8th International IEEE EMBS Conference on Neural Engineering
Shanghai, China, May 25 - 28 , 2017

A Blind Adaptive Stimulation Artifact Rejection (ASAR) Engine for Closed-Loop Implantable Neuromodulation Systems

Sina Basir-Kazeruni, Stefan Vlaski, Hawraa Salami, Ali H. Sayed, and Dejan Marković

***Abstract*— In this work we propose an energy-efficient, implantable, real-time, blind Adaptive Stimulation Artifact Rejection (ASAR) engine. This enables concurrent neural stimulation and recording for state-of-the-art closed-loop neuromodulation systems. Two engines, implemented in 40nm CMOS, achieve convergence of <42µs for Spike ASAR and <167µs for LFP ASAR, and can attenuate artifacts up to 100mV$_{p\text{-}p}$ by 49.2dB, without any prior knowledge of the stimulation pulse. The LFP and Spike ASAR designs occupy an area of 0.197mm^2 and 0.209mm^2, and consume 1.73µW and 3.02µW, respectively at 0.644V.**

I. INTRODUCTION

Closed-loop stimulation is desired to build an effective neuromodulation system. This operation requires concurrent stimulation and recording of neural signals ($<1\,\mathrm{mV}$) in the presence of stimulation artifacts (10s of mV). Conventional systems are primarily concerned with mitigating saturation in neural recording front-ends, but do not allow for recording of signals of interest during stimulation. Recently, front-ends capable of recording Local Field Potentials (LFPs) and Action Potentials (Spikes) with up to $\pm 20\,\mathrm{mV}$ [1] and $\pm 50\,\mathrm{mV}$ [2] linear-input-range have been introduced. These designs overcome the saturation problem in conventional systems. However, large stimulation artifacts remain in the digitized signal. In this work we demonstrate an energy-efficient, implantable, real-time, and blind Adaptive Stimulation Artifact Rejection (ASAR) engine that enables closed-loop neural recording and stimulation, illustrated in Figure 1.

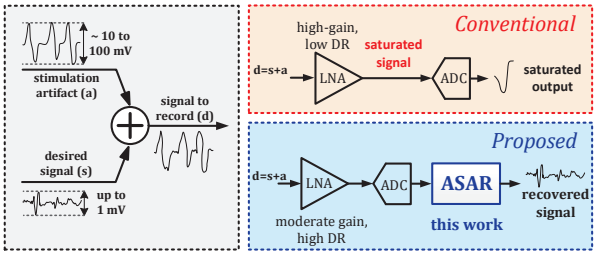

Fig. 1: Conventional versus proposed neuromodulation system.

Existing artifact rejection solutions lack the ability of continuous signal recording during stimulation, thus rendering a critical portion of the data unusable. In [3], an overload recovery technique is employed, where a shunt resistance is used to de-polarize the electrode immediately after stimulation ends. This technique enables quick recovery from saturation, but does not allow recording during stimulation. Self-canceling stimulation electrode configurations have been proposed [4], where recording electrodes are placed differentially around stimulation electrodes, and common-mode passive filters are used to attenuate the stimulation artifact. This solution is very application-specific, and severely restricts the allowed combinations of stimulation and recording sites. The self-canceling technique also requires that artifacts occupy frequencies outside the neural signal band, which is not the case in many practical situations. An adaptive-filtering based method was suggested in [5], [6], with performance that is insufficient for implantable closed-loop systems. The estimated artifact is canceled at the input of the amplifier using a DAC, hence introducing noise at the input of the front-end. This method requires very long convergence times (3–12 seconds), which is impractical since stimulation/recording sites, along with the structural and temporal characteristics of stimulation pulses, can change in real time. Lastly, the feedback nature of artifact subtraction limits the maximum delay between stimulation onset and when the artifact is observed. If the delay is any longer, the method could fail unless the stimulation is periodic.

The proposed ASAR engine is capable of adaptively removing artifacts for varying stimulation characteristics at multiple sites. A blind artifact template detection technique is introduced, eliminating the need for prior knowledge of the temporal and structural characteristics of the stimulation pulse. This technique enables us to avoid estimating the non-linearity through brain tissue and electrode interfaces, and achieves much faster convergence times ($<167\,\mu\mathrm{s}$) and noise power reduction of up to $49.2\,\mathrm{dB}$. Two designs, the Spike ASAR and LFP ASAR (Table I) have been implemented in $40\,\mathrm{nm}$ CMOS technology to demonstrate these capabilities.

TABLE I: Design specifications for two ASAR implementations.

