Mixed-Mode Correlator for Micropower Acoustic Transient Classification

R. Timothy Edwards, *Member, IEEE*, and Gert Cauwenberghs, *Member, IEEE*

***Abstract*— A mixed-mode very large scale integration (VLSI) processor for acoustic transient classification performs a running correlation between a time-frequency decomposed analog input signal and a corresponding template. Differential encoding of the inputs allows simplification of the multiply-and-accumulate operations, operating on binary templates and positive-valued inputs, implemented in current mode with eight MOS transistors per cell including SRAM template storage. The use of a bucket-brigade device (BBD) instead of a charge-coupled device (CCD) for shift-and-accumulate operations allows long integration times necessary for acoustic transients spanning several hundred milliseconds. Correlated double sampling at the output compensates for dark current losses in the BBD and other common-mode effects. The 64 × 16 time-frequency correlator measures 700 μm × 1170 μm in a 1.2-μm CMOS process and dissipates an average of 30 μW in continuous operation.**

***Index Terms*— Acoustic transients, current-mode circuits, mixed-mode circuits, pattern classification, template correlation.**

I. Introduction

TEMPLATE correlation is an essential, yet computationally demanding operation in pattern recognition, signal processing, and communications. Mixed-mode very large scale integration (VLSI) implementations offer high-speed, low-power operation and high density of integration [1]. Several variants on correlation algorithms are embedded in mixed-mode application-specific processors, e.g., for focal-plane image compression using discrete cosine transform coding [2] and for spread-spectrum demodulation for wireless communications [3].

Parallel, distributed, analog correlation-based processing is also the basis of VLSI systems which emulate the function and structure of neural information processing in biological systems. They excel at sensory information processing for applications where low-power operation is especially important [4]. Examples of such systems put onto silicon include vision processors for motion detection [5], stereopsis [6], and character recognition [7], and a range of adaptive matrix processors [8], [9].

Although less publicized, correlation algorithms for acoustic coding and speech recognition have found embedded applications in mixed-mode VLSI such as a chip for auditory localization [10], a hybrid VLSI address-event coded speech processing system [11], and an autocorrelation chip for periodogram analysis [12]. A correlation-based architecture for isolated spoken digit recognition is presented in [13]. While off-the-shelf digital signal procesing (DSP) solutions prevail as the method of choice in the majority of applications, there are applications where extreme low-power operation offered by dedicated VLSI solutions is a necessity, such as hearing aids and implantable biomedical devices [14].

The processor described here is programmed to recognize a particular acoustic transient by performing time-frequency domain template correlation. Relatively short transients (100 ms) can be detected with simple correlation without the need for dynamic time alignment or hidden Markov state decoding [16]. The processor can serve as an acoustic signature detector, or can detect the onset of a longer acoustic event to trigger activation of a more sophisticated (DSP) processor, idle in low-power standby mode. Arrays of template processors can be integrated on a larger die as a frontend to a speech recognition system.

II. Template Correlation

The algorithm for acoustic transient classification on which the correlator is based is described in [17] and [18], demonstrating effective classification on a test set containing prerecorded examples of ten different transient sounds [19]. In its baseline form, the template correlation can be written

$$c_z[t] = \sum_{m=1}^{M} \sum_{n=1}^{N} x[t-n, m]p_z[n, m] \tag{1}$$

where (for acoustic tasks) M is the number of frequency channels of the input, N is the maximum number of time bins in the window, x is the array of time-sampled input signals split into frequency bands, p_z is the matrix of template pattern values for pattern z, and t is the current time. This describes a running correlation $c_z[t]$ of the input array with the template z. A signal may be classified as belonging to class z when the output $c_z[t]$ exceeds a threshold, or by evaluating some function of the vector of outputs over all classes. The topic of how to determine the optimal template for each class is beyond the scope of this paper, but examples may be found in [13] and [17].

The input vector x to the correlator is the output of a frontend system producing a measure of short-term energy across a bank of M bandpass filters spanning the audio range. Micropower systems such as cochlear models [22] are

Manuscript received October 27, 1998; revised May 10, 1999. This work was supported by NSF MIP-9702346 and ONR N00014-99-1-0654.

R. T. Edwards was with the Electrical and Computer Engineering Department, Johns Hopkins University, Baltimore, MD 21218 USA. He is now with the Applied Physics Laboratory, Johns Hopkins University, Laurel, MD 20723 USA (e-mail: tim@bach.ece.jhu.edu).

