Low-Power 128-Channel Neural-Signal Processor with One-Time Programming Memory for High-Density Implantable Neural-Sensing Microsystems

Ting-Wei Sheng1, Po-Tsang Huang2, Shang-Lin Wu1, Chun-Pin Lin3, Nien-Shang Chang1, Hung-Lieh Chen3, Chi-Shi Chen1, Wei Hwang1, Ching-Te Chuang1

1 Department of Electronics Engineering, National Chiao Tung University, Hsinchu, Taiwan
2 International College of Semiconductor Technology, National Chiao Tung University, Hsinchu, Taiwan
3 Nation Chip Implementation Center, Taiwan

ABSTRACT

Highly integrated & miniaturized neural sensing microsystems are crucial for brain function investigation for capturing high spatiotemporal resolution neural signals. In this paper, a low-power neural-signal processor is proposed to extract and classify high-density neural features for an implantable neural-sensing microsystems. The proposed neural-signal processor is designed by ARM Cortex-M0, one-time programmable (OTP) memory with an integrated voltage regulation module (VRM), self-reset circuits and 8 configurable discrete wavelet transform processing elements (DWT PEs). The on-chip OTP memory and self-reset circuits conquer the critical design challenges of implantable neural sensing Microsystems, such as small form factor and ease of the longevity. Additionally, the 8 DWT PEs are designed to extract 128-channel features by filtering the neural signal into different frequency bands. The 128-channel neural-signal processor is implemented using TSMC 40nm CMOS technology, and integrated to a heterogeneous wireless neural-sensing microsystem. The overall power consumption of the proposed high-density neural-signal processor is only 1.92mW for 128 channels.

1. INTRODUCTION

High-density neural-sensing microsystems enable monitoring and mapping of brain activities by sensing high spatiotemporal resolution neural signals [1]. To obtain high-resolution neural signals, neural implant microsystems should meet some critical constraints such as small form factor, ease of implantation and longevity. Thus, several highly miniaturized neural-sensing Microsystems have been presented for acquiring local field potentials (LFPs) or electrocorticogram (ECoG) [2-4]. These Microsystems typically compose of three main parts 1) invasive probes/electrodes for biopotential collection, 2) low-power circuits for brain signal feature acquisition, extraction, classification and transmission, 3) low-noise electrical link wires between electrodes and circuits for signal integrity. In these approaches, however, the sensing density for high spatiotemporal resolution is limited by the shape, pitches and impedance of the probes and the routing density of link wires.

For obtaining high spatiotemporal resolution neural data, feature extraction and classification is an essential function to reduce the required bandwidth of wireless data transmission for implant neural-sensing Microsystems. In view of this, the overall power consumption is also reduced with the decreasing data rate. In this paper, a low-power neural-signal processor is proposed to extract and classify high-density neural features for an implantable wireless neural-sensing microsystems. The proposed neural-signal processor is designed and implemented by an ARM Cortex-M0 core, one-time programmable (OTP) memory with an integrated voltage regulation module (VRM), self-reset circuits, 8 configurable discrete wavelet transform processing elements (DWT PEs) and AHB-lite wrappers for connecting to the neural-signal acquisition dies and wireless power/data (P/D) transmission die. Additionally, this neural-signal processor is realized in a 2.5D heterogeneously integrated bio-sensing microsystem.

The rest of this paper is organized as follow. Section II and III present the 2.5D heterogeneously integrated bio-sensing microsystem and the architecture of the neural-signal processor, respectively. The design of the configurable DWT PE is described in Section IV. Section V describes the program and read modes of OTP memory and the design of self-reset circuits. Section VI summaries the implementation and post-simulation results. Finally, conclusions are given in Section VII.

