Low-power System-on-chip Acoustic Localizer

Shuo Li, Xiao Yun and Milutin Stanacevic
Department of Electrical and Computer Engineering
Stony Brook University
Stony Brook, NY 11794–2350
Email: shuoli,xyun,milutin@ece.sunysb.edu

***Abstract*— We present the design of system-on-chip acoustic localizer chip that can be directly integrated with miniature MEMS microphone array. Continuous-time bandpass filtering and amplification at interface with microphone array enable low-power and small-form factor implementation. Time delays between acoustic signals observed over a planar geometry of four microphones are obtained by relating spatial and temporal differentials and estimated through mixed-signal least-square digital adaptation. Designed $3\ \mathrm{mm} \times 3\ \mathrm{mm}$ chip in 0.5 μm CMOS technology operates with estimated 150 μW power dissipation at 16 kHz sampling rate.**

I. Introduction

Precise and robust localization and tracking of acoustic sources using miniature microphone arrays in small form-factor and low power is of interest to a variety of applications, from hearing aids to surveillance and multimedia. There have been a number of VLSI systems for acoustic direction finding reported in the literature [1]–[3], but not a complete system-on-chip solution that can be integrated directly with microphone array. Gradient flow [3] is a signal conditioning technique for source localization designed for arrays of very small aperture, *i.e.,* of dimensions significantly smaller than the shortest wavelength in the sources. Consider a traveling acoustic wave impinging on an array of four microphones, in the configuration of Figure 1. The 3-D direction cosines of the traveling wave $\mathbf{u}$ are implied by propagation delays τ_1 and τ_2 in the source along directions p and q in the sensor plane. Direct measurement of these delays is problematic as they require sampling in excess of the bandwidth of the signal, increasing noise floor and power requirements. However, indirect estimates of the delays are obtained, to first order, by relating spatial and temporal derivatives of the acoustic field [3]:

$$\begin{aligned} \xi_{10}(t) &\approx \tau_1 \dot{\xi}_{00}(t) \\ \xi_{01}(t) &\approx \tau_2 \dot{\xi}_{00}(t) \end{aligned} \quad (1)$$

where ξ_{10} and ξ_{01} represent spatial gradients in p and q directions around the origin ($p = q = 0$), ξ_{00} the spatial common mode, and $\dot{\xi}_{00}$ its time derivative. From least-square estimates of the time delays in (1), we directly obtain bearing estimates of azimuth angle θ and elevation angle ϕ of $\mathbf{u}$ relative to p and q as

$$\begin{aligned} \tau_1 &= \frac{1}{c}|\mathbf{r}_1| \ \cos\theta \sin\phi \\ \tau_2 &= \frac{1}{c}|\mathbf{r}_2| \ \sin\theta \sin\phi, \end{aligned} \quad (2)$$

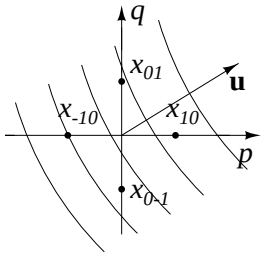

Fig. 1. Configuration of sensors for spatial gradient estimation.

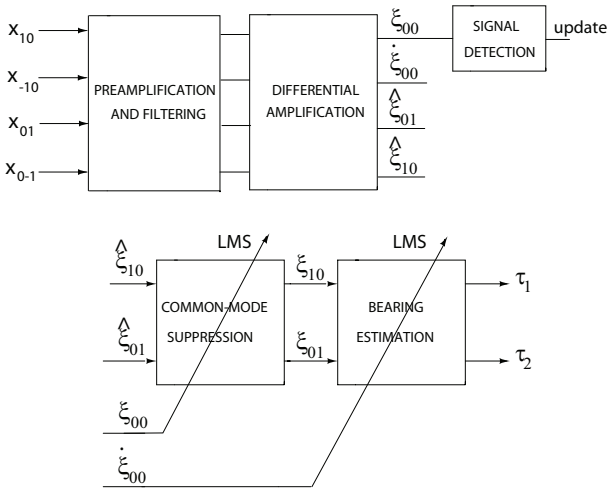

Fig. 2. System block diagram.

