Speech Processing on a Reconfigurable Analog Platform

Shubha Ramakrishnan, Arindam Basu, Leung Kin Chiu, Jennifer Hasler, David Anderson, and Stephen Brink

***Abstract*— We describe architectures for audio classification front ends on a reconfigurable analog platform. Real-time implementation of audio processing algorithms involving discrete-time signals tend to be power-intensive. We present an alternate continuous-time system implementation of a noise-suppression algorithm on our reconfigurable chip, while detailing the design considerations. We also describe a framework that enables future implementations of other speech processing algorithms, classifier front ends, and hearing aids.**

***Index Terms*— Analog signal processing (ASP), noise suppression, reconfigurable, trans-linear.**

I. Introduction

We present a reconfigurable analog chip that can be used as a front end for audio processing and signal enhancement. The possible benefits of an analog chip in terms of power dissipation per unit computation has long been hypothesized by Mead [1]. The popularity of analog signal processing (ASP), however, has remained far less than its digital counterpart owing to difficulty in design and fixed functionality. However, developments in field-programmable analog arrays (FPAAs), which are the analog equivalents of field-programmable gate arrays (FPGAs), show great promise in allowing the end user to easily utilize the power of analog processing [2]. The FPAA provides the user the flexibility to implement circuits for ASP by allowing rapid prototyping and programmability while avoiding the costs of custom analog IC fabrication. Furthermore, the floating-gate (FG) transistor, which enables programmability in the FPAA, allows mismatch compensation, on-chip parameter storage reducing the number of off-chip biases, and quick parameter tuning.

We envision a range of signal processing algorithms that fit into the pathway between speech production (source) and perception (human ear), as seen in Fig. 1. These algorithms are implemented by nonlinear processing of subbanded speech signals for applications such as noise suppression or hearing compensation by proper choice of the nonlinearity. In addition, the outputs of the nonlinear processor can be taken at each subband for speech detection instead of recombining to generate a perceptible signal for the human ear. We demonstrate a few examples of noise suppression algorithms that use nonlinear filtering in different subbands, and examine the tradeoffs in implementing them on a reconfigurable platform.

Section II provides a short review of the reconfigurable hardware that is used. We describe the speech enhancement algorithm used and present measured results from a single channel in Sections III and IV. In Section V we present the measured results from multichannel system implementations of noise suppression.

Manuscript received July 18, 2012; revised December 24, 2012; accepted December 26, 2012. Date of publication February 5, 2013; date of current version January 17, 2014.

The authors are with the Electrical and Computer Engineering Department, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: shubha@gatech.edu; arindamb@ece.gatech.edu; ken.chiu@gatech.edu; jennifer.hasler@ece.gatech.edu; dva@ece.gatech.edu; shneeve@gmail.com).

