Low Power Speech Detector On A FPAA

Sahil Shah and Jennifer Hasler
Department of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332
Email: jennifer.hasler@ece.gatech.edu

***Abstract*—This paper presents a low-power speech detector on a fully reconfigurable Field Programmable Analog Array (FPAA). The entire system is designed and compiled on a FPAA fabricated in 0.35μm CMOS process. The system uses 12 parallel bank of band-pass filters to extract features. The outputs of the filter bank are used by a single layer of 12x2 Vector Matrix Multiplication (VMM) and Winner Take All (WTA). The weights are stored on the VMM using pFET floating gate transistor. The power consumption of the analog system is 155.6μW with a 2.5 V power supply. The low power consumption allows the use of such a system for portable and remote sensing applications. Further, the paper investigates the performance of the system by adding white gaussian noise to the input signal. The system has an accuracy of 99.94% when the input has a SNR of 20dB and of 74% with a SNR of 8dB.**

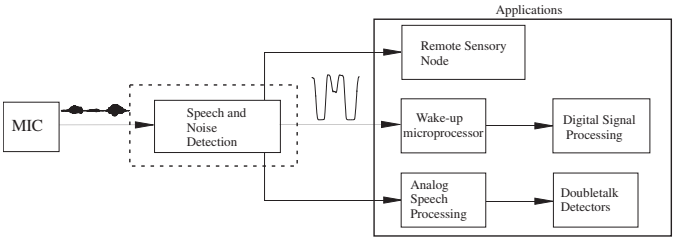

Fig. 1. The figure shows an overview of a low-power speech detector and its potential applications. The system which is implemented in this work is shown by the dashed box. The input to the system could be easily interfaced using a MEMS microphone.

I. Reconfigurable Analog Platform

Onset of internet-of-things has made the need for efficient real time signal processing on an embedded platform a necessity. Most of the digital algorithms require computation on cloud where the sensor data and output have to be transferred wirelessly. Thus, in a bandwidth constrained environment there is a trade-off in the amount of computation done locally as opposed to on a server. Analog computation has been hypothesized to have an efficiency 10,000 times more than custom digital processors [1]. With the advances in digital processors since then a custom analog processing still could achieve an efficiency three orders of magnitude more then a custom digital processors [2]. Reconfigurable and neuromorphic architectures could further increase the efficiency of the computation and reduce the cost.

Figure 1 shows a block diagram of a low-power speech detector and its potential application. The system implemented in this work is shown inside the dashed box in the Fig. 1. A MEMS microphone, similar to the one shown in [3], can directly be interfaced with the system. As shown in the Fig. 1, the system has multiple applications from being used to generate a wake-up signal for a microprocessor for digital signal processing to being used as a startup for analog command word recognition [4] and analog speech classification [5]. Acoustic echo cancellation techniques also involves detecting the near-end speech to halt adaptive filtering [6]. It could also be used as a remote sensory node for continuous speech detection [7].

The system is implemented on a very large scale fully reconfigurable Field Programmable Analog Array (FPAA). FPAAs are poised to revolutionize analog and neuromorphic systems the same way FPGAs revolutionized digital systems, by making prototyping cost effective and shortening the test cycles. A reconfigurable FPAA owing to their low power consumption and efficiency have seen an increase in their use for biomedical systems [8], speech processing [9] [5], image processing [10] and path planning application for robotics [11]. The architecture of the FPAA used in this work is described in detail here [4]. Floating Gate (FG) based FPAA, such as the one used in this work, allows for tuning and calibrating multiple parameters. This allows us to compensate for mismatch and adapt over temperature. FG transistors are used as a biasing element and also as switches as a part of the routing infrastructure. The programming algorithm [12] uses hot-electron injection for tuning the FG and FowlerNordheim tunneling for a global erase of the current design. The system is designed using an open-source Xcos/Scilab tools [13] and can be compiled as an IP block. This would allow for a faster prototyping and adaptation of the system depending on the application.

