A Digital Capacitive MEMS Microphone for Speech Recognition With Fast Wake-Up Feature Using a Sound Activity Detector

Youngtae Yang, *Graduate Student Member, IEEE*, Byunggyu Lee, Jun Soo Cho, *Member, IEEE*, Suhwan Kim, *Senior Member, IEEE,* and Hyunjoong Lee, *Member, IEEE*

***Abstract*—This brief presents a digital capacitive MEMS microphone for speech recognition with fast wake-up feature using a sound activity detector. The proposed microphone wakes up only when an acoustic event within the voice-band occurs using the sound activity detector, which can significantly reduce the average current consumption. All wake-up processes are performed on-chip, which enables the fast wake-up feature to prevent missing keywords. In addition, the proposed deglitching technique is applied to prevent the sound activity detector from responding to non-acoustic signals such as glitch signals. An auxiliary low-dropout regulator is used to further reduce the wake-up time. Our microphone consumes only 16 μA in deep-sleep mode. In active mode, a high-performance readout circuit converts the voice signal into a digital signal. The proposed microphone achieves an A-weighted signal-to-noise ratio (SNR) of 62.8 dBA and an acoustic overload point of 119.4 dB sound pressure level. The readout circuit itself of our microphone, fabricated in a 0.18 μm CMOS process, has an A-weighted SNR of 65.8 dBA.**

***Index Terms*—Speech recognition, MEMS, microphone, readout circuit, wake-up, voice activity detector, sound activity detector, deep-sleep or standby mode.**

I. Introduction

CAPACITIVE MEMS microphones have a lower noise floor and a higher sensitivity than piezoelectric MEMS microphones [1], making them more suitable for speech recognition. A capacitive MEMS microphone with a digital interface [2]–[4] is particularly suitable, but its power consumption is relatively high for speech recognition applications which must be constantly ready for input.

Voice-activated wake-up can significantly reduce power consumption since speech is only being received intermittently. Voice activity detectors are available [5]–[8], and have low power consumption. However, they support only voice triggered wake-up feature, not high-performance readout feature. Therefore, a conventional speech recognition system supporting a wake-up feature is configured using two separate packages. One is a wake-up circuit with a low-power MEMS transducer and the other is a high-performance readout circuit with a MEMS transducer. This significantly increases area and cost, as well as wake-up time. Too long of a wake-up time can cause keywords to be missed in speech recognition.

Integrating the wake-up and readout circuits on a single chip can reduce wake-up time. A piezoelectric analog MEMS microphone which has a built-in wake-up circuit has been proposed [9]. A piezoelectric digital MEMS microphone with a wake-up feature has also been introduced in our previous work [10]. However, their wake-up circuit cannot differentiate between voice and non-voice signals such as glitch signals. Therefore, the wake-up process requires a power-hungry digital signal processor (DSP) to eliminate extraneous sounds.

To overcome the above wake-up time and differentiating voice and non-voice signals issues, we introduce a high-performance, digital capacitive MEMS microphone with an ultra-fast wake-up time from deep-sleep mode in less than several milliseconds. Our microphone monitors sound constantly using a sound activity detector (SAD), which responds only to voice-band signals to reduce the time for which the readout circuit is awake. Non-voice signals such as glitch signals are largely eliminated by a deglitching technique. An auxiliary low-dropout regulator (LDO) is employed to further reduce wake-up time. The readout circuit is switched between deep-sleep and active mode by the SAD. In active mode, the readout circuit converts the incoming speech into a high-resolution digital signal.

The remainder of this brief is organized as follows. In Section II we present the architecture of our system. In Section III we describe the circuit implementation of the proposed architecture. Measurements obtained from a prototype are given in Section IV, and we conclude this brief in Section V.

