

**Time-Frequency Acoustic Processing and Recognition:
Analysis and Analog VLSI Implementations**

by

Robert Timothy Edwards

A dissertation submitted to The Johns Hopkins University in conformity with
the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

March, 1999

© Robert Timothy Edwards 1999

All rights reserved

Abstract

Time-frequency analysis techniques, such as wavelet decomposition and Gabor filtering, are a tool for efficient coding of short-term acoustical features, and so are fundamental to acoustic pattern classification and speech recognition.

This thesis addresses issues of efficiency and robustness in the design and implementation of acoustic signal processors and small-vocabulary speech recognition systems for applications where power dissipation and integration density are primary design constraints. We couple time-frequency signal representations with massively parallel architectures using analog VLSI technology to design compact special-purpose systems with power efficiency surpassing conventional DSPs.

We present the design of, and results from, several prototyped VLSI systems, including processors for time-frequency decomposition and template-correlation-based acoustic transient pattern classification. We present methods for automatically training a template correlator and discuss potential research directions for this architecture, including biological modeling and continuous-speech recognition.

Acknowledgements

In order of appearance, I would like to thank the cast of characters who inspired this thesis, either directly or indirectly:

My parents and sister, who taught me critical thinking, studying, and writing skills, and encouraged me to explore my universe in all forms: Art, music, literature, language, and science. In spite of all their efforts, I was enamored of computers and technology, and became an engineer.

Dr. Hisham Massoud at Duke University, who inadvertently launched me on the path of neuromorphic engineering by recommending and loaning to me a book by Carver Mead called *Analog VLSI and Neural Systems*, and who affected my choice of graduate schools by advising me that the top five electrical engineering programs in the country were, in order, “Stanford, Stanford, Stanford, Berkeley, and Berkeley.”

Dr. Michael Godfrey at Stanford University, who enthusiastically encouraged me to pursue research in Analog VLSI. He and I, Boyd Fowler, and Neal Bhadkamkar put together an Analog VLSI research laboratory in the basement of Durand Building and held our own against the nay-sayers in the department.

My wife, Linglan, who shared with me the occasionally tortuous life of the married graduate student. Under mutual moral support, we kept roughly the same schedule from beginning to end of our respective theses. She beat me to it by a couple of months.

My thesis advisor, Dr. Gert Cauwenberghs, who was forced to bear the brunt of my moments of extreme aggravation that typically accompany thesis research.

All the people whose support, friendship, experiences and discussions have shaped many aspects of my research and this thesis: Among them Pamela Abshire, Andreas Andreou, Marc Cohen, Abbas El Gamal, Paul Furth, Wolfgang Himmelbauer, Mark Martin, Amjad Obeidat, Fernando Pineda, Philippe Pouliquen, Michael Sebert, Shihab Shamma, and many, many others.

Contents

Acknowledgements	ii
Abstract	iii
List of Figures	vii
List of Tables	xii
1 Introduction	1
1.1 Introduction	1
1.2 Mapping the Time-Frequency Plane	6
1.2.1 Current-mode filterbanks	8
1.2.2 Acoustic Transient Recognition	9
1.2.3 Analog VLSI implementation of the transient classifier	10
1.2.4 Overview: Learning and Continuous Speech Recognition	11
2 Continuous Wavelet Transform	14
2.1 Introduction to the 1-Dimensional Continuous Wavelet Transform	14
2.2 CWT vs. DWT	16
2.3 Gabor Logons and Wavelets	18
2.4 Complex Demodulation	20
2.5 Complex Demodulation in the Continuous Wavelet Processor	23
2.6 Post-processing	27
2.7 An Analog CWT Processor	28
2.8 Generating carrier sinusoids	29
2.9 Analog multiplication	30
2.10 Wavelet Gaussian Function	30
2.11 Wavelet chip slice	34
2.12 Chip Specifications	34
2.13 Limitations of the Architecture	36
2.14 Variations on an Architecture	37
2.15 A Mixed-Mode Wavelet Processor	38
2.16 Details of the Bit-Sequence-Finding Algorithm	42
2.17 Sequence generation	44