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/TVLSI.2013.2241089

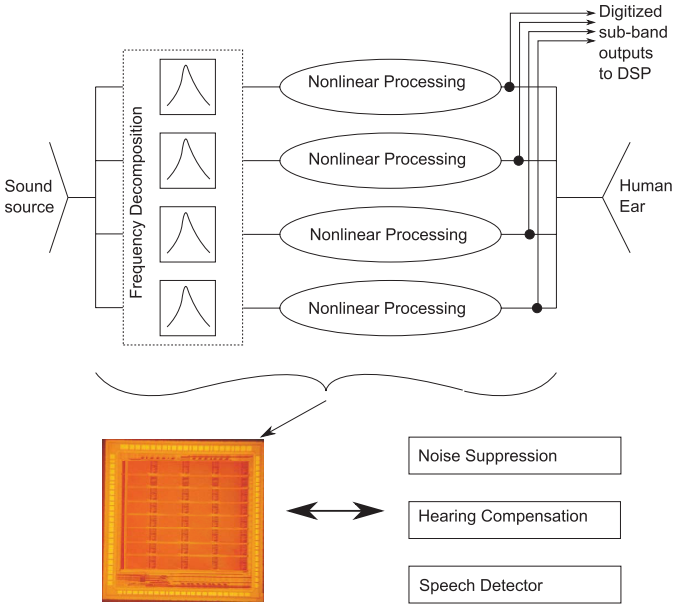

Fig. 1. High-level overview: The frequency decomposition is modeled using a bank of parallel bandpass filters. The choice of nonlinear processing in this framework can result in several applications. The signal is recombined post processing and converted back into an audio signal, or digitized before recombination, providing subbanded input for further digital signal processor. This structure is amenable to easy implementation on the FPAA (die photo pictured), making it an attractive platform for comparing various ASP algorithms.

II. Speech Enhancement

Today, most efforts in audio signal enhancement are concentrated on processing the digitized signal with Wiener filtering, spectral subtraction, and other techniques [4]. We use the algorithm described in [5], where an adaptive gain is applied to the subbanded signal based on the instantaneous signal-to-noise ratio (SNR) estimate. We choose this algorithm because it allows continuous speech enhancement in each subband and is suited for implementation using analog VLSI. A noisy audio signal $x(t)$ can be represented as

$$x(t) = s(t) + n(t) \tag{1}$$

where $s(t)$ is the signal and $n(t)$ is the noise. We assume that the noise is stationary over a longer period relative to the speech signal, resulting in a separation of timescales in $s(t)$ and $n(t)$. We estimate the noise $\hat{n}(t)$ from $x(t)$, and then modulate the gain of the signal in the following stage. When $x(t) > \hat{n}(t)$, our audio signal dominates our noise estimate. Hence, we apply a large gain to the signal, to emphasize speech portions of the signal. When $x(t) \approx \hat{n}(t)$, the audio signal is mostly noise and we reduce the signal gain, so that the noisy portions may be suppressed.

III. Single-Channel System Results

The nonlinear processing block in Fig. 1 can be implemented in a variety of ways using the blocks shown in Fig. 2 to produce processed speech for different applications. A system that expands the dynamic range of the input signal can result in noise suppression [6]. This can be achieved by applying a soft threshold to the subbanded signal envelope to determine the gain for that channel, shown in Fig. 3(a). Fig. 3(b) shows the transient results for such a system for a single tone input. Dynamic range expansion can also be achieved by ensuring a