Design Specifications	LFP ASAR	Spike ASAR
Signal(s) of interest	LFP	Spike + LFP
Bandwidth (Hz)	1-200	1-5K
Sampling rate (KS/s)	6	24
Input resolution	16bits	12bits
Artifact amplitude (mV$_{p\text{-}p}$)	<100	<40

II. PROPOSED BLIND ASAR SOLUTION

As shown in Figure 2, ASAR operates in two phases: (A) statistics calculation (training), and (B) template detection and adaptive filtering.

The authors are with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA. Emails: {sinabk, svlaski, hsalami, sayed, dejan}@ucla.edu.

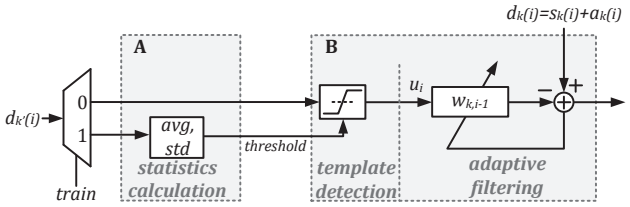

Fig. 2: Blind Adaptive Stimulation Artifact Rejection (ASAR) block diagram.

A. Statistics Calculation

Statistics of the neural signal, from an adjacent recording channel, are calculated in the absence of artifacts during the first N samples, and an appropriate threshold value is set. This is done by recursively updating the values $\boldsymbol{S}(i) = \boldsymbol{S}(i-1) + \boldsymbol{x}(i)$ and $\boldsymbol{T}(i) = \boldsymbol{T}(i-1) + \boldsymbol{x}^2(i)$, where $\boldsymbol{x}(i)$ is the input sample at time i. Mean (avg) and standard deviation (std) at time $i = N$ are then calculated as (for N large enough):

$$\text{avg} = \frac{\boldsymbol{S}(N)}{N} \tag{1}$$

$$\text{std} \approx \sqrt{\frac{1}{N-1}\left(\boldsymbol{T}(N) - N \cdot \text{avg}^2\right)} \tag{2}$$

The number of samples N is chosen as $N = 2^n$ for some n in order to reduce multiply/divide into shift operations, resulting in a more efficient hardware implementation.

B. Template detection and adaptive filtering

To clean the measurement $\boldsymbol{d}_k(i)$ at electrode k, we choose a nearby electrode k' and determine a template $\boldsymbol{u}_i \in \mathbb{R}^{1\times 16}$. Based on the threshold obtained in the previous phase, $\boldsymbol{u}_i(\ell)$, the ℓ-th element of $\boldsymbol{u}_i$ is estimated from $\boldsymbol{d}_{k'}(i)$ through blanking within $\alpha \cdot \text{std}$ of the mean:

$$\boldsymbol{u}_i(\ell) = \begin{cases} \boldsymbol{d}_{k'}(i-\ell), & \text{if } |\boldsymbol{d}_{k'}(i-\ell) - \text{avg}| \geq \alpha \cdot \text{std}, \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

The template is then applied to a Normalized Least Mean Square (NLMS) 16-tap adaptive filter [7], depicted in Figure 3. The clean neural signal $\widehat{\boldsymbol{s}}_k(i)$ is then obtained by subtracting the estimated artifact $\boldsymbol{u}_i\boldsymbol{w}_{k,i}$ from $\boldsymbol{d}_k(i)$:

$$\boldsymbol{w}_{k,i} = \boldsymbol{w}_{k,i-1} + \frac{\mu}{\|\boldsymbol{u}_i\|^2 + \epsilon}\boldsymbol{u}_i^{\mathsf{T}}\left[\boldsymbol{d}_k(i) - \boldsymbol{u}_i\boldsymbol{w}_{k,i-1}\right] \tag{4}$$

$$\widehat{\boldsymbol{s}}_k(i) = \boldsymbol{d}_k(i) - \boldsymbol{u}_i\boldsymbol{w}_{k,i} \tag{5}$$

The implemented filter has latency of 16 sampling clock cycles, and operates in real-time. Additionally, the ASAR is implemented in a fully digital feed-forward manner, which avoids injecting noise at the input of the front-end and does not limit the filter's attenuation as no feedback DAC is required. Due to the feed-forward nature of ASAR, the error signal cannot be used directly as the estimate. As shown in (5), $\widehat{\boldsymbol{s}}_k(i)$ is to be obtained using the most recent coefficients $\boldsymbol{w}_{k,i}$. Most current adaptive filter implementations use a classical Least Mean Squares algorithm, whereas we utilize the normalized LMS variant, which computes a variable step size at every iteration, see (4), and requires 16 additional adders and multipliers but results in faster convergence times.

