Real-time Low-power VLSI Microsystem for Smart Acoustic Interfaces

Milutin Stanaćević
Stony Brook University
www.ece.stonybrook.edu/~milutin
milutin.stanacevic@stonybrook.edu
Outline

- **Motivation**
  - Audio interfaces
  - Directional hearing

- **Gradient Flow**
  - Principle: analysis of acoustic field using miniature arrays
  - Independent Component Analysis and Separation

- **Mixed-Signal VLSI Implementation**
  - Acoustic Localizer
  - ICA implementation for Multiple Source Separation
  - Reverberant Environment:
    - Cochlear Filter Bank
    - Low Form-Factor ICA implementation

- **Conclusion**
Motivation

• “Cocktail Party” Problem
  - The problem of source separation is intrinsically linked to the problem of source localization.
  - Resolving time and amplitude differences between observed sound waves is essential for effective localization and separation.

• Human Auditory System
  - The human auditory system performs remarkably well in segregating multiple streams of acoustic sources.
  - In relatively large animals where the distance between the ears is substantial relative to the wavelength of sound, interaural time, and intensity differences are large enough to be detectable by the central nervous system.
Audio Interfaces

- Far-field noise cancellation for multimedia applications
  - Siri, Google Now, Cortana

- Audience’s earSmart Technology
  - Noise suppression with support for far-field
  - Siri made possible

- Now in 8 of every 10 smartphones
- Hearing aids
Motivation

- **Directional hearing at sub-wavelength scale**
  - Parasitoid fly localizes sound-emitting target (cricket) by a beamforming acoustic sensor of dimensions a factor 100 smaller than the wavelength. 0.6mm ear separation comparing to 170mm of human ears.
  - Tympanal beamforming organ senses acoustic pressure gradient, rather than time delays, in the incoming wave.

Hearing Aid Directionality

- Two microphones allow for one null angle in directionality pattern
- Adaptive beamforming allows to steer the null to noise source
- Presence of multiple noise sources requires source localization and separation with multiple microphones

www.oticon.com
Traveling Wave Source Localization

- Resolving time delays in wave propagation across array of low aperture requires sampling in excess of signal bandwidth, increasing power dissipation and noise bandwidth.
Gradient Flow Localization

- Resolving time delays in wave propagation across array of low aperture requires sampling in excess of signal bandwidth, increasing power dissipation and noise bandwidth.
- Gradient flow obtains time delays at sub-sampling resolution by relating spatial and temporal differentials of the field across the array.
Gradient Flow Localization and Separation

- Resolving time delays in wave propagation across array of low aperture requires sampling in excess of signal bandwidth, increasing power dissipation and noise bandwidth.
- Gradient flow obtains time delays at sub-sampling resolution by relating spatial and temporal differentials of the field across the array.
Separation and Localization
G. Cauwenberghs, M. Stanaćević and G. Zweig, U.S. Patent 6,865,490

Source mixtures are observed with additive sensor noise:

\[ x_{pq}(t) = \sum_{\ell=1}^{L} s^{\ell}(t + \tau^{\ell}_{pq}) + n_{pq}(t) \]

Gradient flow reduces to a static (noisy) mixture problem:

\[
\begin{bmatrix}
\dot{\xi}_{00} \\
\xi_{10} \\
\xi_{01}
\end{bmatrix}
= 
\begin{bmatrix}
1 & \cdots & 1 \\
\tau^1 & \cdots & \tau^L \\
\tau^1 & \cdots & \tau^L
\end{bmatrix}
\begin{bmatrix}
\dot{s}^1(t) \\
\vdots \\
\dot{s}^L(t)
\end{bmatrix}
+ 
\begin{bmatrix}
\dot{\nu}_{00} \\
\nu_{10} \\
\nu_{01}
\end{bmatrix}
\]

\[ x = A \text{s} + n \]

observations
(directions) \hspace{1cm} A \hspace{1cm} s
sources
(time-differentiated) \hspace{1cm} n

UNKNOWN \hspace{1cm} UNKNOWN

direction vectors
noise
(gradients)
Independent Component Analysis

- Blind source separation: separate and recover independent sources from mixed sensor observations.
- The sources and the mixing medium are assumed unknown.