II. Speech Processing On FPAA

Speech processing involves large amount of computation and bandwidth. Thus, a low power solution for speech processing usually involves a low power front end which could detect speech over ambient and white noise and wake the processor when relevant signals are present [14]. Recent efforts in the field of wireless sensors aim at such event driven approaches [15].

In this work the low power front end is composed of 12 parallel filter banks which decimates the input speech signal into different frequency bands. The schematic of the band pass

978-1-4673-6853-7/17/$31.00 ©2017 IEEE

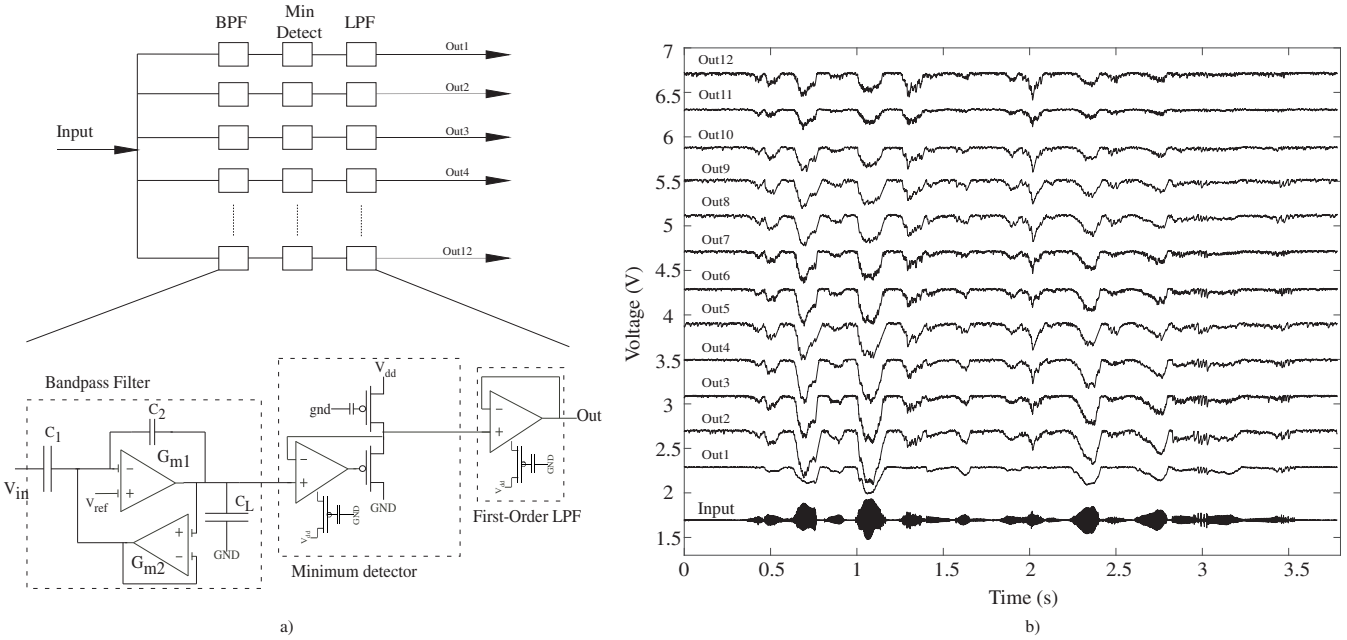

Fig. 2. a) A block diagram of the analog front end is shown in the figure along with the circuit schematic used for building these blocks. Input here is from an external DAC, but could be directly interfaced using a MEMS microphone. b) The input and output waveforms are plotted. The signals have been plotted with an offset to visualize them on the same graphs. The analog front end extracts individual features from the speech. These features could be further processed using analog signal processing.

filter is shown in the Fig. 2a. The band pass filter is a second-order G_m-C filter and uses a FG OTA. FG OTA allows us to compensate the input offset of the OTA and has increased input linearity due to capacitive coupling into the floating node. The biasing current of the FG OTA is a FG pFET which could be programmed between 30 pA to 10 μA and hence allows us to control the quality factor (Q) and the center frequency of the band pass filter. G_{m2} controls the lower frequency pole of the filter whereas G_{m1} controls the higher frequency pole. G_m is the effective transconductance of the FG OTA. In this system, the band pass filters are logarithmically spaced between 100 Hz to 5KHz with a Q of about 2. Noise and THD of such filters have been discussed elsewhere [16].