II. System Overview

The proposed microphone, shown in Fig. 1 (a), consists of a MEMS transducer and a readout circuit with a wake-up circuit. The readout circuit switches between deep-sleep and active mode through the wake-up circuit, which reduces power consumption. Deep-sleep mode is defined as when the readout circuit is completely disabled and consumes only small leakage current. In deep-sleep mode, the auxiliary LDO is active

Manuscript received March 1, 2020; revised June 15, 2020; accepted July 3, 2020. Date of publication July 16, 2020; date of current version September 3, 2020. This work was supported by Samsung Research Funding Center of Samsung Electronics under Project SRFC-IT1701-04. This brief was recommended by Associate Editor A. De Marcellis. *(Corresponding author: Hyunjoong Lee.)*

Youngtae Yang, Byunggyu Lee, and Suhwan Kim are with the Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, South Korea, and also with the Inter-University Semiconductor Research Center, Seoul National University, Seoul 08826, South Korea.

Jun Soo Cho and Hyunjoong Lee are with Research and Development Department, Gwanak Analog Company, Ltd., Seoul 08826, South Korea (e-mail: hyunjoong.lee@gwanakanalog.com).

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

Digital Object Identifier 10.1109/TCSII.2020.3009926

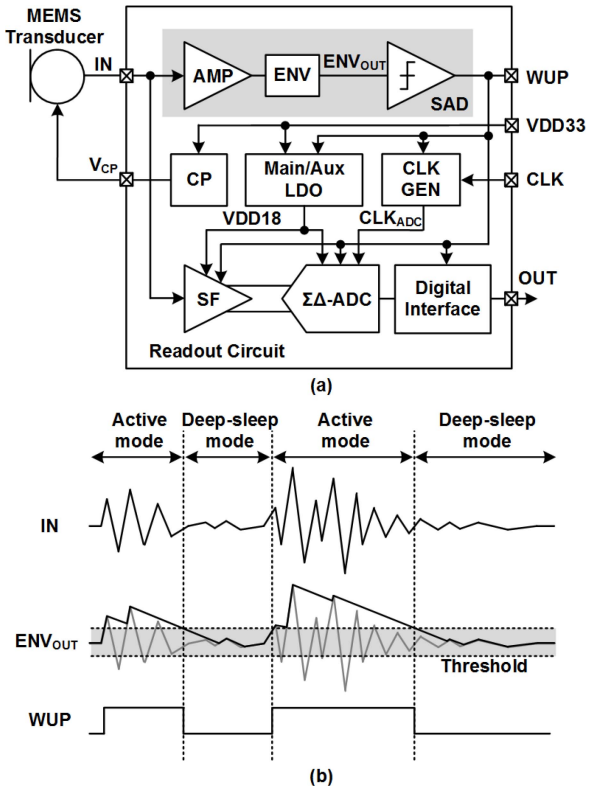

Fig. 1. (a) Architecture of our MEMS microphone and (b) its operation waveforms.

and the main LDO is turned off. In active mode, the main LDO is enabled and the readout circuit produces a 24-bit digital output. The wake-up process can be done without external circuits such as DSP, which reduces system cost and wake-up time.

A. Wake-Up Circuit

The wake-up circuit, which consists of SAD, charge pump (CP), and auxiliary LDO is always active and connected to the MEMS transducer. The SAD consists of an amplifier, an envelope detector, and a comparator. The CP applies a high bias voltage to the MEMS transducer, allowing the MEMS transducer to convert sound into an electrical signal. In deep-sleep mode, the auxiliary LDO maintains a voltage of 1.8 V (VDD18) instead of the main LDO while the readout circuit consumes small leakage current. All of the wake-up circuits operate from an external supply voltage of 3.3 V (VDD33).

The operation of the wake-up circuit is shown in Fig. 1 (b). The SAD amplifies the signal arriving from MEMS transducer (IN), determines its envelope (ENV_{OUT}), and compares it with a preset threshold. If that threshold is exceeded, a wake-up signal (WUP) goes high, which changes the operating mode of the readout circuit from deep-sleep to active. In the absence of an acoustic event, ENV_{OUT} decreases at a constant rate and eventually falls below the threshold voltage after a period of time, which causes the WUP signal to go low. The decay-rate of ENV_{OUT} can be adjusted to ensure that ENV_{OUT} remains above the threshold in between words of a sentence.