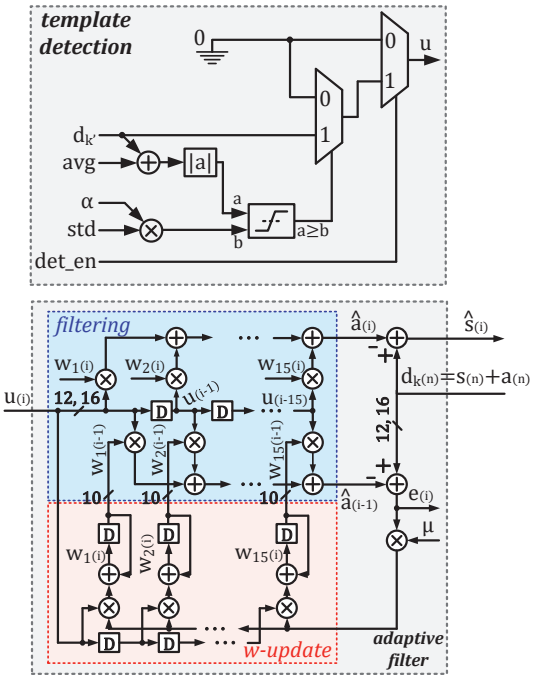

Fig. 3: Detailed implementation of the template detection, 16-tap adaptive filter, and iterative coefficients w_i update blocks.

The choice of the template $\boldsymbol{u}_i$ is critical. The stimulation pulse itself, as employed in [5], [6], is not suitable for this purpose, because: (a) the mapping from stimulator through stimulation electrode, brain tissue and sensing electrode is highly non-linear, resulting in the need for complex filters and long convergence times, and (b) prior knowledge about the structural and temporal shape of the stimulation pulse is required. To remedy both drawbacks, the blind template detection method (1)–(3) was developed, which operates without information on the stimulation waveform. By obtaining a template from an adjacent electrode and learning only the mapping between adjacent recordings, a linear NLMS filter with 16 taps turns out to be sufficient. Most importantly, this enables our implementation to work with any arbitrary stimulation pulse. Figure 3 shows the detailed implementation of the 16-tap adaptive filter, along with the template detection block.

III. Measurement Results

Clinical human patient data were used to test both ASAR designs. Chip measurements with sample data for LFP signals, with input resolution of 16-bits, and sampling frequency of 6 kHz, as well as Spike ASAR, with recordings containing Spike and LFP signals, sampled at 24 kHz, with input resolution of 12-bits are shown in Figure 4. At the time of measurement, recordings with artifacts up to $29\,\text{mV}_{\text{p-p}}$ (LFP) and $36\,\text{mV}_{\text{p-p}}$ (Spike) were available, resulting in artifact attenuation of up to 37 dB and 40 dB, respectively; however, the designs are capable of handling larger amplitudes (Figure 1).

