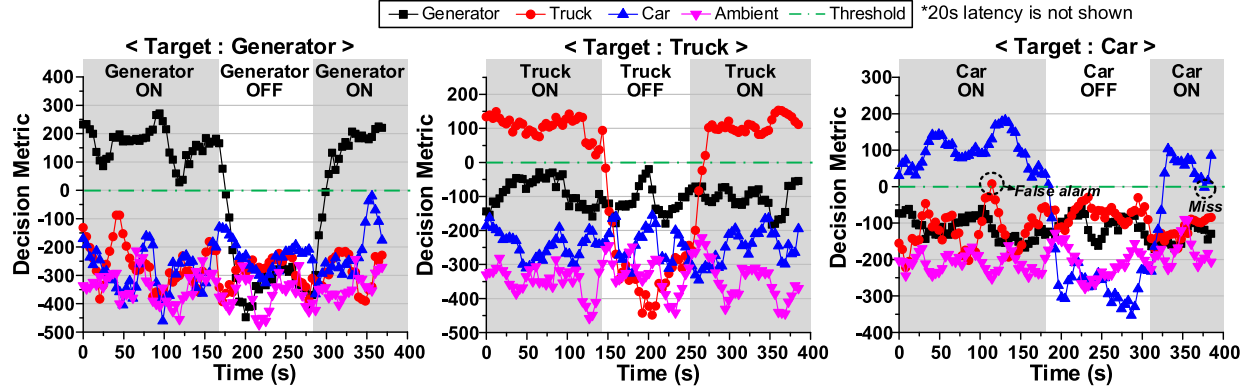
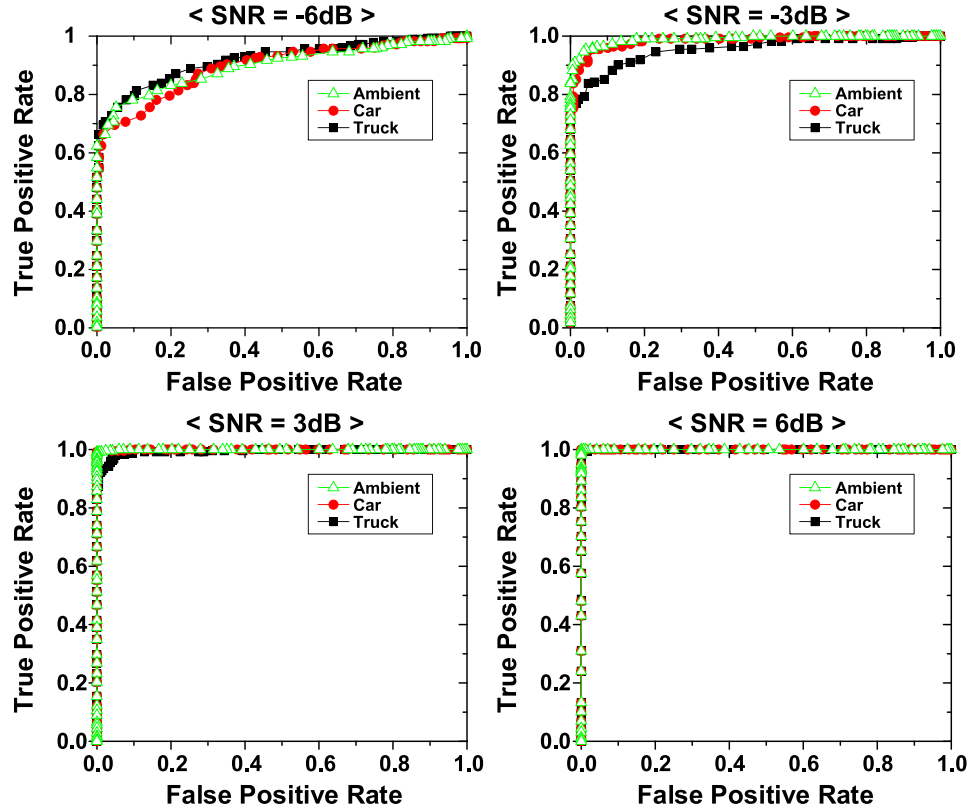
Fig. 27. Identification results for three different target objects ($N = 512$, $K = 8$, and $m = 5$).