2.18	Results and implementation	46
2.19	Modulation Multiplier	49
2.20	Switch-Cap Wavelet Gaussian Function	50
2.21	Output Time Multiplexing	52
2.22	Wavelet chip slice	53
2.23	Experimental Results	55
2.24	Extensions of the Research	58
2.25	Summary	61
3	Current-Mode Filterbank Frontend	62
3.1	Time-Frequency Representations using Filterbanks	62
3.2	Parallel Filterbanks for Transient Classification	63
3.3	Current-Mode Filters for Current-Mode Applications	64
3.4	High-Level Simulations of the Filterbank Frontend	66
3.5	Introduction to Translinear Circuits and Log-domain Filtering	70
3.6	Principles of log-domain synthesis	73
3.7	First-Order Circuit synthesis	76
3.8	Designing second-order sections	78
3.9	Technology limitations for low-frequency filter design	83
3.10	Layout Considerations for VLSI Log-Domain Circuits	86
3.11	Current-Mode Circuits for Non-Filtering Applications	89
3.12	Signal Rectification and Smoothing	89
3.13	Signal Rectifier	90
3.14	Signal Peak-Peak Detector	91
3.15	L-1 Normalization Array	94
3.16	Experimental Results	97
3.17	Summary	105
4	Acoustic Transient Processing	106
4.1	Introduction	106
4.1.1	The Problem of Speech Recognition	106
4.1.2	Acoustic Transients	107
4.2	Algorithms	108
4.2.1	Simplifying The Correlation Equation	112
4.3	Simulations	119
4.3.1	The Hopkins Electronic Ear (HEEAR) Processor	120
4.3.2	Simulating the Acoustic Transient Baseline Algorithm	121
4.3.3	Optimizing Correlation Algorithms	123
4.3.4	Simulations of different zero-mean representations	125
4.3.5	High-Level Simulations of ATP Mixed-Mode VLSI Hardware	128
4.3.6	Optimization of the classifier using per-class gains	128
4.3.7	System Robustness	131
4.3.8	Research Directions	133
4.3.9	Remarks	134
4.4	Hardware Implementation of the Acoustic Transient Processor	135

4.4.1	Current-switching Memory Array	135
4.4.2	Bucket Brigade Device	138
4.4.3	Circuit input section	140
4.4.4	Characterizations of the VLSI Hardware	141
4.4.5	Experimental Results	144
4.4.6	Summary	147
4.5	The Digital ATP	149
4.6	Digital correlator custom VLSI architecture	149
4.7	Digital correlator semicustom FPGA architecture	151
4.8	The Switch-Capacitor Frontend	156
4.9	Experimental results of the trinary-trinary correlation hardware	161
4.9.1	Method of input segmentation	161
4.9.2	Method of template generation	163
4.9.3	Experimental Results	164
5	Learning and Speech Recognition	171
5.1	Automatic Template Learning for Template-Based Correlation	171
5.2	Average-Value Templates	172
5.3	Deterministic Methods: Statistical Component Analysis	174
5.4	Support Vector Machines	177
5.5	Heuristic Methods (Unnikrishnan/Hopfield)	178
5.6	Biologically-Inspired Methods	183
A	Linearity of a Transconductance Amp	187
B	MATLAB Code for Sine Sequence Generation	190
B.1	Commentary	193
C	Correlation with Time Differentiation	195
C.1	Proof of validity of the pipelined architecture	198
D	Simulation of the ATP Frontend	201
D.1	Commentary	207
E	Simulation of the ATP Correlator	209
E.1	Commentary	212
F	Optimizing per-class gains in the ATP	214
F.1	Commentary	219
	Bibliography	220
	Vita	227

