A 128-Channel 6 mW Wireless Neural Recording IC With Spike Feature Extraction and UWB Transmitter

Moo Sung Chae, Zhi Yang, Mehmet R. Yuce, Linh Hoang, and Wentai Liu, Senior Member, IEEE

Abstract—This paper reports a 128-channel neural recording integrated circuit (IC) with on-the-fly spike feature extraction and wireless telemetry. The chip consists of eight 16-channel front-end recording blocks, spike detection and feature extraction digital signal processor (DSP), ultra wideband (UWB) transmitter, and on-chip bias generators. Each recording channel has amplifiers with programmable gain and bandwidth to accommodate different types of biological signals. An analog-to-digital converter (ADC) shared by 16 amplifiers through time-multiplexing results in a balanced trade-off between the power consumption and chip area. A nonlinear energy operator (NEO) based spike detector is implemented for identifying spikes, which are further processed by a digital frequency-shaping filter. The computationally efficient spike detection and feature extraction algorithms attribute to an auspicious DSP implementation on-chip. UWB telemetry is designed to wirelessly transfer raw data from 128 recording channels at a data rate of 90 Mbit/s. The chip is realized in 0.35 μm complementary metal–oxide–semiconductor (CMOS) process with an area of 8.8 × 7.2 mm² and consumes 6 mW by employing a sequential turn-on architecture that selectively powers off idle analog circuit blocks. The chip has been tested for electrical specifications and verified in an ex vivo biological environment.

Index Terms—Digital signal processing (DSP), integrated circuit (IC), low-noise amplifier, neural recording system, ultra-wideband (UWB).

I. INTRODUCTION

MULTICHANNEL neural recording systems are used in neuroscience experiments to study complex neural networks of freely behaving animals [1]. It is also a critical component in brain–computer interface used for cortical-controlled neural prosthetics, which has a wide range of applications such as upper and lower limb prostheses [2]–[5], bladder and bowel movement control for spinal cord injury (SCI) patients [6], [7], respiration control for SCI patients [8], and hand grasping function restoration [9].

To support these applications, a neural recording system has to meet challenging requirements imposed by the environment. First, it should be able to record a large number of channels simultaneously; and high density recording can advance fundamental neuroscience studies and has the potential to improve the performance of neural prosthetic devices. Second, a wireless telemetry that transmits recorded neural data is preferable, because tethering wires impose significant restrictions on the subjects and inhibit free movement in their natural environment. Third, on-the-fly processing of neural data is necessary to enable prosthetic devices functioning in a real-time. In addition, a fast processing capability removes the necessity of storing the large amount of raw data. Fourth, the specifications of the recording system should be designed with programmability and versatility that accommodate a wide range of bio-potentials in different applications. Finally, power consumption and chip area have to be optimized due to the limited space available at the recording site and the system should be powered wirelessly or operated on rechargeable battery.

Several neural recording ICs previously reported in literature can support simultaneous multichannel recording [10]–[12], wireless data telemetry (spike information for 100 channels and raw data for one channel) [13], and on-chip spike detection [13], [14]. However, one of the major limitations of previous systems in the literature is that it allows recording from only a subset of the electrodes to be recorded simultaneously and transmitted to the outside mainly due to the limited bandwidth of the wireless telemetry. Many systems are not optimized and have unnecessary increase in power consumption and chip area. Although there have been several systems equipped with integrated functions of recording, processing (spike detection or spike feature extraction), and wireless telemetry using multiple integrated circuits (ICs) [15], [16], a single IC with simultaneous recording, on-chip spike detection and feature extraction, and low power wireless telemetry that can support raw data from more than 100 channels is advantageous at the miniaturization and versatility point of view.

This paper presents a low-power IC that can record, process, and wirelessly transmit neural signals in real-time. The chip, which supports two different modes of recording and feature extraction, is able to simultaneously record and transmit raw data from 128 channels [17] wirelessly, and serves as the core hardware for applications mentioned above when combined with additional components.