The output of the band pass filter is passed through a amplitude detect and a Low Pass Filter (LPF). The schematic of the minimum detector is shown in the Fig. 2a. The minimum detector output follows the input when it is decreasing and charges up at a rate of I_{bias}/C_L. The I_{bias} is set using a FG transistor biased with a current of 10pA. The LPF further reduces the spikes at the output. LPF is a 9-T OTA configured in a source follower configuration with its bias set at 0.9nA, to have a low F_{-3dB} frequency.

Figure 2b shows the output of these 12 filter banks with its input being a speech signal, taken from the TIMIT dataset. The input from the TIMIT dataset is "She had your dark suit in greasy wash water all year" and the outputs correspond to different features extracted by the analog front end. An analog shift register controlled by the microprocessor is used to scan the outputs. The outputs have the same DC level and have been plotted in Fig. 2b with an offset. These features are used by the algorithm to learn the weights of the VMM.

III. Training Of VMM And Detection With WTA

The extracted features are given to a one layer VMM-WTA system. A VMM on a FPAA is implemented using a pFET FG transistor which stores the weight on the floating node. The output of a single pFET transistor in the VMM is given by (1)

$$I_s = I_{th} e^{\frac{\left(\kappa_p (V_{dd} - V_{fgref} - V_{T0})\right)}{U_T}} W e^{\frac{(-(V_{dd} - V_{in}))}{U_T}} \quad (1)$$

where W is the weight obtained while training and is given by $W = e^{\frac{\left(\kappa_p (-\Delta V_{fg}^{'})\right)}{U_T}}$. Thus the weights of the VMM are adapted using $\Delta V_{fg}^{'}$, the charge on the floating node. $V_{fgref} + \Delta V_{fg}^{'}$ forms the total charge on the floating node. Other parameters in the equation are thermal voltage (U_T), I_{th} is the current at threshold voltage and κ_p is the fractional change in surface potential of the pFET with change in voltage at the gate. The VMMs are implemented in the routing infrastructure of the FPAA to increase the density of computation. Thus, the input to the VMM (V_{in}) is via the source of the pFET transistor. The power supply (V_{dd}) is 2.5 V for the current SoC.

Figure 3a shows the transistor level schematic of a 12x2 VMM. The output of the VMM is a summation of product of 12 inputs and their weights ($I_{out} = \sum W V_{in}$). The weights used for detection are trained off-chip using an algorithm similar to vector quantization. The output of the VMM is fed