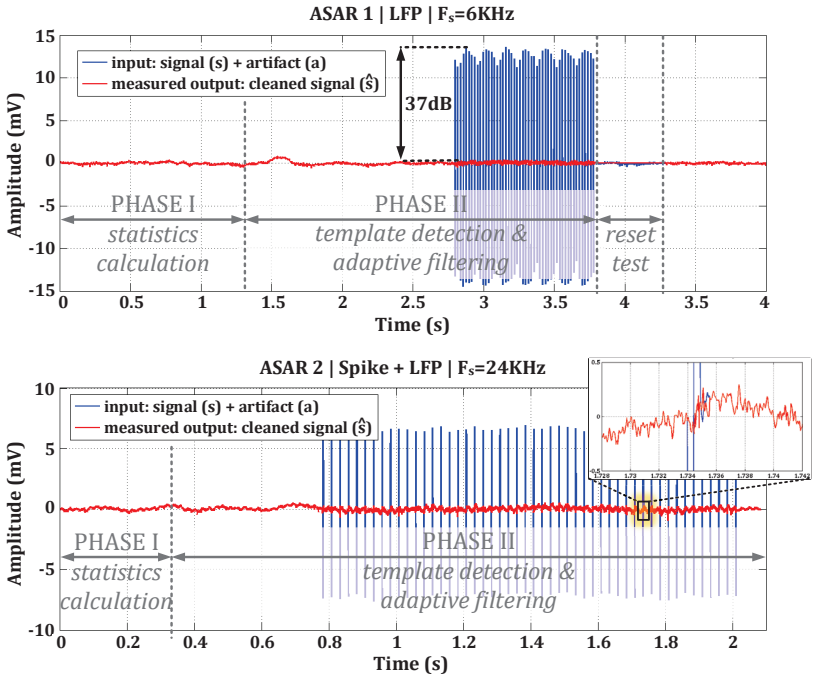

Fig. 4: (top) chip measurement of design 1 (LFP recording, Fs=6 kHz, input resolution=16 bits, stimulation artifact amplitude = 29 $\mathrm{mV_{p-p}}$) (bottom) chip measurement of design 2 (Spike + LFP recording, Fs=24 kHz, input resolution=12 bits, stimulation artifact amplitude = 14 $\mathrm{mV_{p-p}}$).

Figure 5 shows the time-domain recordings and spectrogram of neural signal (Spike and LFP) with 36 $\mathrm{mV_{p-p}}$ stimulation artifacts with and without the ASAR activated.

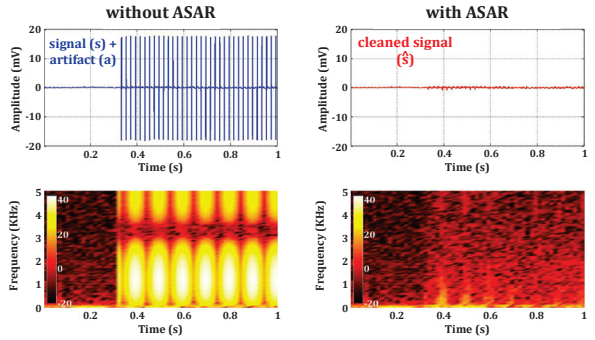

Fig. 5: (left) Time-domain recordings and spectrogram of neural signal (Spike + LFP) with 36 $\mathrm{mV_{p-p}}$ stimulation artifact. (right) Measured ASAR output waveform and spectrogram.

A true measurement of the filter performance is difficult as the cleaned signal $\hat{s}_k(i)$ includes the signal of interest combined with a small residual artifact, which are inseparable, making it impossible to obtain a ground truth. To provide a better measure, we added synthetic artifacts with various amplitudes to recorded data without stimulation artifacts, and ran simulations to calculate the reduction of artifact power (Figure 6). This shows that the ASAR can achieve noise power reduction of up to 49.2 dB.

The designs achieve real-time convergence of $<42\,\mu$s for Spike ASAR and $<167\,\mu$s for LFP ASAR, making them suitable for closed-loop neuromodulation systems. The LFP

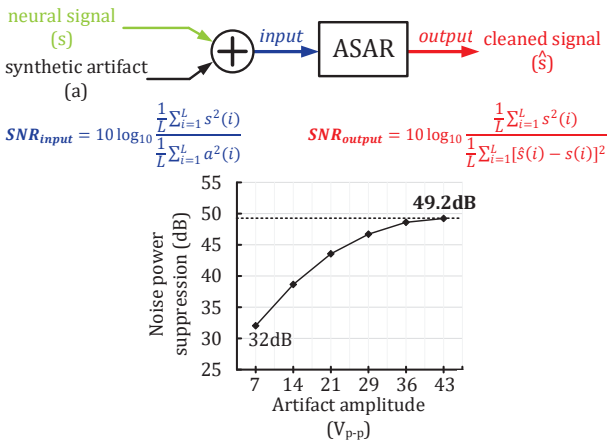

Fig. 6: Calculation of noise (artifact) power suppression based on SNR comparison of added known synthetic artifacts to clinical patient data (input) and simulated cleaned signal by ASAR (output).

and Spike ASAR designs occupy an area of 0.197 $\mathrm{mm^2}$ and 0.209 $\mathrm{mm^2}$, and consume 1.73 µW and 3.02 µW, respectively at a 0.644 V supply. Table II compares our work with the state of the art. Figure 7 shows the chip micrograph implemented in 40 nm CMOS technology.

IV. CONCLUSIONS

This work demonstrates a blind ASAR solution that meets the challenges of modern implantable, closed-loop, neuromodulation systems, and enables concurrent stimulation and recording of neural signals, in real-time, without needing any prior knowledge of the stimulation pulse, and can aid in the investigation of instantaneous response to stimulation.

TABLE II: (top) Comparison of various stimulation artifact rejection methods. (bottom) Comparison with state-of-the-art.