The paper is organized as follows. Section II describes the chip architecture of the proposed multichannel neural recording system and its operations. Section III discusses the circuits of front-end blocks in detail. Section IV introduces on-the-fly spike feature extraction and its hardware implementation. Section V
II. CHIP ARCHITECTURE

The chip has eight 16-channel front-end blocks. Each block consists of amplifiers, a multiplexer and an analog-to-digital converter (ADC). The front-end blocks are followed by a data serializing circuit, a digital signal processor (DSP) unit for spike detection and feature extraction, a digital MUX, an encoder, and an ultra wideband (UWB) transmitter (Fig. 1). On-chip bias generators provide dc bias voltages to the front-end blocks. At front-end, weak and noisy neural signals from electrodes are amplified by self-biased, fully differential preamplifiers, and time-multiplexed by an analog multiplexer (Fig. 2). A second amplifier provides additional gain for the proper operation of the subsequent ADC. A successive-approximation-register (SAR) ADC is used to digitize the 16-to-1 time multiplexed analog signal. The amplifiers are designed to have programmable gain and bandwidth to meet the requirements of various biological experiments. There is a trade-off between power and area because as we increase the number of channel per one ADC, the chip area decreases while the power consumption increases due to the increased multiplexer loading [18]. A careful analysis shows that the power and area product is minimized when 16:1 multiplexer is in the technology used [18]. A sequential turn-on method in the front-end blocks is utilized to save power. Two of the 16 channels are fully turned on at any given time, which leads to 71% additional power reduction when using this technique. For example, only the first and second channels are fully turned on when a sample from the first channel is being digitized (Fig. 2). At the next clock cycle, when the ADC accesses second channel, the first channel is turned off and the third one is turned on instead. In this technique, only the buffer to drive the analog multiplexer and ADC is sequentially turned on, and the preamplifiers which draw very small current compared to the buffers are always turned on.

For an electrode with resistance of 1 MΩ operating at 27 °C with a 20 kHz recording bandwidth, the root mean square (rms) noise voltage is 18.2 μV according to Nyquist’s formula. In general, the magnitude of extracellular spikes is within 1 mV, resulting a 35 dB signal-to-noise ratio (SNR) of the input signal. With this estimated noise figure, we designed our ADC to have a 9-bit resolution after a conservative margin [18]. The sampling rate of one channel is chosen to be 40 ksample/s to avoid aggressive interpolation of spike samples, producing raw data rate of 40ksample/s/channel×128channels×9bits/sample = 46.08Mbit/s.

The 9-bit sampled data from eight front-end blocks are fed into digital data serializing circuits or a DSP engine for spike feature extraction according to the mode setting. The chip can operate in one of the two modes. In streaming mode, all the sampled data from eight front-end blocks are fed into the digital data serializing circuits and serialized by blocks resulting in 9-bit parallel digital data stream. This 9-bit data is expanded...
to 16-bit data to include 7-bit filler data for channel separation purpose at the receiver side and serialized again in the encoder. This final serialized data at the rate of 81.92 Mbps is then Manchester coded at the UWB transmitter to generate UWB pulses, which are transmitted through an off-chip UWB antenna. A pulse shaping filter is used to ensure that the emitted power spectrum of the UWB pulses is under the FCC regulation mask. When operating in DSP mode, a selected channel is connected to the on-the-fly spike feature extraction block and the features are transmitted wirelessly for further processing.

The clock signal of 81.92 MHz is applied to the chip externally and the on-chip internal clock generation circuit provides appropriate clock signals for each circuit blocks as shown in Fig. 1. In this design we chose to use an off-chip crystal oscillator that can supply the required clock signal with a sufficient accuracy. The physical dimension of the commercial available crystal oscillator is small enough to be integrated with this chip in a hermetic sealing to form a higher-level system.

III. FRONT-END BLOCK CIRCUIT DESIGN

The preamplifier (Fig. 3) uses ac-coupling at the input to reject the large dc offset occurring at the electrode-tissue interface [13]. The gain of the preamplifier is 40 dB, which is set by the ratio of feedback capacitances $C_1/C_2$. The second amplifier with a non-inverting resistive negative feedback provides an additional gain of 17–20 dB according to external controls. $C_3$ and...
$C_2$ were chosen to be 20 pF and 200 fF resulting in differential input impedance of 16 MΩ at 1 kHz. The high-frequency roll-off of the preamplifier is configurable from 2 to 20 kHz in 16 steps by varying the load capacitance $C_L$. The low frequency roll-off is tunable from 0.1 to 200 Hz by changing the gate voltage $V_{G3}$ of the NMOS used as bias resistors. The gate voltage provided by the bias generator is adjustable from 600 mV to 1 V by 50 mV steps. The adjusting range and the step size of $V_B$ are designed to achieve the target programmability of low frequency roll-off in the presence of the process variations.

Operational transconductance amplifiers (OTA) have critical effects on the overall performance of the preamplifier. A high common-mode rejection ratio (CMRR) is preferred to suppress the 60 Hz power interference in the neural recordings [19], which makes the preamplifiers saturated and disables signal processing at the later stages of the system. A proposed fully differential self-biased OTA [20] shown in Fig. 3 enables a 90 dB CMRR and a 80 dB power supply rejection ratio (PSRR) with 4.9 $\mu$Vrms input referred noise integrated from 0.1 Hz to 20 kHz. To improve the common mode noise rejection a fully differential output signaling was chosen. A common mode feedback (CMFB) circuit is not required because the output common voltage level is self-biased by the negative feedback leading to low power and small area of the preamplifier.

