A 78.2nW 3-Channel Time-Delay-to-Digital Converter using Polarity Coincidence for Audio-based Object Localization

Daniel de Godoy, Xiaofan Jiang and Peter R. Kinget
Department of Electrical Engineering, Columbia University, New York, NY-10027, USA
{dd2697, xiaofan.jiang, peter.kinget}@columbia.edu

***Abstract*—This paper presents an ultra-low-power intersignal time-delay-to-digital converter. It introduces polarity-coincidence adaptive time-delay estimation, a mixed-signal processing technique that consumes only 78.2nW for a 3-channel delay estimation. A 0.18um CMOS implementation of the converter has been tested and characterized with controlled and real-life stimuli. This analog-to-feature converter has further been used to estimate the time difference of arrival in an audio-based vehicle-bearing IoT system.**

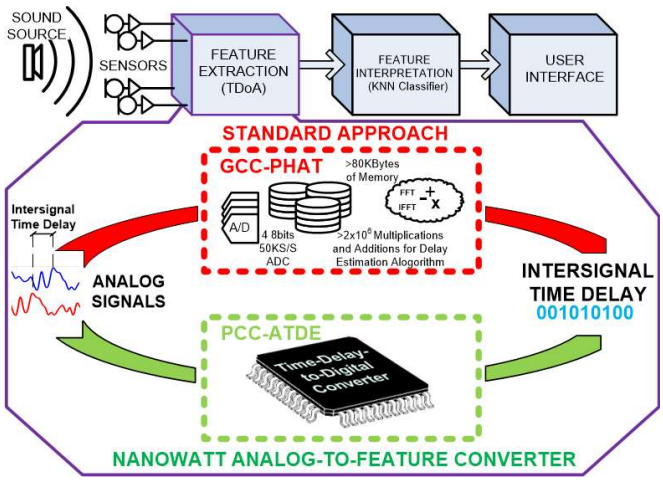
I. Introduction

Audio-based IoT systems promise to provide ultra-low-power solutions for medical, industrial, and societal problems [1]. Acoustic waves carry both spectral and spatial information that can be used to identify and localize sound sources. Most sound-source-localization (SSL) systems use time difference of arrival (TDoA), a.k.a. interaural time difference (ITD) in bio-inspired approaches. Time-delay estimation (TDE) blocks that extract the TDoA from microphone signals typically add significantly to the system power budget and computational requirements [2][6]. In this paper we aim to enhance the performance of the TDoA extraction not by improving the ADC or changing the digital-signal-processing algorithm, but by introducing an analog-to-feature converter that directly encodes the feature information from the analog inputs to digital bits (Fig. 1). Such an ultra-low-power intersignal-time-delay-to-digital converter is a critical block in vehicle aware safety systems for drones or cars [6].

Unlike for spectral features, the sampling frequency of the data converters defines the resolution of the TDE. To support a -1.5ms to 1.5ms delay range with 8-bit resolution for noise-like sources with dominant spectral components up to 250Hz (e.g., noise from approaching automobiles or other vehicles), the signal needs to be sampled at >50KS/s, or 100x the Nyquist rate. The storage of the frames and the arithmetic operations for a TDE using generalized cross-correlation (GCC) then become the bottleneck for a sub-µW implementation [2] as is needed in mobile, wearable, or IoT applications. The low complexity of adaptive time-delay estimation (ATDE) techniques, like the LMSTDE, makes them an attractive approach, but they still require a high-resolution ADC after the sensors. A bio-inspired silicon cochlea [3] estimates the time difference by translating the audio stimulus into asynchronous events, but its power consumption is still in the µW range.

This work was supported in part by the National Science Foundation within the NSF Division of Computer and Network Systems (CNS), Grant No. 1704899 .

We present a 78.2nW 50KS/s time-delay-to-digital converter with four audio input channels and three 8-bit delay outputs. The presented architecture does not require a multi-bit ADC, does not use memory blocks to store frames or intermediate results, and does not execute any computationally expensive algorithm.



