33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

A Programmable and Implantable Microsystem for Multimodal Processing of Ensemble Neural Recordings

Fei Zhang, Mehdi Aghagolzadeh, and Karim Oweiss, *Senior Member, IEEE*

Abstract—**Conditioning raw neural signals recorded through microelectrode arrays implanted in the brain is an important first step before information extraction can take place. This paper reports on the design and implementation of a programmable and fully implantable microsystem that fulfills this purpose. The system design builds on our earlier work that relies on a sparse representation of the neural signals to combat the limited telemetry bandwidth when wireless communication with the external world is sought. The system has a multimodal processing capability to support a wide range of scenarios in real experimental conditions. A transmission link with rate-dependent compression and spike sorting strategy is shown to preserve information fidelity. At 32 channels sampled at 25 kHz, the power consumption of the system is 5.19 mW and has been implemented on a 5mm×5mm nano-FPGA, bringing its performance within the implantable power-size constraints for clinical applications.**

I. INTRODUCTION

NEURAL ensemble recordings with penetrating microelectrode arrays have been shown to yield superior information content about motor intent in subjects with severe motor and communication deficits compared to other signals acquired with noninvasive devices [1-4]. One major obstacle that precludes the extraction of this information in awake, behaving subjects is the need to be tethered to large size, computationally pristine recording equipment that are typically found in laboratory settings. Clinical viability of these devices, however, requires developing fully implantable neural recording microsystems capable of wireless data and power telemetry. Data reduction early in the data stream is one potential solution, provided the information in the neural signals is not compromised.

Our previous work has demonstrated that critical information is preserved when the neural data is sparsely represented, for example, using a discrete wavelet transform (DWT) [5, 6]. A key element to enable rapid translation of these findings to clinical use is to implement this representation on low cost hardware platforms that can be programmed "on the fly", for example, to recalibrate the system in the face of an unreliable wireless telemetry link or changes in neural signal characteristics [6, 7].

Herein, we report some recent results on the development of a system comprising these features. Specifically, the system transmits the most biologically relevant information from 32 channels sampled at 25 kHz per channel over a 1 Mbps wireless link. The system is highly scalable, has the advantage of switching between different modes of operation during run time, and has been fully implemented on an implantable nano-FGPA.

II. SYSTEM ARCHITECTURE

Figure 1 illustrates the system architecture. It constitutes one of three blocks in a Neural Interface Node (NIN) currently under development [8]. The analog front-end amplifies and filters the neural signals before time-to-digital conversion takes place at the input of the neuroprocessor. The output of the neuroprocessor is fed to an RF block that manages the wireless data telemetry of the extracted information to the outside world, and manages the inductive powering of the NIN through pairing the NIN coil with an outside nearby coil.

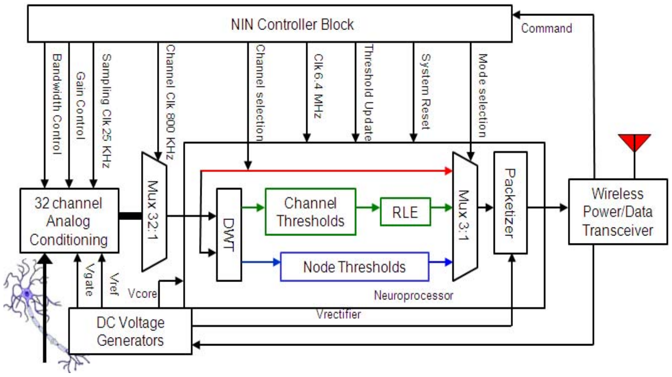


Fig. 1. System Architecture

The neuroprocessor is responsible for extracting information from the neural signals as well as the management of the NIN operation, including selection of the operational mode, the bandwidth and gain of analog conditioning circuits, the selection of channels of interest, power, data communication and parameter updates. As illustrated in Figure 1, the neuroprocessor supports three operational modes through externally programmable registers: 1) Monitoring mode (red), where single channel raw data are transmitted sequentially at full bandwidth to permit estimating compression/spike sorting threshold parameters off chip; 2) Compression mode (green), in which

Manuscript received Apr. 9, 2011. This work was supported by the National Institutes of Health under Grant NS062031.

F. Zhang and M. Aghagolzadeh are with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824 USA (feizhang@msu.edu, aghagolz@msu.edu).