B. Readout Circuit

The readout circuit consists of a pseudo-differential source follower (SF), a delta-sigma analog-to-digital converter (ADC), a main LDO, and a clock generator (CLKGEN). The output signal from the SF is applied directly to the ADC without amplification to achieve a wide dynamic range and a high acoustic overload point (AOP) [4]. The SF has a single-ended input to match the single-ended MEMS transducer. The main LDO supplies a regulated supply voltage to the readout circuit, and CLKGEN generates the ADC operating clock signal using an external clock (CLK). The readout circuit only operates when the WUP signal is high. When the WUP signal is low, the readout circuit is turned off to minimize power consumption.

III. Circuit Design

A. Sound Activity Detector

Fig. 2 (a) is a schematic diagram of the SAD, and Fig. 2 (b) shows how it operates. A preamplifier in [11] is modified to interface with the MEMS transducer. Negative feedback through a Gm-C filter suppresses unwanted DC voltages such as DC offset. The gain of the preamplifier, expressed as R_F/R_S, is designed to be 20 dB. The output from the preamplifier (V_A) is further amplified by a programmable gain amplifier (PGA). The preamplifier and PGA also function together as a band-pass filter, allowing the SAD to respond only to voice-band signals from 600 Hz to 8 kHz while ignoring most other signals such as wind noise. The envelopes of the output signals of the PGA (A_P, A_N) are extracted using a conventional envelope detector (ENV) [12]. The output signals of the envelope detector (E_P, E_N) are compared with their respective threshold voltage. The logical results of E_P and E_N are combined using an AND operation. The output of the AND gate passes through the hold circuit and is output as the WUP signal. Hysteresis is introduced into the comparator by changing the threshold voltage (V_{TH}) between two voltages (V_{TH_H}, V_{TH_L}), determined by the WUP signal. Since the SAD is always on, it is important to minimize its power consumption. The threshold voltage level of sound activity detection is significantly higher than the thermal noise floor, and thus the SAD does not need to dissipate excessive power to minimize the thermal noise. Moreover, the power consumption can be further reduced, as voice signals have a low bandwidth. The total current consumption of the proposed SAD is 2.5 μA.

A glitch in the input signal to the SAD, which reduces the accuracy of the SAD, can be generated for various reasons such as simultaneously switching the operation mode of the readout circuit. Therefore, the SAD incorporates a deglitching technique. Unlike a voice signal, the glitch signal exhibits a unipolar pulse-like behavior, which is easily distinguished from a voice signal. The unipolar glitch signal shown in Fig. 2 (b) causes only E_P to exceed the threshold voltage, thus the output of the AND operation is logical zero and the WUP remains low.

More serious glitches can occur internally when the readout circuit is powered up or down, potentially causing a positive feedback loop that can destabilize the wake-up mechanism. To eliminate the possibility altogether, the SAD has three operation states: reset, hold, and detect. In the detect state, the SAD monitors the input signal and updates the WUP signal. If the WUP signal changes, a glitch signal can be generated by switching the operation mode of the readout circuit; since the readout circuit and the SAD share the same input signal, switching the operation mode such as the power-up operation of the readout circuit could cause the SAD to malfunction. To prevent this problem, the SAD enters the reset state, in which