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are unit orthogonal vectors along p and q directions.

The system block diagram implementing gradient flow for bearing estimation is shown in Figure 2. Filtering and preamplification stage enables direct interface to microphone array and design of low-power system-on-chip matched for hearing aid applications. Spatial gradients are approximated by evaluating finite differences over the four sensors on the planar grid shown in Figure 1. Two stages of mixed-signal adaptation compensate for common mode errors in the differential amplification, and produce digital estimates of delays τ_1 and τ_2 from the spatial and temporal differentials. Signal detection, beside saving in power consumption, improves bearing estimation convergence speed and stability.

II. Circuit implementation

Gradient flow lends itself into efficient implementation in mixed-signal VLSI. The continuous-time $G_m - C$ circuits provide low-power and small-form factor implementation for filtering and preamplification of microphone signals. Switched-capacitor circuits are ideal match for low-noise and low-frequency spatial gradient computation. Digital adaptation for bearing estimation offers flexibility and reconfigurability of the learning rules with estimates in digital format, avoiding need for high-resolution analog-to-digital conversion.

A. Bandpass Filtering and Preamplification of Microphone Signals

The proposed system-on-chip is intended for direct interface with microphone array in order to achieve small-form factor and improve power consumption. Performance of the gradient flow algorithm depends on spatial gradient signal acquisition, which is a small differential signal on large common-mode pedestal. Improved differential sensitivity of gradient sensing allows to shrink the aperture of the sensor array without degrading signal-to-noise ratio. The microphone signals have to be amplified and filtered before the spatial gradient computation. Instead of using discrete implementation of bandpass filter, that covers the bandwidth of the signal of interest, for better matching and improved power consumption, we propose bandpass filter on chip implementation as continuous time filter. Mismatch in amplitude and phase characteristics of filters in different channels leads to leakage of differential signal into common-mode signal and degrades performance of gradient flow algorithm.

The filtering of microphone signals is matched to the speech signal and implemented as second-order bandpass filtering, with low-frequency cutoff set at 100 Hz and high-frequency cutoff set at 8 kHz. The signals are also amplified by a factor of 10. A continuous time $G_m - C$ filter implements the bandpass filtering and preamplification, Design is shown in Figure 3. The transfer function of the filter is

$$H(s) = \frac{G_{m3}}{G_{m2}} \frac{\frac{G_{m1}}{C_1}s + 1}{(\frac{G_{m2}}{C_2}s + 1)^2} \tag{3}$$

G_{m1} and G_{m3} are nominally set to equal values. G_{m1}/C_1 sets the low-frequency cutoff. Since this frequency is small, a large value of capacitance C_1 is required and this capacitance is implemented as off-chip capacitor. The ratio of G_{m2} and G_{m3} sets the gain of the filter, while the ratio of G_{m2}/C_2 sets the high-frequency cutoff. The implementation of OTA used in proposed $G_m - C$ filter implementation is shown in Figure 4.

B. Spatial Gradient Calculation

Estimates of ξ_{00}, ξ_{10} and ξ_{01} are obtained from the sensor observations $x_{-1,0}$, $x_{1,0}$, $x_{0,-1}$ and $x_{0,1}$ as:

$$\begin{aligned} \xi_{00} &\approx \tfrac{1}{4}\big(x_{-1,0} + x_{1,0} + x_{0,-1} + x_{0,1}\big) \\ \xi_{10} &\approx \tfrac{1}{2}\big(x_{1,0} - x_{-1,0}\big) \\ \xi_{01} &\approx \tfrac{1}{2}\big(x_{0,1} - x_{0,-1}\big) \end{aligned} \tag{4}$$

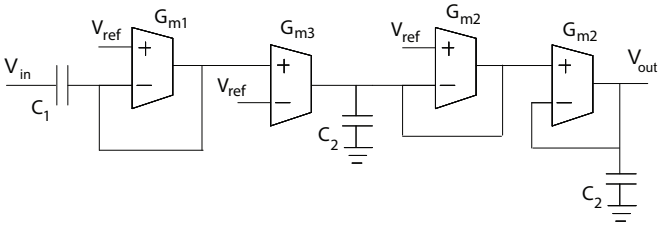

Fig. 3. Continuous-time bandpass filter and pre-amplifier implementation.

