An audio ASIC incorporating a 76dB-A VCO-ADC and a $3\mu W$ time-encoded feature extraction circuit.

Dante Loi, Victor Medina, Javier Fernandez, Luis Hernandez, *Senior Member, IEEE*

***Abstract*—This paper describes a 130 nm CMOS chip demonstrating a new concept in voice recognition at the edge. The chip can operate in two modes. In the active mode, a VCO encodes the audio signal with 76 dB-A of SNDR-peak consuming $175\mu W$. In the feature extraction mode, the biasing of the VCO is reduced to $4\mu A$ and the VCO feeds the frequency-encoded audio signal to a filter bank that calculates the audio energy in 16 frequency bands. The digitized energy information is transmitted off-chip via a low speed serial interface. The filters are implemented as band-pass bi-quads using frequency-encoded integrators based on asynchronous counters. Due to its digital nature, the filter response is independent of PVT variations, does not require calibration and is fully programmable. Each filter consumes 187nW at 0.6V of supply and occupies an active area of 0.007mm^2.**

***Index Terms*—VCO-ADC, Feature Extraction, Time Encoding**

I. Introduction

In recent years, low power voice recognition hardware has been explored to enable edge applications operated with batteries or energy harvesting. Most of the proposed solutions include a feature extraction stage and a signal identification block implementing Voice Activated Detection (VAD) or Key Word Spotting (KWS). Typical hardware implementations pack in a single chip either an analog feature extraction block [1] followed by a neural network classification circuit [2] (see Fig.1.(a)) or use an analog to digital converter and Digital Signal Processing (DSP) for feature extraction [3] (see Fig.1.(b)). Analog solutions excel in power efficiency but require large capacitors and tend to depend on Process, Voltage and Temperature (PVT) variations. DSP solutions feature PVT invariance and low area, but their consumption exceeds that of analog solutions. In many applications, once a voice or keyword are detected, high quality audio data is required, for instance to maintain a conversation. Including in a microphone ASIC the KWS hardware is not very flexible as the detection algorithm or command data set is difficult to update. On top of that, OEM manufacturers may prefer their proprietary sound identification algorithms instead of those built in edge devices.

In this paper, we propose a different system architecture, shown in Fig. 1.(c), where the readout ASIC of a MEMS microphone hosts a full performance audio ADC together with a feature extraction front-end. This architecture has been demonstrated in a proof-of-concept chip. The chip benefits from the combination of a VCO-ADC to implement the audio signal digitization [4], [5] and a bank of frequency encoded filters [6], [2] which seamlessly integrate with the VCO output. When the chip works in the feature extraction mode, the frequency to digital block of the VCO-ADC is powered off and the VCO oscillation is used to clock some counters that operate as integrators, implementing the filtering function. Once the signal energy over the different filter bands is calculated, it is transmitted via a low bit-rate serial interface to an application processor to implement the audio classification. The frequency encoded filters proposed here, are PVT independent due to their digital nature and do not require a full ADC or large capacitors. This solution relieves the application processor from the power consuming task of feature extraction [1], reduces the power of the digital interface and enables a flexible implementation of sound recognition algorithms.

This work was supported by H2020-MSCA-ITN-2020 grant Nr. 956601
The authors are with the Electronics Technology Department, Carlos III University, 28911 Madrid, Spain. (e-mail: dloi@ing.uc3m.es)