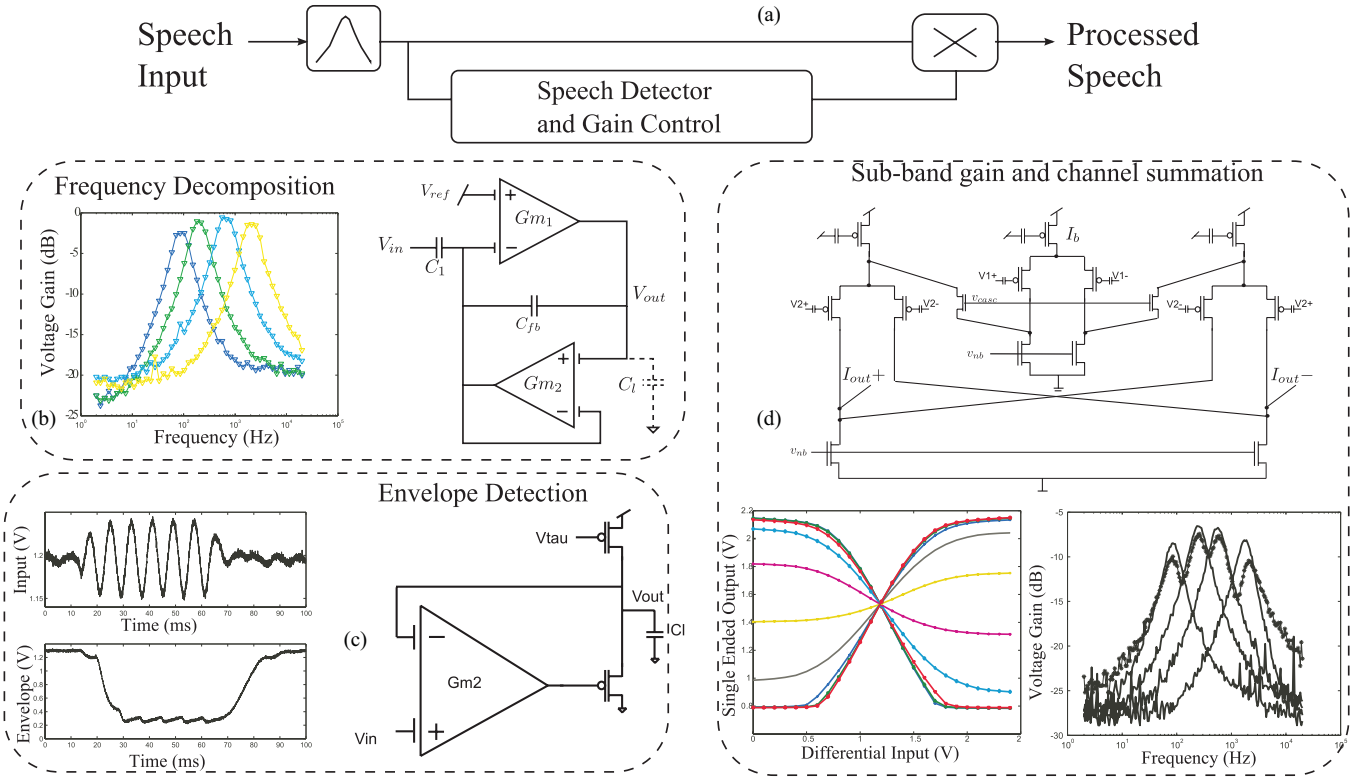

Fig. 2. System components available for signal processing. (a) Following high-level blocks are available for use in audio processing algorithms that take a subbanded time-varying gain approach. (b) Frequency decomposition is implemented using second-order band-pass filters [3] that have exponentially spaced center frequencies. The operational transconductance amplifiers and floating-gate wide input linear range operational transconductance amplifiers (FGOTAs) in our chip enable implementation of Gm-C filters with tunable frequency responses. Their bias currents are set by FG devices that can be precisely programmed. Schematic and measured results are from a bank of four bandpass filters. (c) Envelope detection. Schematic of minimum detector implemented on our chip, and the transient response of the minimum detector. The "attack" and "release" parameters for the envelope detector can be modified by changing the bias current. For the ith channel, we choose the bias current for the peak detector such that the rate of decay of its output does not cause a significant change in output voltage in a time period corresponding to the lowest frequency signal in that band. (d) Gain control and channel summation. Schematic of Gilbert multiplier used for channel amplification and summation. The output of the multiplier versus V_1 is plotted for V_2 between -2.4 and 2.4 V in steps of 600 mV. The summed response from four channels is also plotted.