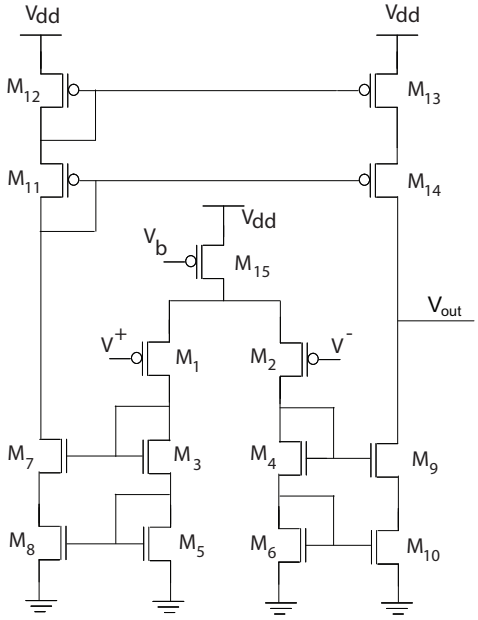

Fig. 4. OTA implementation.

Computation of the gradients is implemented using sampled-data switched-capacitor (SC) circuits. The advantage of this realization is application of correlated-double sampling (CDS) to significantly reduce common-mode offsets and $1/f$ noise [4]. A cascoded inverter, biased in subthreshold regime, is used as high-gain amplifier in these and subsequent SC circuits, supporting high density of integration, and high energetic efficiency.

The common-mode component is decomposed in differential form $\xi_{00} = \xi_{00}^{+}[n] - \xi_{00}^{-}[n]$ with

$$\begin{aligned} \xi_{00}^{+}[n] &= \frac{1}{8}(x_{10}[n - \frac{1}{2}] + x_{-10}[n - \frac{1}{2}] \\ &\quad + x_{01}[n - \frac{1}{2}] + x_{0-1}[n - \frac{1}{2}]) \\ \xi_{00}^{-}[n] &= -\frac{1}{8}(x_{10}[n] + x_{-10}[n] + x_{01}[n] + x_{0-1}[n]). \end{aligned} \tag{5}$$

The contribution ξ_{00}^{+} to ξ_{00} represents the estimate of the average signal at time instance $nT - \frac{T}{2}$, while the contribution ξ_{00}^{-} represents the inverted estimate at time instance nT. The difference between both contributions signals hence produces an unbiased estimate of ξ_{00} centered at time $nT - \frac{T}{4}$. The corresponding switched-capacitor realization is given in Figure 5. Instead of using a differential operational amplifier

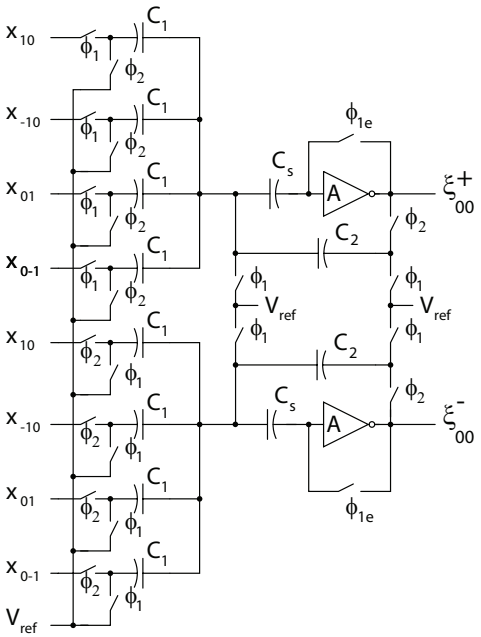

Fig. 5. Pseudo-differential structure for computation of common-mode component.