List of Figures

1.1	Some mappings of the TF plane.	7
2.1	Output sampling. Points marked ‘ \times ’ represent center of the time-frequency area covered by that sampled output.	17
2.2	Frequency-time representation of the input as overlapping Gaussian filters.	17
2.3	Gabor sine and cosine logons, or wavelets.	19
2.4	Demodulation of an input X in the frequency domain with a perfect sine wave S and ideal modulation filter H	23
2.5	Complex demodulation (2 channels shown).	24
2.6	Complex modulation reconstruction (2 channels shown).	25
2.7	Circuit diagram of the frequency-division toggle flip-flops.	29
2.8	Gilbert multiplier with cascodes.	30
2.9	Frequency-domain transfer function of the final output of n cascaded lowpass filters as a function of the number of stages n	32
2.10	Impulse response of the final output of n cascaded lowpass filters as a function of the number of stages n	32
2.11	n cascaded stages of a filter approximating a half-Gaussian function, using continuous-time transconductance-C filters.	33
2.12	Wavelet decomposition block diagram for a single sine/cosine pair.	35
2.13	Wavelet reconstruction block diagram for a single sine/cosine pair.	35
2.14	Wavelet Transform Chip block diagram.	36
2.15	Architecture of Moreira-Tamayo <i>et al.</i>	38
2.16	The analog oscillator architecture of Lu <i>et al.</i> [26] containing a 2nd-order delta-sigma modulator.	40
2.17	<i>Top</i> : Demodulation of an input X in the frequency domain with a perfect sine wave S and ideal modulation filter H . <i>Bottom</i> : Demodulation of an input X in the frequency domain with an oversampled sine wave S' using an ideal smoothing filter G and modulation filter H	45
2.18	The oversampled sine sequence.	47
2.19	Optimized 64-bit oversampled sine sequence, first quadrant.	47
2.20	Frequency domain properties of the raw (\mathbf{o}) and filtered (\mathbf{x}) bit sequences.	48
2.21	Use of Gray code to generate sine and cosine sequences.	49
2.22	Multiplexing <i>vs.</i> Multiplying.	50

2.23	Single discrete-time lowpass filter section.	52
2.24	A scheme for controlling time-multiplexing of the outputs.	53
2.25	Sixteen-channel architecture using the 7-to-5 frequency ratio.	54
2.26	The wavelet chip, block diagram.	55
2.27	Photomicrograph of the mixed-mode continuous wavelet transform processor, a 4 mm×6 mm die size fabricated in a 2 μm CMOS p-well process.	56
2.28	Signal demodulation using the wavelet chip. The top trace is the input signal. The middle trace is the input multiplied by the binary sequence. The bottom trace is the output after filtering.	57
2.29	Gaussian filter magnitude response: predicted and measured.	58
2.30	Gaussian filter phase response: predicted and measured.	59
2.31	Wavelet responses to isolated sine wave input.	59
2.32	An arbitrary tiling of the time-frequency plane.	60
3.1	Frontend filterbank system—block diagram.	64
3.2	Sampled-data input from a recording of an acoustic transient—that of a book being dropped onto a desk.	67
3.3	Bandpass-filtered acoustic transient input using two cascaded second-order filters on each channel.	69
3.4	Parallel filterbank output after rectification and smoothing of the acoustic transient input signal across all frequency bands.	71
3.5	Translinear loop with common-base and common-emitter configurations.	74
3.6	Filter pole formed using a transconductor.	76
3.7	First-order log-domain filter circuit.	77
3.8	Computing a current difference at a log-domain filter input. A) The underlying idea, which is physically unrealizable. B) An equivalent working implementation.	80
3.9	Bandpass structure formed from first-order sections.	81
3.10	An alternative common-emitter circuit generating I'_{DC} (see text for discussion).	82
3.11	Complete circuit schematic for the second-order bandpass filter.	83
3.12	Another BiCMOS Log-domain bandpass filter, using the common-emitter configuration for all feedback circuits.	84
3.13	Base compensation (B) eliminates undesirable behavior due to significant base current which occurs in (A).	85
3.14	Complete circuit schematic for the second-order bandpass filter. Bipolar transistors are minimum size, and MOS dimensions are indicated as W/L in units of $\lambda = 0.6 \mu\text{m}$	86
3.15	Circuit of Figure 3.12, showing base compensation circuits, cascode connections, and the Q -generating circuit	87
3.16	Structure of a bandpass filterbank channel and stacking of channels to form the whole filterbank.	89
3.17	Simplified schematic of the adaptive current full-wave rectifier.	91
3.18	Cascaded log-domain lowpass filters.	92
3.19	Simple diode-based peak detector.	92
3.20	Behavior of the peak-peak detector.	94
3.21	Peak-peak detector circuit, simplified.	95
3.22	Peak-peak detector with cascodes, as fabricated.	96