Fig. 28. Accuracy of generator classification depending on the SNR.

system has been tested with a sound data recorded in rural and urban areas. Fig. 29 shows the classifier output detecting a generator in different environments. When a threshold is

set to have 0 false alarms, the rural and urban conditions show 98% and 96% of detection accuracy, respectively, for the same SNR condition (~ 6 dB). Miss-detection mainly occurs

TABLE I
PERFORMANCE SUMMARY AND COMPARISON TABLE

	This Work	K. Badami, <i>et al</i> JSSC Jan. 2016 [1]	B. Rumberg, <i>et al</i> JETCAS 2011 [2]	Y. Chen, <i>et al</i> JSSC Jan. 2015 [4]	P. Harpe, <i>et al</i> JSSC Jan. 2016 [3]
Target of Interest	Sound	Sound	Sound	ECG	ECG
Technology	0.18 μ m	90nm	0.5 μ m	65nm	65nm
Amplifier	Supply Voltage	1.2V	N/A	Off-chip	0.6V
	Power	5.6nW 3.4nA(LNA), 1.2nA(VGA)	1-4 μ W 0.96 μ W(LNA), 36n-4 μ W(VGA)		16.8nW 22.5nA(LNA), 5.4nA(VGA)
	Gain	31 ~ 59dB	32.5 ~ 83.5dB		51 ~ 96dB
	Bandwidth	470Hz	3kHz		250Hz
	Input referred Noise	17 μ V _{rms}	32.5 μ V _{rms}		6.52 μ V _{rms}
	NEF	1.8 (w/o VGA) 2.0 (w/ VGA)	N/A		2.4 ¹ (w/o VGA) 2.64 (w/ VGA)
ADC	Supply Voltage	0.6V	Mode B: Mixed-signal Classifier Mode C: Off-chip ADC is used	Off-chip	0.6V
	Power	2.7nW			1.8nW
	Sampling Rate	1kS/s			500S/s
	Resolution	8bit			8bit
	ENOB	7.7			7.14
	Linearity	DNL: +0.34/-0.65LSB INL: +0.67/-0.63LSB			DNL: +1.0/-1.0LSB INL: +1.8/-1.8LSB
	FOM [fJ/conv·step]	13.0			25.5
DSP	Supply Voltage	0.6V	N/A	N/A	0.4V
	Power	2nW	2.6 μ W (Mode B)	51 μ W	45nW
	Clock Frequency	1kHz	N/A	N/A	10kHz
	Feature Type	Digital	Analog	Analog	Digital
	Classifier	On-chip	On-chip	Off-chip	On-chip
	Latency	20s Min: 1s, Max:131s	<100ms	100ms	N/A
	Accuracy	>95% for Generator, Car, and Truck detection	HR SP 89% HR Non SP 85%	>90% for Car, and Truck detection	N/A
Total System Power	12.2nW	3.8 μ W (Hybrid mode A+B+C)	51 μ W	64nW	3nW