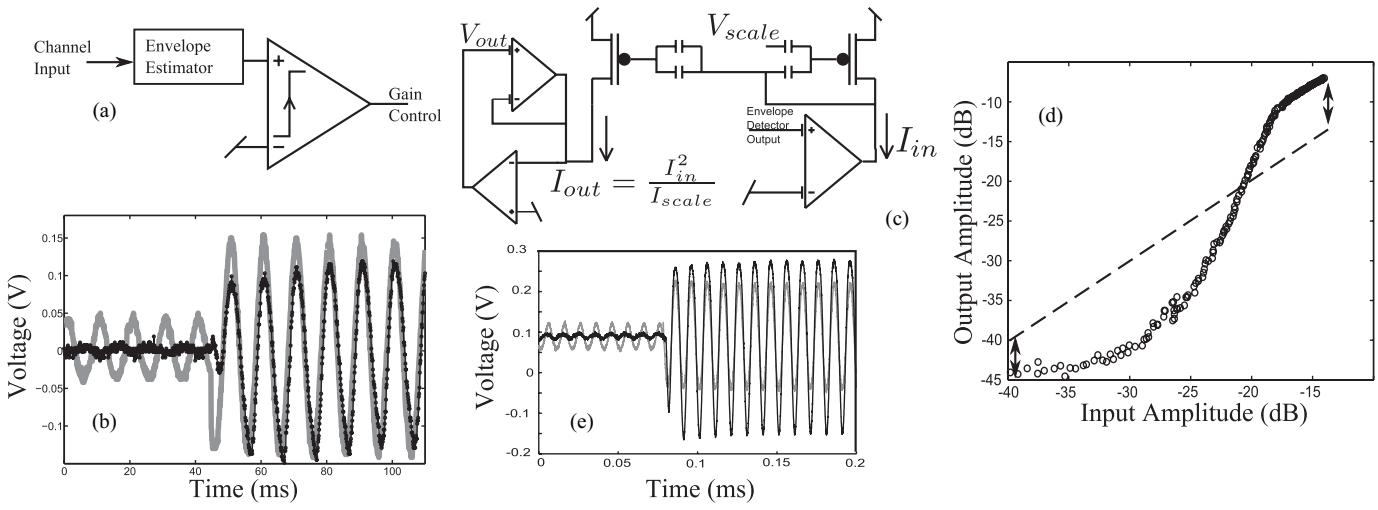

Fig. 3. Single channel results for envelope thresholding system. (a) Channel gain is determined by thresholding the envelope estimate. (b) Transient results for single tone inputs (gray trace is the input) shows attenuation of "small" amplitude inputs, while "large" inputs are unchanged. Squaring nonlinearity. (c) Schematic of squaring block implemented with multiple input trans-linear elements in the FPAA. (d) Transfer function of single-channel system implementing the squaring nonlinearity, resulting in an expansion of the dynamic range by 12 dB. (e) Transient response of system for a single tone input (gray trace). Large input amplitudes are amplified while smaller inputs are attenuated.

power law relationship with an exponent >1 between the input and output. The nonlinear function can be implemented in current mode employing the trans-linear principle using multiple-input floating gate (MIFG) transistors [7]. The voltage output from the envelope detector is converted to current using an FGOTA block which has a linear range of about 600 mV. Fig. 3(c) depicts a circuit with a squaring