2. HETEROGENEously INTEGRATED WIRELESS NEURAL-SENSING MIRCOSYSTEM

A miniaturized implantable 128-channel (128-CH) wireless neural-sensing microsystem is realized using TSV-embedded dissolveable μ-needle array, a flexible interposer and 4 dies as shown in Fig. 1 [5]. The high-density wireless neural-sensing microsystem is integrated by TSV 2.5D/3D heterogeneous system-in-package (SiP) technology. The TSV-embedded μ-needle array and 4 dies are integrated on the bio-compatible Cu-Via flexible interposer using chip-to-chip ENIG bonding and TSV 2.5D/3D integration technologies [6]. The 4 dies are designed for high-density neural-signal acquisition, feature extraction and wireless power/data (P/D) transmission. Fig. 2 presents the block diagrams of the heterogeneous 128-CH wireless neural-sensing microsystem composed of 1 neural-signal processor (die-1, 40hn), 1 wireless P/D transmission circuitry (die-2, 0.18μm) and two 64-CH neural-signal acquisition dies (die-3 & die-4, 90nm). Each 64-CH neural-signal acquisition circuitry is composed of 16 4-CH low-noise chopper-stabilized neural amplifiers and 16 area-power-efficient hybrid ADCs. The detail of the 64-CH neural-signal acquisition circuitry has been presented in [7]. For battery-less biomedical implants, power and data telemetry are realized and
transmitted through an on-interposer inductor. The neural-signal processor is designed to control this microsystem and to extract and classify neural features for reducing the required bandwidth of data transmission. The details of the neural-signal processor are described in the following sections.

3. ARCHITECTURES OF NEURAL-SIGNAL PROCESSOR

The neural-signal processor utilizes ARM Cortex-M0 and 8 configurable DWT PEs to manage the data flow of 128 channels and cluster the data for medical diagnosis by filtering the features into different frequency bands, respectively. Fig. 2 also presents the block diagram of the neural-signal processor. All the parameters of the configurable datapath are determined and compiled off-line in the boot code. Thus, the boot code is stored in a 320kB on-chip OTP memory. Before implanting this microsystem in a brain, the boot code is programmed to the OTP memory via a 12-pin connector. After programming, this connector would be removed and the OTP memory is changed to the execution to control the data flow of 128 channels. The system clock controller generates the corresponding clocks for all the blocks in the neural-signal processor and 2 neural-signal acquisition dies. All the clocks are relative to the clock of Cortex-M0, 50MHz, which is the fastest frequency in the digital domain of the overall microsystem.

The ARM Cortex-M0 core is utilized to control the overall microsystem and implements the ARMv6-M architecture by Thumb instruction set (Thumb-2) providing the exceptional performance expected of a modern 32-bit architecture, with a higher code density than other 8-bit and 16-bit microcontrollers. This Cortex-M0 core is extensively optimized for low-power and area that offers the significant benefits of simple architecture, easy-to-use programmer model, low-power operation, excellent code density, deterministic interrupt handling and upward compatibility. Thus, this core delivers exceptional power efficiency through its efficient instruction set.

With the low-power Cortex-M0 core, AMBA AHB-Lite bus is utilized as the on-chip interconnect architecture of the neural-signal processor. AMBA AHB-Lite addresses the requirements of high-performance synthesizable design, and supports a single master providing high-bandwidth operation. Thus, the bus interconnect logic consists of one address decoder and a slave-to-master multiplexer. The decoder monitors the address from the master so that the appropriate slave is selected, and the multiplexer routes the corresponding slave output data back to the master. The AHB-Lite master, Cortex-M0, provides address and control information to initiate a read or write operation. Thus, AHB-Lite responds to bus transfers initiated by Cortex-M0. Additionally, a slave uses the HSE1x signal from the decoder to control when it responds to a bus transfer. The slave signals back to the master includes Success, Failure, and Waiting transfer. The decoder decodes the address of each transfer and provides a selected signal for the slave that is involved in the transfer. Hence, the decoder provides a control signal to the multiplexer. The slave-to-master multiplexer is required to multiplex the read data bus and response signals from the slaves to the master.