as high gain element in SC circuit, we have chosen to use a lower power, single ended inverting amplifier in pseudo-differential implementation. In this implementation, the offset between virtual grounds of two branches directly contributes to the differential signal. In the previously reported structure [3], the virtual ground was defined by the inverting point of the cascoded amplifier and was prone to mismatch errors. The proposed structure contains the virtual ground capacitor C_s, that defines the virtual ground. C_s is precharged to the mid-point voltage V_{ref} in clock phase ϕ_1, and voltage V_{ref} is the value of virtual ground in both branches, eliminating dependance on input transistor parameters. The clocks ϕ_1 and ϕ_2 are nonoverlapping, and ϕ_{1e} replicates ϕ_1 with its falling edge slightly preceding the falling edge of ϕ_1.

An estimate of the temporal derivative signal $\dot{\xi}_{00}$ centered at same time instance $nT - \frac{T}{4}$ is computed differentially in similar manner, by differencing signal averages (or the reference level) at time instances $nT - \frac{T}{2}$ and nT,

$$\begin{aligned}\dot{\xi}^{+}_{00}[n] &= 0\\ \dot{\xi}^{-}_{00}[n] &= -\frac{1}{8}(x_{10}[n]+x_{-10}[n]+x_{01}[n]+x_{0-1}[n])\\ &\quad +x_{10}[n-\frac{1}{2}]+x_{-10}[n-\frac{1}{2}]\\ &\quad +x_{01}[n-\frac{1}{2}]+x_{0-1}[n-\frac{1}{2}].\end{aligned} \tag{6}$$

The spatial gradients are computed in fully differential mode, to provide increased clock and supply feedthrough rejection. The first-order spatial gradient ξ_{10}, likewise centered at time $nT - \frac{T}{4}$, is computed by differencing estimates of ξ_{10} at time instances $nT - \frac{T}{2}$ and nT

$$\begin{aligned}\xi^{+}_{10}[n] &= \frac{1}{4}(x_{10}[n-\frac{1}{2}]-x_{-10}[n])\\ \xi^{-}_{10}[n] &= \frac{1}{4}(x_{-10}[n-\frac{1}{2}]-x_{10}[n]).\end{aligned} \tag{7}$$

The first-order spatial gradient in the q direction, ξ_{01}, is computed in identical fashion.

C. Common-mode suppression

Gain mismatch in the microphones and preamplification stage leads to contribution of common mode signal into the spatial gradient estimate. To estimate the leakage coefficient ε_1 of common-mode signal into spatial gradient estimation, a digital sign-sign LMS (SS-LMS) adaptation rule is used [3]. ε_1 is stored as digital value in a 12-bit counter and it is represented in two's complement. The update is performed by incrementing or decrementing the counter based on sign of spatial gradient and average signal

$$\begin{aligned}\varepsilon^{+}_{1}[n+1] &= \varepsilon^{+}_{1}[n]\\ &\quad +\mathrm{sgn}(\xi^{+}_{10}[n]-\xi^{-}_{10}[n])\mathrm{sgn}(\xi^{+}_{00}[n]-\xi^{-}_{00}[n])\\ \varepsilon^{-}_{1}[n+1] &= 2^{12}-1-\varepsilon^{+}_{1}[n+1].\end{aligned} \tag{8}$$

The 8 most significant bits are presented to a multiplying D/A capacitor array to construct the LMS error signal, for the case of ξ_{10}

$$\begin{aligned}\xi^{+}_{10}[n] &= \hat{\xi}^{+}_{10}[n]-(\varepsilon^{+}_{1}\xi^{+}_{00}[n]+\varepsilon^{-}_{1}\xi^{-}_{00}[n])\\ \xi^{-}_{10}[n] &= \hat{\xi}^{-}_{10}[n]-(\varepsilon^{-}_{1}\xi^{+}_{00}[n]+\varepsilon^{+}_{1}\xi^{-}_{00}[n]).\end{aligned} \tag{9}$$