\[
\begin{align*}
\text{Source signals} & \quad s(t) \quad \rightarrow \quad A \
\text{Sensor observations} & \quad x(t) \\
\text{Mixing model} & \\
\text{Sensor observations} & \quad y(t) \quad \rightarrow \quad W \
\text{Reconstructed source signals} & \\
\text{Solution (unmixing)} & \\
\text{Feedforward} & \\
(\text{Feedback}) & \\
\end{align*}
\]

\[
\begin{align*}
x &= A s \\
\text{if } A \text{ is a constant matrix: linear static ICA problem} \\
y &= W x \\
(W + I) y &= x
\end{align*}
\]
Independent Component Analysis

- Independent component analysis (ICA) adaptively minimizes higher-order statistical dependencies between reconstructed signals to blindly estimate the unmixing matrix.

Unmixing model:  \[
(W + I)y = x
\]

Update rule:  \[
\Delta W = \mu f(y)g(y^T)
\]

- Jutten-Herault (1991)

- InfoMax (Bell and Sejnowski, 1995)

- Natural gradient (Amari 1996, …)

- and many extensions…
Differential Sensitivity

- Gradient flow bearing resolution is fundamentally independent of aperture
  - Assumes interference noise dominates sensor/acquisition noise

- In practice, aperture is limited by differential sensitivity in gradient acquisition
  - Enhanced through differential coupling and common-mode suppression
Aperture Selection

The beampattern for the impinging 8 kHz acoustic source as a function of the unit grid size (0.5cm, 1cm, 2cm) of the microphone array. The canceled interfering directional source signals originate from azimuth angles of $\theta_2 = -30^\circ$ and $\theta_3 = 100^\circ$.

The noise sensitivity, the transfer function of the signal-to-noise ratio from the observed signals to the integrated reconstructed sources, shown for three different apertures of the microphone array, 0.5cm, 1cm and 2cm.
Separation Performance

S. Li and M. Stanaćević, "Analysis of Gradient Flow Technique for Blind Source Localization and Separation using Miniature Microphone Arrays" TASSP (in review)

The separation of the first speech source signal from the two directional interfering speech signals as a function of incidence angle $\theta_1$ for two different signal-to-noise ratios at the microphones. The incidence angles of the second and third source are $\theta_2 = 150^\circ$ and $\theta_3 = 165^\circ$.

Photo of the printed circuit board with the microphone array and signal conditioning circuitry.
Mixed-Signal VLSI Implementation: Single Source Localization

Analog inputs

Spatial gradients

Estimated delays

\[ \xi_{10} \approx \tau_1 \xi_{00} \]

\[ \xi_{01} \approx \tau_2 \xi_{00} \]
Mixed-Signal VLSI Implementation: Multiple Source Separation and Localization

Spatial gradients

Estimated delays

\[
\begin{bmatrix}
\dot{\xi}_{00} \\
\dot{\xi}_{10} \\
\dot{\xi}_{01}
\end{bmatrix} =
\begin{bmatrix}
1 & 1 & 1 \\
\tau_1^1 & \tau_2^2 & \tau_3^3 \\
\tau_1^1 & \tau_2^2 & \tau_3^3
\end{bmatrix}
\begin{bmatrix}
\dot{s}_1^1(t) \\
\dot{s}_2^2(t) \\
\dot{s}_3^3(t)
\end{bmatrix} +
\begin{bmatrix}
\dot{v}_{00} \\
\dot{v}_{10} \\
\dot{v}_{01}
\end{bmatrix}
\]
Spatial Gradient Computation

Analog inputs

\[ \dot{\xi}_{00} = \frac{1}{4} (x_{-1,0} + x_{1,0} + x_{0,-1} + x_{0,1}) \]
\[ \dot{\xi}_{10} = \frac{1}{2} (x_{1,0} - x_{-1,0}) \]
\[ \dot{\xi}_{01} = \frac{1}{2} (x_{0,1} - x_{0,-1}) \]

Temporal derivative of average and estimated first-order spatial gradients
CDS Differential Sensing