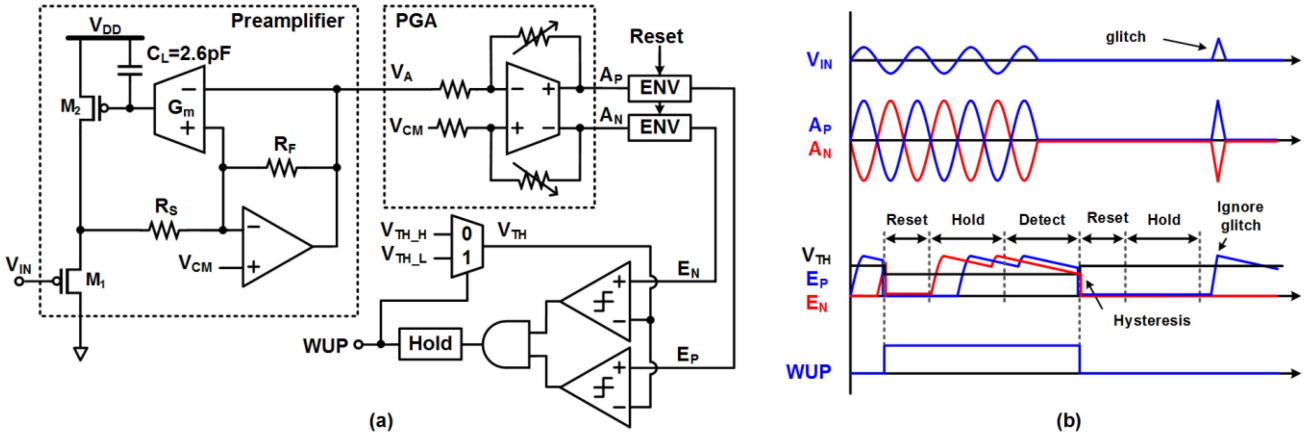

Fig. 2. (a) The sound activity detector schematic and (b) its operation waveforms.

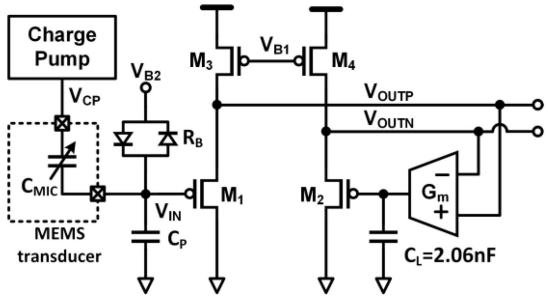

Fig. 3. The low-noise pseudo-differential source follower schematic.

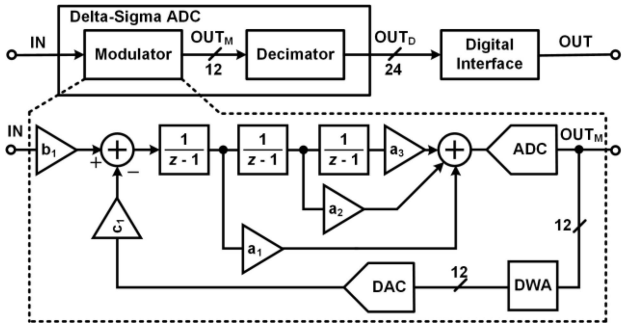

Fig. 4. The block diagram of delta-sigma ADC.

it ignores the input signal. After a short period, it enters the hold state. In this state, SAD operates, but the WUP signal holds the previous value. The hold state prevents the SAD from switching to detect mode immediately after reset mode. Without a hold state, the WUP signal could be changed to zero because either E_P or E_N would be smaller than V_{TH}, even though a voice signal existed. The reset state lasts 0.5 ms and the hold state lasts 2.5 ms.

B. Pseudo-Differential Source Follower

Source followers are widely used as MEMS readout interfaces due to their high input impedance and excellent driving capacity. Fig. 3 shows a schematic diagram of a source follower with a pseudo-differential structure, which allows it to act as an interface between a single-ended MEMS transducer and a differential delta-sigma ADC [2], [3]. This structure also suppresses common-mode and supply noise. The source follower used in previous designs [2], [3] is subject to a DC offset due to mismatch between M_1 and M_2. In our design, this offset is canceled by a feedback loop composed of a Gm-C filter. The unity-gain bandwidth of this filter is set to be less than 20 Hz so that there is no signal loss in the audio band. The input bias voltage is set by a high-impedance back-to-back diode (R_B). A capacitive MEMS transducer for a microphone can be modeled as a variable capacitor (C_{MIC}) whose value changes with sound pressure [13]. A large parasitic capacitor (C_P) can reduce the sensitivity of MEMS transducers. However, reducing the size of M_1 to minimize C_P causes performance degradation due to increased flicker noise. Therefore, the size of M_1 is chosen to maximize the ratio of sensitivity to noise power. The charge pump applies a high bias voltage to the MEMS transducer to detect sound. The output voltage of the charge pump (V_{CP}) is adjustable from 10 V to 16 V.