D. Bearing estimation

For implementation of bearing time-delay estimation, digital SS-LMS differential on-line adaptation is used similar to common-mode error correction. Bearing estimates are represented as 12-bit values in two's complement [3]

$$\begin{aligned}\tau^{+}_{1}[n+1] &= \tau^{+}_{1}[n]\\ &\quad +\mathrm{sgn}(e^{+}_{10}[n]-e^{-}_{10}[n])\mathrm{sgn}(\xi^{+}_{10}[n]-\xi^{-}_{10}[n])\\ \tau^{-}_{1}[n+1] &= 2^{12}-1-\tau^{+}_{1}[n+1],\end{aligned} \tag{10}$$

with the 8 most significant bits used for computation of LMS error signal

$$\begin{aligned}e^{+}_{10}[n] &= \xi^{+}_{10}[n]-(\tau^{+}_{1}\dot{\xi}^{+}_{00}[n]+\tau^{-}_{1}\dot{\xi}^{-}_{00}[n])\\ e^{-}_{10}[n] &= \xi^{-}_{10}[n]-(\tau^{-}_{1}\dot{\xi}^{+}_{00}[n]+\tau^{+}_{1}\dot{\xi}^{-}_{00}[n]).\end{aligned} \tag{11}$$

E. Signal Detection

For acoustic localization, detection of the presence of the signal is critical for bearing estimation process, specially in the case of speech like signals, that have a long pauses in the signal. It is also important for power saving, as certain parts of the chip can be shut-off if the signal is not present. The simple thresholding function of the common-mode signal is chosen for signal detection. As the sign of common-mode signal is already necessary for update computation in common-mode suppression, the thresholding function is incorporated in the same comparator structure. Implementation of comparator

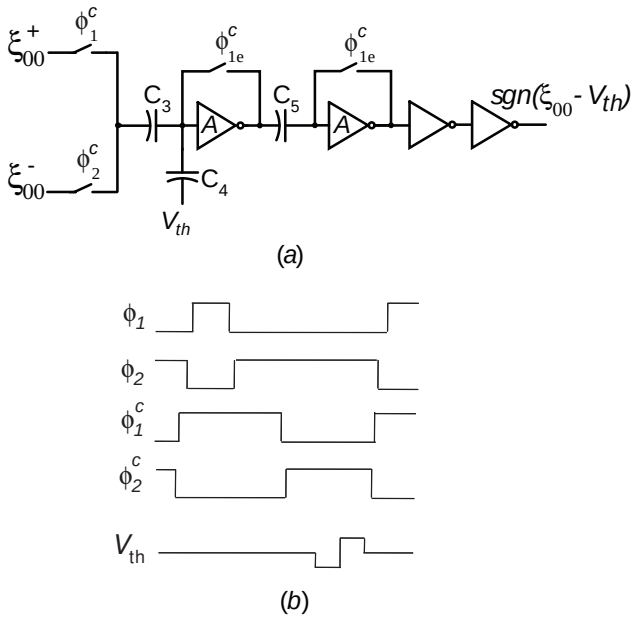

Fig. 6. (a) Comparator design for sign and threshold comparison. (b) Illustration of relative position of clock signals and variable comparison level V_{th}.

that is able to compare signal with variable level, zero for the sign and positive and negative thresholding value, is shown in Figure 6(a). While the output signals are valid, ξ_{00}^{+} is sampled in phase ϕ_1^c on capacitor C_3. The sign of the comparison of the common-mode signal ξ_{00} with variable comparison level V_{th} is computed in the evaluate phase ϕ_2^c, through capacitive coupling into the amplifier input node. Sign of ξ_{00} would be computed in standard comparator implementation if capacitor C_4 was omitted. The addition of C_4 allows the computation of multiple level comparisons in a single cycle. Sign of ξ_{00} is computed first for reference level of V_{th} and the comparison with threshold level is enabled by the change in voltage V_th in phase ϕ_2^c in a single clock cycle. The clocks ϕ_1^c and ϕ_2^c are non-overlapping and their relative timing with respect to clocks ϕ_1 and ϕ_2 is shown in Figure 6(b), as well as the changes in voltage V_{th} for the sign and threshold comparison. ϕ_{1e}^c replicates ϕ_1^c with its falling edge slightly preceding the falling edge of ϕ_1^c.

