Time-Frequency Acoustic Processing and Recognition: Analysis and Analog VLSI Implementation

R. Timothy Edwards

Johns Hopkins University
Department of Electrical and Computer Engineering

March 26, 1999

Advisor: Dr. Gert Cauwenberghs

Time-Frequency Acoustic Processing and Recognition:
Analysis and Analog VLSI Implementation

- Introduction
- Template Correlation
- An Acoustic Transient Processor in VLSI:
  - Frontend Filterbank Systems
  - Template Correlation
  - Experimental Results
- A Trinary-Trinary Correlator
- Conclusion

Research Goals:

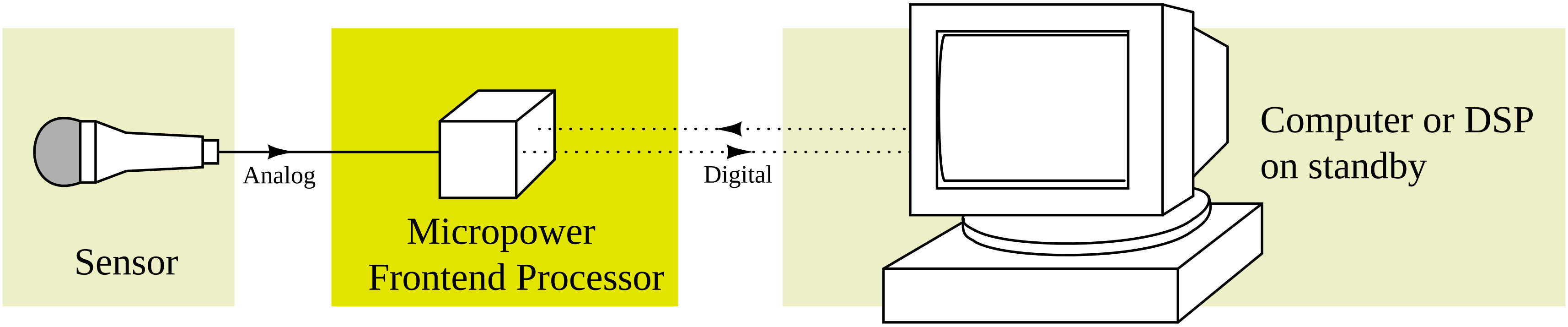
Investigate signal processing systems for which primary constraints are **power consumption** and **physical area**.

- Develop these systems into efficient analog and digital hardware
- Maintain high accuracy with respect to microprocessor and DSP solutions
- Use circuits with ultra-low (µW) power consumption
- Exploit parallel architectures to achieve robustness in the presence of noise and mismatch

Performance depends critically on the choice of **algorithm** and **hardware**.

Applications:

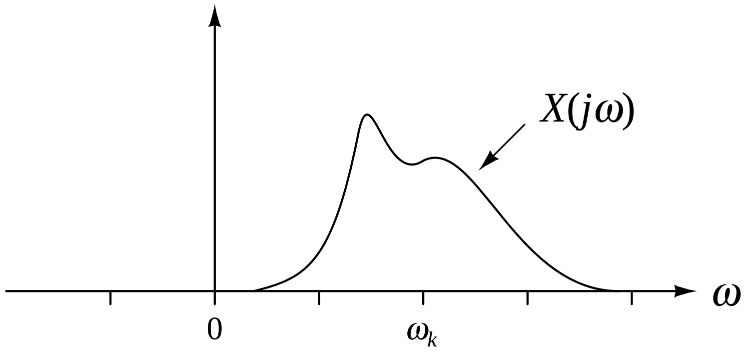
- Handheld or mobile (battery-operated) environments
- Can be used to monitor environment and wake up systems in “standby” mode
- “Smart” A/D conversion

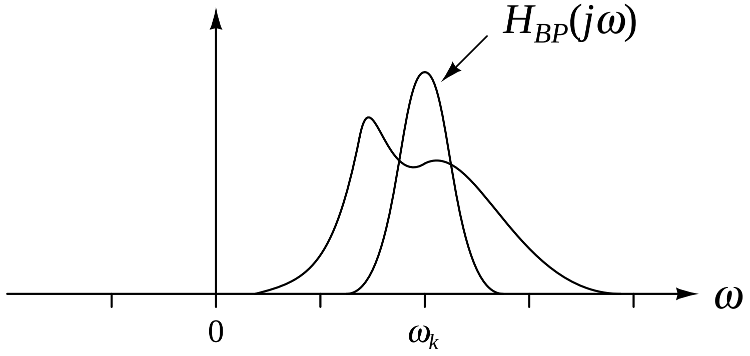


Microelectronic Systems Laboratory Research: Systems and Algorithms

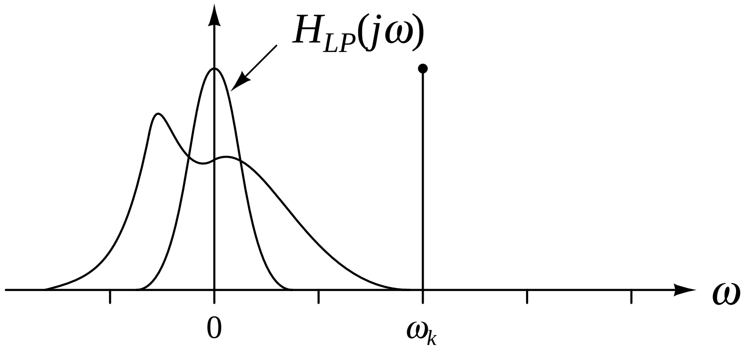
- Continuous wavelet transform (CWT) processor
- Log-domain filters for audio-frequency applications
- Micropower mixed-mode template correlation acoustic pattern classifier
- Model-free adaptive correction of optical aberrations

Bandpass Filterbank vs. Complex Demodulation





$X(j\omega)\ H_{BP}(j\omega)$

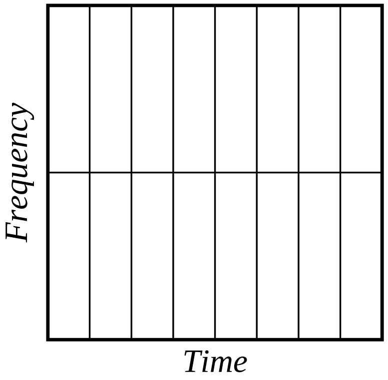


$(X(j\omega) * \delta(j\omega - j\omega_k))\ H_{LP}(j\omega)$

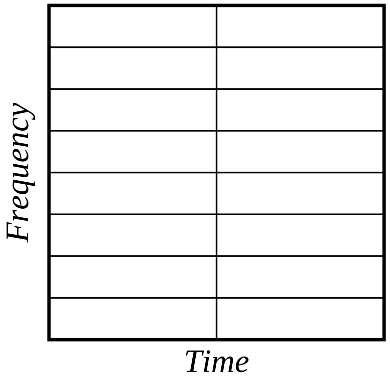
Tilings of the Time-Frequency Plane

Area of uncertainty in time and frequency: $\Delta f \Delta t \geq \frac{1}{2}$

Gives good time resolution but poor frequency resolution



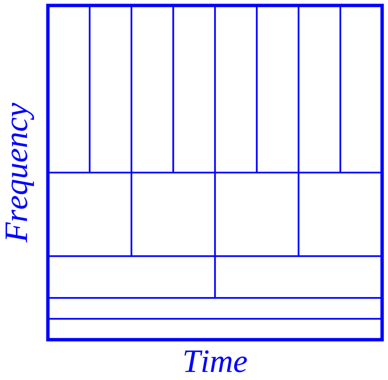
(a) FFT, short window



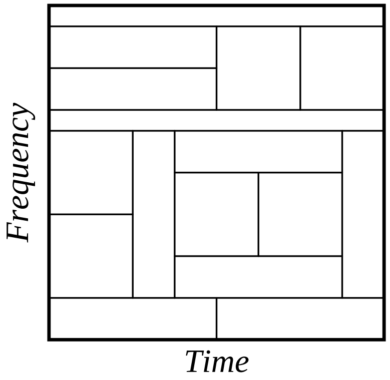
(b) FFT, long window

Gives good frequency resolution but poor time resolution

Most efficient time/frequency tradeoff for arbitrary inputs



(c) Wavelet Basis



(d) Wavelet Packets

The most efficient description of a specific signal; difficult to compute on the fly

Time-Frequency Acoustic Processing and Recognition:
Analysis and Analog VLSI Implementation

- Introduction
- Template Correlation
- An Acoustic Transient Processor in VLSI:
  - Frontend Filterbank Systems
  - Template Correlation
  - Experimental Results
- A Trinary-Trinary Correlator
- Conclusion

Acoustic Event Classification Using Correlation

Goal: Try to classify an acoustic event by matching the time-frequency decomposed input against a set of templates. Each template has $M \times N$ values and represents a “typical” example of one class.