C. Delta-Sigma ADC

As shown in Fig. 4, the delta-sigma ADC consists of a 12-level delta-sigma modulator and decimator, and the output of the ADC is transferred to digital interface. The 3rd-order delta-sigma modulator has cascade-of-integrators with feedforward (CIFF) structure. The CIFF structure reduces the burden on the amplifiers, as the outputs of amplifiers are smaller than the cascade-of-integrators with feedback (CIFB) structure. A high signal-to-quantization-noise ratio is achieved through the use of an oversampling ratio of 256 and a 12-level quantizer. The delta-sigma modulator is implemented as a switched-capacitor circuit. To achieve target noise performance, the value of the sampling capacitor of the 1st integrator is 15 pF. Noise shaping enables the 2nd and 3rd integrators to be implemented with small capacitors. Data-weighted averaging (DWA) compensates for inaccuracies introduced by the non-linearity of a digital-to-analog converter (DAC) in this modulator. The output of the modulator (OUT_M) is converted to a 24-bit signal (OUT_D) by the decimator. A digital interface transmits the final output (OUT) following Sony/Philips Digital Interface (S/PDIF) standard.

D. Main and Auxiliary Low-Dropout Regulator

Fig. 5 is a schematic diagram of the main and auxiliary LDO, which both are designed as capacitor-less LDOs.

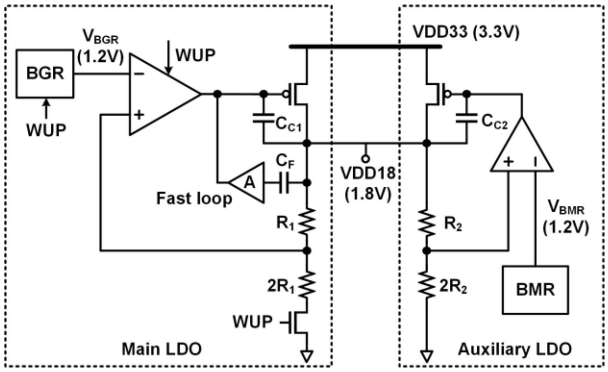

Fig. 5. The main and auxiliary low-dropout regulator.

The outputs of two regulators are connected, the auxiliary LDO is present to reduce wake-up time. Each LDO receives the reference voltages (V_{BGR}, V_{BMR}) from a bandgap reference (BGR) and a beta-multiplier (BMR), respectively, and produces a 1.8 V output voltage (VDD18). The output response of the main LDO is improved by adding a fast transient path [14]. The main LDO which supplies current to the readout circuit, can be turned on and off by the WUP signal. The auxiliary LDO is always on and keeps VDD18 at 1.8 V, virtually eliminating settling time so that the readout circuit can be woken up quickly.

IV. Measurement

The readout circuit of our MEMS microphone was fabricated in a 0.18 μm CMOS process with an area of 3.71 mm^2. The microphotograph of the readout circuit, as well as the unpackaged microphone and measurement setup, are also shown in Fig. 6. Measurement was conducted in an anechoic chamber, where the prototype board was fixed to the holder and the acoustic signal was applied through the speaker. We tested the wake-up function by measuring the WUP and OUT signals using an input signal containing the spoken word "hello" and a glitch signal. The digital output signal was passed to an external DAC and its input and output signals were displayed on an oscilloscope. As shown in Fig. 7, in response to the "hello" signal, WUP goes high and turns on the readout circuit, which starts to digitize the voice input. When there is no input, the WUP signal goes low and the readout circuit goes into deep-sleep mode. The SAD does not respond to the glitch signal, indicating that the deglitching technique is effective. Fig. 8 shows the magnified waveform of Fig. 7 at wake-up. The time to the first valid output is 5.7 ms, of which 1.7 ms is the latency of the system. Wake-up time is 4 ms excluding system latency. When the waveform of Fig. 7 was played back through a speaker, the effect of the wake-up time on the spoken word was barely noticeable.