Reference	[3]	[4]	[5-6]	This work
Method	Blanking	Self-canceling	Adaptive Filtering	Adaptive Filtering
Enable recording during stimulation	No	Yes	Yes	Yes
Handle in-band artifacts	Yes	No	Yes	Yes
Application specific	No	Yes	No	No
Agnostic to recording delay	Yes	Yes	No	Yes
Tolerate large artifact amplitudes	Yes	Yes	No	Yes
Need prior knowledge of stimulation pulse	No	No	Yes	No

		JSSC '16 [6]	This work	
			ASAR Design 1	ASAR Design 2
Technology (nm)		180	40	
Area (mm^2)		0.17	0.197	0.209
Power (µW)		0.33	1.73 [a,b]	3.02 [a,b]
BW (Hz)		1-2K	1-200	1-5K
Signals of interest		ECoG	LFPs	Spikes + LFPs
Stimulation Artifact Rejection	Architecture	Mixed-signal Feedback	Digital Feedforward	
	Adaptive filter	Signed LMS, 8-tap, fixed step size	Full LMS, 16-tap, adaptive step size	
	Sampling rate (KS/s)	4	6	24
	Attenuation (dB)	24 [c]	up to 37	up to 40
	Noise power suppression (dB)	-	up to 49 [d]	
	Operating Frequency (KHz)	4	96 (=16·6) [a]	288 (=12·24) [a]
	Tolerable amplitude (mV_{p-p})	<10	<100	<40
	Convergence time (µs)	>3,000,000	<167	<42
	Prior knowledge of stimulation	Yes	No	

[a] due to limited IO pads, Serial-to-Parallel and Parallel-to-Serial circuits were added and ASAR input and outputs were serialized. This resulted in operating frequencies 12x and 16x higher than the respective sampling frequency of each design, and higher power.
[b] power calculations assume stimulations to be ON for 5% of the total duration of the test.
[c] artifact attenuation of up to 42dB is based on the resolution of the 8-bit DAC and not a true measure of the filter attenuation (reported measured attenuation of 24dB was chosen for comparison).
[d] emulated by adding varying synthetic artifacts to clinical human patient neural signal.

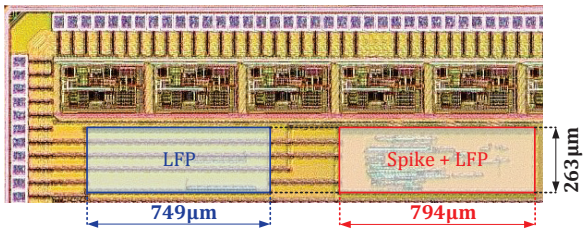

Fig. 7: Chip micrograph.

ACKNOWLEDGMENT

The authors thank Hariprasad Chandrakumar, Wenlong Jiang, Vahagn Hokhikyan, Dejan Rozgic, and Yuta Toriyama for technical discussions, Dr. Zahra Aghajan and Dr. Nanthia Suthana for discussions about neural recordings and the intended use of the system in a clinical setting, and Dr. Itzak Fried and Dr. Nanthia Suthana for human LFP and spike data.

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

REFERENCES

[1] H. Chandrakumar et al., "A 2µW 40mVpp linear-input-range chopper-stabilized bio-signal amplifier with boosted input impedance of 300MΩ and electrode-offset filtering," in *Proc. IEEE ISSCC*, San Francisco, CA, Feb 2016, pp. 96–97.

[2] W. Jiang et al., "A ±50mV linear-input-range VCO-based neural-recording front-end with digital nonlinearity correction," in *Proc. IEEE ISSCC*, San Francisco, CA, Feb 2016, pp. 484–485.

[3] E. Brown et al., "Stimulus-artifact elimination in a multi-electrode system," *IEEE Trans. Biomed. Circuits Syst.*, Vol. 2, No. 1, pp. 10–21, Mar 2008.

[4] S. Stanslaski et al., "Design and validation of a fully implantable, chronic, closed-loop neuromodulation device with concurrent sensing and stimulation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, Vol. 20, No. 4, pp. 410–421, Jul 2012.

[5] A. Mendrela et al., "Enabling closed-loop neural interface: A bi-directional interface circuit with stimulation artifact cancellation and cross-channel CM noise suppression," *2015 Symposium on VLSI Circuits (VLSI Circuits)*, Kyoto, 2015, pp. C108-C109.

[6] A. Mendrela et al., "A bidirectional neural interface circuit with active stimulation artifact cancellation and cross-channel common-mode noise suppression," *IEEE J. Solid-State Circuits*, Vol. 51, No. 4, pp. 955–965, Apr 2016.

[7] A. H. Sayed, *Adaptive Filters*, John Wiley & Sons, NJ, 2008.