II. Proposed Method

A. Working Principle and Analysis

We introduce the *polarity-coincidence correlation adaptive time-delay estimation* (PCC-ATDE) method (Fig. 2). PCC-ATDE, similar to GCC, estimates the delay between the inputs by finding the argument of the maximum value of the cross-correlation function. But, instead of calculating all the points of the cross-correlation function, the *PCC-ATDE uses a negative feedback loop to constantly track the argument of the peak*.

Let's assume that the delay difference between two inputs is D. The analog input signals from the microphones are first connected to a zero-crossing comparator acting as a 1-bit ADC. Then each signal goes through a variable-delay cell τ_{var} controlled by the converter's output offset by τ_{offset} to avoid negative delays. The differential delay experienced by the inputs is defined as $\Delta := \tau_{var1} - \tau_{var2}$ (Fig. 2). Next, the two signals are multiplied to create V_{MIXER1}. The average value of V_{MIXER1} is the polarity-coincidence correlation (PCC) function with argument Δ, i.e. PCC(Δ). PCC is a well-known alternative

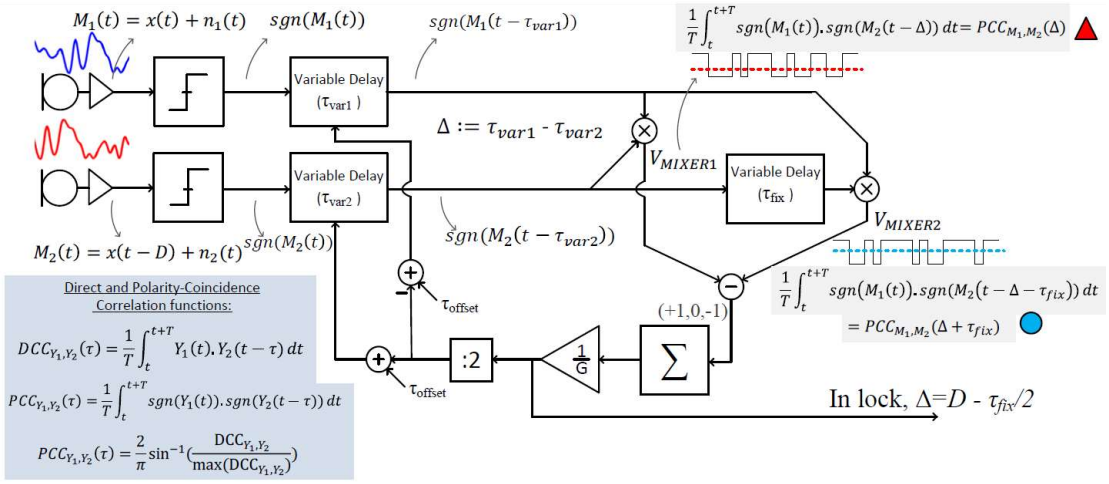


Fig. 2. Block diagram and analysis of the Polarity-Coincidence-Correlation Adaptive Time-Delay Estimation (PCC-ATDE).

to direct cross-correlation (DCC) for time delay estimation [4], and $PCC(\tau) = \frac{2}{\pi}\sin^{-1}(\frac{DCC(\tau)}{\max(DCC)})$ [5]. Hence, the argument of the max. of the PCC(τ) is the same as the argument of the max. of the DCC(τ). Note, however, that the computation of PCC only needs 1-bit signals, in contrast to DCC which requires multi-bit signals. The red triangles (Fig. 3) mark three possible cases of the mapping of the average of V_{MIXER1} onto the PCC(τ) for $\tau = \Delta$. The signal in the lower branch is further delayed by a fixed value, τ_{fix}, and then multiplied with the upper branch to create V_{MIXER2}. The average V_{MIXER2} can also be mapped onto the PCC(τ), now for $\tau = \Delta + \tau_{fix}$, shown as blue circles (Fig. 3). If the current Δ value is sufficiently close to the peak, the difference between PCC(Δ) and PCC$(\Delta+\tau_{fix})$ indicates whether Δ is smaller or greater than the argument of the peak. V_{MIXER1} and V_{MIXER2} are then subtracted and accumulated. The output of the accumulator will continuously increase or decrease until PCC(Δ) and PCC$(\Delta+\tau_{fix})$ have equal values, locking the loop at $\Delta = D - \frac{\tau_{fix}}{2}$, which is a measured of the desired intersignal delay. Note that the averages of the multiplier's output, V_{MIXER1} and V_{MIXER2}, do not need to be explicitly calculated since they arise from the infinite DC gain provided by the accumulator and the low-pass behavior of the loop.