The power spectrum of the microphone system is shown in Fig. 9. An audio precision AP2722 analyzer was used to supply a 1 kHz sine wave to a speaker arranged to produce a standard calibration sound pressure of 94 dB sound pressure level (SPL) at the microphone. The AP2722 analyzer also acquired the digital output of the microphone. At low frequencies, a 1/f spectrum appears due to flicker noise, and 2nd harmonic distortion is observed at 2 kHz due to the single-ended MEMS transducer. The measured microphone sensitivity is −37.2 dBFS, and its A-weighted signal-to-noise ratio (SNR) is 62.8 dBA. The noise performance of the readout

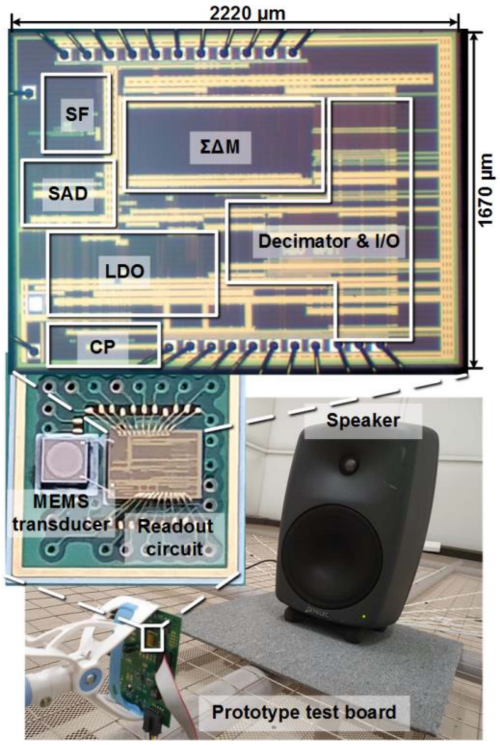

Fig. 6. Measurement setup and a microphotograph of the readout circuit.

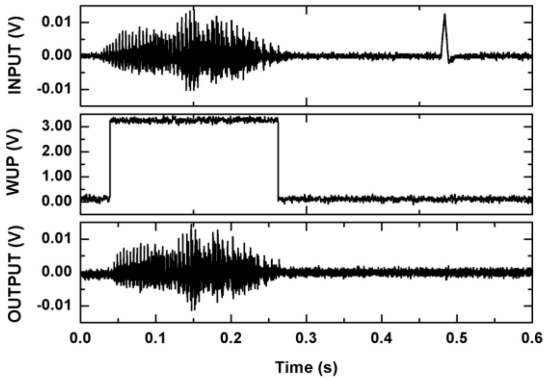

Fig. 7. Waveforms output by microphone in response to the test input.

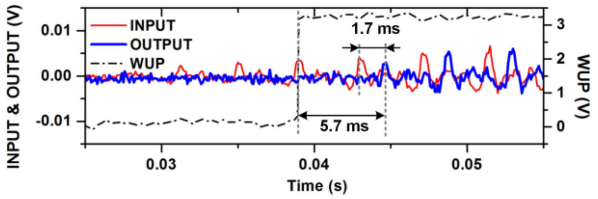

Fig. 8. Measured wake-up time of the microphone.

circuit alone was measured by applying a 1 kHz signal with a voltage corresponding to the output of the microphone when it is receiving an acoustic input of 94 dB SPL. The A-weighted SNR of the readout circuit was found to be 65.8 dBA. Fig. 10 shows the variation of the A-weighted SNR and signal-to-noise and distortion ratio (SNDR) of the microphone and readout circuit with the amplitude of the input signal. The SNR of the microphone is approximately 3dB lower than the SNR of the readout circuit, which can be attributed to the noise generated by the MEMS transducer. Above 94 dB