Switched-capacitor, discrete-time analog signal processing
- Correlated Double Sampling (CDS)
  - Offset cancellation and $1/f$ noise reduction
- Fully differential
  - Clock and supply feedthrough rejection

\[
\frac{d}{dt} \left[ \begin{array}{c} + \bigcirc \bigcirc \bigcirc \bigcirc + \\
+ \bigcirc \bigcirc \bigcirc \bigcirc + \\
- \bigcirc \bigcirc \bigcirc \bigcirc + \\
- \bigcirc \bigcirc \bigcirc \bigcirc + \\
\end{array} \right]
\]
First-order spatial gradient computation

\[ \xi_{10}^+[n] = x_{10}[n - \frac{1}{2}] - x_{-10}[n] \]

\[ \xi_{10}^-[n] = x_{-10}[n - \frac{1}{2}] - x_{10}[n] \]
Adaptive Common-Mode Suppression

Systematic common-mode error in finite-difference gradients:
due to gain mismatch across sensors in the array:

\[ \hat{\xi}_{00} \approx \frac{1}{4} (x_{-1,0} + x_{1,0} + x_{0,-1} + x_{0,1}) \approx \xi_{00} \approx \sum_\ell s^\ell (t) \]

\[ \hat{\xi}_{10} \approx \frac{1}{2} (x_{1,0} - x_{-1,0}) \approx \xi_{10} + \varepsilon_1 \xi_{00} \approx \sum_\ell \tau_1^\ell \hat{s}^\ell (t) + \varepsilon_1 \sum_\ell s^\ell (t) \]

\[ \hat{\xi}_{01} \approx \frac{1}{2} (x_{0,1} - x_{0,-1}) \approx \xi_{01} + \varepsilon_2 \xi_{00} \approx \sum_\ell \tau_2^\ell \hat{s}^\ell (t) + \varepsilon_2 \sum_\ell s^\ell (t) \]

can be eliminated using second order statistics only:

\[ E[\hat{s}^\ell (t)s^m (t)] = 0, \quad \forall \ell, m \quad \Rightarrow \quad \begin{cases} E[\xi_{00}\xi_{10}] = 0 \\ E[\xi_{00}\xi_{01}] = 0 \end{cases} \]

Adaptive LMS calibration:

\[ \hat{\xi}_{10} \approx \hat{\xi}_{10} - \frac{E[\hat{\xi}_{00}\hat{\xi}_{10}]}{E[\hat{\xi}_{00}^2]} \hat{\xi}_{00} \]

\[ \hat{\xi}_{01} \approx \hat{\xi}_{01} - \frac{E[\hat{\xi}_{00}\hat{\xi}_{01}]}{E[\hat{\xi}_{00}^2]} \hat{\xi}_{00} \]
Spatial Gradient Computation with Common-mode Suppression

Analog inputs

Average, temporal derivative and estimated spatial gradients

Spatial gradients with suppressed common-mode
Source Localization: Mixed-Signal LMS Adaptation

\[ \xi_{10} \approx \tau_1 \dot{\xi}_0 \]

- **Sign-sign LMS differential on-line adaptation rule**
  - Delay parameter estimation:
    \[ e_{10}^+[n] = \xi_{10}^+[n] - (\tau_1^+ \xi_{00}^+[n] + \tau_1^- \xi_{00}^-[n]) \]
    \[ e_{10}^-[n] = \xi_{10}^-[n] - (\tau_1^- \xi_{00}^-[n] + \tau_1^+ \xi_{00}^+[n]) \]
    \[ \tau_1^+[n+1] = \tau_1^+[n] + \text{sgn}(e_{10}^+[n] - e_{10}^-[n]) \text{sgn}(\xi_{00}^+[n] - \xi_{00}^-[n]) \]
    \[ \tau_1^- = 2^n - 1 - \tau_1^+ \]

- **Digital storage and update of parameter estimates**
  - 12-bit counter
  - 8-bit multiplying DAC to construct LMS error signal
Mixed-signal LMS Implementation

SS-LMS adaptation

12-bit estimated parameter

Multiplying DACs

16 x 16 DAC array

16 x 16 DAC array

12-bit Counter

d<11:8> d<7:4> d<3:0>

Bin2Thermocoder Bin2Thermocoder

b0 b0 b1 b15 b15 b0 b0 b1 b15 b15

\( \phi_2^+ \phi_1^+ V_{\text{ref}} \cdots \phi_2^- \phi_1^- V_{\text{ref}} \cdots \)