The $\frac{1}{G}$ attenuator controls the speed of this convergence and bandwidth of the feedback loop. The output of the attenuator is split differentially to control the variable delay cells τ_{var}.The

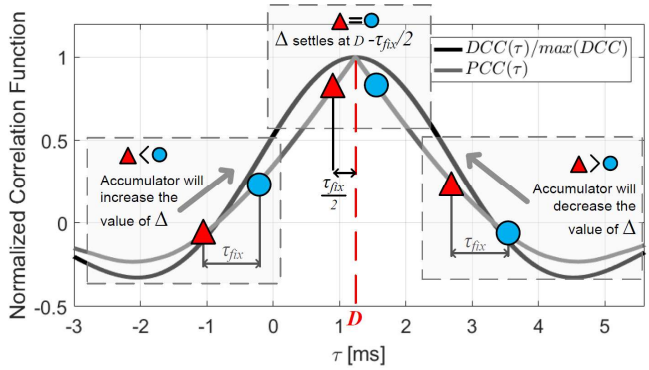


Fig. 3. Direct Cross-Correlation (DCC); Polarity-Coincidence Correlation (PCC); and 3 possible mapping of PCC-ATDE multiplier's output.

bandwidth of the PCC-ATDE loop must be kept smaller than the input signal's lowest frequency so that the continuous changes in the intersignal delay are still tracked in the PCC function. The first order nature of the feedback loop guarantees stability for the system.

B. Time-Discrete Implementation

Figure 4 shows the time-discrete implementation of the proposed PCC-ATDE. Four microphones interface with the chip; one channel is the reference, and the chip computes the delay of the three other signals w.r.t. this reference. The zero-crossing comparators (Fig. 4) do not consume power when in reset mode, leading to a total power consumption of 3.1nW per

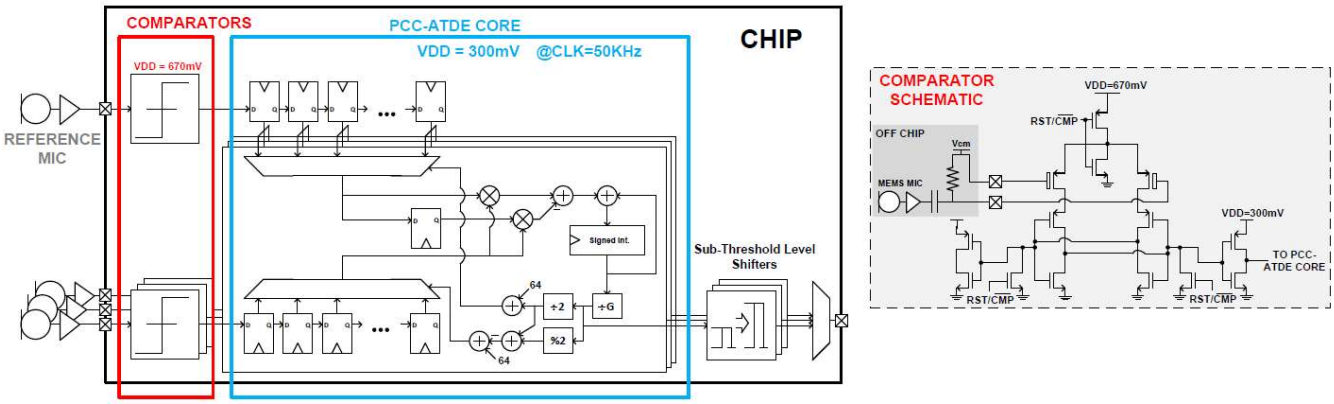


Fig. 4. Mixed-signal implementation of the Time-Delay-to-Digital Converter using PCC-ATDE;

comparator from 670mV when operated at 50KHz. Thick oxide transistors are used in the front end for ESD protection. The core of the PCC-ATDE was synthesized to work with sub-threshold CMOS logic [7]. The variable delay cells are implemented with chains of 128 flip-flops and multiplexers. The multiplications can be easily computed with XOR logic gates given the use of one-bit signals. An extra flip-flop in the lower path provides the fixed delay τ_{fix}. A 10-bit register and adder are used as an accumulator, and the $\frac{1}{G}$ attenuation is realized with arithmetical shifts of 0, 1 or 2 bits. The chip's core consumes 65.9nW from 300mV at 50KHz. Sub-threshold level shifters are used to convert the logic signals before sending them off-chip. The PCC-ATDE core together with the comparators occupy an area of 421µm x 391µm.