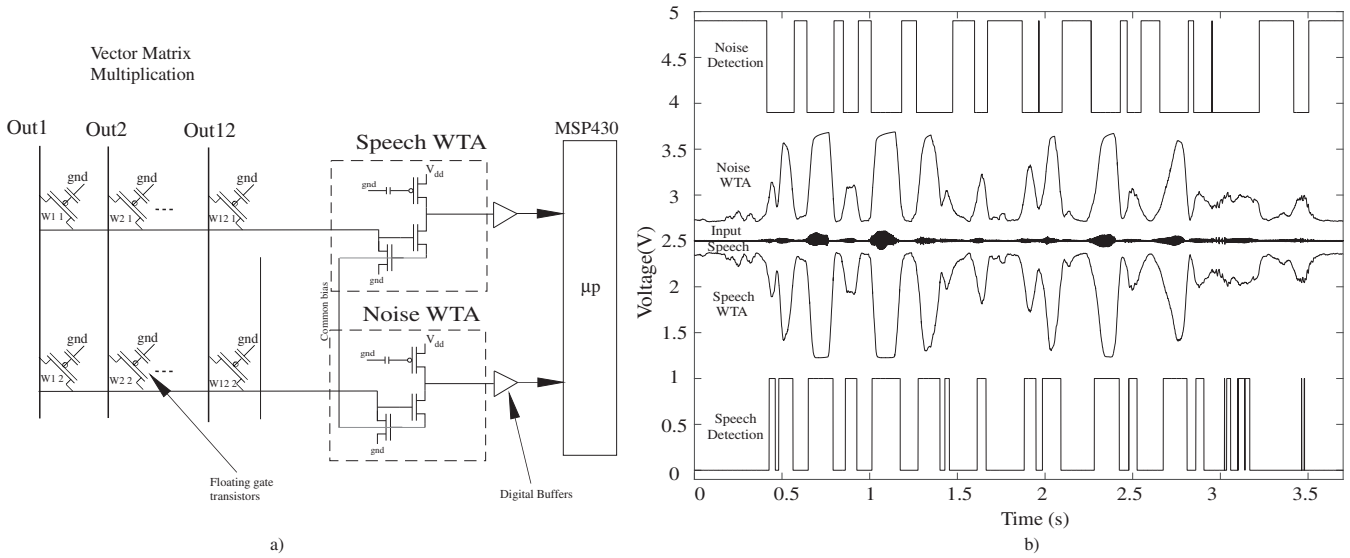

Fig. 3. a) Circuit schematic for a 12x2 VMM and WTA are shown here. The output of the WTA are routed to the microprocessor. This signal could eventually be used as an interrupt to the microprocessor for further speech analysis and classification. b) Analog output of the WTAs are plotted along with the digitized output. Here the digital signal are inverted, to signify a detection when the output is one, as opposed to the WTA output where the winner has a low output.

to the WTA. The schematic of the WTA is shown in the Fig. 3a. The biasing of the WTA is such that it allows only winner at a time. The architecture of the WTA circuit is such that the output is low when it wins and high when it loses. The output of the WTA is routed to the microprocessor using a digital buffer and thus it compares the output with 1.25 volts. The analog output could also be used as the confidence level of the classification when more than few classes are present. The discussion on confidence level of the classification and multi-class classification is beyond the scope of this paper.

Figure 3b shows the output of the WTAs and the corresponding digital outputs are also plotted. The digital outputs are inverted, with respect to the WTA outputs, so as to obtain a digital one when speech or noise is detected. The input given to the system here has a Signal-to-Noise Ratio (SNR) of 20dB, above which a human auditory system cannot perceive the difference [17]. The VMM of noise WTA has its weights set such that it has winning output in absence of speech. The speech VMM-WTA is trained to win only when relevant features are observed, i.e in presence of speech. Certain inputs result in a higher level of confidence in detection compared to others.

IV. Accuracy With SNR

One of the important metric for real time speech processing is to measure the accuracy of the system in noisy environment. To test the system under different noise condition gaussian white noise was added to the input. The signal was then given as an input to the system using an external arbitrary waveform generator. Figure 4a shows eight different SNR conditions being tested on the system. Input having different SNR are overlayed with corresponding outputs of the WTA used for speech detection.

The accuracy of the system is measured not only by checking if the WTA detects the speech but also by measuring the accuracy with which the second WTA detects the noise. Thus, the plot in Fig. 4b considers both as a factor for accuracy. Since the input data is labeled the accuracy was measured by comparing the output with an ideal output. In case of speech, the error rate was calculated if the WTA is able to detect it or not. For noise the delay in detection is also considered, so as to reduce false positives. The system performs with an accuracy of 99.94% for input having a SNR of 20 dB. The accuracy of the system stays above 70% for SNR above 7 dB. Below 7 dB of SNR the noise and signal are almost indistinguishable and hence the accuracy falls to almost 2%. The accuracy of the system in low SNR cases could be improved by having the output of the WTA compared to a different threshold or having several null clusters.