K. Oweiss is with the Department of Electrical and Computer Engineering and Neuroscience Program, Michigan State University, East Lansing, MI 48824 USA (koweiss@msu.edu).

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

only the sparse coefficient representation of 32 channel data is transmitted after run length encoded (RLE); and 3) Sensing mode (blue), where only spike time stamps extracted on active channels are transmitted after DWT-based spike sorting is implemented with an alternative threshold selection scheme [8-9].

A. *VLSI Architecture*

As shown in Figure 2, the neuroprocessor [9] includes a DWT module (green dashed frame), a global controller (blue dashed frame) and a communication module (red dashed frame). The DWT module has a Finite State Machine (FSM) based controller for controlling timing and sequence operations, a lifting DWT based Computation Core (CC), five memory modules for incoming data, coefficients, intermediate CC products and intermediate values for multichannel multilevel interleaved DWT computations [10-11].

The global controller consists of a command decoder, one threshold memory and four registers, and is used to decode the received command into analog bandwidth and gain control parameters and update the thresholds used for both the compression and sensing modes. It also selects the channel used for the monitoring mode, manages selection of operational mode and the half-duplex wireless communication with external module.

The communication module is mainly composed of two alternant communication buffers to ensure no neural information loss over the bandwidth limited wireless link. It organizes the different data outputs for three operational modes, and also packetizes the closed-loop power status to update the level of externally supplied power so that the voltage level in NIN is kept constant, even in case of coil misalignment or mismatch in the inductive power link.

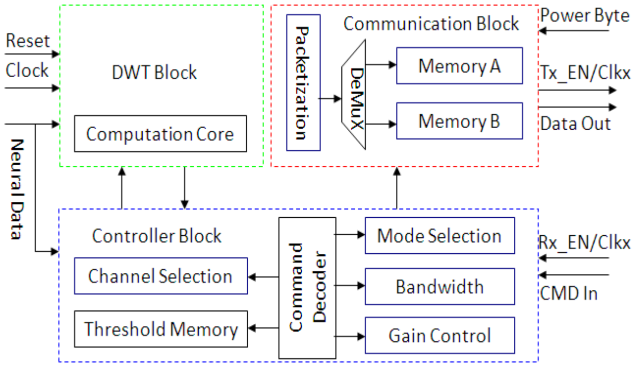


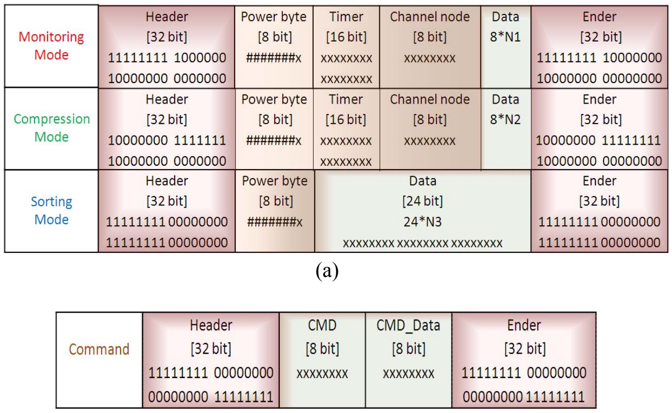
Fig. 2. VLSI architecture of the neuroprocessor

B. *Data and Command Communication Protocol*

Bidirectional communication and low power consumption is a key design specification to optimize the data and command communication protocol. A half-duplex communication protocol is used to wirelessly transfer neural information and power status data from the NIN, and receive clock, power, and command from outside to conserve power/bandwidth and reconfigure system operations.

As shown in Figure 3.a, to realize multi-modal functionality, the uplink data packetizer organizes the processed neural data in three different structured frames with overheads for synchronization and error detection, where every frame length is 840 bits (*N1*=*N2*=3*N3*-3=93, where *N1*, *N2* and *N3* are the byte number of neural data for three modes, respectively). This amounts to 8.45 % of frame overhead, with 7.62 % contributed by the header and ender. The power byte here is used to monitor and relay the power level received by the NIN, and close the loop of the wireless power supply from outside to the NIN to obtain a stable power supply, regardless of the misalignment or distance change between the coils [10]. The two timer bytes in monitoring and compression modes are used to record the timestamp of the first data in the packet frame. Similarly, the channel node byte in monitoring and compression modes is used to mark the source of the first data sample in this packet frame. The timer together with channel node information makes recovering the neural information feasible in case of packet loss.