III. Measurements

A 3-channel time-delay-to-digital converter prototype has been fabricated in 180nm CMOS and characterized in three different setups: first on a lab test bench with controlled inputs for precise characterization; then with speaker and microphones in a closed room, where the performance was compared to a traditional TDE technique; and finally in a street deployment, where the prototype is tested under realistic dynamic stimuli.

A. Controlled Inputs using Arbritary Waveform Generators

A 600mVpp 60Hz-200Hz band-limited white noise signal is used as analog inputs to simulate the sound of approaching vehicles. The delay relative to the reference channel is carefully controlled in the AWG source for accurate measurements. The -1ms to 1ms step response varies from 514ms to 2.05s depending on the choice of G (Fig. 5). The ENOB varies from 5.41bits to 6.06bits. Static linearity tests have also been conducted (Fig. 6). The clocked nature of the algorithm helps it achieve a linear operation with peak INL of -1.57/1.33LSB and peak DNL of -0.85/97LSB with T_{LSB}=20µs without the need of calibration. Using an FOM for the performance of the PCC-ATDE similar to the one used to compare ADCs, namely $FOM = \frac{Power}{\#channels.2^{ENOB}.Fs}$, the system operating with a 50KHz clock reaches an FOM of 7.84fJ/conv.-step. It was further tested with clock frequencies from 10KHz to 800KHz (Fig. 7). Operating with higher clock frequencies reduces T_{LSB} and can be used to enable TDoA calculation in higher frequency applications like ultrasound.

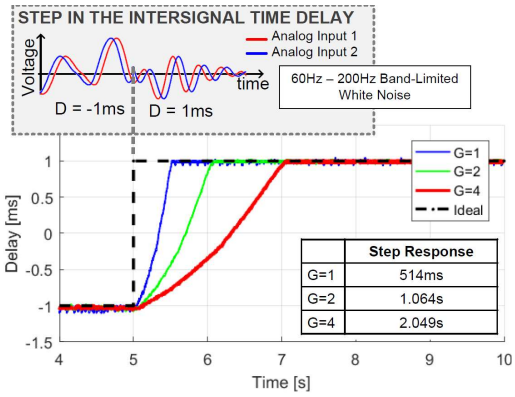


Fig. 5. -1ms to 1ms Step Response of the protype.

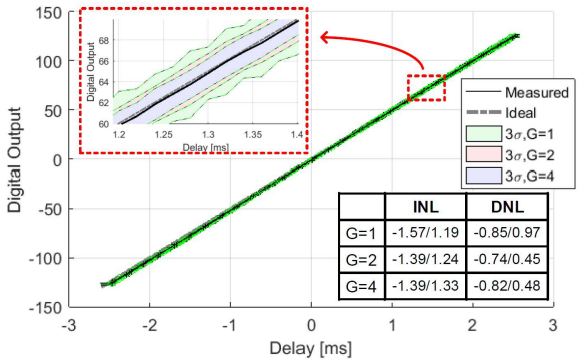


Fig. 6. Linearity Plot and peak INL and DNL Values.

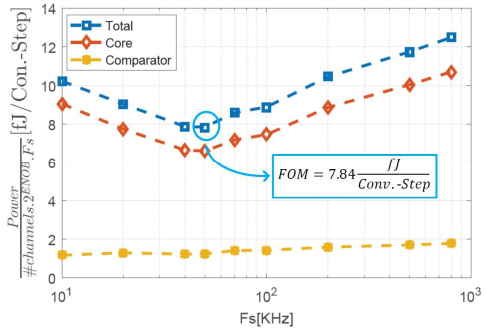


Fig. 7. Figure-of-Merit measurement over Sampling Frequency.