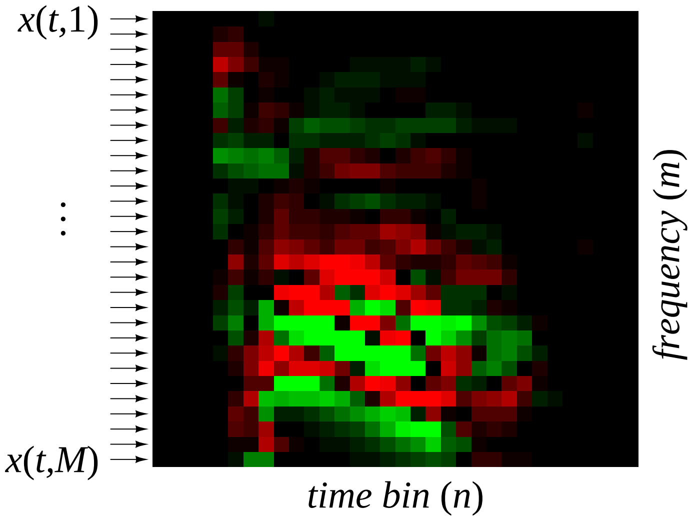
Signal (x) characteristics (output of frontend filtering system):

- M frequency channels (log-spacing)
- Continuous-time (or Discrete-time)
- Continuous-valued
- Normalized energy envelope

Template (p_z) characteristics:

- One template per class (z)
- M frequency channels (log-spacing)
- N time-bins (1 ms per bin for transients)

Template

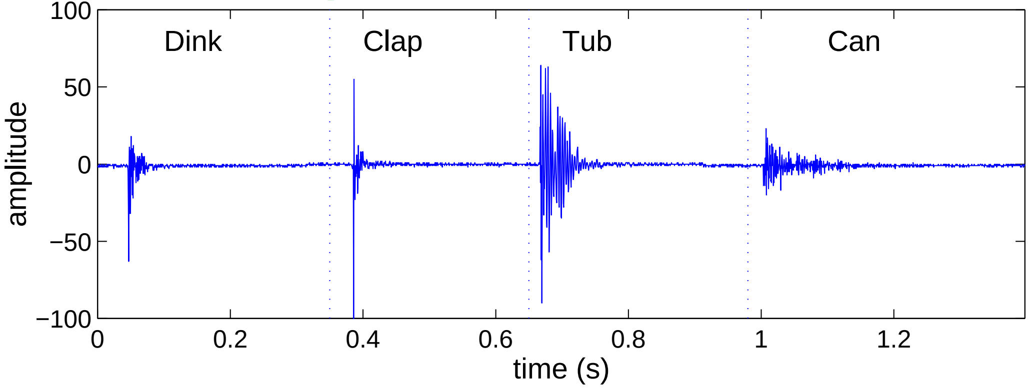


Basic Correlation equation:

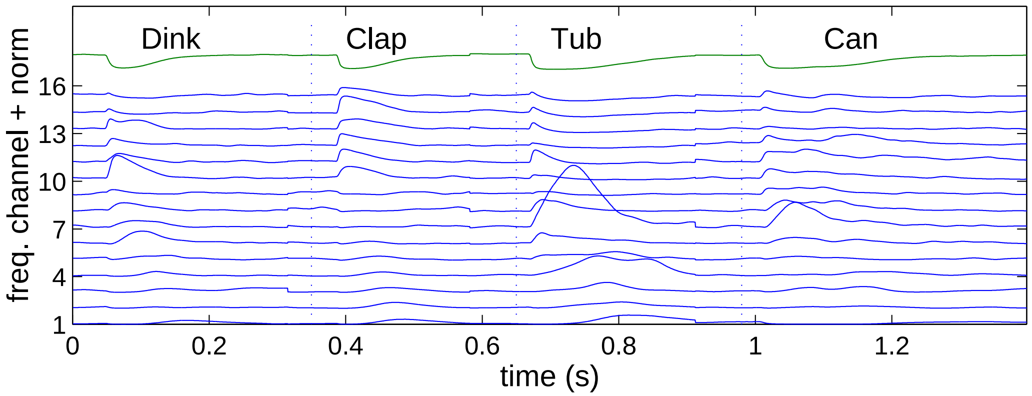
$$\boxed{c_z[\,t\,] = \sum_{m=1}^{M} \sum_{n=1}^{N} x[t-n,m]\,p_z[n,m]}$$

Acoustic Event Recognition: Example

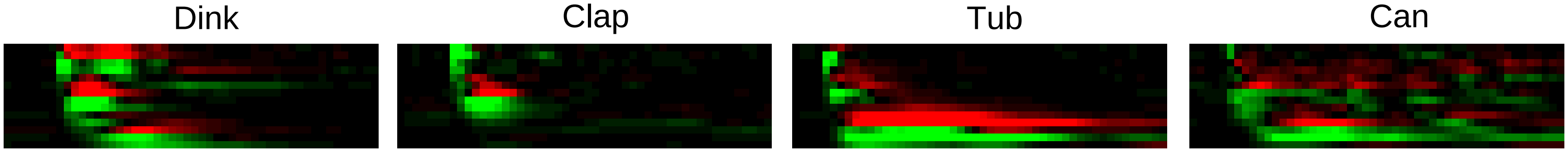
Inputs: acoustic data



Frontend processing



Classification via template correlation



Acoustic Signal Classification

Three main components:

1. **Filterbank**: Time-frequency decomposition converts the input signal into an efficient representation for subsequent processing.
2. **Template Correlator**: Simple correlation is sufficient to recognize complicated acoustic events.
3. **Digital Post-processing**: Uses correlator outputs as **features** for tasks such as speech recognition.

This talk will focus on current research (parts 1 and 2) and conclude with comments about (3).

Time-Frequency Acoustic Processing and Recognition:
Analysis and Analog VLSI Implementation

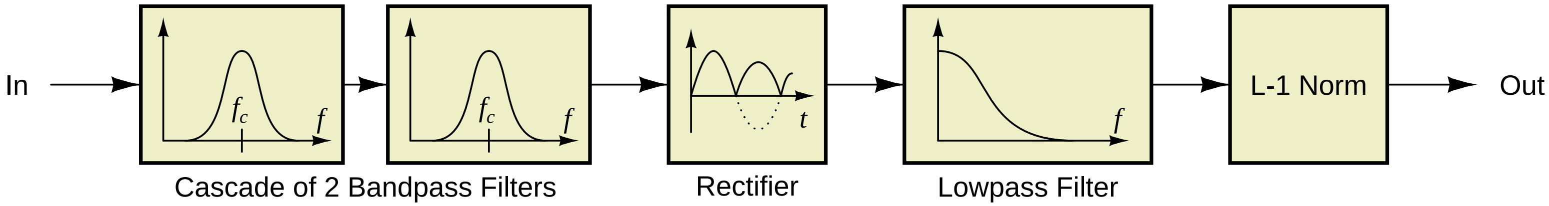
- Introduction
- Template Correlation
- An Acoustic Transient Processor in VLSI:
  - Frontend Filterbank Systems
  - Template Correlation
  - Experimental Results
- A Trinary-Trinary Correlator
- Conclusion