III. Simulation results

The proposed architecture was implemented in 0.5 μm CMOS technology. Frequency response of the designed bandpass filter and preamplifier is shown in Figure 7. The simulated dynamic range of implemented OTA is 51 dB. The bearing estimation process is demonstrated using sine wave input signal at frequency of 1 kHz, with sampling frequency of 16 kHz. The signal was presented to x_{10} and x_{01}, and under 5 μs delay to x_{-10} and x_{0-1}. The update of time-delay estimation process is shown in Figure 8. The total power consumption of the proposed system-on-chip is estimated at total of 150 μW, where 100 μW represents the power consumption of the bandpass filter, while 50 μW represents

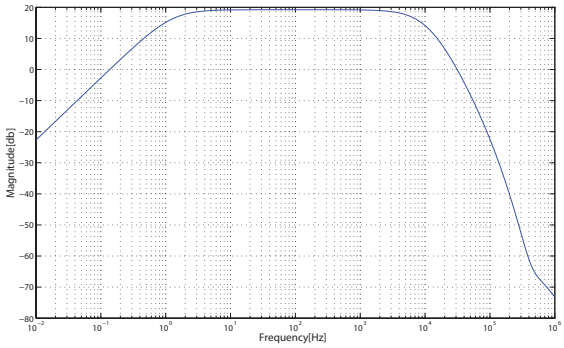

Fig. 7. Frequency response of the designed bandpass filter and preamplifier.

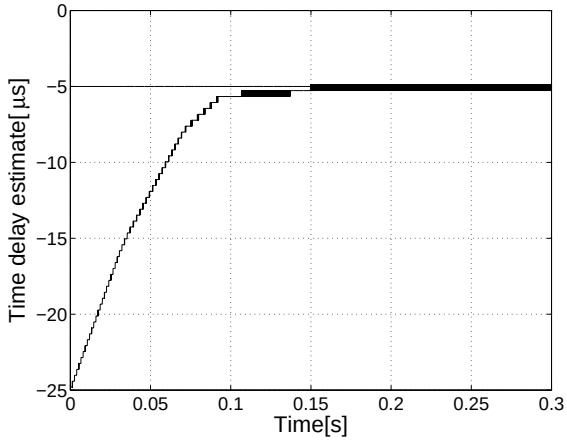

Fig. 8. Bearing estimation adaptation for estimation of delay between two sine wave input signals.

the power consumption of the spatial gradient computation, signal detection and bearing estimation.

IV. Conclusion

The proposed system-on-chip solution for acoustic localization is implemented in small-form factor and operates at microwatt power consumption, making it suitable for direct integration with microphone array. The chip well matches requirements for hearing aid and acoustic surveillance applications.

References

[1] J. Lazzaro and C.A. Mead, "A silicon model of auditory localization", *Neural Computation*, vol. **1**, pp. 47-57, 1989.

[2] P. Julian, A.G. Andreou, P. Mandolesi and D. Goldberg, "A low-power CMOS integrated circuit for bearing estimation", *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS'2003)*, Bangkok, Thailand, 2003.

[3] M. Stanacevic and G. Cauwenberghs, "Micropower Gradient Flow VLSI Acoustic Localizer", *IEEE Transactions on Circuits and Systems I : Regular Papers*, vol. **52** (10), pp. 2148 2157, 2005.

[4] C.C. Enz and G.C. Temes, "Circuit Techniques for Reducing the Effects of Op-Amp Imperfections: Autozeroing, Correlated Double Sampling, and Chopper Stabilization", *IEEE Proceedings*, vol. **84** (11), pp. 1584-1614, 1996.