B. Audio Inputs in a Crontrolled Environment

Experiments with speakers and microphones were conducted to test the PCC-ATDE with real-life audio inputs, using recordings of band-limited white noise to simulate the noise from vehicles (Fig. 8). A single source rotating around a microphone pair was used to compare the PCC-ATDE to a

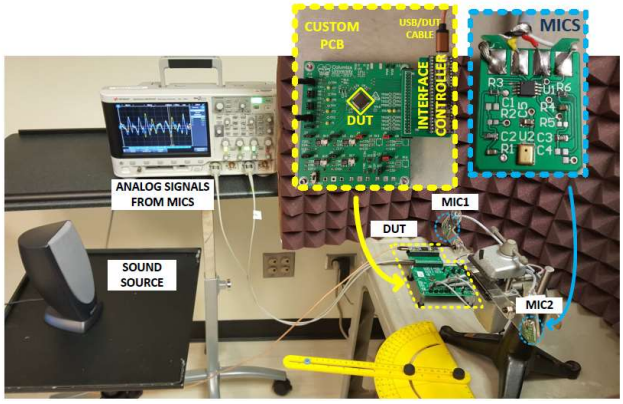


Fig. 8. In-Lab Measurement with Speakers and Microphones Setup.

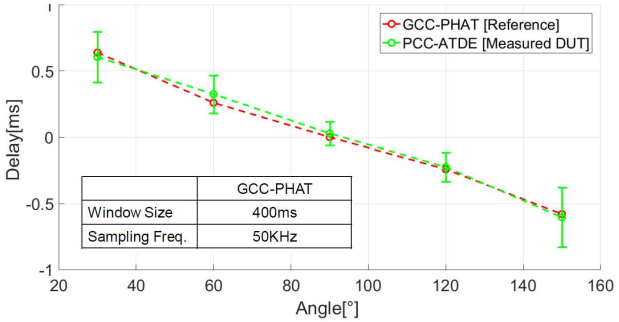


Fig. 9. Comparison of a standard TDE techinque (GCC-PAT) and the proposed PCC-ATDE.

standard GCC-PHAT approach (Fig.9). The experiment was conducted in a 4m by 6m closed room without acoustic isolation. The PCC-ATDE measurement matches the reference delay estimation closely, with a RMS error of 37.2µs (or 2.3%), demonstrating that the technique can be a reliable alternative to the GCC-PHAT of current system designs.

C. *Estimating the Bearing of an Approaching Vehicle*

To further validate the PCC-ATDE capability to extract the TDoA of moving vehicles, an experiment was conducted in the streets of New York City. A setup with 4 microphones was placed on the sidewalk and connected to the chip. As the cars crossed the setup the 4 analog inputs from the microphones were used to extract 3 intersignal delays and their values were sent to a host running a machine-learning K-Nearest Neighbor (KNN) classifier. The classifier was trained indoors to map the extracted TDoA onto incidence angles in real-time. A picture of the experiment setup is shown in Fig. 10. The extracted TDoA from a car crossing from right to left of the microphone array is plotted in Fig. 11. The output of the KNN classifier (Fig. 12) changes from 0° when the car is on the right-hand side of the microphone array to 180° after it crosses to the left.

IV. Comparison to the State of the Art

Cross-correlation based time-delay estimation are the conventional approach for TDoA extraction. In [2], a GCC-PHAT implementation was presented in 0.18um CMOS. This solution requires 20KBytes of memory, FFT, and iFFT cores, in contrast with the 257 DFF, 2 XORs and 1 accumulator required by the PCC-ATDE. As a result, the area of [2] is more than 6x larger than the PCC-ATDE, and the FOM in [2] 10^5x higher. In [3], a silicon cochlea presented can extract both spectral information and the intersignal delay from audio inputs, but its power consumption is 10^4x higher than the PCC-ATDE.

