NeuralCLIP: A Modular FPGA-Based Neural Interface for Closed-Loop Operation

Vaishnavi Ranganathan¹, Jared Nakahara¹, Soshi Samejima², Nicholas Tolley², Abed Khorasani², Chet T. Moritz² and Joshua R. Smith¹

¹Dept. of Electrical Engineering and ²Dept. of Rehabilitation Medicine, University of Washington, Seattle, USA.

Abstract—

The need for a miniaturized device that can perform closedloop operation is imminent with the growing interest in braincontrolled devices and in stimulation to treat neural disorders. This work presents the Neural Closed-Loop Implantable Platform (NeuralCLIP), a modular FPGA-based device that can record neural signals, process them locally to detect an event and trigger neural stimulation based on the detection. Specifically, the NeuralCLIP is designed to record and process different neural signals in the frequency range between 20 Hz and 1 kHz. It is a flexible platform that can be reconfigured to optimize parameters like channel count and operation frequency based on the processing requirements. The signalagnostic feature is demonstrated by testing the device with calibration signals from standard bio-signal emulators. The application focus for this device is a brain-computer-spinal interface (BCSI) which is demonstrated based on local field potential (LFP) signals recorded from a rat motor cortex. This work demonstrates recording and on-device processing of LFP signals to decode action intent and determine stimulation timing. The FPGA implementation of the device also targets development of low power algorithms for closed-loop operation.

I. INTRODUCTION

Neural interface devices enable brain-controlled technology and provide tools for studying the brain and treating neural disorders. The next generation of such devices must be miniaturized and implantable to record neural signals and stimulate neurons [1]. Using the recorded signals, these devices should enable real-time detection and treatment of neural disorders. Hence, the devices must perform local computation on the recorded signals to identify triggers for closed-loop neural stimulation.

Research in this field can be classified into three broad efforts. First, neural signal acquisition, including the development of state-of-the-art recording devices and electrodes [2] [3] [4]. Second, neural stimulation circuits capable of activating and blocking neuron function [5] [6]. Third, processing the recorded signals to either detect events (like epilepsy seizures) or decode intent (correlate neural signals to action intent) for closed-loop operation. While recording and stimulation are moving from bench top circuits to integrated circuits (ICs), signal analysis is yet to be miniaturized. Moreover, tying the three efforts together to make a small closed-loop device is still at an early stage. Statistical tools have provided us methods like Discrete Wavelet Transform and Support Vector Machines to analyze recorded data. The recent work in [5] [7] implement processing of Spikes and Electroencephalogram (EEG for epilepsy detection) and are early attempts at fully closing the loop. However, these systems are limited to depending on external devices for processing or operate with signals at higher amplitudes (mV). Brain-machine interface applications require the ability to detect spike and local field potential (LFP) signals that lie in the 20 μ V to 200 μ V range. A detailed study is presented in [8] which presents closed-loop neural recording and stimulation in primates. This system, however, uses a rack-mounted test setup.

In this work, we present a miniaturized FPGA-based platform that combines the ability to record and process neural signals in the frequency range of spikes, LFP, ECoG and EEG. The platform is developed as a modular COTS printed circuit prototype that can scale processing in terms of system frequency, sampling rates and number of channels based on power availability and the neural signal of interest. The vision for this modular device is to enable research in low power closed-loop algorithms as well as to provide a platform to study and develop treatment for neural disorders. The FPGA-based processing makes it a useful development platform for future neural-interface ASIC development. The device is tested with a bio-signal calibrator as well as prerecorded LFP signals from a rodent. The test application for this platform is a brain-computer-spinal interface where closed-loop operation triggers stimulation in the spinal cord to bypass an injury and reanimate paralyzed limbs. The concept of stimulation for limb reanimation is explored in [9]. An illustration of the test application, description of the recording signal space and the NeuralCLIP platform are shown in Fig.1. The design, features and results from testing this device are described in the following sections.

II. SYSTEM DESIGN

A. Design for re-configurable operation

The key capabilities of the NeuralCLIP (shown in Fig.1) are recording, local processing and stimulating on a small form-factor device. The platform uses a record/stimulate front end from Intan technologies (RHS2116), which has 16 unipolar channels that can be configured as low noise amplifiers or as constant-current stimulators. A four-wire serial peripheral interface (SPI) bus is used to configure and poll recording data from the Intan. The on-chip ADC provides 16-bit samples from 16 channels at more than 44 ksps each. The configuration architecture used in our device supports on-demand channel disabling to reduce power consumption by unused channels.



Fig. 1: Image shows the functional block diagram and application of the NeuralCLIP. The application includes recording LFP signals from the motor cortex, processing them locally on the implantable device to identify intention and trigger stimulation in the spinal cord to reanimate an arm paralyzed due to spinal cord injury. The device block diagram shows the pipeline processing on FPGA. This includes modular synchronized blocks to configure the recording front-end to acquire the data from it, followed by filtering and correlation of the recorded signals to decode the motor intent. Decoded intent is then used to trigger stimulation. The device also provides a development feature where the secondary SPI module can be used to record data at each processing stage and communicate it to an external computer for training, study and analysis.