G. Cauwenberghs is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: gert@bach.ece.jhu.edu).

Publisher Item Identifier S 0018-9200(99)08250-5.

appropriate to the task, although a current-domain approach using log-domain filters [20], [21] is better suited to the current-mode correlator described here. The frontend system performs spectral decomposition of the input, followed by rectification and smoothing to obtain an energy envelope for each band. The final stage of the frontend system involves normalization across all bands using the L-1 normalization function

$$x[t, m] = \frac{y[t, m]}{\theta + \sum_{k=1}^{M} y[t, k]} \tag{2}$$

where θ is a threshold (auxiliary constant input) introduced to suppress noise at the output during quiet intervals in the input. Spectral decomposition, rectification, smoothing, and normalization are all performed on a separate chip in the present implementation. The output of one filterbank system can be used as the input to several template processors.

The hardware complexity of the correlator is determined by the multiply-accumulate function performed at each template value. We have established that the correlation can be reduced to multiplication of an analog input with a binary template with no major degradation in classification performance [17]. The key to the simplification is a differential encoding of the analog inputs, much like single-bit delta modulation encoding of analog signals. In particular, an input in differential form

$$\Delta x[n, m] = x^{+}[n, m] - x^{-}[n, m] \tag{3}$$

and the corresponding template values also in differential form but additionally thresholded

$$q_z[n, m] = \text{sign}(\Delta p_z[n, m]) \tag{4}$$

$$= \text{sign}\big(p^{+}[n, m] - p^{-}[n, m]\big) \tag{5}$$

are correlated in analog-binary form

$$c_z'[t] = \sum_{m=1}^{M} \sum_{n=1}^{N} \Delta x[t - n, m] q_z[n, m]. \tag{6}$$

Although $q_z[n, m]$ may be bipolar (± 1), the use of L-1 normalization (2) allows $q_z[n, m]$ to be binary $(0, 1)$ format without affecting classification performance. Representing templates as binary values stored locally in RAM greatly reduces the circuit complexity. In addition, simple transformation of variables allows the differential encoding at the input to be transferred to a differencing operation at the output. Equation (6) expands to

$$c_z'[t] = c_z^{+}[t] - c_z^{-}[t] \tag{7}$$

where

$$c_z^{+}[t] = \sum_{m=1}^{M} \sum_{n=1}^{N} x^{+}[t - n, m] q_z[n, m] \tag{8}$$

$$c_z^{-}[t] = \sum_{m=1}^{M} \sum_{n=1}^{N} x^{-}[t - n, m] q_z[n, m]. \tag{9}$$

The advantage of performing differencing at the correlation output c_z, and not at the level of the inputs x or correlation products, is additional simplification of the hardware, allowing analog inputs to be strictly nonnegative and eliminating differencing circuitry from the array. The combined simplifications result in a design in which each correlator cell performs a multiply-and-accumulate operation on a positive-valued analog input and a binary template, with straightforward implementation in analog VLSI.

In [17], the differential encoding was implemented with correlation performed on the time derivative of the input pattern, transforming time-frequency domain features to accentuate transients. That is

$$x^{+}[n, m] = x[n, m] \tag{10}$$

$$x^{-}[n, m] = x[n - 1, m] \tag{11}$$

with corresponding template values

$$p^{+}[n, m] = p[n, m] \tag{12}$$

$$p^{-}[n, m] = p[n - 1, m]. \tag{13}$$

This leads to an especially efficient implementation since

$$c_z^{-}[t] = c_z^{+}[t - 1]. \tag{14}$$

An alternate correlation results from computing pairwise channel differences rather than time differences, accentuating spectral features of the input

$$x^{+}[n, m] = x[n, m] \tag{15}$$

$$x^{-}[n, m] = x[n, m - 1]. \tag{16}$$

Combinations of time and frequency differential encoding are also possible. Effectiveness of the encoding depends on the nature of the input and is determined through simulations of a classification task. Hardware architectures for these alternate encodings are very similar to the time-differential architecture presented here.

Fig. 1 shows the system as described by (3)–(14), where each template value is a single bit controlling a switch (multiplexer) which adds either zero or the unidirectional current input to the column sum. The summed current is accumulated over time, with the difference between successive samples taken at the output.