\( 16C \quad C \quad V_0^+ \quad V_0^- \)
Common-mode compensated spatial gradient

\[ \hat{x}_{10}^+ [n] = x_{10} [n - \frac{1}{2}] - x_{-10} [n] \]
\[ \hat{x}_{10}^- [n] = x_{-10} [n - \frac{1}{2}] - x_{10} [n] \]

Multiplying DAC for common-mode compensation

Uncompensated first-order spatial gradient computation along p-direction

T-cell attenuates output swing

Stony Brook University

Integrated Microsystems Lab

CAS 2016
Gradient Flow Localizer


- Digital LMS adaptive 3-D bearing estimation
  - Analog microphone inputs
  - Digital bearing outputs
  - Analog gradient outputs
- 3 mm x 3 mm in 0.5 μm 3M2P CMOS
- 8-bit effective digital resolution
  - 2 μs at 2 kHz
  - 0.25 μs at 16 kHz
- Power dissipation
  - 32 μW at 2 kHz
  - 54 μW at 16 kHz

Stony Brook University
Localization Performance

- One directional source in room environment: *Band-limited (100-1000Hz) Gaussian signal presented through loudspeaker*
  - distance between microphones: 2.5 cm
  - distance between loudspeaker and sensor array: 1 m
  - sampling frequency: 16 kHz
  - SNR around 35 dB

Source localization with
- 5 degree increments

Source localization with
- 0.5 degree increments
Open-field Vehicle Tracking

- **Acoustic Surveillance Unit (ASU) with integrated GradFlow ASIC**
- **Aberdeen Proving Grounds field tests**
  - Sensor network with 3 ASUs
  - 5 degree bearing accuracy in tracking ground vehicles over 600m range
  - Detection of overflying aircraft, tracking azimuth & elevation
ICA Implementation

- **HJ learning rule**

\[(W + I)y = x \quad y = (W + I)^{-1} x \approx (I - W)x\]

\[\Delta W = \mu f(y)g(y^T)\]

\[\Delta W = \begin{cases} -\mu f(y_i)g(y_j) & i \neq j \\ 0 & i = j \end{cases}\]

- **NG learning rule**

\[y = Wx\]

\[\Delta W = \mu (I - f(y)y^T)W = \mu W - \mu f^*(y)z^T \quad z = W^T y\]

\[\Delta W_{ij} = \mu w_{ij} - \mu f(y_i)z_j\]
ICA System Diagram

- Digitally reconfigurable ICA update rule

\[ \Delta w_{ij} = \mu w_{ij} - \mu f(y_i)z_j \]

- \( f(y_i) = \text{sign}(y_i) \), and 3-level quantization of \( z_j \)
ICA SC implementation

- Digital storage and update of weight coefficients
  - 14-bit counter
  - 8-bit multiplying DAC to construct output signal
ICA VLSI Processor


- 3 inputs – sensor signals or gradient flow signals
- 3 outputs – estimated sources
- 14-bit digital estimates of unmixing coefficients
- 3mm x 3mm in 0.5μm CMOS
- 192μW power consumption at 16kHz
Gradient Flow Source Separation

Time waveforms and spectrograms of spatial gradients, input signals to the ICA processor, and reconstructed source signals, output signals of the ICA processor for the case of two source signals with azimuth angles of $\theta_1 = 30^\circ$ and $\theta_2 = 105^\circ$.

SIR for two reconstructed speech sources played through a loudspeakers to a miniature microphone array in a conference room. The azimuth angle of the first source was fixed at $\theta_1 = 30^\circ$ and the azimuth angle of the second source $\theta_2$ was varied from $45^\circ$ to $135^\circ$. The elevation angles $\phi_1$ and $\phi_2$ for both sources were $8^\circ$. 
Separation in Reverberant Environment

• In a reverberant environment, the multi-path wave propagation contributes delayed mixture components to the observations.

\[ x_i(t) = \sum_{j=1}^{N} h_{ij} \ast s_j(t) \]

• Frequency domain techniques are attractive for solving the convolutive mixtures, as the convolution becomes product in the frequency domain.

\[ X_i(\omega) = \sum_{j=1}^{N} H_{ij}(\omega)S_j(\omega) \]

• The mixing matrix are dependent on the frequency, but can be considered constant in a narrow band.