V. Discussion

The system designed in this work has a power consumption of 155.6μW, not taking into account the power consumption of the processor MSP430. The table I summarizes the power consumption for each components of the analog signal processing system. The majority of the power is consumed in biasing the G_m-C filters. These can be further optimized by considering the frequency spectrum of the input speech.

In this system the VMM-WTA were trained with an input having an SNR of 20 dB. The accuracy was tested using an input with several different inputs. The accuracy of the system needs to be evaluated for several different inputs, input with different phonemes, which is beyond the scope of this paper. The low power consumption of the system enables it to operate over several days and processes the speech signal in real time. This allows it to be deployed for various applications like echo

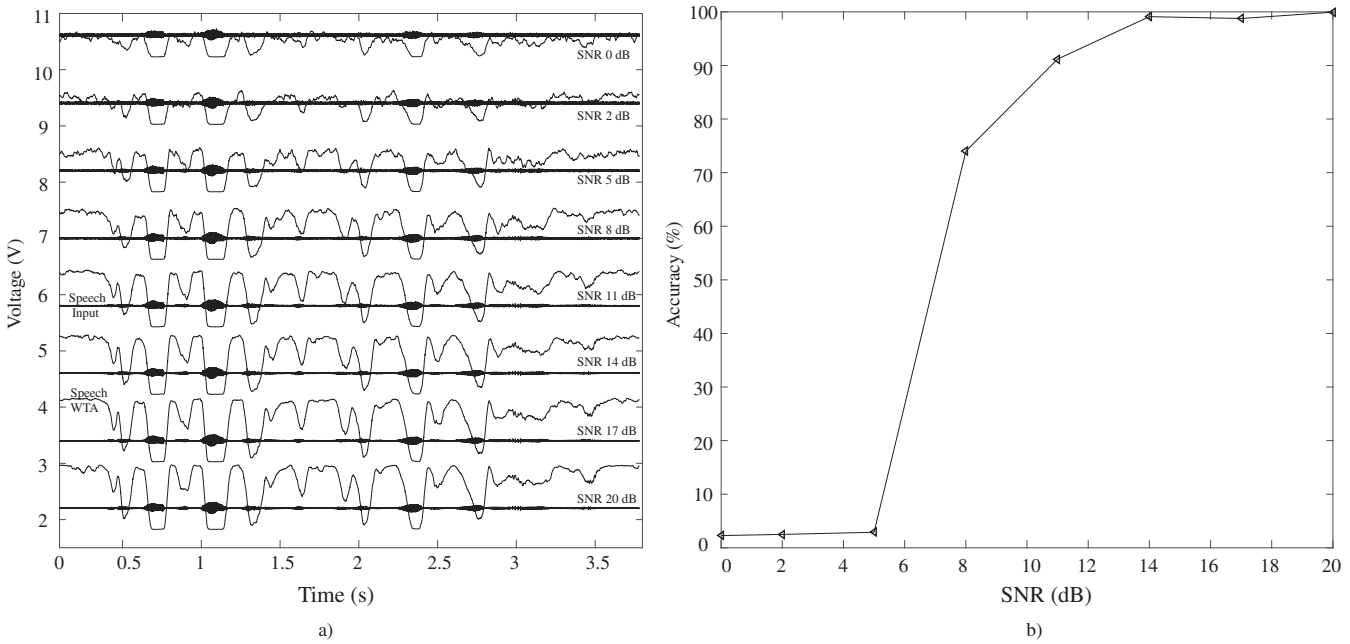

Fig. 4. a) Shows the output of the WTA detecting speech for inputs having different SNR. The signals have been plotted with an offset to visualize them on the same graph. b) Accuracy of speech and noise detection with respect to the SNR. A white gaussian noise is added to the clean signal from the TIMIT dataset.