III. Mixed-Signal VLSI Implementation

All correlation multiply-accumulate operations are performed in current mode in parallel across the array of template cells. In the correlator array shown in Fig. 2, each template cell is implemented with an SRAM storage element which contributes to the column sum by switching the mirrored channel m input onto a shared wire. Each array position in the template is individually addressable to allow simple chip-in-the-loop learning under computer control. The currents I_{in} represent the channel inputs x, presented either as a mirrored input current or directly by the gate voltage, selected by $\overline{I}/V$ *select*. Multiple templates are used in a master–slave configuration in which one template correlator takes current inputs from the frontend filterbank and generates voltages to be used as inputs by the remaining template correlators in the system. Fanout is large due to the small currents and long time periods involved.

A bucket brigade device (BBD) [14], shown at the bottom of Fig. 2, performs the time shift and accumulate operations.

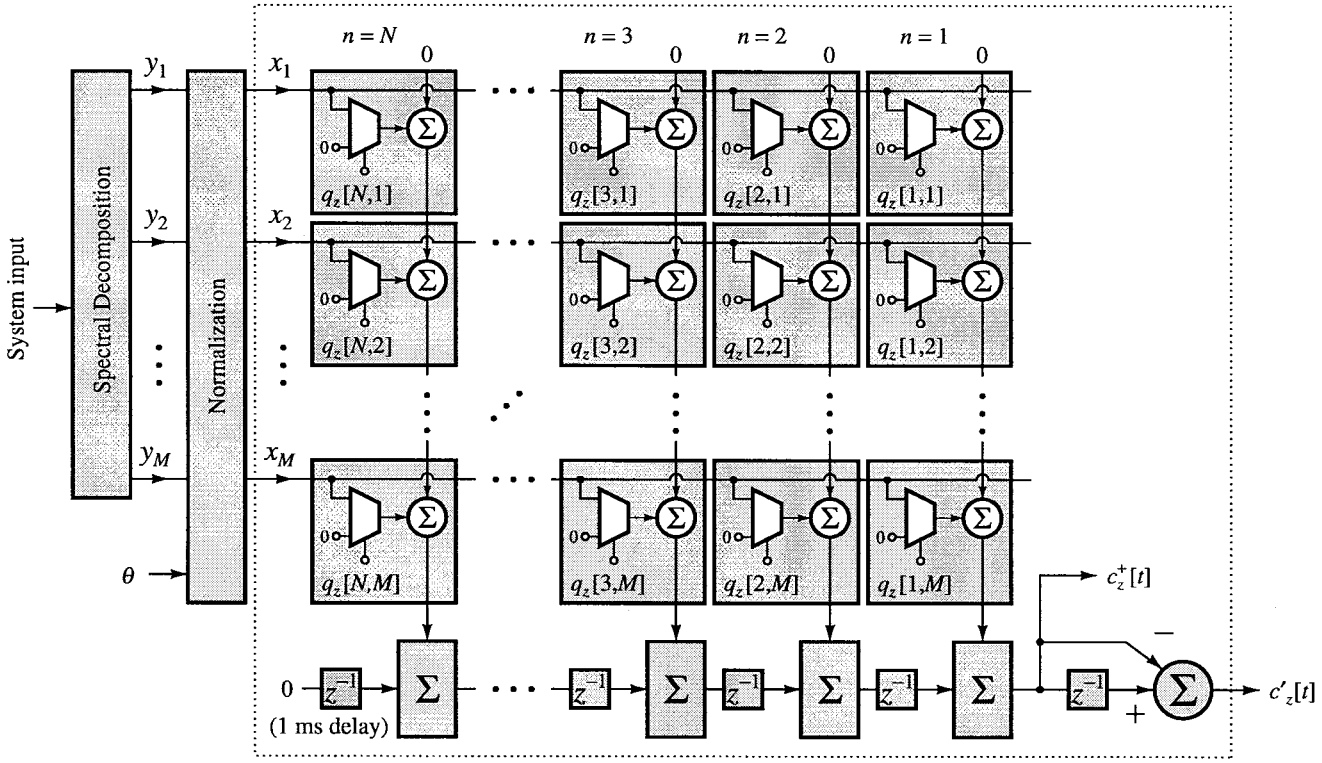


Fig. 1. Block diagram of the temporal current correlator.