1. Estimated number.

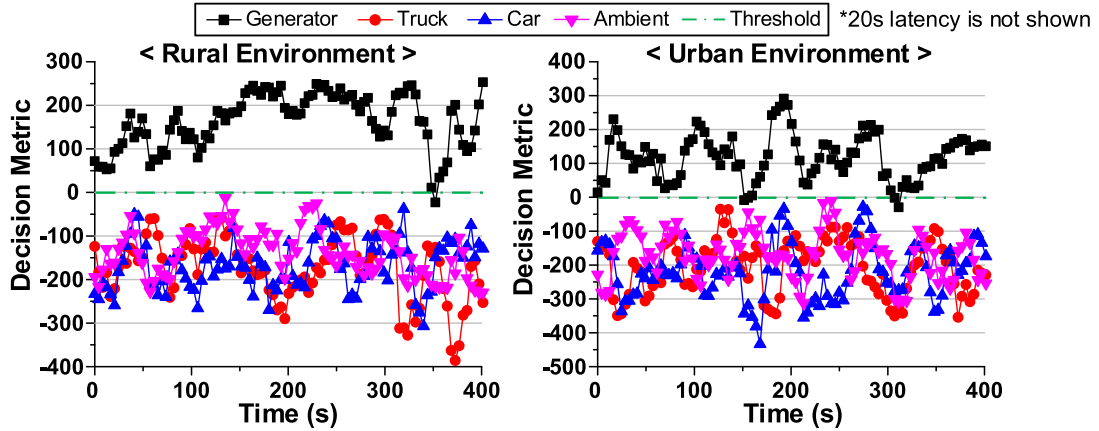


Fig. 29. Identification of the generator in different environments ($N = 512$, $K = 8$, and $m = 5$).

when the target signal sound is masked by various ambient noise sources (e.g., airplane) that are not present in the testing lab environment.

Overall, the proposed ToI-based SVM exhibits reliable performance for our target objects whose features are stationary in time and sparse in the frequency domain. The

system demonstrates a consistent >95% detection rate for all tests.

VII. CONCLUSION

An ULP always-on acoustic sensing and object recognition microsystems are proposed. The system exploits the stationary

and sparse target signal characteristics to reduce the required SNR and signal bandwidth. An 8-bit SAR-ADC with unique DAC structure is introduced to significantly decrease the size of sampling capacitance which reduces power consumption of the preceding amplifier. Serialized ToI DFT feature extraction is proposed for the digital backend, replacing a high-power/area-consuming conventional FFT. The overall system successfully identifies target objects with >95% accuracy while operating continuously with 12 nW of power consumption. The proposed work focused on detecting targets sounds with relatively low-frequency content (<500 Hz). However, as the key idea of this paper is the overall system architecture including the co-optimization of the analog frontend and the DSP, sounds of interest that have higher frequency content would still benefit from the overall architecture as long as they can be characterized via the TOI DFT approach, even though the power level itself would be necessarily higher.

ACKNOWLEDGMENT

The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

REFERENCES

- [1] 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.
- [2] B. Rumberg, D. W. Graham, V. Kulathumani, and R. Fernandez, "Hibernets: Energy-efficient sensor networks using analog signal processing," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 1, no. 3, pp. 321–334, Sep. 2011.
- [3] P. Harpe, H. Gao, R. van Dommele, E. Cantatore, and A. H. M. van Roermund, "A 0.20 mm² 3 nW signal acquisition IC for miniature sensor nodes in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 51, no. 1, pp. 240–248, Jan. 2016.
- [4] Y. P. Chen *et al.*, "An injectable 64 nW ECG mixed-signal SoC in 65 nm for arrhythmia monitoring," *IEEE J. Solid-State Circuits*, vol. 50, no. 1, pp. 375–390, Jan. 2015.
- [5] S. Bouwstra *et al.*, "Silicon microphones—A danish perspective," *J. Micromechan. Microeng.*, vol. 8, no. 2, pp. 64–68, 1998.
- [6] *Acoustics—Description, Measurement and Assessment of Environmental Noise—Part 2: Determination of Environmental Noise Levels*, document ISO 1996-2, Geneva, Switzerland, 2007.
- [7] *Electroacoustics—Sound Level Meters—Part 1: Specifications*, document IEC 61672-1, Geneva, Switzerland, 2002.
- [8] J. Čitaković *et al.*, "A compact CMOS MEMS microphone with 66dB SNR," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2009, pp. 350–351.
- [9] S. Ersoy, R. H. M. van Veldhoven, F. Sebastiano, K. Reimann, and K. A. A. Makinwa, "A 0.25mm² AC-biased MEMS microphone interface with 58dB SNR," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2013, pp. 382–383.
- [10] S. Song *et al.*, "A 430nW 64nV/vHz current-reuse telescopic amplifier for neural recording applications," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Nov. 2013, pp. 322–325.
- [11] Y. Nakagome *et al.*, "An experimental 1.5-V 64-Mb DRAM," *IEEE J. Solid-State Circuits*, vol. 26, no. 4, pp. 465–472, Apr. 1991.
- [12] K. Kimura, "Low voltage techniques for bias circuits," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 44, no. 5, pp. 459–465, May 1997.
- [13] T. Jang, M. Choi, S. Jeong, S. Bang, D. Sylvester, and D. Blaauw, "A 4.7nW 13.8ppm/°C self-biased wakeup timer using a switched-resistor scheme," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Jan. 2016, pp. 102–103.
- [14] Y. P. Chen, Y. Lee, J.-Y. Sim, M. Alioto, D. Blaauw, and D. Sylvester, "45pW ESD clamp circuit for ultra-low power applications," in *Proc. IEEE CICC*, Sep. 2013, pp. 1–4.
- [15] D. Jeon, M. Seok, C. Chakrabarti, D. Blaauw, and D. Sylvester, "A super-pipelined energy efficient subthreshold 240 MS/s FFT core in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 23–34, Jan. 2012.
- [16] J. E. Volder, "The CORDIC trigonometric computing technique," *IRE Trans. Electron. Comput.*, vol. EC-8, no. 3, pp. 330–334, Sep. 1959.