3.23	L-1 normalization circuit, after Gilbert [50].	98
3.24	Measured magnitude response of the log-domain first-order lowpass filter from the test chip.	99
3.25	Measured magnitude response of one bandpass channel in the filterbank system, made of two cascaded second-order log-domain bandpass filters, over three tunings of the center frequency.	101
3.26	Measured magnitude response of all bandpass channels in the filterbank system, as measured at the output after peak-detection and smoothing.	102
3.27	Measured center frequencies of all bandpass channels in the filterbank system (circles), compared to the ideal exponential spacing (solid line).	103
3.28	Photograph of the fifteen-channel bandpass filterbank fabricated in 1.2 μm technology inside a 2.2 mm \times 2.2 mm padframe.	104
4.1	Template correlator, as a direct implementation of the baseline algorithm, Equation (4.1).	110
4.2	Template correlation.	111
4.3	Filterbank output after applying an L-1 normalization across all channels. A single channel has been added to the system (top trace), containing the result of a constant value less the instantaneous sum of the remaining channel values.	113
4.4	Template correlator with pipelined architecture and multiplexors replacing the multipliers.	115
4.5	Values in the pipelined delay registers in the correlation algorithm simulation.	116
4.6	Template correlator with [0,1] encoding of template values.	117
4.7	Block diagram of the temporal current correlator.	119
4.8	Example transient recorded data.	120
4.9	Example HEEAR-processed transients.	122
4.10	Example templates learned by the ATP algorithm.	123
4.11	Example of a real-valued template.	124
4.12	The same template reduced to trinary values.	124
4.13	The same template reduced to binary values.	125
4.14	Effect of decreasing the number of time-bins.	132
4.15	Effect of white noise added to the correlator inputs.	133
4.16	Block diagram of the temporal current correlator.	136
4.17	Efficient dynamic memory cell for the correlator array.	138
4.18	Details of the bucket brigade device (BBD).	139
4.19	A switched capacitor circuit to compute channel differences at the output.	141
4.20	The complete correlator array including both bucket brigade devices. Representative MOS device sizes are given as W/L in units of $\lambda = 0.6 \mu\text{m}$	142
4.21	Linearity between output voltage and input current.	143
4.22	Impulse response at each tap of the bucket brigade device.	144
4.23	Matching between devices throughout the template at 1 μA input.	145
4.24	Matching between devices throughout the template at 100 nA input.	145
4.25	Photomicrograph of the acoustic transient correlator, a 2.2 mm \times 2.2 mm die fabricated in a 1.2 μm CMOS technology.	146