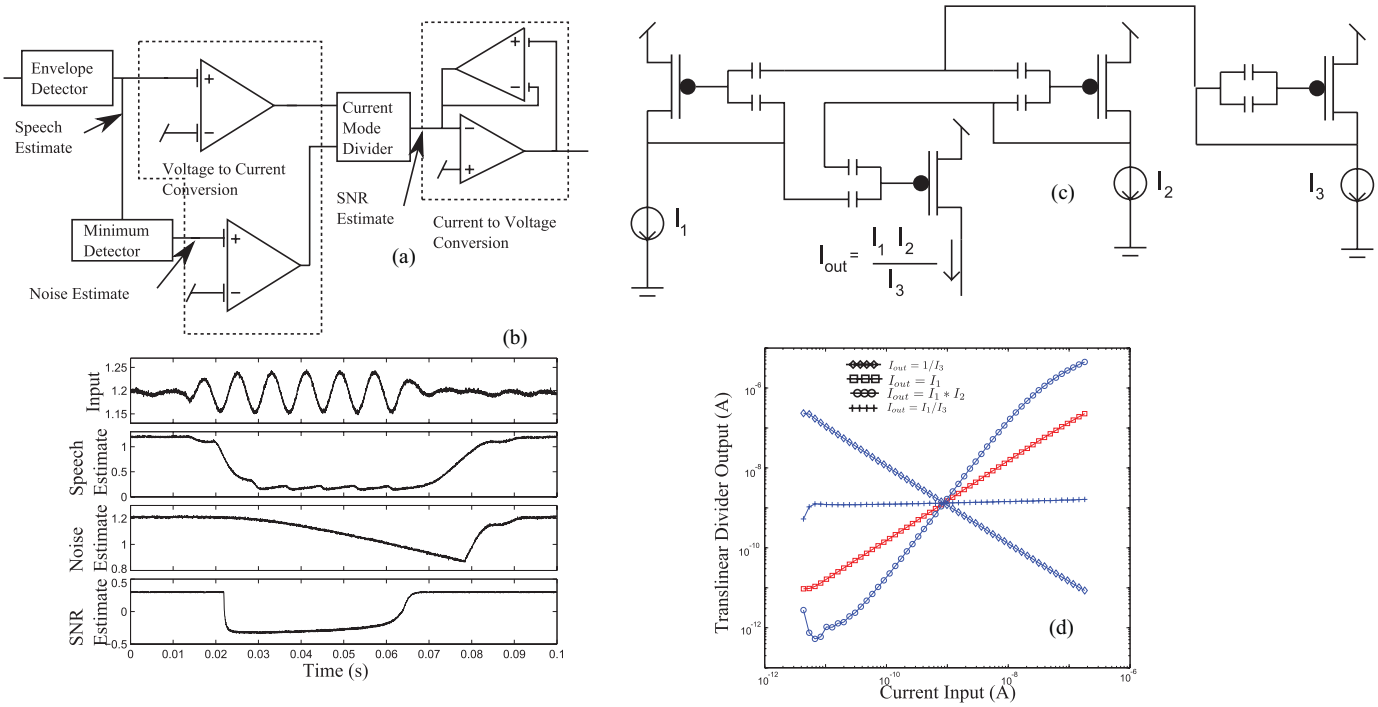

Fig. 4. SNR estimation. (a), (b) SNR estimation block. Block diagram and measured output from the SNR estimation circuit. The SNR estimate is computed using the speech estimate and the noise estimate. The current-mode divider requires conversion of the estimates from voltage to current domain. The divider output is converted back into voltage using a trans-impedance amplifier. Since the current-to-voltage converter provides an inversion, the divider voltage output is low for high SNR estimates and high for low SNR. (c) Implementation of the current mode multiplier/divider circuit on the FPAA using MIFG transistor elements. (d) Characterization of the trans-linear multiplier/divider circuit with the relation $I_{\text{out}} = I_1 * I_2/I_3$. The slope when I_1 and I_2 are swept together is double that of when I_1 is swept alone. When I_3 is swept with I_1 and I_2 kept constant, the slope is -1.

nonlinearity. The output of the squaring circuit is converted back to voltage using a trans-impedance amplifier. Fig. 3 (d) and (e) plots the transfer function and the transient response of the system for a single-tone input. While the circuits described previously may be suitable for enhancement of speech signals with moderately high SNR, they do not perform well for inputs with SNR $\leq$10 dB and an SNR estimation approach yields better results. To do real-time estimation of the noise floor, we detect the minimum of the noisy subband envelope [8]. The integration time constant of the minimum detector is chosen to be slower than the envelope detector for that channel, but large enough that the speech signal in that band is too fast to cause a significant change in the noise estimate. The voltage output from the envelope estimator and the noise estimator are then converted into currents, as shown in Fig. 4(a). The MIFG transistors can be configured to do current mode multiplication and division using the circuit shown in Fig. 4(c) [7]. The output of the divider circuit is plotted in Fig. 4(d). The relevant signals in the SNR estimator circuit are shown in Fig. 4(b).