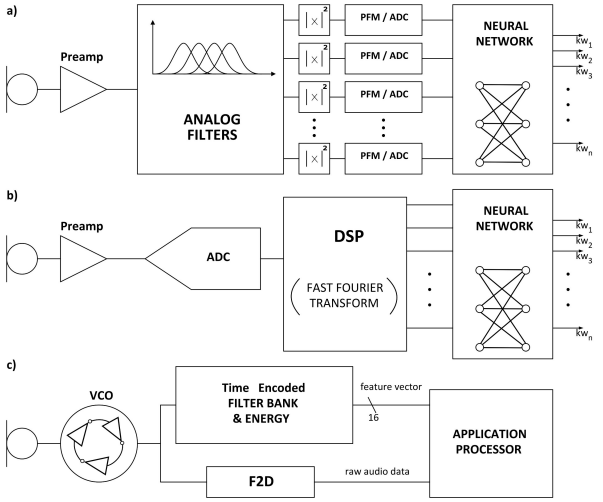

Fig. 1: Implementation options for edge voice recognition

II. System level description

The block diagram of the chip is displayed in Fig. 2. The system features an open loop VCO-ADC implemented with a pseudo differential architecture [5]. Two identical VCO Ring Oscillators (VCRO) encode the signal of a constant-charge capacitive MEMS microphone [4]. The frequency-encoded signal generated by the VCROs can be forwarded to a standard Frequency to Digital (F2D) circuit based on a sampler and a first order differentiator. Alternatively, the VCRO signal can drive a time-based 16-channel filter bank capable of extracting the energy of the input signal in different bandwidths. The filters are the BPF blocks in Fig. 2 and are implemented as biquadratic band pass filters with a digitally programmable center frequency f_c. Each channel is equipped

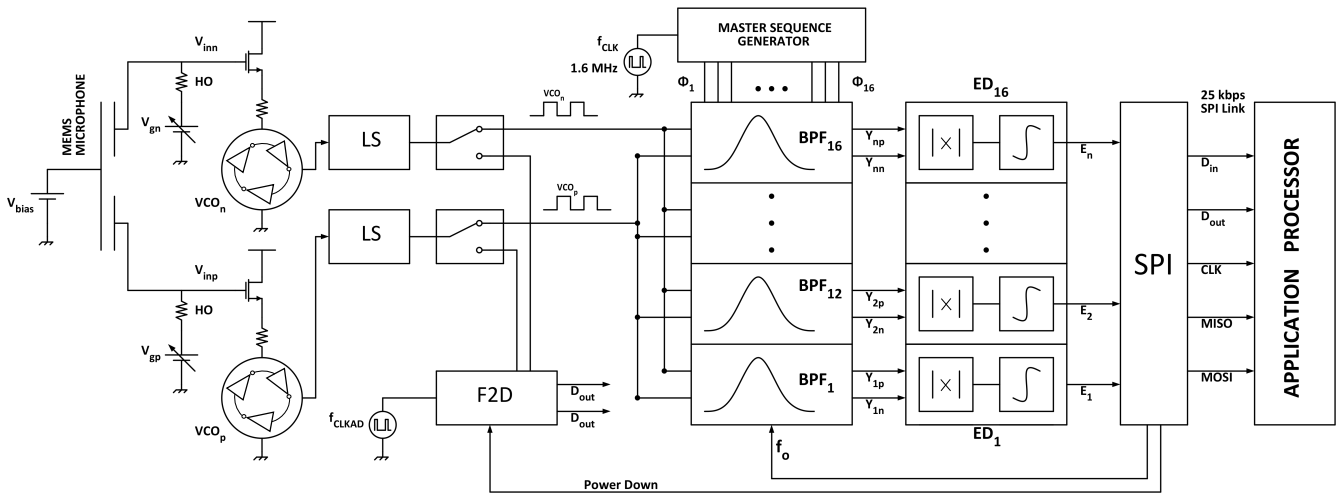

Fig. 2: Block diagram

with a frequency-encoded rectifier and an integrate-and-dump circuit to estimate the energy of the input signal corresponding to the bandwidth of the channel (Energy Detector blocks, ED). The time window used for energy integration is also programmable. The computed energy is represented with a 16-bit digital value, which can be accessed via a serial communication module (SPI block). The speed of the serial communication depends on the integration time window, because the previous calculated energy value must be read before a new integration starts.

A. Frequency encoded biquadratic filter

Typically, feature extraction circuits use second-order biquadratic filters [1], where the input is assumed a voltage signal and the filter is built with amplifiers and integrators (see Fig. 3(a)). We can also encode an analog signal using Pulse Frequency Modulation (PFM) [6] with a VCO. Integration can be directly performed on PFM encoded signals, by counting pulses over time. This feature of PFM signals provides a mostly digital primitive that can be used to implement filters. Fig. 3(b) shows a biquadratic filter with PFM integrators, which uses digital up/down counters and Digitally Controlled Oscillators (DCOs) to re-encode in PFM the intermediate state variables. To implement a differential filter we can connect two sections as displayed in Fig. 3(c).