The downlink command frame to NIN is 80 bits and includes command to switch between different NIN modes, control the analog conditioning circuits such as bandwidth and gain, select desirable output channel for monitoring mode, and update the threshold values for either compression or sorting mode. At a transmission frequency of 1 Mbps, the 840-bit data packet takes 0.84 ms for transmission, and the 80-bit command packet takes 0.080 ms. Thus, assuming that the packet propagation delay and the idle time between receipt and transmission are negligible, the shortest time it can afford to wait for the incoming data to be packetized and filled into the communication buffers is 0.92 ms. In the current design, two 840-bit communication buffers are used in order to avoid data overflowing. At any given time, only one buffer is active for receiving incoming data, and the other acts as a reserve buffer after sending the data received during its active period.



Monitoring Mode	Header [32 bit] 11111111 1000000 10000000 0000000	Power byte [8 bit] #######x	Timer [16 bit] xxxxxxxx xxxxxxxx	Channel node [8 bit] xxxxxxxx	Data 8*N1	Ender [32 bit] 11111111 10000000 10000000 00000000
Compression Mode	Header [32 bit] 10000000 1111111 10000000 0000000	Power byte [8 bit] #######x	Timer [16 bit] xxxxxxxx xxxxxxxx	Channel node [8 bit] xxxxxxxx	Data 8*N2	Ender [32 bit] 10000000 11111111 10000000 00000000
Sorting Mode	Header [32 bit] 11111111 00000000 11111111 00000000	Power byte [8 bit] #######x	Data [24 bit] 24*N3 xxxxxxxx xxxxxxxx xxxxxxxx			Ender [32 bit] 11111111 00000000 11111111 00000000

(a)

Command	Header [32 bit] 11111111 00000000 00000000 11111111	CMD [8 bit] xxxxxxxx	CMD_Data [8 bit] xxxxxxxx	Ender [32 bit] 11111111 00000000 00000000 11111111

(b)

Fig. 3. Packet format for data and command frame, (a) Uplink data transmission from NIN; (b) Downlink command transmission to NIN

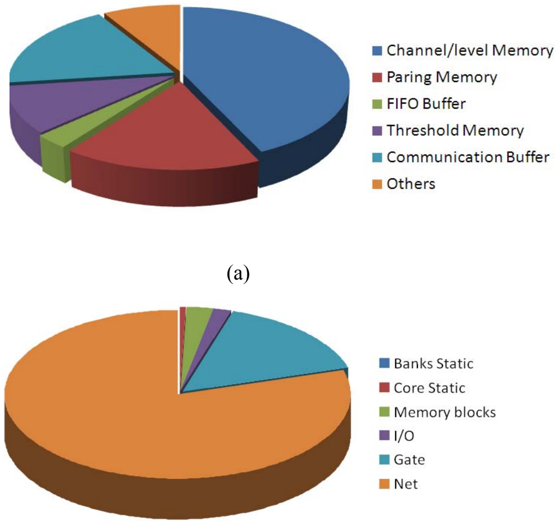
III. NANO-FPGA PROTOTYPING

The AGLN 250 nano-FPGA from Actel was chosen to implement the designed neuroprocessor due to its superior features including programmability, low-power and small size (5mm×5mm). Most importantly, it also includes memory blocks that reduce unnecessary usage of system gates. Figure. 4.a summarize the memory demands needed, while Table I summarizes the resource allocation of the entire neuroprocessor on this FPGA.

TABLE I. HARDWARE RESOURCES OF FINAL DESIGN ON AGLN 250

Type of Resources	Resource of AGLN250 Nano-FPGA		
	Total	Used	Percentage
Embedded RAM Blocks	8	8	100 %
Versa Tile (D-flip-flops)	6144	4328	70.44 %
PLL	1	1	100 %
Chip Global	6	6	100 %

For the design of 32-channel, 4-level DWT, and sampling rate of 25 ksps per channel, the total power consumption of the neuroprocessor was 5.14 mW, evaluated with the Smart Power tool in Actel Designer, which matched closely with the measured 5.19 mW. The detailed distribution of power budget is plotted in Figure 4.b. The time analyzer in Actel Designer reported 53.50 ns of critical path.