4. CONFIGURABLE DWT PE

In addition to increase the energy-efficiency of the neural-signal processor, 8 16-CH configurable DWT PEs are realized to accelerate the high-resolution feature extraction. The power consumption of each DWT PE would increase rapidly when the number of channel is larger than 16 due to the complex wire routing. Fig. 3 shows the datapath of 1 configurable DWT PE which is implemented by a lifting-based DWT accelerator [2]. The timing windows, channel numbers, frequency bands and mother wavelets can all be adjusted by setting the configuration status register.

The input buffer is a register to temporarily store input data because the DWT does not need input data immediately. The channel counter and level counter record the level and the channel which are calculated currently. The neural-signal processor can monitor the status of both level and channel. The iteration counter is utilized to control the DWT accelerator and the distributed SRAM, and calculates the corresponding addresses to store the output data in a 512x10 SRAM array. Therefore, Cortex-M0 can arrange multi-channel data and send the successive data to the wireless transmission circuits. Fig. 4 presents the timing control of different channels for 1 DWT PE. Moreover, the configuration status register records the states of the DWT, which provides all the observers including the input timing controller, channel/level counters and iteration counter. Multi-channel configurable lifting-based DWT is used to extract neural features. Both the area and power consumption can be reduced by reducing the computation circuits using...
lifting-based DWT algorithm [8]. For 1-level iteration of each channel, 10 cycles are required with two multipliers and two adders in the computation core. Additionally, both the time window and mother wavelets (Haar, Db2, Symlet4, Symlet6) can be adjusted for different neural sensing applications in this configurable lifting-based DWT. The lifting-based DWT can be realized by a single computation core to calculate the equations in 1-level iteration [2]. Thus, data accesses for multi-channel and multi-level DWT is the critical design challenge for the configurable datapath. In view of this, five shift registers, \( T_1 \) to \( T_5 \), are utilized to store the intermediate values for the next sample from the same channel. Additionally, in the datapath of the multi-level and multi-channel DWT, numerous storages are required as the coefficient memory for the quantized filter coefficients, input buffers (multi-channel and CC) and other buffers (CC and multi-channel/multi-level).

**5. OTP MEMORY AND SELF-RESET CIRCUITS**

Before implanting this microsystem in a brain, the boot code is programmed to the OTP memory via a 12-pin connector. Fig. 5 presents the control and datapath of the boot OTP memory in the neural-signal processor. The OTP memory with the integrated VRM is designed using a patented split-channel non-volatile anti-fuse memory cells, and is optimized for configuration flexibility supporting a wide range of high density and high performance applications. Apart from the standard digital core supply, \( V_{DD} \), three additional voltage levels are required for OTP memory, including a program voltage \( V_{PP} \), a pre-charge voltage \( V_{QG} \) and a read voltage \( V_{RR} \). Therefore, an integrated VRM is also utilized to generate the three voltage levels, and a ready signal is asserted to high while these three voltages are stable. Since this integrated VRM is at the end of the wireless power topology, this ready signal also indicates the stable condition of the overall system.

The programming mode is controlled by the program pulse width as defined in Fig. 6. All memory cells in a macro are initially manufactured as ‘0’ and their state may have changed a ‘0’ to a ‘1’ using the program operation. Once programmed, it is not possible to reverse ‘1’ to a ‘0’. Memory cells are programmed by gate oxide breakdown while selectively exposed to the elevated voltage level \( V_{PP} \). A write cycle begins by precharging the internal circuitry to the \( V_{QG} \) and \( V_{PP} \) levels when \( WE \) is HIGH and the clock, \( CK \), is low. We must be held HIGH for the entire program cycle. A program pulse which means \( CK \) HIGH to the addressed memory location is initiated on the rising edge of \( CK \), when both \( WE \) and \( SEL \) are HIGH. The address inputs \( A[5:0] \), data inputs \( D[7:0] \) and \( SEL \) are latched by the rising edge of \( CK \). During sequential programming of multiple bits of the selected address, the \( WE \) signal should be kept HIGH between the programming pulses to ensure the \( V_{PP} \) voltage level is maintained in the internal circuitry, minimizing power dissipation and internal voltage switching. \( V_{PP}, V_{QG} \) and \( WE \) should only be changed when \( CK \) is LOW.