B. p-DCO based Integrator

In Fig. 3(b), we need to convert the multilevel digital signal at the output of a counter into a frequency encoded pulse stream. This function is conventionally implemented by cascading a DAC to a VCO, (DCO) [6], [2], [5]. Such solution requires a significant area and is PVT dependent. An alternative to the DAC + VCO solution is given by the numerically controlled oscillators used in Direct Digital Synthesizers (DDS), which are scalable and fully digital although they require a sampling clock. In this chip we introduce a simplification of a DDS that will be designated as pseudo-DCO or p-DCO [7]. The p-DCO is a finite state machine

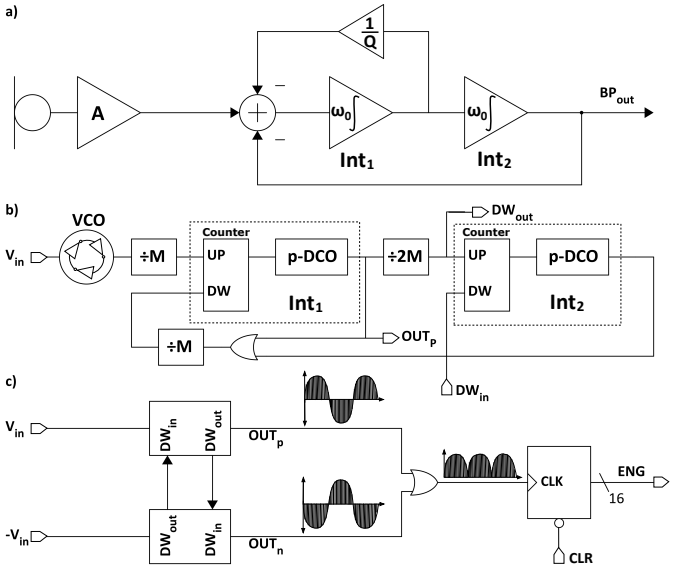

Fig. 3: a) analog biquad, b) frequency-encoded biquad c) differential configuration with energy estimation.

capable of encoding a frequency control input into a pulse sequence that, on average, has a frequency proportional to the frequency input. Compared to a DDS, the p-DCO is simpler at the cost of adding some phase error. This phase error manifests as in-band quantization noise. However, the SNR required for energy measurement in voice feature extraction tolerates such in band quantization errors. The idea behind the p-DCO is to generate a set of non-overlapping pulse signals with base frequencies to be cast around the chip, and subsequently combine them to generate the desired output rate. In this chip, the set of frequencies is generated by the Master Sequence Generator block in Fig. 2. A simplified example of this block is depicted in Fig. 4(a), where a ring counter with clock frequency f_{clk} is connected to a set of edge detectors to produce non overlapping pulse sequences p_i (see Fig. 4(b)). These pulse sequences are combined with logic

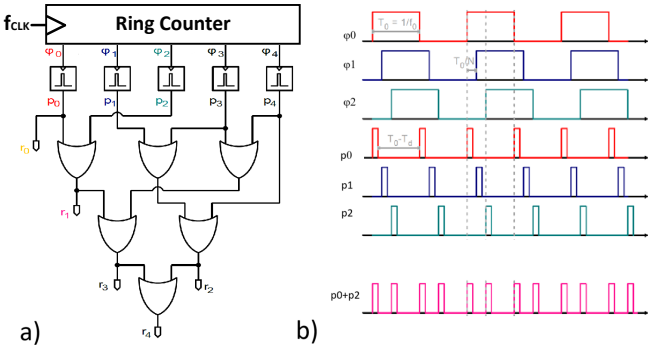

Fig. 4: Master Sequence Generator example.

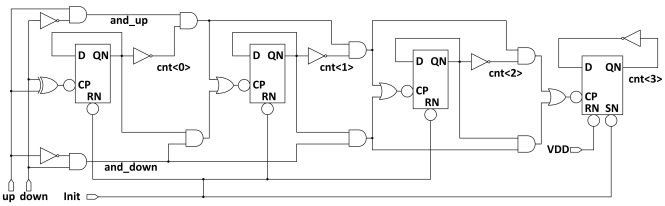

Fig. 5: Asynchronous up/down counter.

gates to produce an output signal with a digitally controllable pulse rate. For instance, in Fig. 4(b) we show how the logic OR of p_0 and p_2 produces a double frequency signal r_1.