• So we propose the subband ICA architecture that decomposes gradients into subbands and applies static ICA separately in each frequency band.
Subband ICA Separation Architecture

1. Localization results from full-band static ICA are obtained as directional cues.

2. 16-channel filterbank is used to decompose the spatial gradient signals.

3. Static ICA algorithm is used in each frequency band to obtain the unmixing matrix and signal estimation.

4. Signals from each band is aligned and synthesized to reconstruct full-band estimations.

Block diagram of the proposed subband gradient flow ICA architecture.
Subband ICA

• **Subband deposition with 16 filterbanks in mel scale (0-1kHz in linear scale, 1kHz-8kHz in log scale)**

• **Initial Condition**
  - Use fullband unmixing matrix as initial condition in each bin to ensure fast and robust convergence

• **Permutation Ambiguity**
  - Directional information in each band is compared with fullband localization to align estimations from the same source
  - If directional information strongly deviates from the fullband result in a certain band, then fullband unmixing matrix overwrites

• **Scaling Ambiguity**
  - Scaling resolver from the estimated mixing matrix is reapplied to ensure the uniformity of amplitude across all frequency bands
Subband Gradient Flow Source Separation


<table>
<thead>
<tr>
<th></th>
<th>RT&lt;sub&gt;60&lt;/sub&gt; = 200ms</th>
<th>RT&lt;sub&gt;60&lt;/sub&gt; = 300ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT&lt;sub&gt;60&lt;/sub&gt;</td>
<td>200ms</td>
<td>300ms</td>
</tr>
<tr>
<td>SIR&lt;sub&gt;1&lt;/sub&gt;</td>
<td>25.30dB</td>
<td>26.91dB</td>
</tr>
<tr>
<td>SIR&lt;sub&gt;2&lt;/sub&gt;</td>
<td>28.99dB</td>
<td>10.12dB</td>
</tr>
<tr>
<td>Static ICA</td>
<td>6.95dB</td>
<td>12.80dB</td>
</tr>
<tr>
<td>Subband ICA</td>
<td>24.65dB</td>
<td>15.75dB</td>
</tr>
</tbody>
</table>
Subband Decomposition: VLSI Implementation

- 16 channels from 150Hz to 10kHz (from 150Hz to 1kHz 8 channels in linear space and from 1kHz to 10kHz 8 channels in logarithmic space)
- High linearity and wide dynamic range are required due to the linear mixing model
- Small chip area
- Low power consumption
Band-pass Filter Structure

• Gm-C filter with capacitive attenuation, attenuation ratio: $A_1=19$, $A_3=4$
• Gain = 1, $Q = 4$, $|V_x| = |V_y|$
• Fully differential OTA, $G_{m1}=G_{m2}=4G_{m3}$
Band-pass Filter Structure

- Gm-C filter with capacitive attenuation, attenuation ratio: $A_1=19$, $A_3=4$
- Gain = 1, $Q = 4$, $|V_x| = |V_y|$  
- Fully differential OTA, $G_{m1}=G_{m2}=4G_{m3}$
Band-pass Filter Structure

- Gm-C filter with capacitive attenuation, attenuation ratio: $A_1=19$, $A_3=4$
- Gain = 1, $Q = 4$, $|V_x| = |V_y|$
- Fully differential OTA, $G_{m1}=G_{m2}=4G_{m3}$
OTA Structure

- Source degeneration
- Current division
- \( M_M \) is \( M \) times wider than \( M_1 \)
- Tuning method
  - fix capacitor value to keep SNR stable
  - \( M_R \) overdrive voltage
  - channel length of \( M_R \)
  - current division ratio \( M \)

\[
G_m = \frac{g_{dsMR}}{g_{mM1}} \frac{g_{dsMR}}{1 + (M + 1)} \approx \frac{g_{dsMR}}{1 + M}
\]

when \( \frac{g_{dsMR}}{g_{mM1}} \ll (M + 1) \)
OTA Simulation

- Simulation environment: Cadence
- MR overdrive voltage: 350mV
- Input peak to peak voltage: 50mV to 300mV
- $\text{HD}_3=-55\text{dB} @ 100\text{mV V}_{pp}$
Band-pass Filter Simulation