IV. Multichannel System Results and Discussion

A four-channel noise suppression system with an SNR estimation approach was implemented on the Rasp 2.8a chip. The maximum number of channels that can be implemented on this chip is 8, limited by the number of Gilbert multiplier blocks. For the multiband system, the time constants for the envelope detector and noise estimator blocks were set independently using on-board digital-to-analog converters. The envelope detector for the lowest frequency channel is tuned to have the highest time constant. We used speech samples from the NOIZEUS database, which provides acoustically recombined noisy speech samples with fixed SNR [9]. Noisy speech with 0 and 5 dB SNR was played, and the output from our system

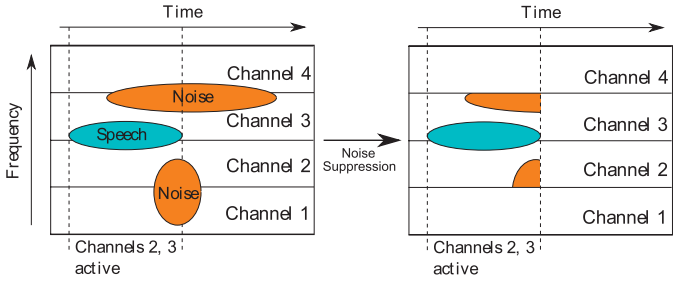

Fig. 5. Effect of limited channels on performance: The gain control system is active when speech activity is detected, transmitting the entire channel for that duration. Any noise present in the active channel in that duration is also transmitted to the output.

was recorded. Listening tests on the processed speech revealed that the system effectively reduced noise in between speech portions, but failed to do so during the speech [10]. This behavior is expected and is illustrated by Fig. 5. Channels 2 and 3 are active when speech is detected and the SNR is high enough. However, this allows the noise present in these channels to also leak through to the output. Increasing the number of bands will increase the ability of the system to resolve speech and noise into multiple bands, thereby allowing it to suppress the noise better. The spectrogram of the processed speech for selected inputs is plotted in Fig. 6, and shows that our system suppresses noise considerably for moderately low input SNR.

A. Power Consumption

Increasing the number of bands allows separation of speech and noise activity, resulting in better noise suppression, but also increases the power consumption. The power consumed in individual blocks

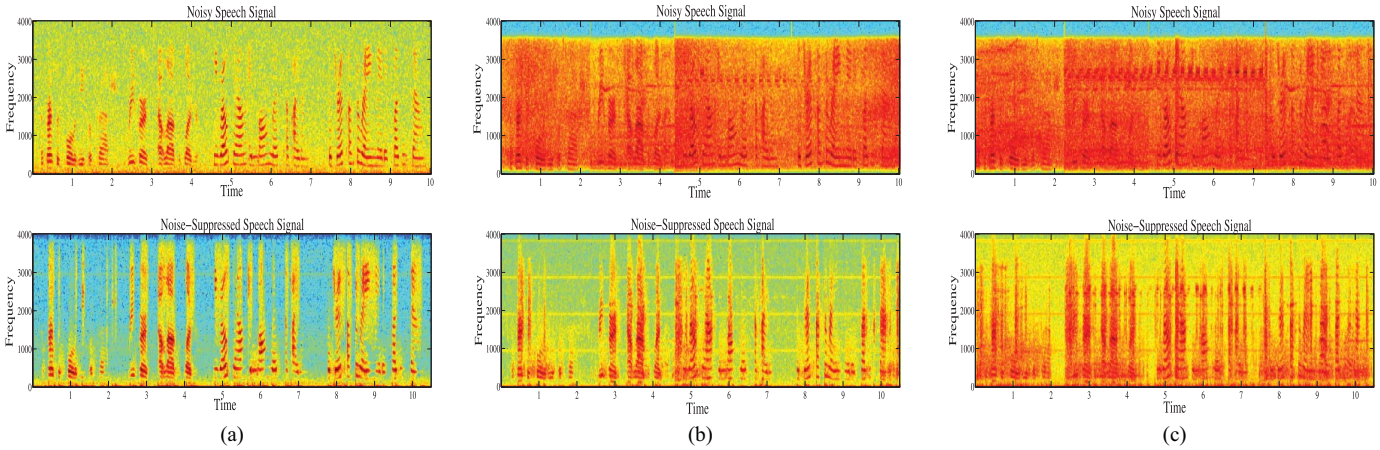

Fig. 6. Spectrogram of noisy and processed speech. Comparison of noisy and processed speech for (a) speech with added pink noise at 13-dB SNR, (b) with street noise at 5-dB SNR, and (c) with added pink noise, at 0-dB SNR.