C. Frequency-encoded Energy Detector

The biquad filters of Fig. 2 provide a frequency-encoded differential output. To compute the energy per filter band, the circuit of Fig. 3(c) is used [8], [2]. This circuit calculates the absolute value of the differential signal by selecting which of the outputs has a larger frequency (logic-OR). Afterwards, energy is estimated by an integrate-and-dump counter at programmable time intervals.

III. Circuit Design

The input VCOs are implemented as a 21-tap ring oscillator driven by an NMOS source follower (see Fig. 2). The circuit of this input stage is identical to that in [5]. A resistive degeneration at the source of the transistor improves its linearity by providing negative feedback at the inverter control nodes. Each VCO output phase is equipped with a level shifter (LS in Fig. 2) to comply with the downstream digital electronics voltage levels. The rest frequency is set with an external bias DC voltage. The VCO signals are cast to the filter bank, with rest frequencies centered at 1.6 MHz.

The integrators in the biquad filters are constructed around the asynchronous 4 bit up/down counters displayed in Fig. 5. The output of every counter is re-encoded in frequency by the p-DCOs, which share a common master sequence generator similar to that in Fig. 4(a) but with 16 different phases Φ_i. The sensitivity of the p-DCOs present in the biquad filters can be digitally programmed by a division factor M, which also serves to average the phase error introduced during the rate combination process (see Fig. 4(b)). We have adopted a

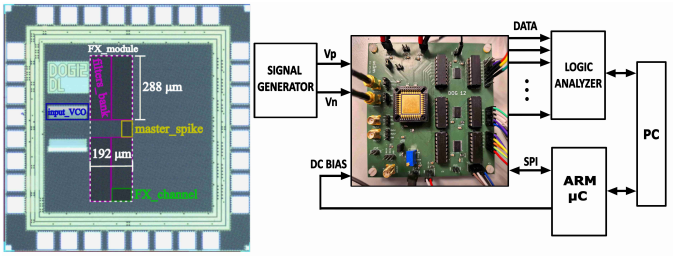

Fig. 6: Micrograph and Test setup.

pseudo-differential topology employing a cross-coupled configuration of two PFM branches. Considering 16 possible codes at the counters (4 bit), the operating point should be set at the middle (code 8). Given $f_{clk} = 1.6MHz$ for the master sequence generator, the resulting p-DCO gain K_{DCO} is 100 kHz/LSB and the DC operating point corresponds to an input rest frequency of 800kHz. The center frequency f_c of the filter can be expressed as $f_c = K_{DCO}/(2\pi M)$. The filter dividers have 8 bits, therefore M can range between 1 and 255. This results in 255 possible center frequencies ranging from $f_c = 15.9kHz$ down to $f_c = 62Hz$. In general, to ensure proper operation and particularly due to the double feedback loop present at the first integration stage, the input rest frequency provided by the VCOs must be 1.6 MHz, twice the rest frequency defined by the in-loop p-DCOs. All filters are configured to have a Q factor of 1.

The energy detectors (ED in Fig. 2) are equipped with a 16 bits counter that allows integration times of up to 100 ms, for a 100 kHz/LSB p-DCO gain. In absence of signal, a baseline energy is given by the DC level of the integrators. This DC level is digitally removed off-chip prior to data representation.