Filterbank Frontend Systems

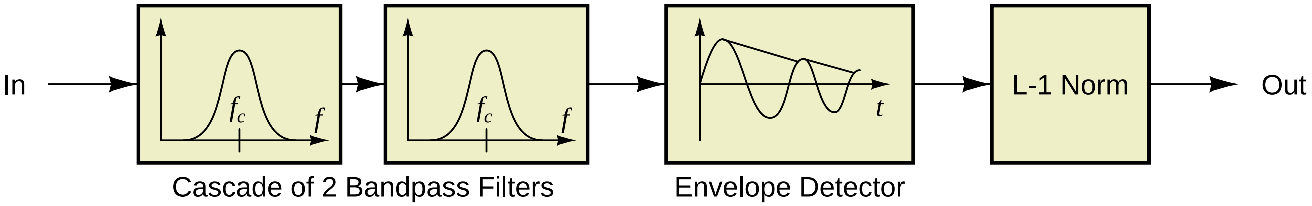
- Cochlear-model filters:
  - Mead and Lyon (1989)
  - Weimin Liu et al. “HEEAR” (1994)
  - . . . and many others
- Unnikrishnan, Hopfield, and Tank (1991)
- Shihab Shamma (1994)
- Pineda, Edwards, and Cauwenberghs (1995)

Our frontend architecture: two versions

(original)


In
f_c f
Cascade of 2 Bandpass Filters
t
Rectifier
f
Lowpass Filter
L-1 Norm
Out

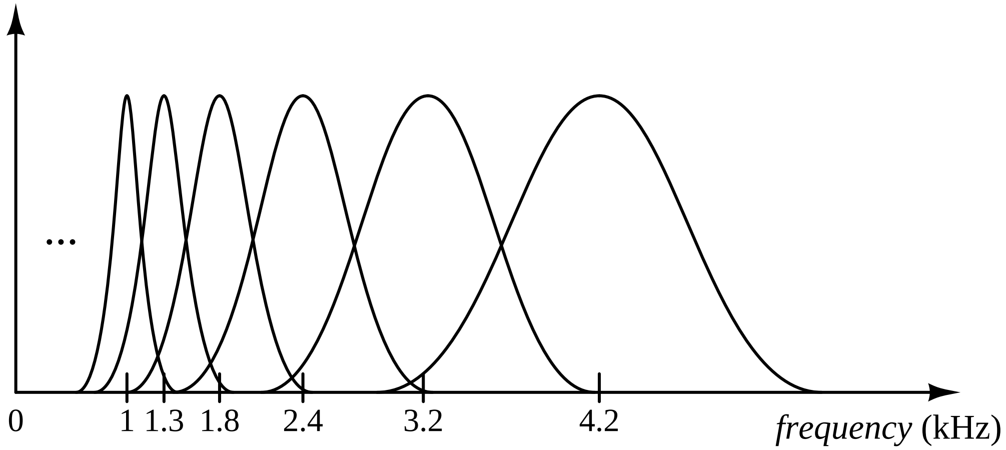
(modified)


In
f_c f
Cascade of 2 Bandpass Filters
t
Envelope Detector
L-1 Norm
Out

Bandpass Filterbank Frontend Processor

System characteristics:

- Parallel bank of 2nd- or higher order bandpass filters with constant Q
- Logarithmically-spaced center frequencies across the bank



- Rectification and smoothing of output to get energy envelope of signal
- Automatic gain control through *L-1* normalization across outputs (or other method)

A BiCMOS Current-Mode Filterbank Frontend

System implementation:

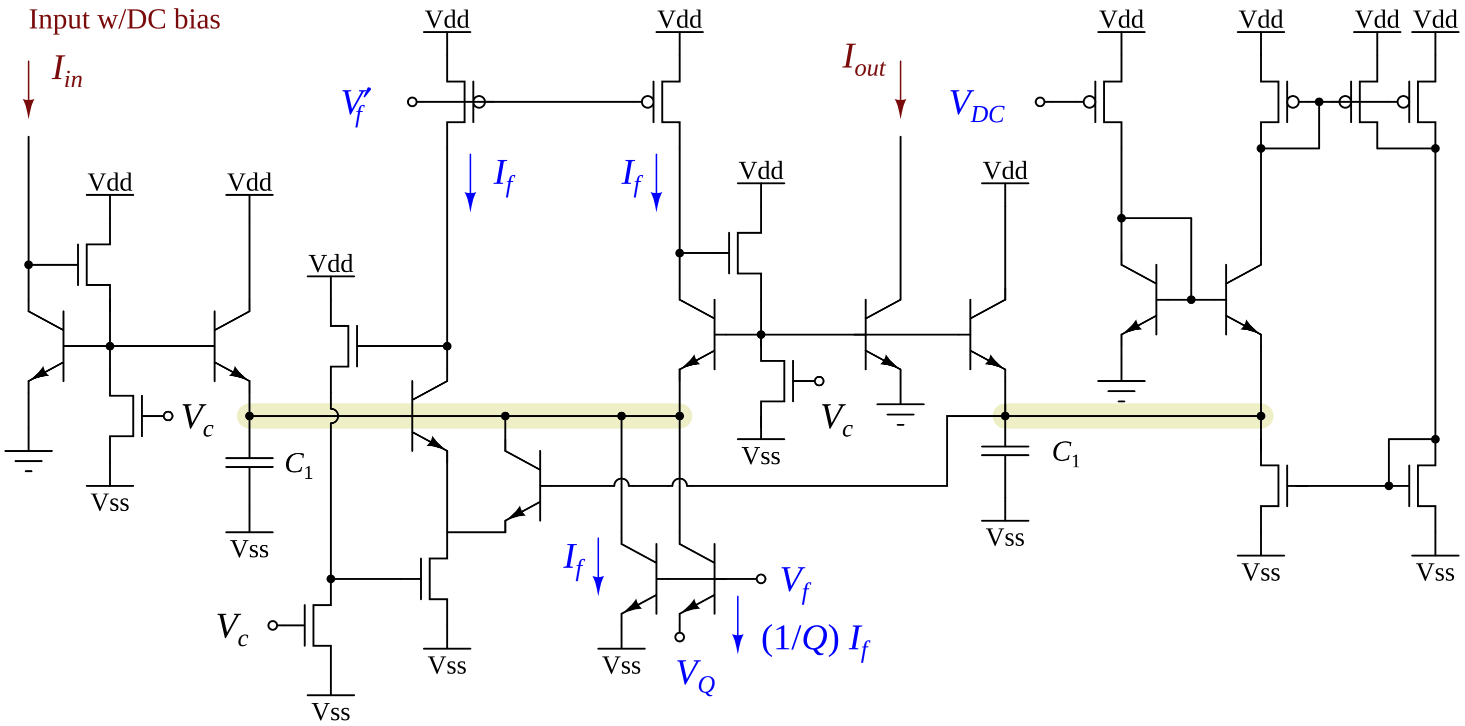
- *Log-domain filters*[*] based on *translinear circuits*
- BiCMOS design for greatest dynamic range and linearity

Why a Current-Mode VLSI Implementation?

- Large dynamic range compared to voltage-mode circuits
- Dense circuit layouts
- Filter parameters (f_c, Q) are tunable over multiple decades
- µW-range power consumption: good linearity at low power
- Most convenient format to interface to current-mode backend processors

[*] B. Gilbert (1975), D. Frey (1996), Y. Tsividis (1997), R. T. Edwards and G. Cauwenberghs (1998), and others