TABLE I
POWER CONSUMPTION OF ANALOG SYSTEM

Components	Power
Band Pass filter(100 Hz - 5 KHz)	150.7μW
Minimum Detector	3μW
Low Pass Filter	27nW
VMM and WTA biasing	1.8μW
Total	155.6μW

cancellation systems as a double talk detector, remote sensing applications, and as a low power front end for digital signal processing.

REFERENCES

[1] C. Mead, "Neuromorphic electronic systems," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct 1990.

[2] J. Hasler and H. B. Marr, "Finding a roadmap to achieve large neuromorphic hardware systems," *Frontiers in Neuroscience*, vol. 7, no. 118, 2013.

[3] S. Y. Peng, M. S. Qureshi, P. E. Hasler, A. Basu, and F. L. Degertekin, "A charge-based low-power high-snr capacitive sensing interface circuit," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, no. 7, pp. 1863–1872, Aug 2008.

[4] S. George, S. Kim, S. Shah, J. Hasler, M. Collins, F. Adil, R. Wunderlich, S. Nease, and S. Ramakrishnan, "A programmable and configurable mixed-mode FPAA SoC," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 6, pp. 2253–2261, June 2016.

[5] S. Ramakrishnan, A. Basu, L. K. Chiu, J. Hasler, D. Anderson, and S. Brink, "Speech processing on a reconfigurable analog platform," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 2, pp. 430–433, Feb 2014.

[6] J. H. Cho, D. R. Morgan, and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 718–724, Nov 1999.

[7] G. Cauwenberghs, A. Andreou, J. West, M. Stanacevic, A. Celik, P. Julian, T. Teixeira, C. Diehl, and L. Riddle, "A miniature low-power intelligent sensor node for persistent acoustic surveillance," pp. 294–305, 2005.

[8] S. Shah, H. Treyin, O. T. Inan., and J. Hasler, "Reconfigurable analog classifier for knee-joint rehabilitation," *IEEE Engineering in Medicine and Biology Society*, 2016.

[9] S. Y. Peng, P. E. Hasler, and D. V. Anderson, "An analog programmable multidimensional radial basis function based classifier," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 10, pp. 2148–2158, Oct 2007.

[10] C. Schlottmann, S. Nease, S. Shapero, and P. Hasler, "A mixed-mode fpaa soc for analog-enhanced signal processing," in *Proceedings of the IEEE 2012 Custom Integrated Circuits Conference*, Sept 2012, pp. 1–4.

[11] S. Koziol, P. Hasler, and M. Stilman, "Robot path planning using field programmable analog arrays," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, May 2012, pp. 1747–1752.

[12] S. Kim, J. Hasler, and S. George, "Integrated floating-gate programming environment for system-level ICs," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. PP, no. 99, pp. 1–9, 2016.

[13] M. Collins, J. Hasler, and S. George, "An open-source tool set enabling analog-digital-software co-design," *Journal of Low Power Electronics and Applications*, vol. 6, no. 1, p. 3, 2016.

[14] T. Delbruck, T. Koch, R. Berner, and H. Hermansky, "Fully integrated 500uw speech detection wake-up circuit," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, May 2010, pp. 2015–2018.

[15] R. Olsson, R. Bogoslovov, and C. Gordon, "Event driven persistent sensing: Overcoming the energy and lifetime limitations in unattended wireless sensors," *IEEE Sensors*, 2016.

[16] D. W. Graham, P. E. Hasler, R. Chawla, and P. D. Smith, "A low-power programmable bandpass filter section for higher order filter applications," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 6, pp. 1165–1176, June 2007.

[17] P. C. M. Wong, A. K. Uppunda, T. B. Parrish, and S. Dhar, "Cortical mechanisms of speech perception in noise," *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 4, pp. 1026–1041, 2008.