IV. Implementation and Measurements

The chip, implemented in a 130nm CMOS process, is shown in Fig. 6 and occupies an active area of 0.138 mm^2 of which 0.112 mm^2 are the feature extraction circuitry. To test the chip (see Fig. 6), a differential chirp signal has been applied with an arbitrary waveform generator. The ADC output and the digital outputs of the filters can be captured with a logic analyzer. The energy data is read at 25 kb/s using the built-in serial interface of the chip (see Fig. 2), connected to a microcontroller board. The microcontroller board generates the DC biasing voltages for the input VCOs to adjust the center frequency and compensate DC offset. Fig. 9 shows the FFT of a transistor simulation of the VCOs biased at 0.73V and with a 25mVp input signal, yielding a SNDR of 50dB in a 10 kHz bandwidth, sufficient for VAD or KWD. Under these conditions, the VCO signal could drive the filter bank at 1.6MHz, keeping the power of the VCOs below $4\mu W$. When the VCOs are biased to its nominal value (1.21V), they reach the full ADC dynamic range and peak SNDR of 76dB-A, at the expense of consuming $175\mu W$. See [5] for detailed measurements in this mode. Due to practical limitations in the level shifters in [5] (LS in Fig. 2), the measurements have been done with the nominal biasing and a frequency divider. Fig. 7 shows the time-domain chirp response of the filter centered at

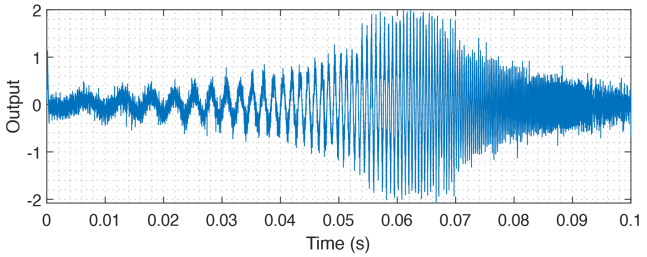

Fig. 7: Response of filter at $f_c = 1kHz$ to a frequency chirp

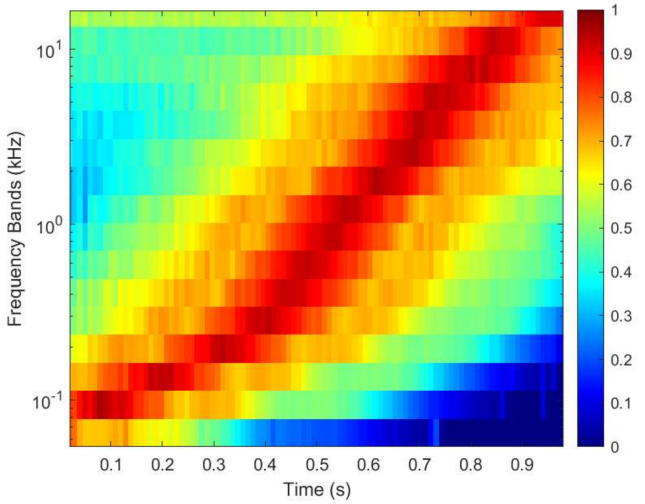

Fig. 8: Spectrogram.

1kHz (M=16). This signal is the post processed PFM output (signal OUT_p in Fig. 3(b)) captured with the logic analyzer. The dynamic range for a 1kHz tone located in the center frequency has been measured on the OUT_p output as the input singal range with positive SNDR, reaching 43dB. Fig. 8 shows the spectrogram corresponding to a linear frequency chirp, measured from the energy output and after DC offset removal and normalization. Power consumption is 187nW per filter powered at 0.6V. Considering that the input VCOs would consume $4\mu W$ biased at 0.73V, the total power of the chip in feature extraction mode would be below $7\mu W$.

V. Conclusions

This paper has presented a novel 130 nm CMOS chip designed for efficient voice feature extraction at the edge, integrating a high-performance VCO-ADC and a digitally programmable, frequency-encoded 16-channel filter bank. Table I shows a comparison of our chip with other chips showing similar functionality (FoM as in [2]). Chip results demonstrate exceptional power efficiency (187 nW at 0.6 V supply) and compactness (0.007 mm^2 per filter), being suitable for battery-powered or energy-harvesting IoT applications. Furthermore, the fully digital nature of the filter bank ensures robustness against process, voltage, and temperature variations, eliminating calibration requirements and enhancing programmability.