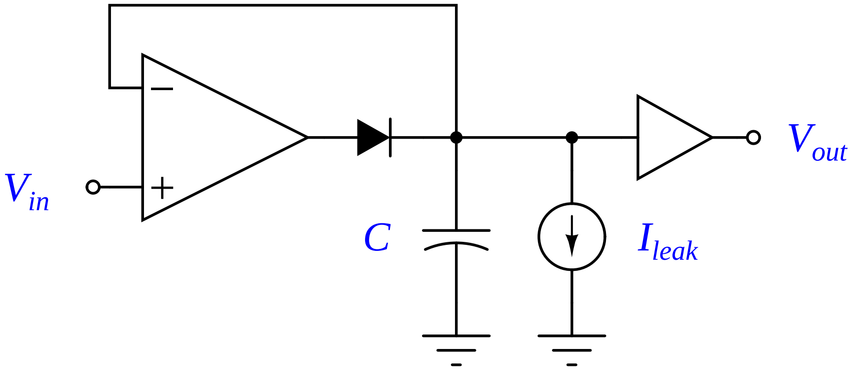
The Log-domain Bandpass Filter



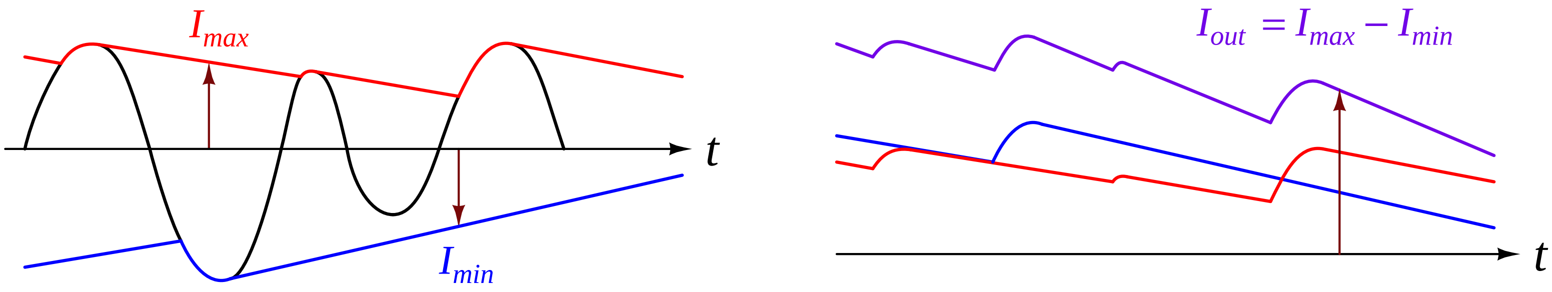
$$I_{out} = I_{in}\left(\frac{\tau s}{1+(1/Q)\tau s+\tau^2 s^2}\right) + I_{DC} \qquad \text{where} \qquad \tau = \left(\frac{V_t C_1}{I_f}\right) \qquad \text{and} \qquad Q = e^{V_Q/V_t}$$