The central controller and processing are implemented on a low power FPGA from Microsemi (AGLN250). Specifically, the state machines for configuring and data acquisition from the Intan, the processing algorithm to provide stimulation trigger, and a secondary SPI debug channel are all implemented as synchronous modular blocks in the FPGA. These blocks are part of a pipeline controlled by the system clock. This modular implementation allows parallel data acquisition, processing and debugging, and makes adding or removing blocks easy. Each block can scale flexibly in frequency or channel count to optimize the system performance without affecting the processing pipeline. This modular architecture also allows tailoring of the processing blocks to specific signals such as LFP or spikes. The parameters that allow for scaling and flexibility are as follows:

- Processing clock: This parameter can be set to either 0.5, 1, 2, 4, 8, or 16 MHz to scale the overall power consumed by the system.
- 2) Channel Count: The 16 available channels can either record, stimulate, or be disabled to save power.
- 3) Sampling frequency: This is a parameter that scales with the system clock. It is set by the rate at which the Intan is queried for data.
- Debug interface: The platform provides an optional secondary SPI block for debugging and transferring data off-device for additional post-recording analysis.

The data acquisition and processing blocks are implemented in hardware description language and used to configure the FPGA through the Libero SoC IDE.

B. Data acquisition and processing pipeline

A pipeline diagram for the data flow description is provided in Fig.2. The first stage in the pipeline is a state machine that communicates with the Intan to configure it and acquire data. All data is in 2's complement format to simplify arithmetic operations on the signed data. The second stage is an optional common average reference (CAR) filter that can be used to remove common noise. The third stage is a band pass filter (BPF) which typically consumes a large amount of resources. For example, a single instance of a 16x16 bit multiplier takes up 17 percent of the resources available on the FPGA. The filter implementation on the NeuralCLIP, however, is an approximate computing block. The filter coefficients are first generated for the frequency range of interest using MATLAB. By normalizing these coefficients to their nearest fraction of 2, we implement the divide operation as arithmetic right shifts. Since we handle 2's complement data, the division of a k-bit Sample (S) by a coefficient (C) to produce result (f_O) is reduced to the following form, where "k-bit-ext" is bit extension by k bits: If (C[sign] is 1):

$$f_O = \{\{k - bit - ext(C[sign]) \bigoplus S[sign]\}, \overline{S} + 1\} <<< C$$
else :

$$f_O = \{\{k - bit - ext(C[sign]) \bigoplus S[sign]\}, S\} <<< C$$

The fourth stage is a canonical correlation analysis (CCA) block that scales the channels with correlation coefficients [10]. The offline training to determine these coefficients is done with recorded data from N channels and the corresponding ground truth data in Matlab. The FPGA implementation is similar to that of the band-pass filter to optimize for available resources.

C. Device Power and Control

The power supply for this platform is derived from a 3.3 V line that is used to generate a variable ± 3.3 V to ± 12 V supply for the Intan stimulator and a 1.2 V core supply for the FPGA. The baseline static power for the NeuralCLIP, which includes power for regulators and idle systems, was measured to be an average of 58 mW. The power consumption measured for active recording and



Fig. 2: To record from N channels, the processing pipeline begins with the N Intan queries and the data is buffered for the N channels. This buffered data is then fed to the first processing block in parallel. The output of each block is buffered by N channels as input for the next block. A single system clock controls the transfer of data from each buffer to the next block and hence makes the functional blocks on the FPGA independent and synchronous. The data from N channels are thus processed in parallel. Processing blocks also function in parallel allowing optional addition or removal of blocks and channels.

processing of four channels of LFP data across six system frequencies are provided in Table.I. For neural signals, the maximum required sampling rate per channel is about 20 ksps. However, for LFP (signals below 500 Hz) the sampling rate can be reduce further. This allows the system clock to be lowered in relation to the sampling rate. To ensure this power scaling with respect to system frequency, the sampling and data processing blocks are driven by subclocks derived from the system clock. Thus, the overall device power consumption can be decreased by reducing the system frequency, effectively lowering the data sampling and processing rate.

System Frequency (MHz)	0.5	1	2	4	8	16
Sampling Rate kS/s/ch	2.4	4.8	9.6	19.2	38.5	77.1
Record and Processing Power (mW)	0.99	1.32	1.98	4.29	6.93	13.2

TABLE I: Power Consumption vs. System Frequency

III. TEST SETUP

To validate the NeuralCLIP operation, a ground truth study was first performed with calibrations signals. The signals used for testing are sine waves distributed in the bands of interest at 30 HZ, 200 Hz (Coulbourn Biosignal calibrator) and 800 Hz(Tucker Davis Technology (TDT)). To verify the full processing pipeline, testing was done with prerecorded data from the motor cortex of a rat. The rat was trained to push a lever in order to receive a reward. Neural signals were recorded from the motor cortex during this period. The CCA coefficients were extracted from a window of this recording in correlation to the lever push. A TDT neural interface setup was used to emulate the rat and play back this neural data. Supply voltage of 3.3 V was derived from an external DC source. A digital logic analyzer was used to extract data from each block. The following signals were logged; raw recorded signal, the BPF output and the CCA output from N channels.