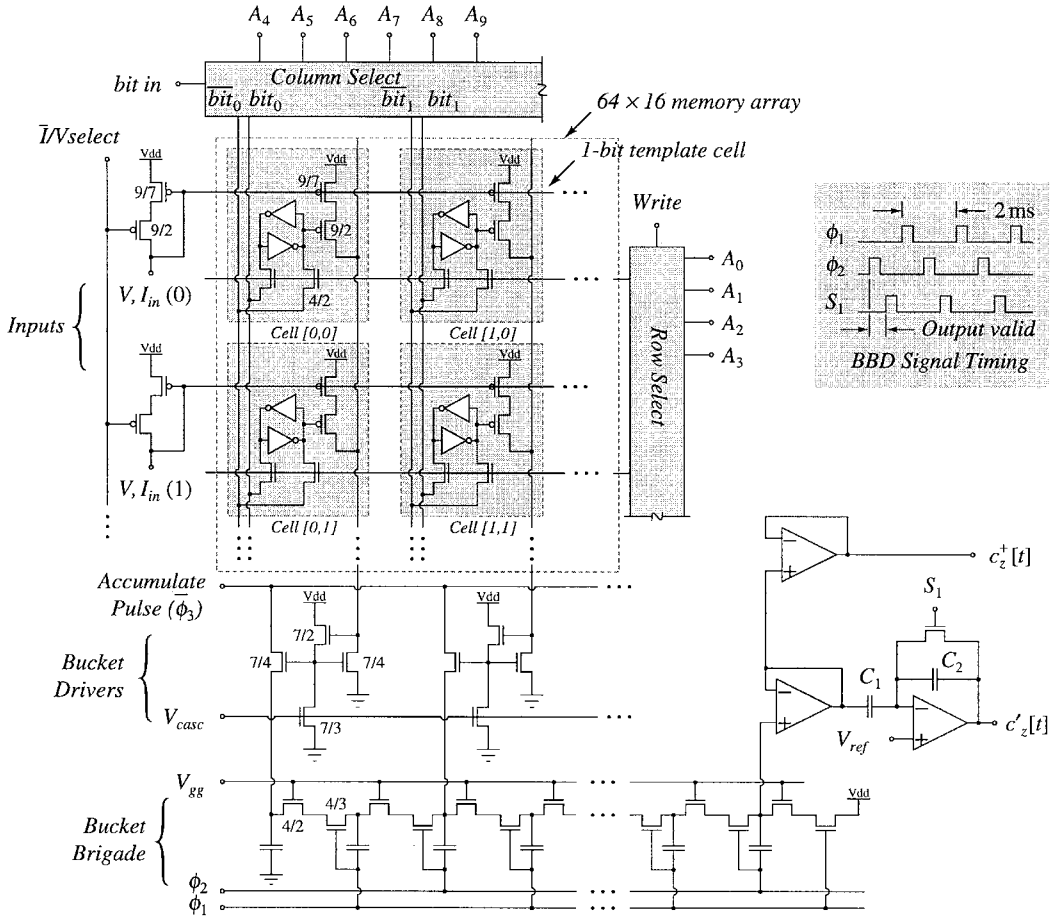


Fig. 2. The correlator array, input, and programming circuitry. Representative MOS device sizes are given as W/L in units of $\lambda = 0.6\ \mu$m.

TABLE I
CORRELATOR CHARACTERISTICS

Frequency channels	16
Time bins	64
Layout size	700 × 1170 μm
Technology	1.2 μm CMOS, double metal, double poly
Power supply	5 V
Max. input current	1 μA (typical)
Avg. power dissipation	30 μW

The BBD uses double-poly capacitors and thus does not suffer the significant loss of charge due to dark currents seen in charge-coupled devices (CCD's) over the relatively long integration times of up to several milliseconds needed for acoustic transients.

In our bucket brigade accumulator, values are accumulated by integrating the summed current from each column of the correlator array off of the bucket brigade capacitor. The integration occurs over a short, constant time determined by the pulse ϕ_3 (Fig. 2). Since the input currents vary slowly with respect to the pulse width of ϕ_3, the integration of current closely approximates a current-to-voltage conversion. The current-conveyor driver circuits ensure a quick response to the ϕ_3 pulse even for very low-level input currents.

The voltage values in the BBD are shifted with a two-phase nonoverlapping clock (ϕ_1, ϕ_2). A constant voltage bias V_{gg} improves the charge transfer efficiency (CTE) of the BBD device [15]. The bucket brigade is fully pipelined, yielding one full correlation at every time step. The switched-capacitor circuit shown in Fig. 2 performs time differencing of the output of the bucket brigade. The output is measured relative to the voltage $V_{\rm ref}$, and scaled by the ratio C_1/C_2. The system output is valid between the ϕ_2 and *Reset* clocks.

We designed and fabricated a chip containing a single template correlator in a 2.2-mm die, the characteristics of which are given in Table I. Fig. 3 is a photograph of the chip.