- Center frequency: 150Hz
- Input signal peak to peak amplitude: 1.92V
- HD3 = -59.6dB
- SNR = 59.9dB
Cochlear Filter Bank


- 0.5µm 3M2P CMOS technology
- 16 channels filter bank and 1 test channel
- Single channel area: 0.16mm²
- Power consumption: 375 µW
ICA Circuit Architecture

• Digitally reconfigurable natural gradient ICA update rule

\[ \Delta w_{ij} = \mu w_{ij} - \mu f(y_i)z_j \]

• \( f(y_i) = \text{sign}(y_i) \), and 3-level quantization of \( z_j \) time encoded

• \(<x_i>\) and \(<y_i>\) denote the pulse-width modulated signal of the input signal \( x_i \) and the output signal \( y_i \).
ICA Update Rule Adaptation

- The 3-level staircase function $q(z)$ is approximated with the presence/absence of the voltage pulse and by the relative position of the pulse.
- The function $f(y)$ is coded as a two-level signal, with the $\text{sign}(y)$ determining the order of the levels $V_{lo}$ and $V_{hi}$.
- The resulting change on the capacitor is given by

$$V_{ij}^{+}[n + 1] = V_{ij}^{+}[n] + \frac{C_p}{C_w + C_p} (V_{Ai j}^{+}[n] - V_{ij}^{+}[n])$$
Vector-Matrix Multiplication

\[ y = Wx \quad z = W^T y \]

• The vector-matrix multiplications are implemented by integrating switched currents controlled by a PWM signal.

• To minimize the chip area, the two multiplications and the quantization of the signal \( z \) are implemented in three phases using the same the integration and voltage-to-time conversion circuitry.

• A comparison of the decreasing voltage ramp at the output node of the amplifier with a reference voltage \( V_{comp} \) generates a pulsed signal with a width proportional to the input.
Quantization

- As the pulse-width modulated output signals $< y_{i+} >$ and $< y_{i-} >$ are available, with a single D-latch the sign of the $y_i$ is determined.
- To generate the quantized signal $q(z_j)$, a comparison with a positive and a negative threshold voltage $V_{th}$ is required.

- The comparison with the threshold voltage is performed by delaying one of these pulses before the connection to the input of the D-latch.
- Voltage $V_b$ controls the threshold voltage by controlling the delay time.
Simulation Results

- The adaptation cell is simulated with a constant sign of update.
- The incremental values of the unmixing coefficient as the current of transistor $M_0$ are shown.
- The output voltage of the integrator $y_1$ is shown for three different values of the unmixing coefficient $w_{11}$ while the other current sources representing unmixing coefficients are switched off.
- Measured linearity of matrix-vector multiplication is 0.05%.
ICA VLSI Processor


• Technology 0.5 µm 2P3M CMOS
• Area: 0.49 mm²
• Supply: 5V
• Power dissipation: 80µW
• Separation: 8dB – 13dB
Conclusions

• Wave gradient “flow” converts the problem to that of static ICA, with unmixing coefficients yielding the direction cosines of the sources.

• The technique works for arrays of dimensions smaller than the shortest wavelength in the sources.

• Localization and separation performance is independent of aperture, provided that differential sensitivity be large enough so that ambient interference noise dominates acquisition error noise.

• High resolution delay estimation for source localization using miniature sensor arrays and blind separation of mixed signals with reconfigurable adaptation, have been experimentally demonstrated.

• System allows integration with sensor array for small, compact, battery-operated “smart” sensor applications in surveillance and hearing aids.
Acknowledgment

• IML students Shuo Li (now with Second Sight), Yingkan Lin (now with OmniVision) and Xiao Yun (now with Synopsys).

• This work was supported by the National Science Foundation CAREER Award 0846265.

• The chips were fabricated through the MOSIS foundry service.