We use N = 4 due to a limitation on the number of output channels from the TDT.

IV. RESULTS AND DISCUSSION

This section presents the results from testing and discusses future directions. The device was first tested to verify its ability to record and process signals in the neural frequency range (Fig. 3). The recorded signals, exported through the debugging interface, show the raw and BPF outputs for 30 Hz, 200 Hz, and 800 Hz test signals. Removal of the high frequency noise is evident on comparing the two rows in each case. In addition, the insets also show the difference in the sinusoidal signal quality before and after filtering.

Next, the processing block was tested using LFP data recorded from a rat while the rat was performing a lever push task. The plots in Fig. 4 show the extracted signal after BPF processing from one of the channels, where the amplitude change corresponding to the lever push intent is present. The CCA block then correlates the multiple channels, based on pre-determined coefficients, to provide a single output that is used to identify intent (Fig. 4). On comparing the CCA output and the lever push plots, we see a correlated increase in amplitude of the CCA signal that is used to trigger stimulation on windowed-threshold crossing. Thus the closed-loop operation that is typically limited to benchtop test setups can be enabled on an implantable platform. Following this validation of recording and processing to trigger closed-loop stimulation, the future objective is to implant the NeuralCLIP and test for long-term ability to provide closed-loop stimulation in the spinal cord.

V. CONCLUSION

This work presents the NeuralCLIP, a device capable of recording neural signals and performing local computation on an FPGA to trigger stimulation. By implementing modular synchronous blocks the device achieves reduced resource



Fig. 3: FFT plots for the three calibration signals from the NeuralCLIP. Buffered data from Record SPI block is provided on the top panel and the corresponding BPF buffer output in the bottom panel. Insets show filtering of high frequency signal and improved signal quality.



Fig. 4: Validation with LFP signals recorded from the motor cortex of a rat during a lever push task. First panel: signal from one channel after CAR and BPF processing. Second panel: CCA block output (amplitude correlation to lever push). Third panel: ground truth lever push recording. The BPF and CCA outputs are provided as ADC steps which have a resolution of 0.195 μ V/LSB. Comparing panel two and three, the output on the CCA shows the intent as an increase in amplitude correlated to the ground truth lever push, marked by the horizontal orange bars in the second panel.

and power consumption for data processing. In addition, these processing blocks can be easily modified for different algorithms. This work also demonstrates decoding of LFP signals to enable closed-loop operation. Our application is a brain-computer-spinal interface that records LFP signals from the motor cortex of a rat, processes them to detect a lever-push intent to trigger stimulation in the spinal cord for reanimation of the limb. The NeuralCLIP also allows access to data at each stage of processing for either training or postrecording analysis to enable neuromodulation research. With the implementation on FPGA, this device facilitates development of low power algorithms for closed-loop operation.

REFERENCES

- [1] M. A. L. Nicolelis, "Actions from thoughts," *Nature*, vol. 409, no. 6818, pp. 403–407, 2001.
- [2] R. R. Harrison and C. Charles, "A low-power low-noise cmos amplifier for neural recording applications," *IEEE JSSC*, vol. 38, no. 6, pp. 958– 965, June 2003.
- [3] A. C. Patil and N. V. Thakor, "Implantable neurotechnologies: a review of micro- and nanoelectrodes for neural recording," *Medical*

& Biological Engineering & Computing, vol. 54, no. 1, pp. 23-44, Jan 2016.

- [4] X. Liu *et al.*, "A fully integrated wireless compressed sensing neural signal acquisition system for chronic recording and brain machine interface," *IEEE TBCAS*, vol. 10, no. 4, pp. 874–883, Aug 2016.
- [5] D. J. Yeager *et al.*, "A 4.78mm2 fully-integrated neuromodulation soc combining 64 acquisition channels with digital compression and simultaneous dual stimulation." in *VLSIC*. IEEE, 2014, pp. 1–2.
- [6] V. Ranganathan et al., "A high-voltage compliant neural stimulator with hf wireless power and uhf backscatter communication," *IEEE Wireless Power Transfer Conference (WPTC)*, pp. 1–4, 2016.
- [7] W.-M. Chen *et al.*, "A fully integrated 8-channel closed-loop neuralprosthetic cmos soc for real-time epileptic seizure control." *IEEE JSSC*, vol. 49, no. 1, pp. 232–247, 2014.
- [8] M. Capogrosso *et al.*, "A brainspine interface alleviating gait deficits after spinal cord injury in primates," *Nature*, p. 284, nov.
 [9] S. E. Mondello *et al.*, "Therapeutic intraspinal stimulation to generate
- [9] S. E. Mondello *et al.*, "Therapeutic intraspinal stimulation to generate activity and promote long-term recovery," *Frontiers in Neuroscience*, vol. 8, no. 8 FEB, pp. 1–7, 2014.
- [10] A. Khorasani *et al.*, "Brain control of an external device by extracting the highest force-related contents of local field potentials in freely moving rats," *IEEE TNSRE*, vol. 26, pp. 18–25, 2018.