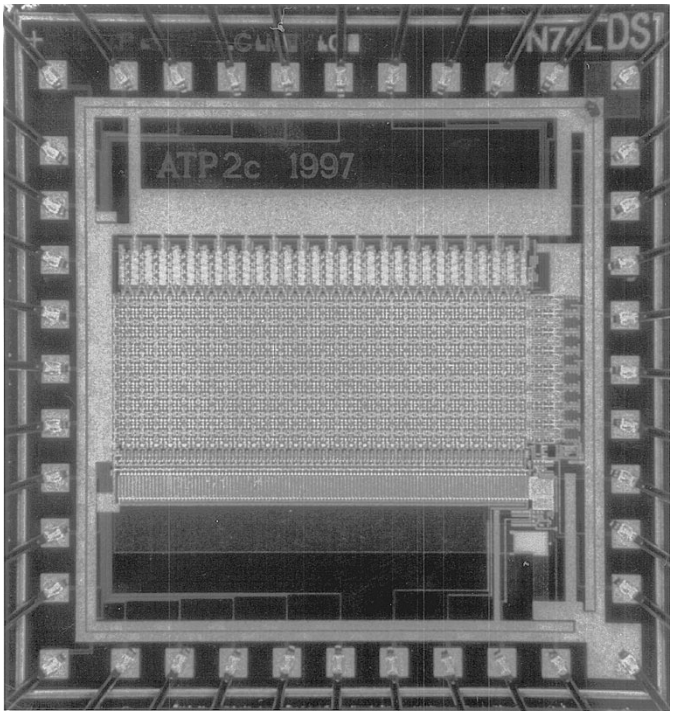


Fig. 3. Photograph of the correlator chip.

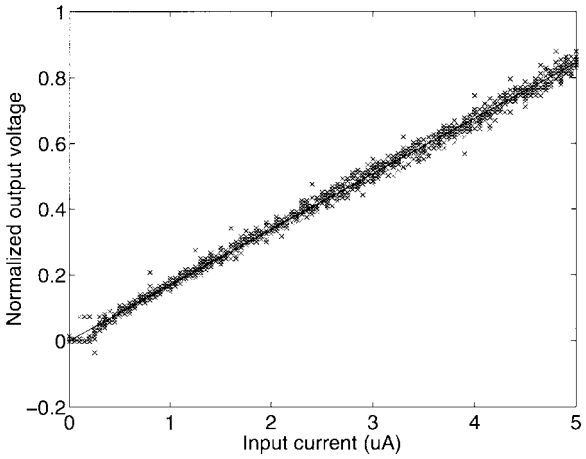


Fig. 4. Input current and correlation output transfer characteristics for ten channels along one column.

IV. SYSTEM RESPONSE CHARACTERIZATION

To characterize performance of the transient acoustic classifier, we first measured the response characteristics of the correlator accounting for nonidealities in the correlator array and the bucket brigade. The results are represented in Figs. 4–6.

The current-to-voltage conversion on the bucket brigade through a single element in the array is shown in Fig. 4. The slight nonlinearity in the response at the output is due partly to residual current switching transients in the current conveyor during the ϕ_3 pulse.

Fig. 5 records the impulse response of one channel of cells in the correlator, illustrating the charge transfer efficiency (CTE) of the bucket brigade device. A fixed current input of 1 μA is integrated onto one of the 64 taps in the BBD over a fixed duration (width of the ϕ_3 pulse) of 3.4 μs, then relayed to the output. We repeat the measurement for each of 64 BBD taps. The randomly distributed variation in outputs (14.5%) is typical for device mismatch in the MOS transistors which mirror the input current onto the bucket brigade capacitor (see Fig. 2). The measured CTE of 99.7% is normal for a BBD although less than what can be achieved with a buried-channel CCD. Charge transfer inefficiency affects all inputs and all templates in the same manner; the principle effect is to increase the relative weight of acoustic events temporally far away from the transient onset, which are more susceptible to temporal skew. The distributed nature of the correlation ameliorates this effect. In simulation, classification accuracy degrades almost linearly with decreasing charge transfer efficiency, and is indistinguishable from the ideal case (91.44% correct classification on our transient test set) for the measured CTE.

The peak impulse response over all channels yields an indirect measurement of transistor current mismatch across the template array, shown in Fig. 6 for an input amplitude of 3 μA.