Each preamplifier and buffer draws 2 and 20.3 $\mu$A, respectively, to drive the analog multiplexer when it is turned on by the sequential turn-on control signals, leading to an average current of 2.54 $\mu$A. The second amplifier draws 40.6 $\mu$A to drive 10 pF input sampling capacitance of the SAR ADC.

Care was taken to minimize the coupling noise from adjacent channels. All input signal lines that connect the inputs of the preamplifiers to bonding pads were shielded by quiet ground lines. In addition, the space between input signal line and shielding line was determined so that the metal parasitic capacitance is small enough not to affect the input impedance of the preamplifiers.

SAR ADC is well suited for low-power and small-area applications because it requires minimal amount of analog circuitry [21]. The resolution of the ADC can be adjusted from 6 to 9 bits by external control signals. The total sampling capacitance of the SAR ADC is 10 pF and the power consumption is 14 $\mu$W, which is very small compared to other analog circuits. To reject common mode noise, the ADC is designed to digitize differential signal directly. The block diagram of the SAR ADC is given in Fig. 4. The reference voltages are supplied by the on-chip DC voltage generators and the voltage levels controlled by the external control signals are variable from 100 to 500 mV by 50 mV steps. The comparator of the ADC is based on track-and-latch comparator and its schematic is also shown in Fig. 4. The sizes of the transistors N1, N2, N3, N4, P1, and P2 can directly affect the offset of the ADC and therefore are chosen large enough to guarantee that offset is only a few least significant bits (LSBs) when minimum reference voltage is applied.

IV. ON-THE-FLY SPIKE FEATURE EXTRACTION

The recorded neural signals by a single electrode may contain spikes from multiple firing neurons. Spike sorting is the procedure to attribute spikes to individual neurons and recover
the encoded information in the neural signals. Spike sorting is challenging due to several factors, such as the presence of unresolved neuronal activity, and similarity in recorded shapes. Training may be required when the electrodes move, and this could be frequent in certain applications [22]. The frequent training over a large amount of recording channels significantly increases the computational power and memory required [23]. Thus, for an implantable neural recording system where the hardware needs to be integrated and miniaturized, power and area efficient algorithms are advantageous.

Fig. 5 shows the overview of the algorithm that enables on-chip spike feature extraction. The 9-bit ADC output containing the time multiplexed neural signals are fed to the spike detector. A quantitative evaluation of different spike detection was reported in [24] using synthesized spikes. The results indicated that all the surveyed detection algorithms performed comparably, except an implementation of template marching is coasty. In this work, we chose a nonlinear energy operator (NEO) [27], [28] due to its efficiency in separating spikes from background activities, which was reported to exhibit a low frequency fashion through power spectrum measurement of data from in vivo experiments [25], [26]. Realization of NEO using analog circuit was also reported [29], however, the involved design complexity and circuit size exceed those of digital implementation.

According to [25], a derivative based frequency-shaping filter significantly attenuates the low frequency noise and help differentiating similar spikes from different neurons. As a complementary approach to principle component analysis, spike feature extraction algorithm based on informative sample set was first reported in [26] to identify uncorrelated local features. This concept requires only a subset of samples containing the necessary information to cluster the data. Intuitively, a sample is considered to be informative if the superimposed spikes can be classified into multiple clusters by evaluating the sample alone. Combining derivative operation and sample selection, improved sorting results were demonstrated in [26]. As a preliminary implementation to our feature extraction algorithm [30], the peaks of the original spike waveforms and maximum and minimum values of its first derivatives are used as the features to classify spikes. The choice of this simplified sample set for implementation is based on three reasons. First, it requires small computation and little memory requirement [30]. Second, samples during the fast transition period frequently exhibit high information score [26]. Third, obtaining these three features requires no training.

Fig. 6 shows the block diagram of the proposed on-the-fly spike detection and feature extraction engine. The 9-bit serial ADC output containing the multiplexed neural signals is fed to the spike detection and sorting hardware units. The serial output data after the sorting consist of 64 bits of three features and time information of the spikes. Each spike is processed, encoded and transmitted on-the-fly within 41 cycles (1.025 ms). After the system reset, the coefficients register array is loaded with 32 9-bit filter coefficients and a 16-bit threshold value before the recording procedure. These values are derived by off-chip computation based on few seconds of initial test recordings. If a spike is recognized, the feature scores are calculated by the min–max detector, and then sent to the wireless telemetry module through coder/packer. These operations, including detection, filtering, feature extraction, and coding, are performed in parallel to meet the real-time requirement.