(a)

(b)

Fig. 4. Distributions of: (a) hardware resource and (b) power consumption

IV. RESULTS

To investigate the optimal bit precision that preserves information fidelity in the neural signals as a wired system, the Receiver Operating Characteristics (ROC) for different bit precisions of the neural data is shown in Figure 5. Spike sorting threshold parameters are selected to maximize the area under the ROC graphs. An 8-bit resolution has a very similar performance to a 10-bit precision and hence was chosen.

Figure 6.a qualitatively demonstrates some examples of original and reconstructed waveforms. Only 20% and 50 % of the coefficients were used to obtain the reconstructions shown. In Figure 6.b, this is also quantified by the degree of separability between the different neuronal clusters in the feature space. The class separability is defined as the Euclidean distance between spike waveforms of two neurons represented in the compression domain by a user-defined number of coefficients [6].

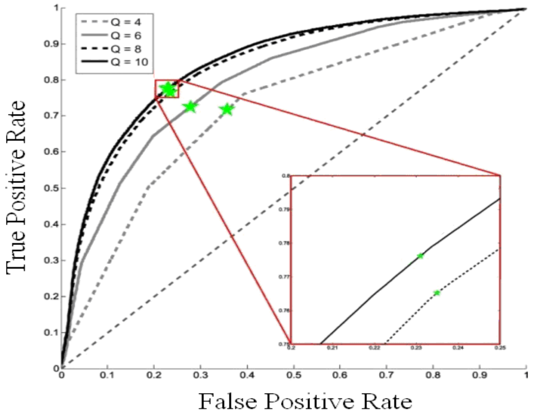
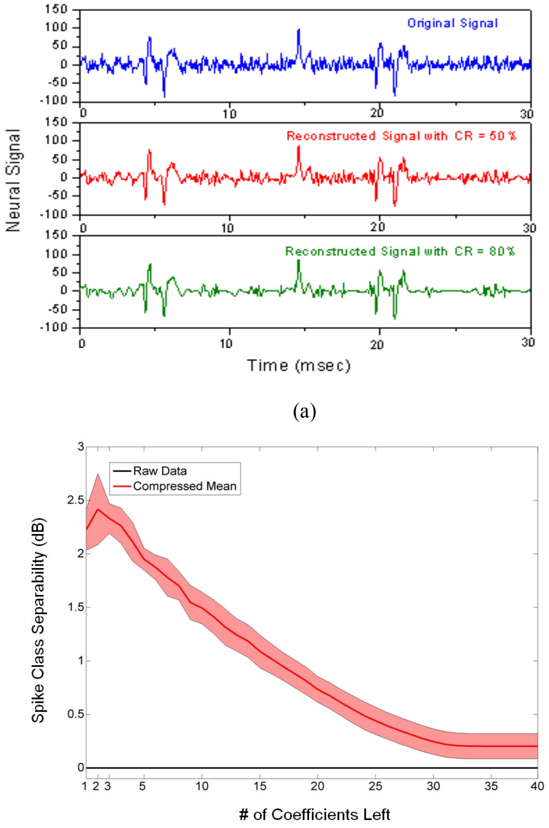


Fig. 5. ROC curves for different bit precisions



(a)

(b)

Fig. 6. (a) Neural signals at different compression rates; (b) Spike class separability vs. compression rate (modified from [6]).

In the compression mode, Figure 7 shows the statistical distribution of the time needed for 50 data frames to fill the communication buffer, with an average value of value of 4.85 ms. The minimal time taken to fill a packet was recorded to be 1.25 ms, which is within the allowed time limit of 0.92 ms described in Section II.B, thereby effectively avoiding data overflow.

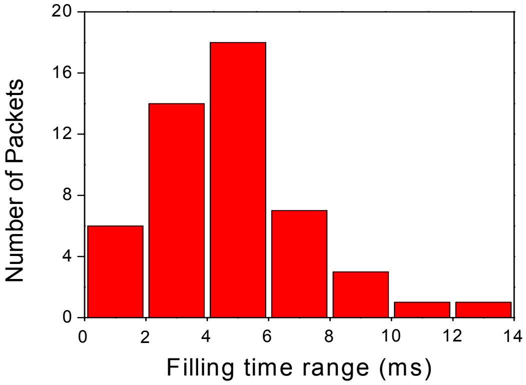


Fig. 7 Statistical distribution of the filling time of 50 data frames in the compression mode

For the sensing mode, the minimum filling time of the communication buffer was found to be 3.15 ms and the average was measured to be 26.02 ms as shown in Figure 8, which are much larger than the filling time of the compression mode. This demonstrates that compression and spike sorting on chip with this neuroprocessor design is feasible and desirable, minimizes system latency and results in orders of magnitude savings in transmission bandwidth as expected.