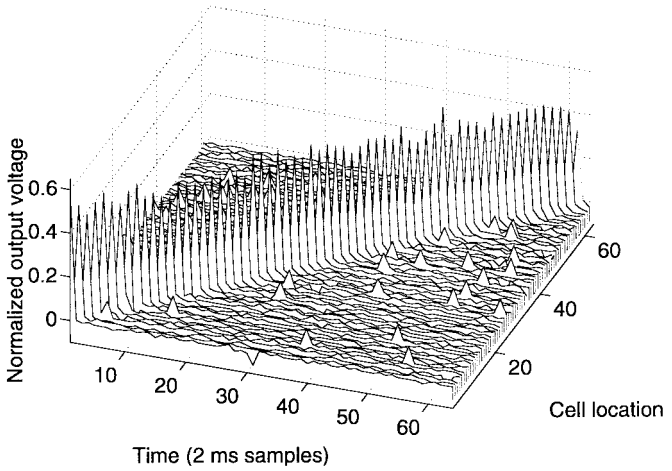


Fig. 5. Impulse response, in time and space, of cells selected along one channel of the correlator.

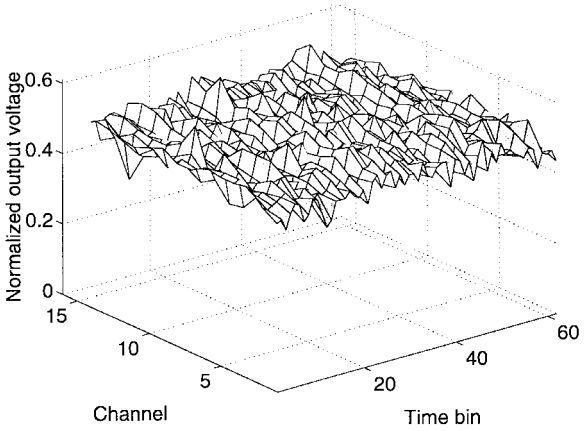
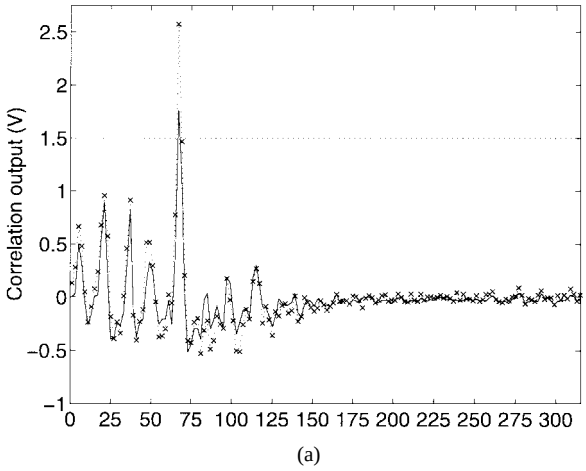
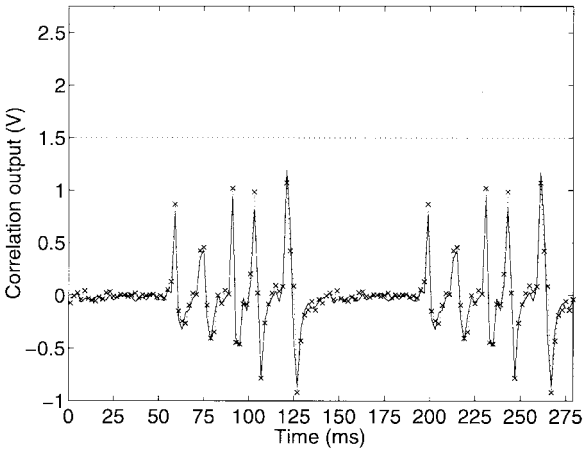


Fig. 6. Matching between cells across the template array at 3-μA input.



(a)



(b)

Fig. 7. Measured ($\times$) versus simulated (solid line) correlation of repeated sounds (a) "can" and (b) "snap" with the "can" template loaded into chip memory. The dotted line is the threshold for classification.

Systematic offsets exist in both channel number and time bin, following typical spatial patterns of mismatch as reported in [23]. While mismatch varies significantly with current level, by nature the correlation computation is a distributed process in which individual element mismatch is relatively unimportant. Furthermore, adaptive schemes for template formation are capable of removing residual effects of component mismatch.

V. Classification Results

We tested the transient correlator system using an experimental setup which allows us to evaluate the correlator independently of the frontend system. We have also designed and fabricated an analog current-mode frontend filterbank chip [20], [21] which interfaces directly to multiple correlators, that is able to take analog acoustic data directly from a microphone or other source.