The Envelope (Peak) Detector

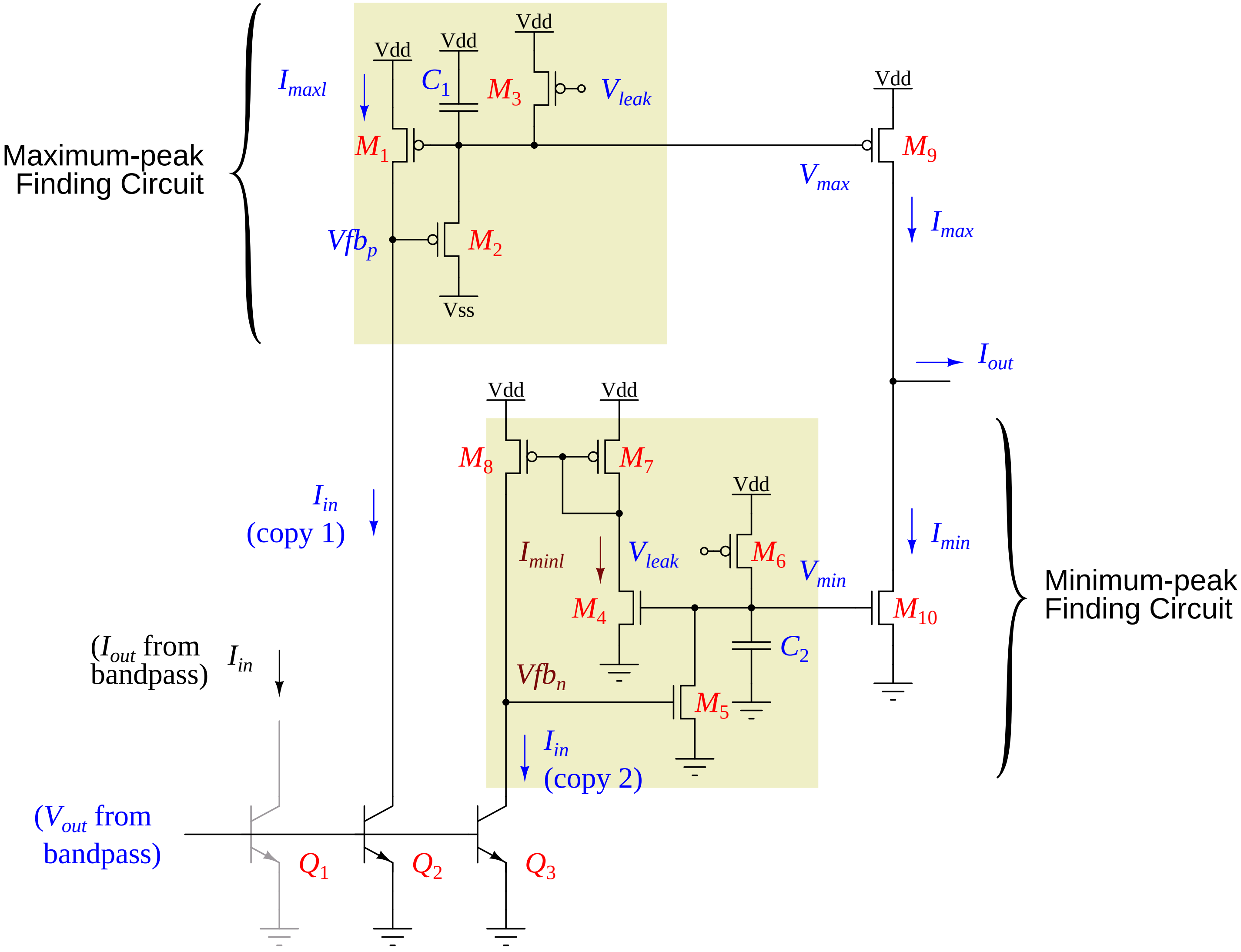
Simple Op-Amp Peak Detector



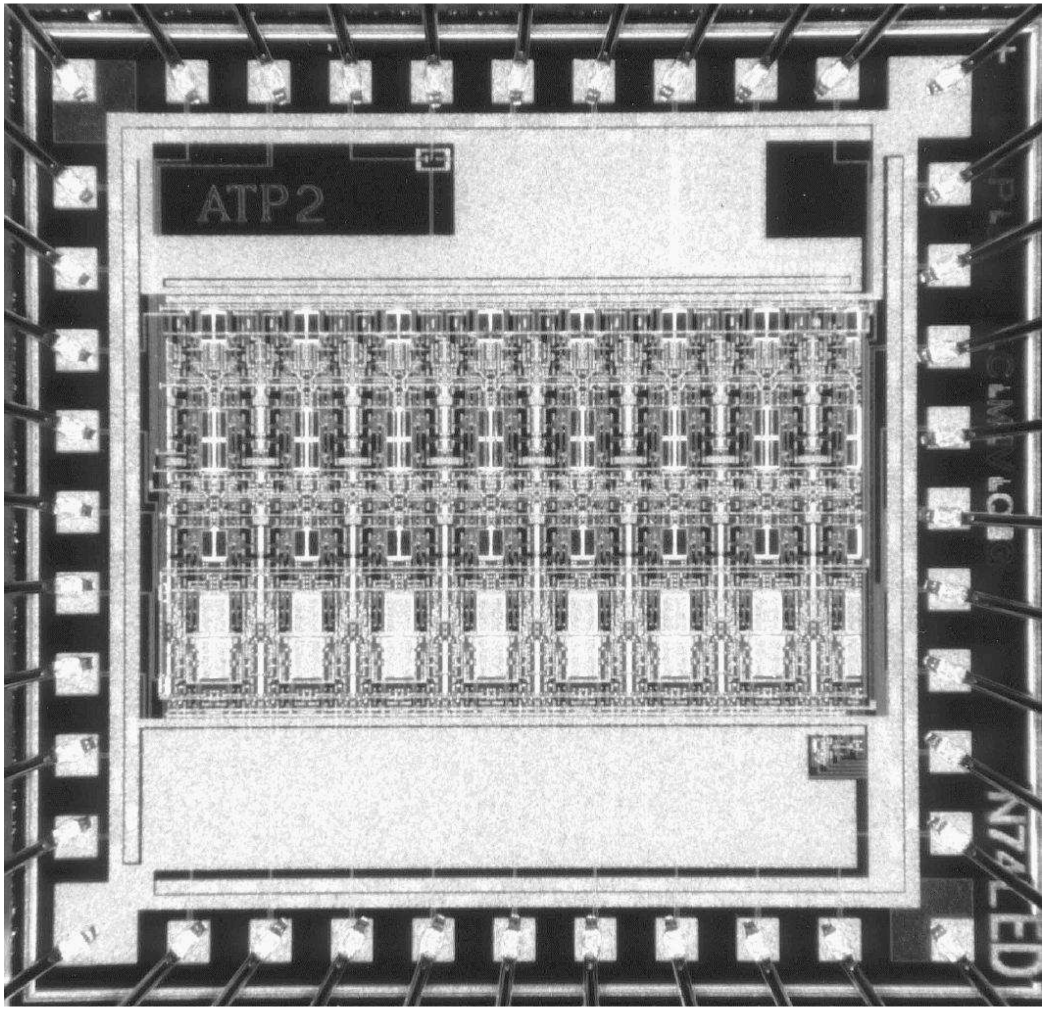
Behavior of a Peak-to-Peak Detector



The Envelope Detector

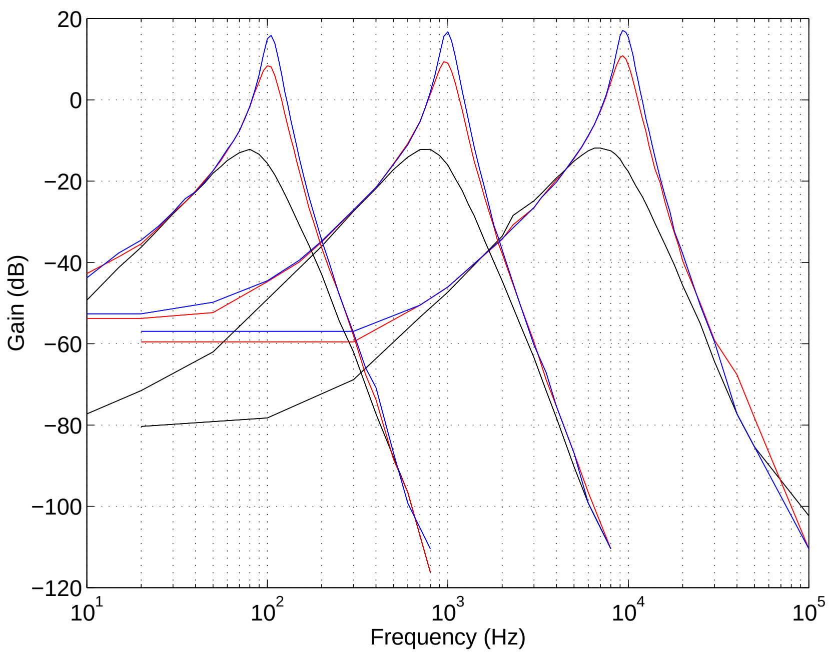


Log-Domain Frontend Filterbank—Photomicrograph

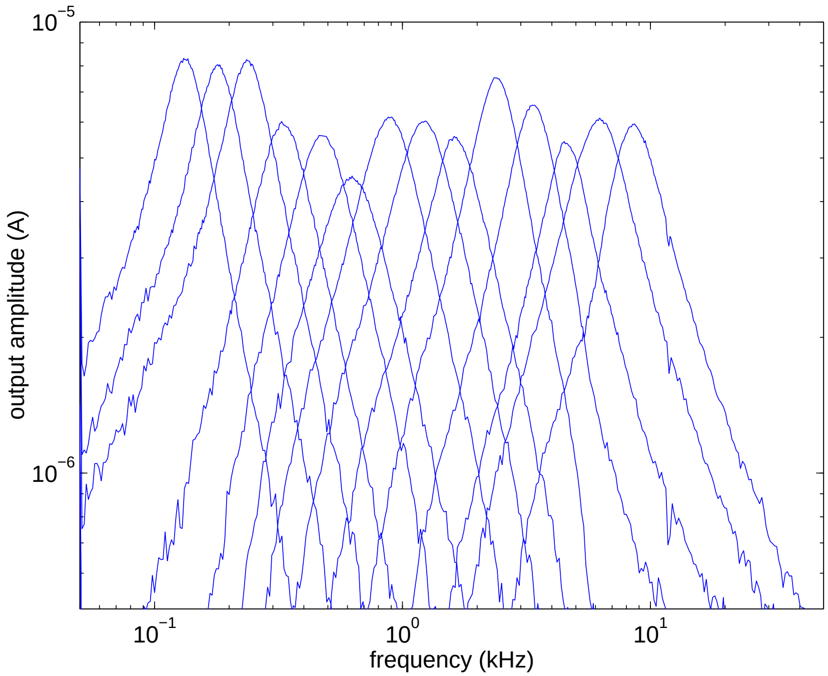


The fifteen-channel bandpass filterbank fabricated in 1.2 µm technology inside a 2.2 mm × 2.2 mm padframe.

Measured results of the Log-domain filters



Measured frequency response of one filter (4th-order bandpass) measured at three different center frequencies and three different Q-values.



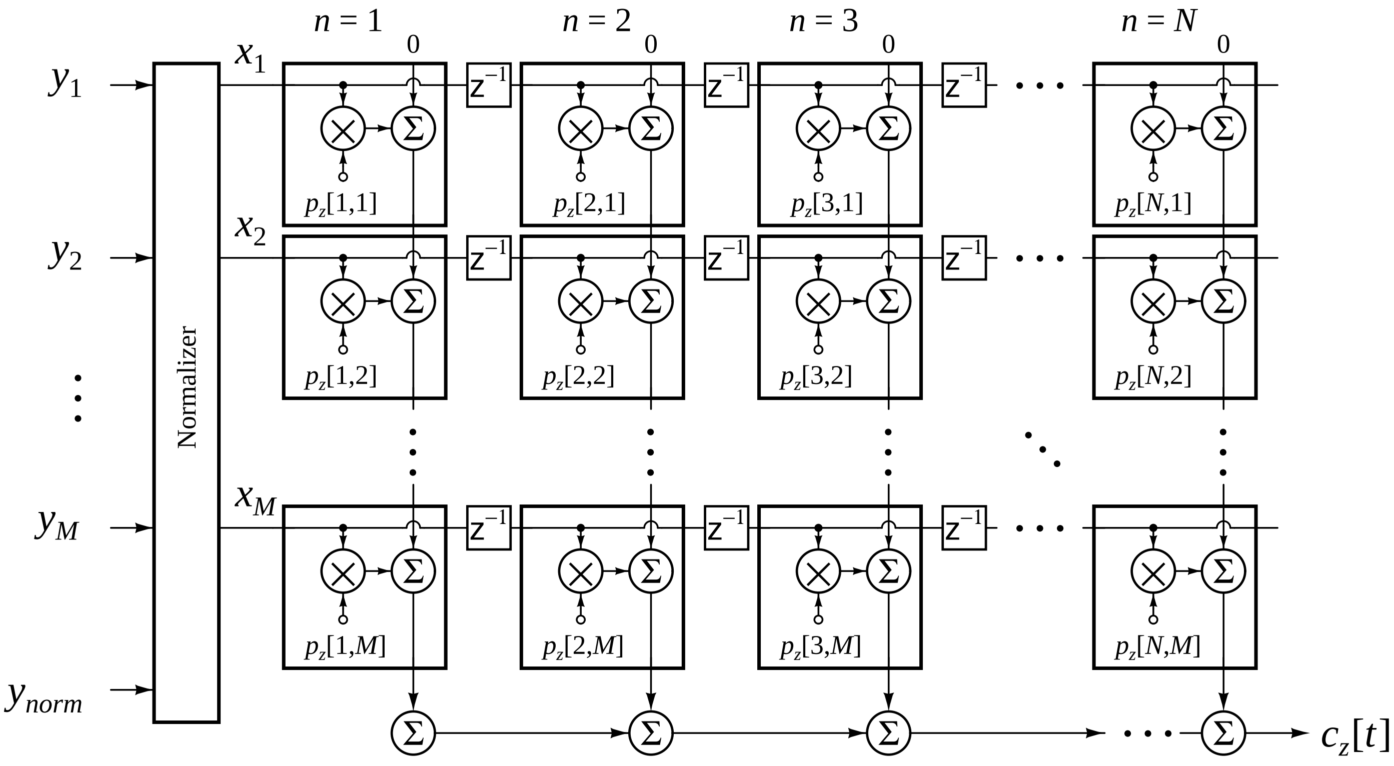
Measured output of the envelope detector of all filter outputs, showing mismatch between channels.