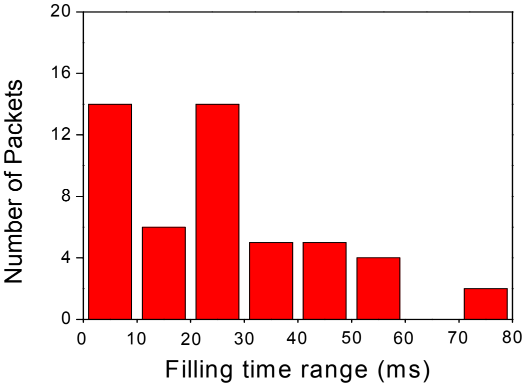


Fig. 8 Statistical distribution of the filling time of 50 data frames in the sensing mode

V. CONCLUSION

In this paper, we reported on a fully implantable and multimodal neuroprocessor with bidirectional communication capability for full programmability. The implementation of the neuroprocessor consumes 5.19 mW power to process 32 channels of electrode data at 25 ksps and 8-bit resolution on a 5mm×5mm nano-FPGA, which brings its power density to 20.76 mW/cm^2. This conforms to the power density limits for clinical grade implants, estimated to be ~62 mW/cm^2 [11]. The system is highly scalable, programmable and cost effective, making it well suited for basic neuroscience research as well as clinical Brain Machine Interface applications.

REFERENCES

[1] M. A. Lebedev and M. A. L. Nicolelis, "Brain-Machine Interfaces: Past, Present and Future," *Trends Neurosci.*, vol. 29, pp. 536-546, 2006.

[2] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A Low-power Integrated Circuit for a Wireless 100-electrode Neural Recording System," *IEEE J. Solid-State Circuits*, vol. 42, pp. 123-133, 2007.

[3] A. M. Sodagar, G. E. Perlin, Y. Ying, K. Najafi, and K. D. Wise, "An Implantable 64-channel Wireless Microsystem for Single-unit Neural Recording," *IEEE J. Solid-State Circuits*, vol. 44, pp. 2591-2604, 2009.

[4] M. S. Chae, Z. Yang, M. R. Yuce, H. Linh, and W. Liu, "A 128-channel 6 mW Wireless Neural Recording IC with Spike Feature Extraction and UWB Transmitter," *IEEE Trans. Neural Systems and Rehabilitation Eng.*, vol. 17, pp. 312-321, 2009.

[5] K. G. Oweiss, A. Mason, Y. Suhail, A. M. Kamboh, and K. E. Thomson, "A Scalable Wavelet Transform VLSI Architecture for Real-time Signal Processing in High-density Intra-cortical Implants," *IEEE Trans. Circuits and Syst. I*, vol. 54, pp. 1266

[6] M. Aghagolzadeh and K. G. Oweiss, "Compressed and Distributed Sensing of Neuronal Activity for Real Time Spike Train Decoding," *IEEE Trans. Neural Syst. and Rehabilitation Eng.*, vol. 17, pp. 116-127, 2009.

[7] K. G. Oweiss, Statistical Signal Processing for Neuroscience and Neurotechnology, Academic Press, Elsevier, pp. 15-74, 2010

[8] F. Zhang, M. Kiani, M. Aghagolzadeh, M. Ghovanloo, K. Oweiss, "WIMNIS 1.0: Wireless Intracortical Multichannel Neural Interface System for Neural Recording in Freely Behaving Subjects," *in preparation*.

[9] F. Zhang, M. Aghagolzadeh and K. Oweiss, An Implantable Neuroprocessor for Multichannel Compressive Neural Recording and On-the-fly Spike Sorting with Wireless Telemetry, *The 2010 IEEE International Conference of Biomed. Circuits and Syst., Paphos, Cyprus*, 2010

[10] M. Kiani and M. Ghovanloo, "An RFID-based closed-loop wireless power transmission system for biomedical applications," *IEEE Trans. Circuits and Syst. II*, vol. 57, pp. 260-264, 2010.

[11] Y. Sun, S. Huang, J.J. Oresko and A.C. Cheng, “Programmable Neural Processing on a Smartdust for Brain-Computer Interfaces”, *IEEE Trans. Biomed. Circuit Syst.*, vol. 4, pp.265-273,2010