4.26	Measured correlation of repeated sounds “can” (left) and “snap” (right) with the “can” template loaded into chip memory.	148
4.27	Measured correlation of repeated sounds “can” (left) and “snap” (right) with the “snap” template loaded into chip memory.	148
4.28	Block diagram of the sequential digital correlator architecture.	151
4.29	Memory allocation in the digital correlator SRAM.	153
4.30	SRAM address generator for the digital correlator.	154
4.31	Digital correlator controller state diagram, simplified.	155
4.32	Serial digital correlator structure for one template.	155
4.33	Digital ATP correlator system, configurable for up to sixteen templates.	157
4.34	Switched-capacitor bandpass filterbank, single channel.	160
4.35	Audio input circuit for the frontend system.	160
4.36	One board of the switched capacitor frontend filterbank, bandpass filtering and encoding the audio input into sixteen parallel channels.	161
4.37	A mixed-signal method for detecting the onset of a transient event from the output of an analog frontend filterbank.	162
4.38	A simple finite state machine which detects the onset of transient events in the trinary-valued frontend filterbank output.	163
4.39	Templates for twelve transient classes, determined from data obtained from the frontend hardware. Frequency is on the x -axis, with the lowest frequency channel on the right, and time is on the y -axis, with the transient onset at the bottom of the template. Colors correspond to trinary values as follows: black = -1 , white = $+1$, gray = 0	165
4.40	System input and output for sound “bar” (left) and “book” (right). Top graph is the frontend system output (32 channels); middle graph is the correlation output over 12 classes; bottom graph is the output of the segmenter showing both the segmented transient and the optimal point of alignment.	166
4.41	System input and output for sound “shelf” (left) and “tub” (right). Top graph is the frontend system output (32 channels); middle graph is the correlation output over 12 classes; bottom graph is the output of the segmenter showing both the segmented transient and the optimal point of alignment.	167
5.1	Unnikrishnan, Hopfield, and Tank correlator architecture, block diagram.	179
5.2	Unnikrishnan <i>et al.</i> correlator simulation results on isolated digit recognition.	180
5.3	Response of architecture in simulation to the same input as Figure 5.2 but with white noise added to the input. The system remains robust in the presence of noise.	180
5.4	Gaussian kernels which diffuse the input in the Unnikrishnan <i>et al.</i> model.	181
5.5	Example of a biologically-inspired template encoding a falling tone of specific frequency and rate. <i>left</i> : Continuous-valued encoding. <i>right</i> : Trinary encoding.	184
A.1	Simple differential pair transconductance amplifier.	188
A.2	Linear error in a simple transconductance amplifier as a function of the differential input voltage.	189

C.1	Block diagram of the temporal current correlator, using time-differencing operations on the input.	197
C.2	Switch-cap time-differencing circuit.	199
C.3	Time-differencing correlation architecture.	200

List of Tables

2.1	Technologies used for the Wavelet processors.	35
3.1	Measured filterbank characteristics.	100
4.1	Classes and descriptions of the recorded transient dataset.	121
4.2	Confusion matrix for leave-one-out cross-validation loop, using the baseline algorithm.	124
4.3	Simulation results with different architectures.	126
4.4	Confusion matrix for leave-one-out cross-validation loop, using binary (1, 0) templates and channel differencing.	126
4.5	Simulation results for different methods of computing channel differences	127
4.6	System accuracy with and without per-class normalization.	131
4.7	Transistor, cell, and array sizes corresponding to Figure 4.16.	137
4.8	Left: logic for the binary-binary correlation operation. Right: logic for the trinary-trinary correlation. “X” represents a don’t-care condition.	150
4.9	2-bit Sign-amplitude trinary representation of input and template values in the digital ATP. “X” represents a don’t-care condition.	159
4.10	Offline recognition task: Cross-validation on the transient dataset using a non-real-time segmenter.	168
4.11	Real-time recognition task: Cross-validation on the transient dataset using the simple finite state machine segmenter.	169
4.12	Real-time recognition task: Cross-validation on the transient dataset using the simple finite state machine segmenter. In this transient set, spurious transient events were hand-tagged as “unknown” classes (denoted by a question mark).	170

Vita

R. Timothy Edwards received the degree of Bachelor of Science in Electrical Engineering (BSEE) from Duke University, Durham, North Carolina in 1990, graduating Summa Cum Laude and with Honors. In 1992, he received the degree of Master of Science in Electrical Engineering (MSEE) from Stanford University, Stanford, California. In 1993, he came to Johns Hopkins University to pursue a Ph.D. degree in the department of Electrical and Computer Engineering, Whiting School of Engineering, under the direction of Professor Gert Cauwenberghs. He is currently employed as a senior staff researcher in the Space Department at the Johns Hopkins University Applied Physics Laboratory.