Time-Frequency Acoustic Processing and Recognition: Analysis and Analog VLSI Implementation

- Introduction
- Template Correlation
- An Acoustic Transient Processor in VLSI:
  - Frontend Filterbank Systems
  - Template Correlation
  - Experimental Results
- A Trinary-Trinary Correlator
- Conclusion

The Acoustic Transient Processor Correlation Algorithm

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



Original Software Trials of the Transient Classifier

Confusion Matrix for Cross-Validation Test

Baseline Algorithm Results

Event	Bar	Book	Can	Dink	Door	Finger	Hand	Mallet	Shelf	Tub
Bar	28	0	0	0	0	0	0	0	0	0
Book	0	19	0	0	0	0	0	0	0	1
Can	0	0	27	0	2	0	0	0	0	0
Dink	1	0	0	25	0	0	1	1	0	0
Door	0	0	0	0	9	0	0	1	0	0
Finger	0	0	0	0	0	17	1	0	0	0
Hand	0	0	0	0	0	0	21	0	0	0
Mallet	0	0	0	0	0	0	0	12	0	0
Shelf	0	0	0	0	0	0	0	0	28	0
Tub	0	0	0	0	0	0	0	0	0	28

Statistics:

Total instances presented:	222
Correct:	214
Incorrect:	8
Accuracy:	96.4%

Template Correlation: Computational Requirements

- M = 32 frequency channels
- timestep = 1 ms
- N = 100 bins (timespan of 1/10 second)
- one correlation per timestep

$>$3 million multiply-accumulate operations per second per template

---

Implementation goals:

- minimize operations per timestep (modify algorithm)
- minimize power per operation (analog design)

Different Correlation Architectures

Method	Both Cont.	Binary Input	Both Binary	Binary $(1,-1)$ Template	Binary $(1,0)$ Template
One-to-One	$\boxed{\text{96.40\%}}$	—	—	—	—
Time Difference	85.59%	65.32%	59.46%	82.43%	81.98%
Channel Difference	90.54%	53.60%	$\boxed{\text{95.05\%}}$	94.59%	$\boxed{\text{94.14\%}}$
Center-Surround	92.79%	53.60%	95.05%	92.34%	92.34%

Conclusion:

Best architectures use channel difference computation and binary template values and either binary or continuous-valued input.

Software Trials of the Transient Classifier

Confusion Matrix for Cross-Validation Test

Channel-differenced input and binary-valued template

Event	Bar	Book	Can	Dink	Door	Finger	Hand	Mallet	Shelf	Tub
Bar	27	0	0	0	1	0	0	0	0	0
Book	0	19	0	0	0	0	0	0	0	1
Can	1	0	22	0	4	0	0	0	1	0
Dink	1	0	0	26	0	0	1	0	0	0
Door	0	0	0	0	10	0	0	0	0	0
Finger	0	0	0	0	0	18	0	0	0	0
Hand	0	0	0	0	0	3	18	0	0	0
Mallet	0	0	0	0	0	0	2	10	0	0
Shelf	0	0	0	0	1	0	0	0	27	0
Tub	0	0	0	0	0	0	0	0	0	28

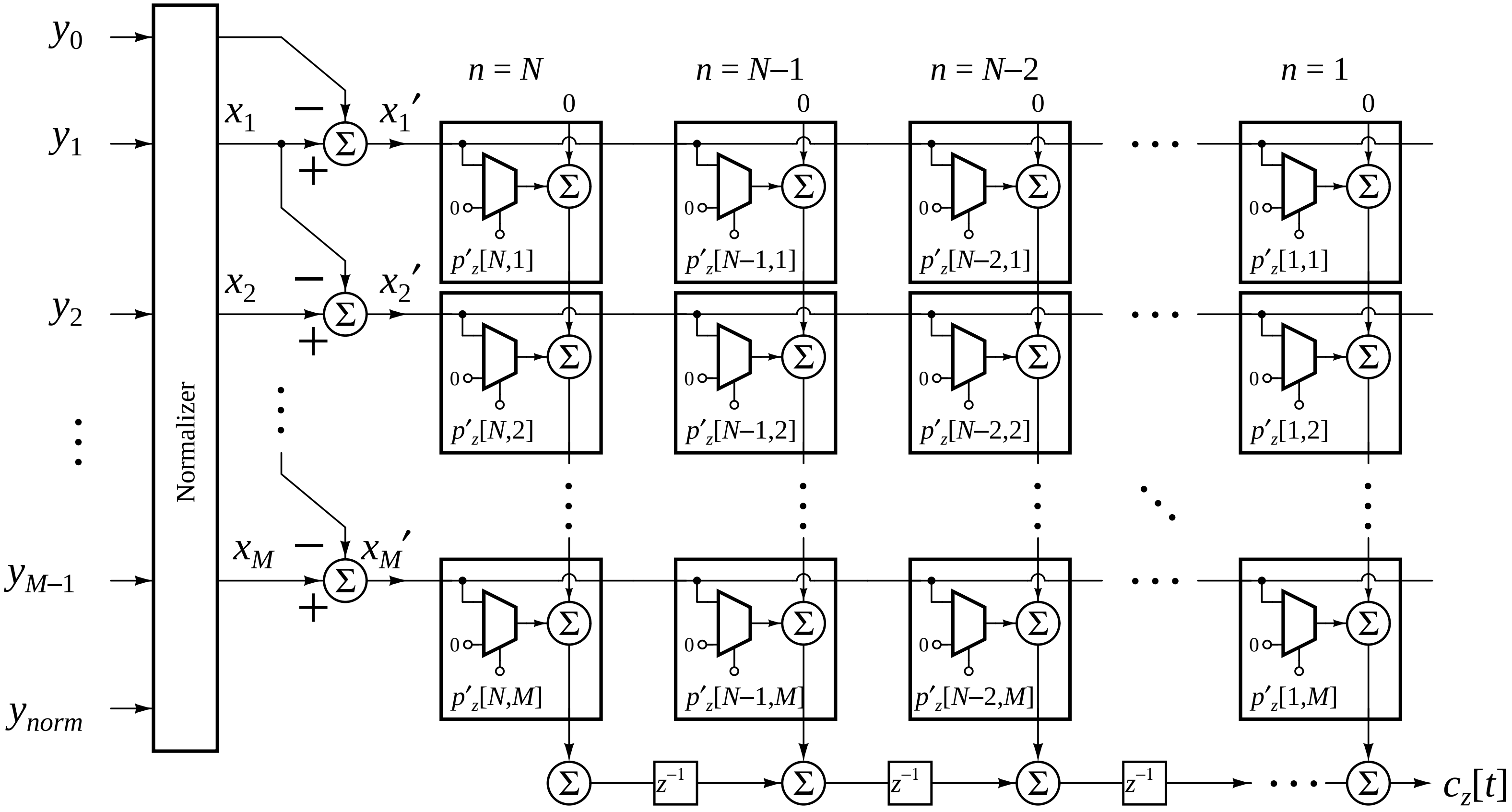
Statistics:

Total instances presented: 221
Correct: 205
Incorrect: 16
Accuracy: 92.7%

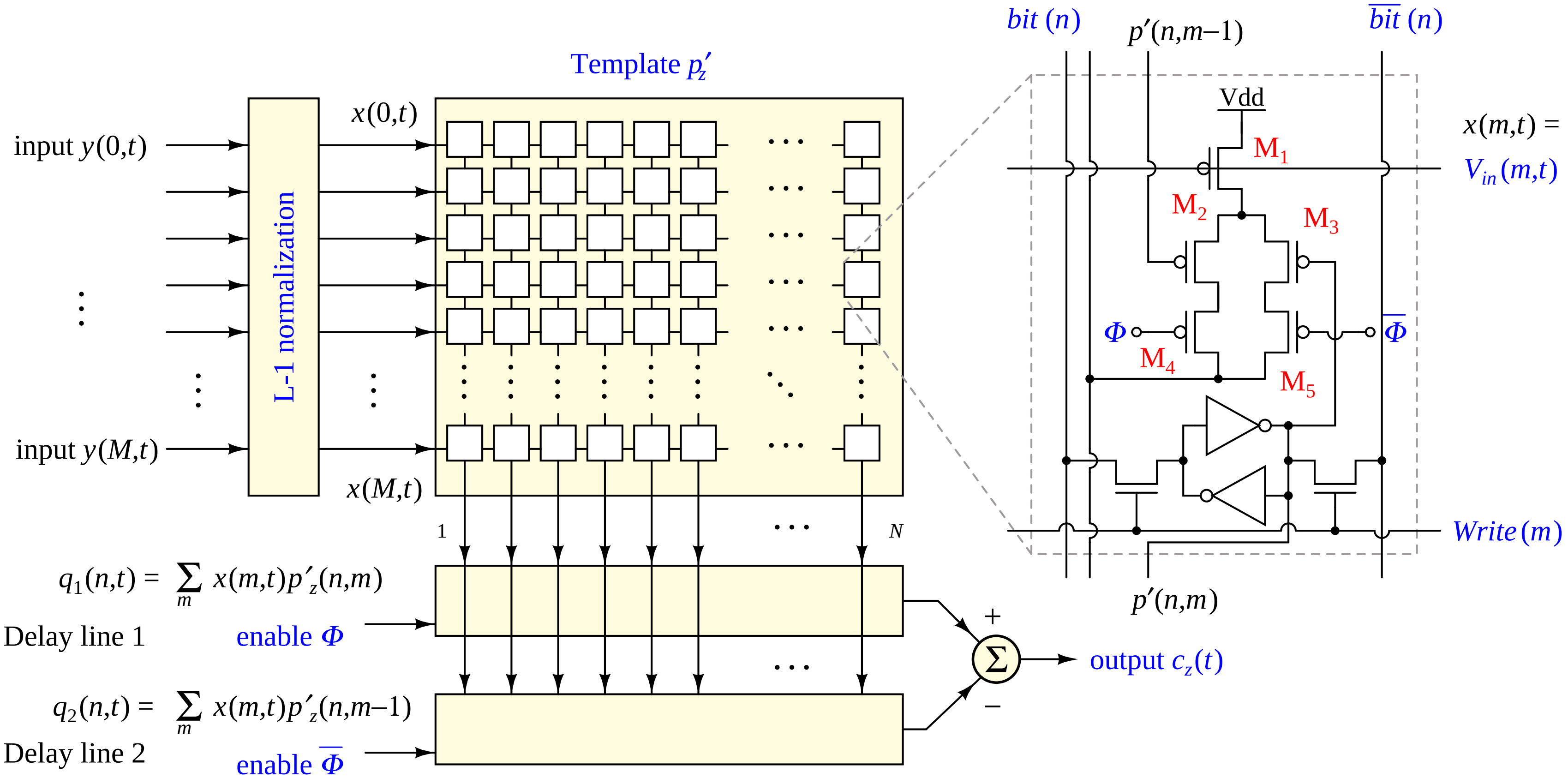
An Efficient Architecture for the Acoustic Transient Classifier

$$c_z[t] = \sum_{m=1}^{M} \sum_{n=1}^{N} \left(x[t-n,m] - x[t-n,m-1]\right) p'_z[n,m]$$

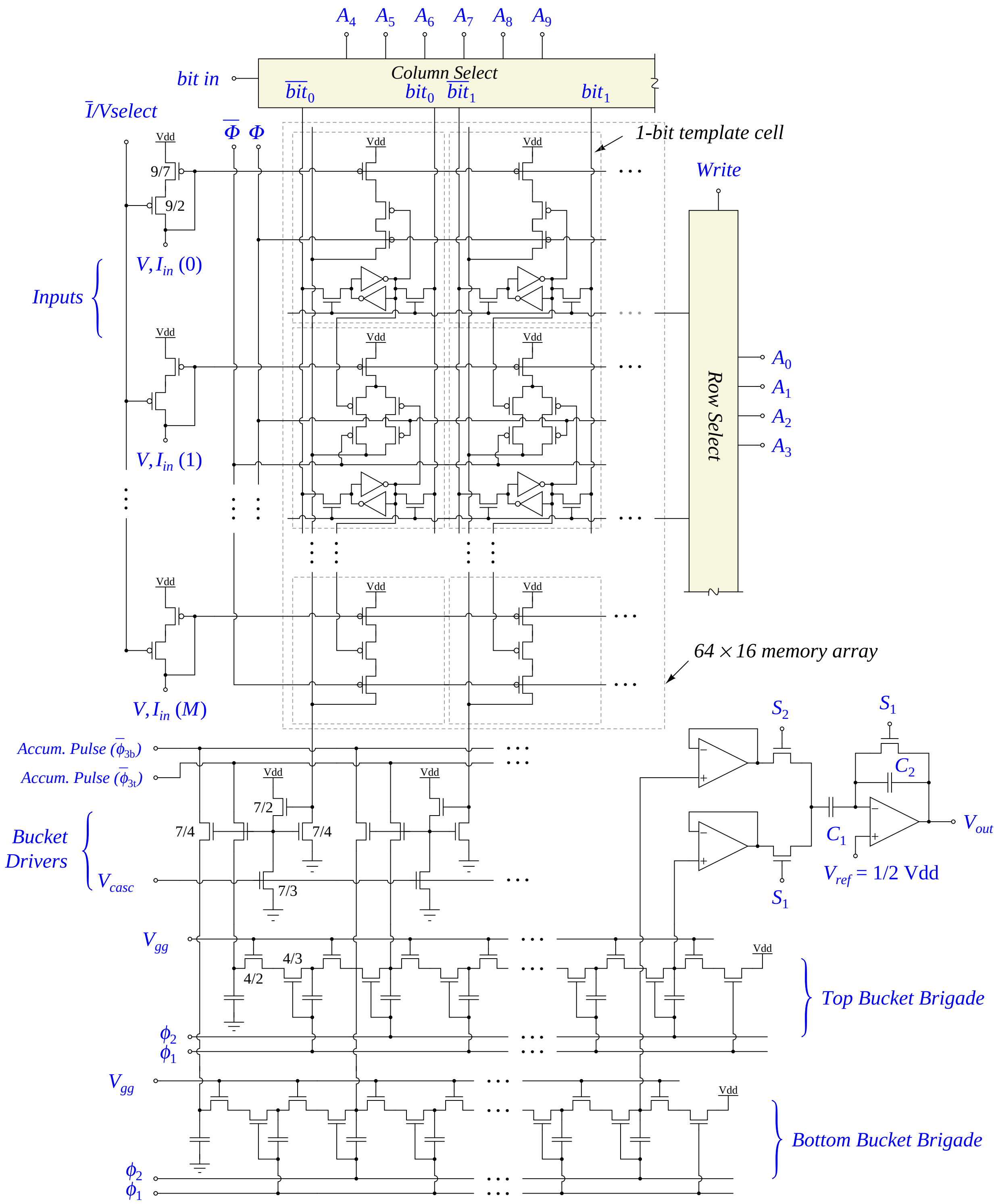
$$p'_z[n,m] = \begin{cases} 1 & (p_z[n,m] - p_z[n,m-1]) > 0 \\ 0 & (p_z[n,m] - p_z[n,m-1]) \leq 0 \end{cases}$$