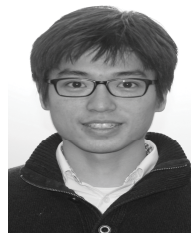
Seokhyeon Jeong (S'12) received the B.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2011, and the Ph.D. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2017.

He is currently with CubeWorks, Ann Arbor, researching and developing interface circuits for ultra-low power sensor nodes. His current research interests include ultra-low power temperature sensor, A/D converter, and the design of millimeter-scale computing systems.



Yu Chen received the M.S. degree in electrical and computer engineering from the University of Michigan, Ann Arbor, MI, USA, in 2016. He is currently pursuing the Ph.D. degree with the University of Michigan.

His current research interests include multimedia signal processing, machine learning, computer vision, and ultra-low power systems.



Taekwang Jang (S'06) received the B.S. and M.S. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, South Korea, in 2006 and 2008, respectively. He is currently pursuing the Ph.D. degree with the University of Michigan, Ann Arbor, MI, USA.

In 2008, he joined the Samsung Electronics Company Ltd., Yongin, South Korea, where he was involved in mixed signal circuit design including charge-pump and all-digital phase-locked loops for communication systems and mobile processors fabricated in 20 to 45 nm CMOS processes. His current research interests include clock generation, data converters, and ultra-low power system design.

Mr. Jang was the co-recipient of the 2009 IEEE Circuits and Systems Society Guillemain-Cauer Best Paper Award.



Julius Ming-Lin Tsai received the B.S. and Ph.D. degrees in power mechanical engineering from National Tsing Hua University, Hsinchu, Taiwan, in 1998 and 2004, respectively.

From 2003 to 2004, he was a Visiting Scholar with Carnegie Mellon University, Pittsburgh, PA, USA, where he demonstrated mechanical-noise dominated CMOS accelerometers. From 2004 to 2009, he was with VIA Technologies, New Taipei City, Taiwan, where he was involved in work on high-frequency small-signal modeling. From 2009 to 2013, he was a Principal Investigator with the Institute of Microelectronics, Singapore, and an Adjunct Assistant Professor with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. From 2013 to 2016, he was with InvenSense Inc., San Jose, CA, USA, working on advanced technology development, where he integrated AlN MEMS with CMOS for commercial applications. He is currently with Nextinput, Inc., San Jose. He has authored over 100 technical papers and holds more than 40 patents. His current research interests include RFMEMS, NEMS switches, MEMS ultrasound transducers (cMUT), and ruggedized MEMS sensors utilizing the advantage of silicon and AlN.



David Blaauw (M'94–SM'07–F'12) received the B.S. degree in physics and computer science from Duke University, Durham, NC, USA, in 1986 and the Ph.D. degree in computer science from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1991.