After programming, this connector would be removed and the OTP memory is changed to the execution mode (or read mode) for reading instructions by welding to the ground on the interposer. The read operation of the OTP memory is similar to
Fig. 8. Layout view of the proposed neural-signal processor.

Fig. 9. Area breakdown of the neural-signal processor.

Fig. 10. Power breakdown of the neural-signal processor.

that of SRAM or other on-chip ROMs. The overall microsystem is controlled by the neural-signal processor. Thus, a self-reset signals generated based on the stable condition of the wireless power. In the execution mode, the reset signal initializes the implanted microsystem by a self-reset circuitry as shown in Fig. 7. The integrated VRM in the OTP memory generates a ready signal (at the end of the critical path in the wireless power deliver network) when the boosted voltage reaches the operation voltage. To enhance the timing tolerance to the total settling time of the power network, the reset signal is generated behind the ready signal by around 41μs using a positive edge-detector and a 12-bit counter to increase the safety timing margin.

6. IMPLEMENTATION AND RESULTS

The proposed neural-signal processor is designed and implemented using TSMC 40nm GP CMOS technology, and developed by ARM Cortex-M System Design Kit (CMSDK) which provides the software/hardware co-design environment. Fig. 9 presents the layout view of this neural-signal processor. The blocks at the upper sides are the integrated VRM and OTP memory, respectively. The block at the lower right corner is the internal SRAM used by Cortex-M0. Additionally, the 8 small blocks are the 512x10 SRAMs in the DWT-AMBA wrapper. Fig. 10 shows the area breakdown of each block. The neural processor is divided into four parts. First, the AHB-Lite wrapper includes GPIO bi-directional circuits and 3 wrappers connected to the 2 neural-signal acquisition dies and wireless P/D transmission die. It occupies less than 0.1% of the total area. The second part is the internal SRAM accounted for 20%. Third, the MCU system accounted for 38% area includes Cortex M0, 8 16-CH DWT PEs, serializer, AHB and APB circuits. The last part is the OTP memory with the integrated VRM, and it occupies around 42% of the total area. The area occupation of this part is dominated by the integrated VRM and high-performance fuse interface. The power breakdown of each block in the execution mode is as shown in Fig. 10. The power breakdown is also divided into the same four parts as those of the area breakdown. First, the AHB-Lite wrappers consumes less than 1% of the overall power dissipation. The second part is the internal SRAM accounted for 14%. Third, the OTP memory dissipates 22% power consumption by the read operations. The major power consumption of the neural-signal processor is dominated by the MCU system and up to 63% of the overall power consumption but only 1.211mW.

7. CONCLUSION

With the development of high-density neural-signal sensing, the low-power neural-signal processor is one of the critical designs to reduce neural features. In this paper, a low-power 128-CH implantable neural-signal processor with OTP memory is presented and realized in a 2.5D heterogeneously integrated wireless neural-sensing microsystem. This neural-signal processor is composed of four main parts, including Cortex-M0 core, OTP memory with an integrated VRM, AHB interconnection architecture and AHB wrappers, and 8 16-CH configurable DWT PEs. The on-chip OTP memory and self-reset circuits conquer the critical design challenges of implantable neural sensing microsystems, such as small form factor and ease of the longevity. Additionally, the 8 DWT PEs are designed to extract 128-channel features by filtering the neural signal into different frequency bands. The 128-CH neural-signal processor is implemented using TSMC 40nm CMOS GP process. The total power consumption and area occupation are 1.921mW and 977548μm², respectively. Moreover, this proposed 128-CH neural signal processor is integrated in a high-density neural-sensing microsystem with 2 neural-signal acquisition dies and one wireless P/D transmission die.

REFERENCES