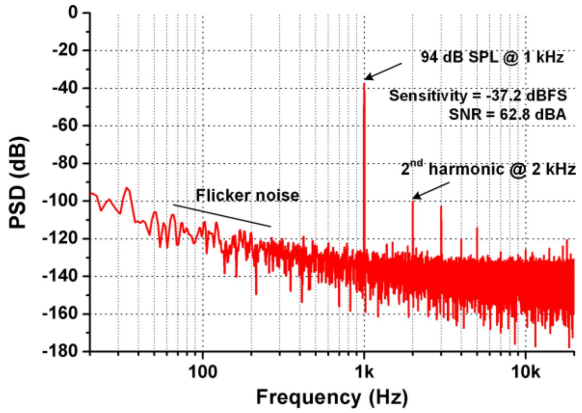

Fig. 9. Measured power spectrum density of the MEMS microphone, for a 1 kHz sine wave input.

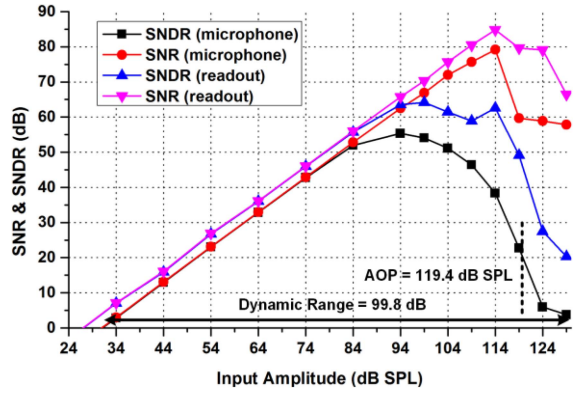

Fig. 10. Measured A-weighted SNR/SNDR versus amplitude of the sound input.

TABLE I
COMPARISON WITH OTHER DIGITAL CAPACITIVE MEMS MICROPHONES

Parameter	**This Work**	[2]	[3]	[4]
MEMS type	**Single-Ended**	Single-Ended	Single-Ended	Differential
Wake-up feature	**Yes**	No	No	No
Wake-up time [ms]	**4**	N/A	N/A	N/A
Technology [μm]	**0.18**	0.35	0.25	0.13
Supply voltage [V]	**3.3**	1.8	1.8	1.8
Sensitivity @ 94 dB SPL [dBFS]	**-37.2**	-42	-26	-46
SNR @ 94 dB SPL [dBA]	**62.8**	63 [a]	63 [a]	67
Dynamic range [dB]	**99.8**	80 [a]	83 [a]	113
AOP [dB SPL]	**119.4**	N/A	N/A	136
Active current [mA]	**2.75**	0.46	0.47	1.2
Deep-sleep current [μA]	**16**	N/A	N/A	N/A

[a] Readout circuit only

SPL, the SNDR of the microphone and the readout circuit drop off due to non-linearity. However, a comparison of the SNDR of the microphone and the readout circuit shows that the readout circuit does not limit the distortion performance. The acoustic overload point, defined as the maximum sound pressure level where the total harmonic distortion exceeds 10 %, was found to be 119.4 dB SPL. The measured current consumption, excluding digital interfaces and test circuitry, was 2.75 mA and 16 μA in active and deep-sleep modes, respectively. In deep-sleep mode, the SAD draws 2.5 μA, the charge pump draws 12 μA, and the remaining circuit, including the auxiliary LDO, draws 1.5 μA. Table I summarizes the performance of our microphone and compares it with that of other digital capacitive MEMS microphones [2]–[4], which have low active current consumption but no wake-up feature. Our microphone and [4] have higher dynamic ranges compared to [2] and [3]. The active current consumption of our microphone is higher than [4], but the average current consumption can be lower than [4], depending on the ratio of deep-sleep mode and active mode.