Until 2001, he was with Motorola, Inc., Austin, TX, USA, where he was the Manager of the High Performance Design Technology Group. Since 2001, he has been a Professor with the Faculty of the University of Michigan, Ann Arbor, MI, USA. He

has authored or co-authored over 500 papers and holds 60 patents. His research has a threefold focus. He has an extensive research in ultra-low-power computing using subthreshold computing and analog circuits for millimeter sensor systems and for high-end servers. His current research interests include cognitive computing using analog and in-memory neural-networks.

Dr. Blaauw has received numerous best paper awards and nominations. He won the Motorola Innovation Award. He has investigated adaptive computing to reduce margins and improve energy efficiency using a new approach he pioneered, called Razor, for which he received the Richard Newton GSRC Industrial Impact Award and the IEEE Micro annual Top-Picks Award. He received the 2016 SIA-SRC Faculty Award for lifetime research contributions to the U.S. semiconductor industry. His research group and collaborators introduced so-called near-threshold computing, which has become a common concept in semiconductor design. This work led to a complete sensor node design with record low power consumption, and was selected by the MIT Technology Review as one of the year's most significant innovations. He was a General Chair of the IEEE International Symposium on Low Power, a Technical Program Chair for the ACM/IEEE Design Automation Conference, and serves on the IEEE International Solid-State Circuits Conference's Technical Program Committee.



Hun-Seok Kim (S'10–M'11) received the B.S. degree in electrical engineering from Seoul National University, Seoul, South Korea, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles, CA, USA.

He is currently an Assistant Professor with the University of Michigan, Ann Arbor, MI, USA. He holds nine granted patents and has over ten pending applications in the areas of digital communication, signal processing, and low-power integrated circuits.

His current research interests include system novel algorithms and efficient VLSI architectures for low-power/high-performance signal processing, wireless communication, computer vision, and machine learning systems.

Dr. Kim received multiple fellowships from the Ministry of Information and Telecommunication, South Korea, Seoul National University, and UCLA. He was a Technical Staff Member of Texas Instruments Inc. from 2010 to 2014, while serving as an industry liaison for multiple university projects.



Dennis Sylvester (S'95–M'00–SM'04–F'11) received the Ph.D. degree in electrical engineering from the University of California, Berkeley, CA, USA, where his dissertation was recognized with the David J. Sakrison Memorial Prize as the most outstanding research in the UC-Berkeley EECS Department.

He is currently a Professor of electrical engineering and computer science with the University of Michigan, Ann Arbor, MI, USA, and the Director of the Michigan Integrated Circuits

Laboratory, a group of ten faculty and 70+ graduate students. He has held research staff positions with the Advanced Technology Group of Synopsys, Mountain View, CA, USA, Hewlett-Packard Laboratories, Palo Alto, CA, USA, and visiting professorships with the National University of Singapore, and Nanyang Technological University, Singapore. He is also a Consultant and a Technical Advisory Board Member for electronic design automation and semiconductor firms in his research areas. He co-founded Ambiq Micro, Austin, TX, USA. He has authored over 450 articles along with one book and several book chapters. He holds 37 U.S. patents. His current research interests include the design of millimeter-scale computing systems and energy-efficient near-threshold computing.

Dr. Sylvester was a recipient of an NSF CAREER Award, the Beatrice Winner Award at ISSCC, an IBM Faculty Award, an SRC Inventor Recognition Award, and ten best paper awards and nominations. He received the University of Michigan Henry Russel Award for distinguished scholarship. He was named as one of the top contributing authors at ISSCC, and most prolific author at the IEEE Symposium on VLSI Circuits. He was with the Technical Program Committee of the IEEE International Solid-State Circuits Conference and was with the Executive Committee of the ACM/IEEE Design Automation Conference. He was also a Distinguished Lecturer for the IEEE Solid-State Circuits Society from 2016 to 2017. He was an Associate Editor for the IEEE JOURNAL OF SOLID-STATE CIRCUITS, the IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, and the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION SYSTEMS.