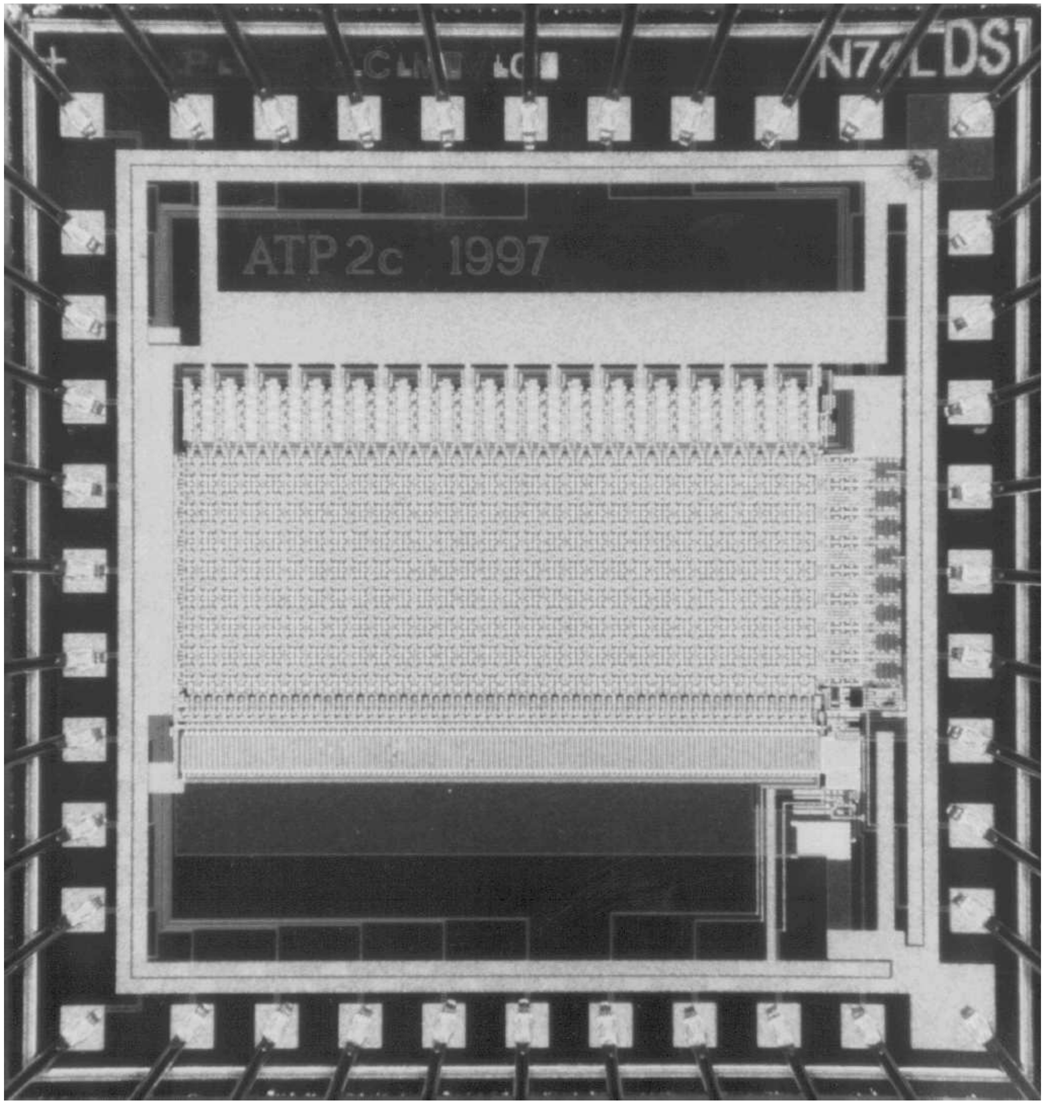
ATP Chip Architecture—Static Memory Core Cell



ATP—Complete Chip Architecture

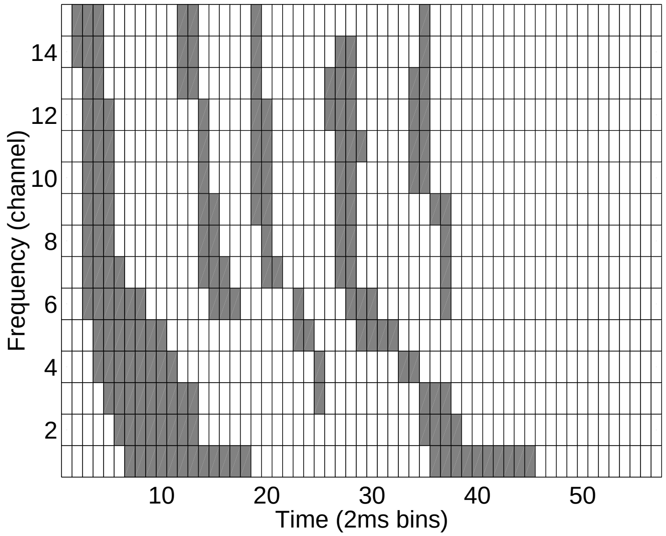


ATP—Photomicrograph



The acoustic transient correlator, a 2.2 mm × 2.2 mm die fabricated in a 1.2µm CMOS technology.

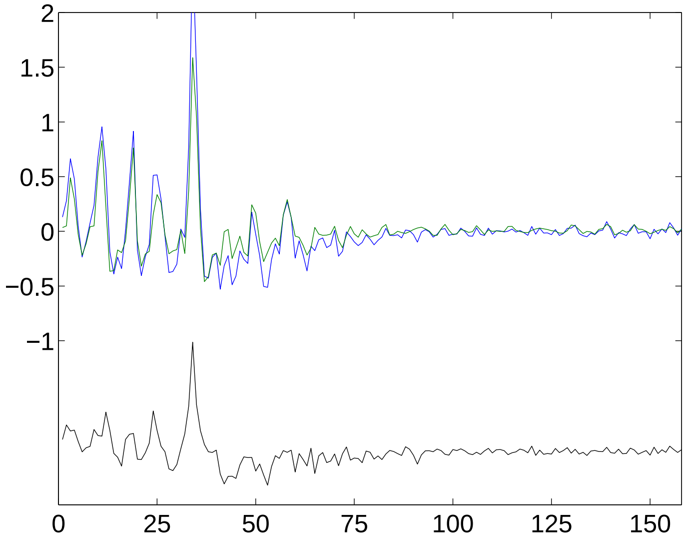
ATP Chip—Results

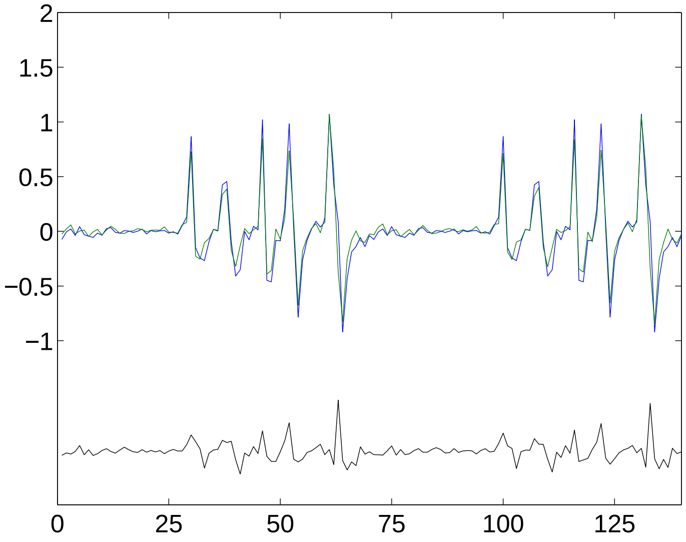


Simple template (left) detects the transient “can” (bottom left) but rejects the transient “snap” (bottom right).

Blue: numerical simulation.
Green: measured chip response.
Black: Residual error.

Chip measured power consumption: 50 μW





Time-Frequency Acoustic Processing and Recognition:
Analysis and Analog VLSI Implementation

- Introduction
- Template Correlation
- An Acoustic Transient Processor in VLSI:
  - Frontend Filterbank Systems
  - Template Correlation
  - Experimental Results
- A Trinary-Trinary Correlator
- Conclusion