V. Conclusions

In this paper, an ultra-low-power analog-to-feature converter is presented to estimate the TDoA of audio signals for sound-source-localization systems. The PCC-ATE technique is presented which uses a negative-feedback architecture and 1-bit front-end ADCs to implement a 78.2nW 7.84fJ/conv.-step time-delay-to-digital converter prototype in 0.18um CMOS. In controlled experiments with audio inputs the PCC-ATDE measurements present a 37.2us RMS error to a reference GCC-PHAT TDE while consuming 100,000 times less energy. The prototype was integrated into a real-time vehicle bearing system to validate the applicability of the technique with real-world signals from urban areas.

Acknowledgment

Thanks to Prof. M. Seok and João Pedro Cerqueira from Columbia U. for their assistance with the ASIC integration.

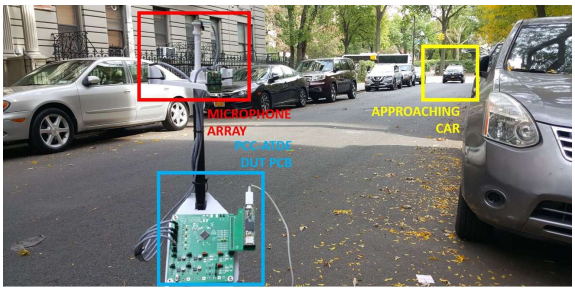


Fig. 10. Street deployment setup.

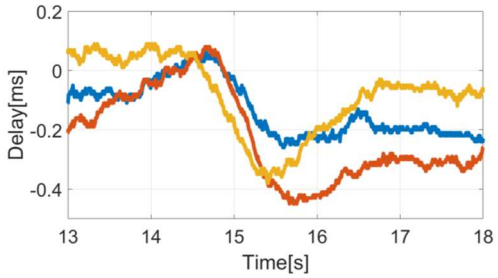


Fig. 11. Extracted Time Difference of Arrival of an approaching car.

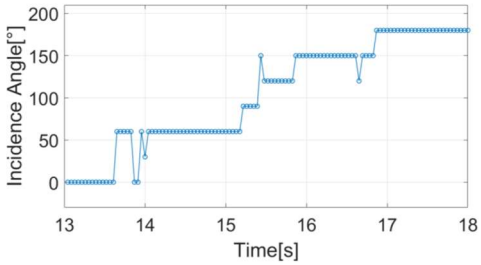


Fig. 12. Output of the KNN Classifier for the incidence angle

PERFORMANCE SUMARY	
Technology	180nm CMOS
Sampling Frequency	50KS/s
Range	[-2.52ms,2.52ms]
T_{LSB}	20µs
Peak INL	-1.57/1.33 LSB
Peak DNL	-0.85/0.97 LSB
TDE ENOB	6.06bits
Number of TDE Channels	3
Comparator Power (Unit)	3.1nW
PCC-ATDE Core Power	65.9nW
Total Power	78.2nW
TDE Energy per Conversion Step per Channel	7.84fJ/Conv.-Step
Active Area	0.28mm²
Die Area	1mm²

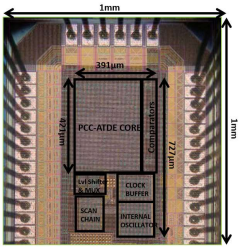


Fig. 13. Performance summary and die photo.

References

[1] S. Jeong et al., "A 12nW Always-On Acoustic Sensing and Object Recognition Microsystem Using Frequency-Domain Feature Extraction and SVM Classification." *ISSCC,* 2017

[2] D. Halupka et al., "Robust Sound Localization in 0.18 um CMOS," IEEE TSP, 2005

[3] M. Yang et al., "A 0.5V 55uW 64x2-Channel Binaural Silicon Cochlea for Event-Driven Stereo-Audio Sensing," ISSCC, 2016

[4] A. Fertner et al., "Comparison of Various Time Delay Estimation Methods by Computer Simulation," IEEE TASSP, 1986

[5] J. Vleck et al., "The spectrum of clipped noise," Proceedings of the IEEE, 1966

[6] R. Chandrasekaran et al., "SEUS: A Wearable Multi-Channel Acoustic Headset Platform to Improve Pedestrian Safety: Demo Abstract.", SenSys, 2016

[7] M. Seok et al., "Optimal technology selection for minimizing energy and variability in low voltage applications." ISLPED, 2008.