Among the tests we performed was to take a template for a prerecorded acoustic transient from our simulation program, download it to the chip, then compare its output to that of the simulation program, where both are given the same input. We used two different transients for the input, one corresponding to the template class and the other a different class example. Fig. 7 shows both measured and simulated output of the correlation between the template for "can" and a "can" sound and a "finger snap" sound as input. Residual error is systematic and can be traced to the nonlinearity in the current-to-voltage conversion (Fig. 4).

Total power dissipation of the acoustic transient processor depends primarily on the static dissipation of the amplifiers in the switched-capacitor circuit and the bias current in the current conveyors driving the bucket brigade inputs. At minimal values of these biases allowing correct system operation ($V_{\text{casc}} = 0.4$ V and amplifier bias $= 0.6$ V), average power dissipation is 30 μW with a peak dissipation of approximately 50 μW during the onset of a transient input.

VI. Conclusions

We have designed and fabricated a chip in standard CMOS technology intended for use as a classifier of transient (short-term) acoustic signals using template correlation. The chip uses analog current-mode circuits to produce the running correlation between an input signal, spectrally decomposed by a filterbank frontend into an array of normalized energy envelopes, and a stored template of binary values. The input is differentially encoded for robust classification using binary-valued templates. The use of binary template values reduces

the correlation computation to one-quadrant multiplication, effectively implemented by current switching. The core cell of only eight transistors contains the SRAM template storage and performs the multiply-and-accumulate operation. A BBD performs time shift-and-accumulate operations over cycle times of several milliseconds.

Test results show that the chip output voltage closely follows predicted values from simulations. The chip can be used to classify acoustic transient events quickly, accurately, and with a high degree of robustness due to a parallel, pipelined architecture. The power consumption of 30 μW per template is significantly lower than that for a DSP or microprocessor performing the same computation, and the layout area of 700 μm $\times$ 1170 μm per template allows a high integration density. The efficiency is obtained by combining an analog input with mixed-mode processing and integrated digital template storage.

Applications for the acoustic transient processor include acoustic signature detection, frontend processing to speech recognition systems, and a low-power continuously operating monitor used to trigger a DSP system on standby.

ACKNOWLEDGMENT

The authors would like to thank F. J. Pineda of the Johns Hopkins University Applied Physics Laboratory for his original research, collaboration, and helpful discussions.

REFERENCES

[1] S. Munroe, D. Arsenault, K. Thompson, and A. Lattes, "Programmable 4-channel, 128-sample, 40-MS/s analog-ternary correlator," *IEEE J. Solid-State Circuits*, vol. 25, pp. 425–430, Apr. 1990.

[2] S. Kawahito, M. Yoshida, M. Sasaki, K. Umehara, D. Miyazaki, Y. Tadokoro, K. Murata, S. Doushou, and A. Matsuzawa, "A CMOS image sensor with analog 2-dimensional DCT-based compression circuits for one-chip cameras," *IEEE J. Solid State Circuits*, vol. 32, pp. 2030–2041, Dec. 1997.

[3] K. Onodera and P. Gray, "A 75-mW, 128-MHz DS-CDMA baseband demodulator for high-speed wireless applications," *IEEE J. Solid-State Circuits*, vol. 33, pp. 753–761, 1998.

[4] E. Vittoz "Analog VLSI signal processing—Why, where, and how," *J. VLSI Signal Processing*, vol. 8, no. 1, pp. 27–44, July 1994.

[5] R. Sarpeshkar, J. Kramer, G. Indiveri, and C. Koch, "Analog VLSI architecture for motion processing—From fundamental limits to system applications," *Proc. IEEE*, vol. 84, pp. 969–987, July 1996.

[6] G. Erten and R. Goodman "Analog VLSI implementation for stereo correspondence between 2-D images," *IEEE Trans. Neural Networks*, vol. 7, pp. 266–277, Mar. 1996.

[7] G. Bo, D. Caviglia, and M. Valle, "An analog VLSI implementation of a feature extractor for real-time optical character recognition," *IEEE J. Solid-State Circuits*, vol. 33, pp. 556–564, Apr. 1998.

[8] H. Card, D. McNeill, and C. Schneider, "Analog VLSI circuits for competitive learning networks," *Analog Integr. Circuits Signal Processing*, vol. 15, no. 3, pp. 291–314, Mar. 1998.