TABLE I
POWER CONSUMPTION OF INDIVIDUAL BLOCKS FOR A FOUR-CHANNEL SYSTEM

Functional Block	Power Consumption
Bandpass filter	0.4 μW
Noise estimator	3.84 μW
Speech estimator	4.03 μW
Multiplier	5.52 μW
Divider and gain control	27.84 μW
Buffers	86.4 μW
Total	128.03 μW

is listed in Table I. The total power consumption for the four-channel system, without including the FG programming circuitry and the amplifiers to drive the audio ports, is 128.03 μW. A significant portion of this power is consumed in the buffers that are necessary for driving sensitive analog signals over routing lines with large parasitic capacitance. The projected power consumption for a 32-channel system is 1.02 mW. The power consumption of all blocks except the bandpass filter can be linearly scaled. The filter power consumption is dependent on its center frequency, but since it is a small fraction of the total power, we assume it to be constant.

V. CONCLUSION

In this brief, we described how the FPAA can be used for implementing ASP algorithms. The framework developed in this brief also supports other applications such as voice activity detection, hearing compensation, and classifier front ends. While we demonstrated using expansive nonlinearities for noise suppression, a compressive nonlinearity can be used in hearing-aid applications. One approach to noise suppression, discussed in this brief was previously published in [11]. However, we provided measured data for the first time from an integrated system compiled on a single reconfigurable chip. The goal of this brief was not to develop the best noise suppression system, but to make ASP accessible to a wide audience.

REFERENCES

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

[2] A. Basu, S. Brink, C. Schlottmann, S. Ramakrishnan, C. Petre, S. Koziol, F. Baskaya, C. M. Twigg, and P. Hasler, "A floating-gate-based field-programmable analog array," *IEEE J. Solid-State Circuits*, vol. 45, no. 9, pp. 1781–1794, Sep. 2010.

[3] 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 Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 6, pp. 1165–1176, Jun. 2007.

[4] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.

[5] D. Anderson, P. Hasler, R. Ellis, H. Yoo, D. Graham, and M. Hans, "A low-power system for audio noise suppression: A cooperative analog-digital signal processing approach," in *Proc. IEEE 10th Digit. Signal Process. Workshop, 2nd Signal Process. Educ. Workshop*, Oct. 2002, pp. 327–332.

[6] D. N. Parikh, S. Ravindran, and D. V. Anderson, "Gain adaptation based on signal-to-noise ratio for noise suppression," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2009, pp. 185–188.

[7] B. A. Minch, "Synthesis of static and dynamic multiple-input translinear element networks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 51, no. 2, pp. 409–421, Feb. 2004.

[8] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.

[9] Y. Hu and P. Loizou, "Subjective evaluation and comparison of speech enhancement algorithms," *Speech Commun.*, vol. 49, nos. 7–8, pp. 588–601, Jul. 2007.

[10] S. Ramakrishnan, A. Basu, L.-K. Chiu, P. Hasler, D. Anderson, and S. Brink, "Speech processing on a reconfigurable analog platform," in *Proc. IEEE Subthreshold Microelectron. Conf.*, Oct. 2012, pp. 1–3.

[11] R. Ellis, H. Yoo, D. W. Graham, P. Hasler, and D. V. Anderson, "A continuous-time speech enhancement front-end for microphone inputs," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 2. May 2002, pp. 728–731.