V. CONCLUSION

We have presented a digital capacitive MEMS microphone with a sound activity detector that activates the readout circuit using voice-band input. The wake-up procedure is fast because it takes place on-chip. A deglitching technique is introduced to prevent the SAD from responding to glitch noise generated during changes in operating mode. The wake-up time of the microphone is 4 ms and the total current consumption in deep-sleep mode is 16 μA. Our microphone with an AOP of 119.4 dB SPL achieves an A-weighted SNR of 62.8 dBA, and our readout circuit achieves an A-weighted SNR of 65.8 dBA.

REFERENCES

[1] R. J. Littrell, "High performance piezoelectric MEMS microphones," Ph.D. dissertation, Dept. Doctor Philos., Univ. Michigan, Ann Arbor, MI, USA, 2010.

[2] S. A. Jawed, D. Cattin, M. Gottardi, N. Massari, A. Baschirotto, and A. Simoni, "A 828μW 1.8V 80dB dynamic-range readout interface for a MEMS capacitive microphone," in *Proc. IEEE 34th Eur. Solid-State Circuits Conf. (ESSCIRC)*, Edinburgh, U.K., Sep. 2008, pp. 442–445.

[3] A. Barbieri and G. Nicollini, "A 470μA direct readout circuit for electret and MEMS digital microphones," in *Proc. 20th Int. Conf. Electron. Circuits Syst. (ICECS)*, Abu Dhabi, UAE, Dec. 2013, pp. 341–344.

[4] E. Bach *et al.*, "9.5 A 1.8V true-differential 140dB SPL full-scale standard CMOS MEMS digital microphone exhibiting 67dB SNR," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2017, pp. 166–167.

[5] T. Delbruck, T. Koch, R. Berner, and H. Hermansky, "Fully integrated 500μW speech detection wake-up circuit," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Paris, France, May 2010, pp. 2015–2018.

[6] K. M. H. Badami, S. Lauwereins, W. Meert, and M. Verhelst, "A 90 nm CMOS, 6μW power-proportional acoustic sensing frontend for voice activity detection," *IEEE J. Solid-State Circuits*, vol. 51, no. 1, pp. 291–302, Jan. 2016.

[7] M. Yang, C.-H. Yeh, Y. Zhou, J. P. Cerqueira, A. A. Lazar, and M. Seok, "A 1μW voice activity detector using analog feature extraction and digital deep neural network," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2018, pp. 346–348.

[8] M. Cho *et al.*, "17.2 A 142nW voice and acoustic activity detection chip for mm-scale sensor nodes using time-interleaved mixer-based frequency scanning," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2019, pp. 278–279.

[9] *Low-Noise Bottom Port Analog Piezoelectric MEMS Microphone With Wake on Sound*, document VM1010, Vesper, Boston, MA, USA, Feb. 2019.

[10] Y. Yang, J. S. Cho, B. Lee, and S. Kim, "A sound activity detector embedded low-power MEMS microphone readout interface for speech recognition," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design (ISLPED)*, Lausanne, Switzerland, Jul. 2019, pp. 1–6.

[11] M. Baker and R. Sarpeshkar, "A low-power high-PSRR current-mode microphone preamplifier," *IEEE J. Solid-State Circuits*, vol. 38, no. 10, pp. 1671–1678, Oct. 2003.

[12] M. W. Kruiskamp and D. M. Leenaerts, "A CMOS peak detect sample and hold circuit," *IEEE Trans. Nuclear Sci.*, vol. 41, no. 1, pp. 295–298, Feb. 1994.

[13] S. A. Jawed, D. Cattin, N. Massari, M. Gottardi, B. Margesin, and A. Baschirotto, "A simplified modeling approach for a MEMS capacitive sensor," in *Proc. IEEE 18th Eur. Conf. Circuit Theory Design (ECCTD)*, Seville, Spain, Aug. 2007, pp. 112–115.

[14] R. J. Milliken, J. Silva-Martinez, and E. Sanchez-Sinencio, "Full on-chip CMOS low-dropout voltage regulator," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 9, pp. 1879–1890, Sep. 2007.