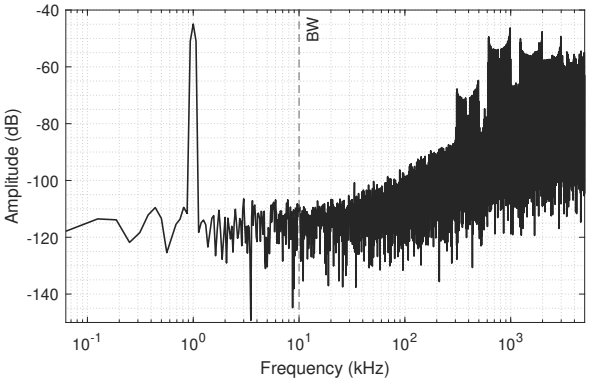

Fig. 9: FFT of VCO with reduced biasing

TABLE I: Comparison with other works

	[9]	[10]	[11]	[2]	**This w.**
Process (nm)	180	90	180	65	130
mm^2/Channel	0.26	0.13	0.1	0.1	0.007
Nr. Channels	64×2	16	16	16	16
F. Range (Hz)	8-20k	75-5k	100-5k	111-10.4k	62-16k
Supply (V)	0.5	-	0.6	0.5	0.6
Power (μW)	55	6	0.38	9.3	7
Frame Shift (ms)	-	31.25	10	16	10
DR (dB)	55	45	40	54.89	43
FoM$_{S,DR}$(dB)	-	82.3	91.5	93.11	114
Supports Mic	No	✓	No	✓	✓

References

[1] K. Kim and S.-C. Liu, "Continuous-time analog filters for audio edge intelligence: Review on circuit designs," *IEEE CASM*, vol. 23, no. 2, pp. 29–48, 2023.

[2] K. Kim, C. Gao, R. Graça, I. Kiselev, H.-J. Yoo, T. Delbruck, and S.-C. Liu, "A 23-μW Keyword Spotting IC With Ring-Oscillator-Based Time-Domain Feature Extraction," *JSSC*, vol. 57, no. 11, pp. 3298–3311, 2022.

[3] L. Zhu, W. Shan, J. Xu, and Y. Lu, "AAD-KWS: a sub-μW keyword spotting chip with a zero-cost, acoustic activity detector from a 170nW MFCC feature extractor in 28nm CMOS," in *ESSCIRC*, 2021, pp. 99–102.

[4] C. Perez, R. Garvi, G. Lopez, A. Quintero, F. Leger, P. Amaral, A. Wiesbauer, and L. Hernandez, "A vco-based adc with direct connection to a microphone mems, 80-db peak sndr and 438-w power consumption," *IEEE Sensors J.*, vol. 23, no. 8, pp. 8466–8477, 2023.

[5] V. Medina, R. Garvi, J. Granizo, P. Amaral, and L. H. Corporales, "Second-order vco-adc architecture with low-area and high dynamic range using internal binary encoding," *TCAS-I*, pp. 1–12, 2025.

[6] L. Hernandez, E. Gutierrez, and F. Cardes, "Frequency-encoded integrators applied to filtering and sigma-delta modulation," in *ISCAS 2016*, pp. 478–481.

[7] D. Loi, V. Medina, and L. H. Corporales, "A second-order true-vco adc employing a digital pseudo-dco suitable for sensor arrays," *MDPI Sensors*, vol. 24, no. 24, 2024.

[8] D. Loi, N. Zbida, L. Hernandez, and A. Wiesbauer, "A vco-based voice activity detection system for iot applications," in *IEEE Austrochip*, 2022, pp. 13–16.

[9] M. Yang, C.-H. Chien, T. Delbruck, and S.-C. Liu, "A 0.5 V 55 μW times 2 Channel Binaural Silicon Cochlea for Event-Driven Stereo-Audio Sensing," *JSSC*, vol. 51, no. 11, pp. 2554–2569, 2016.

[10] K. Badami, S. Lauwereins, W. Meert, and M. Verhelst, "A 90 nm CMOS, 6μW Power Proportional Acoustic Sensing Frontend for Voice Activity Detection," *JSSC*, vol. 51, no. 1, pp. 291–302, Jan. 2016.

[11] M. Yang, C.-H. Yeh, Y. Zhou, J. P. Cerqueira, A. A. Lazar, and M. Seok, "Design of an Always-On Deep Neural Network-Based 1μW Voice Activity Detector Aided With a Customized Software Model for Analog Feature Extraction," *JSSC*, vol. 54, no. 6, pp. 1764–1777, 2019.