[9] A. Kramer, "Array-based analog computation," *IEEE Micro*, vol. 16, pp. 20–29, Oct. 1996.

[10] J. Lazzaro and C. Mead, "A silicon model of auditory localization," *Neural Comput.*, vol. 1, no. 1, pp. 47–57, 1989.

[11] J. Lazzaro and J. Wawrzynek, "Speech recognition experiments with silicon auditory models," *Analog Integr. Circuits Signal Process.* vol. 13, no. 1/2, pp. 37–51, May/June 1997.

[12] T. Lande, J. Nesheim, and Y. Berg, "Autocorrelation in micropower analog CMOS," *Analog Integr. Circuits Signal Process.* vol. 7, no. 1, pp. 61–68, Jan. 1995.

[13] K. P. Unnikrishnan, J. J. Hopfield, and D. W. Tank, "Connected-digit speaker-dependent speech recognition using a neural network with time-delayed connections," *IEEE Trans. Signal Processing*, vol. 39, pp. 698–713, 1991.

[14] R. Coggins, M. Jabri, B. Flower, and S. Pickard, "A hybrid analog and digital VLSI neural network for intracardiac morphology classification," *IEEE J. Solid-State Circuits*, vol. 30, pp. 542–550, May 1995.

[15] Panasonic, Inc., *Audio Signal Delay BBD Series.* Data Book, 1994.

[16] J. Lazzaro, J. Wawrzynek, and R. Lippmann, "A micropower analog circuit implementation of hidden Markov model state decoding," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1200–1209, Aug. 1997.

[17] F. J. Pineda, G. Cauwenberghs, and R. T. Edwards, "Bangs, clicks, snaps, thuds, and whacks: An architecture for acoustic transient processing," in *Proc. Neural Information Processing Systems (NIPS)*, Denver, CO, 1996.

[18] R. T. Edwards, G. Cauwenberghs, and F. J. Pineda, "A mixed-signal correlator for acoustic transient classification," in *Proc. ISCAS '97*, vol. 1, pp. 621–625.

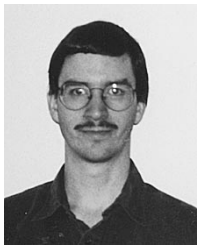
[19] F. J. Pineda, K. Ryals, D. Steigerwald, and P. Furth, "Acoustic transient processing using the Hopkins electronic ear," in *Proc. World Conf. Neural Networks*, Washington, DC, 1995.

[20] R. T. Edwards and G. Cauwenberghs, "A second-order log-domain bandpass filter for audio frequency applications," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, Monterey, CA, June 1998.

[21] ——, "Synthesis of log-domain filters from first-order building blocks," submitted for publication.

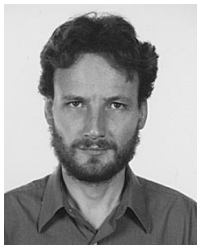
[22] A. van Schaik, E. Fragnière, and E. Vittoz, "A silicon model of amplitude modulation detection in the auditory brainstem," in *Proc. Neural Information Processing Systems (NIPS)*, Denver, 1996.

[23] A. Pavasović, A. G. Andreou, and C. R. Westgate, "Characterization of subthreshold MOS mismatch in transistors for VLSI systems," *J. VLSI Signal Processing*, 1994.



R. Timothy Edwards (M'94) received the B.S. degree in electrical engineering from Duke University, Durham, NC, in 1990, the M.S.E.E. degree from Stanford University, Stanford, CA, in 1992, and the Ph.D. degree from Johns Hopkins University, Baltimore, MD, in 1999.

His interests include analog and mixed-mode VLSI systems, acoustic signal processing, nonlinear circuits and systems, field-programmable analog array design, and computer-aided design (CAD) software development. He is currently employed in the Space Department, Johns Hopkins University Applied Physics Laboratory, Laurel, MD.



Gert Cauwenberghs (M'89) received the Engineer's degree in applied physics from the University of Brussels, Belgium, in 1988 and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, in 1989 and 1994, respectively.

In 1994, he joined Johns Hopkins University, Baltimore, MD, where he is now an Associate Professor in Electrical and Computer Engineering. His research covers analog and digital VLSI circuits, systems, and algorithms for parallel signal processing and adaptive neural computation. He edited the book *Learning on Silicon* (Norwell, MA: Kluwer, 1999).

Dr. Cauwenberghs received the National Science Foundation Career Award in 1997 and the Office of Naval Research Young Investigator Award in 1999.