V. UWB TELEMETRY

Although there are several communication standards for biomedical applications, most of them cannot provide enough bandwidth for simultaneous recording from more than 100 channels. For example, MICS band, which is allocated for the unlicensed use of implantable devices, allows only 300 kHz for data transmission. Other wireless technologies such as Wi-Fi and Bluetooth cannot be used directly due to the power and area constraint of the implanted device. Therefore, there is a need for higher-bandwidth data transmission telemetry that consumes low-power and occupies less physical area.

UWB is recently formed wireless technology that is used for low/high data rate wireless personal area network (WPAN)
and short-range applications. It has the widest bandwidth (3.1–10.6 GHz) among all technologies and the smallest emission power density (−41.3 dBm/MHz), providing an opportunity for a wideband wireless telemetry for neural recording systems. Although the reported receiver designs of UWB systems have been consuming more power than those of narrow band communications [31], in our neural recording applications the implanted device requires only a transmitter and the receiver is outside the body, unlike the other short range applications. Hence the transmitter power consumption and complexity have been traded off with that of the receiver as the receiver is located outside and its power consumption and size are not crucial. This greatly simplifies the complexity of the implanted telemetry design leading to reduction in power and area.

In this system, impulse radio based UWB (IR-UWB) is employed as the targeted application does not require a multiple access communication, which requires complex transmitters consuming high power [32], [33]. In IR-UWB, short pulses are generated for sending data, and both the center frequency and bandwidth of the pulse do not have to be very accurate. Such a process is simple to design in complementary metal–oxide–semiconductor (CMOS) technology resulting in a very simple, small-area, and low-power transmitter design while providing enough data bandwidth because of its wideband nature.

Fig. 7 shows the block diagram of the UWB transmitter. The sampled and serialized data is Manchester encoded first and then either on-off keying (OOK) or pulse-position modulation (PPM) is employed to generate short pulses. The redundant fillers are
Fig. 8. Data format of the UWB transmitter.

Fig. 9. A high-speed, USB2.0, data acquisition system designed to support 80 Mbit/s neural recording. Trace 4 depicts original serial data stream containing long logic-lows as separators and short transition bits as recorded samples. A packet marker monitors this serial data stream and generates markers for the start and end of a data sample as depicted by Trace 2 and 3, respectively. Trace 1 presents decoded bits streams.

set to all “1” for channel (0), and all “0” for other channels in case of PPM, and the fillers for channel (0) are set to 7-bit “logic low” of Manchester coding in case of OOK (Fig. 8). Therefore, the final maximum pulse repetition rate at the UWB transmitter output is twice the output data rate, 163.84 M pulses/s in case of using PPM and the same as the output data rate, 81.92 M pulses/s for OOK. A short pulse generator is based on a simple edge detector [34] and the circuit implementation is shown in Fig. 7. The pulse width can be controlled from 180 to 980 ps by an external control voltage for adjusting the transmitting power. The generated short pulses are then passed through a high pass filter to remove low frequency component of the pulses and for the transmitting power to fit under the FCC emission mask. The filtered pulses are finally fed into the off-chip UWB antenna. Because of the high-frequency nature of the UWB, the size of the antenna is small enough to be implantable or to be carried by the animals.

The UWB receiver was built from off-the-shelf components (Fig. 7) since it is outside the biological objects. The signal received by an UWB antenna is passed through a 1 GHz band-pass filter (BPF), whose center frequency is 4 GHz, to remove the interfering signals from other narrowband wireless devices. The signal is then amplified by the low-noise amplifier (LNA) stages, which plays a critical role in determining the maximum distance between the implanted transmitter and receiver outside the biological subject. A RF diode and a low-pass filter (LPF) down converts the UWB signal to low frequency one by performing envelope detection and the digital data is finally recovered by FGPA (Fig. 7).

Since the maximum data rate of the recording exceeds 80 Mbit/s, an interface between the receiver and a PC must be able to support that bandwidth. Among several standard input/output interface buses for PCs, USB2.0 and FireWire (IEEE 1394) are the two most popular buses that can support the required bandwidth. Fig. 9 shows a block diagram of the PC interface realized to support USB 2.0. The interface consists of phase-locked loop (PLL) for clock recovery, an FPGA for decoding and deserializing, and USB2.0 chips to store recording data into a standard personal computer.

VI. TEST RESULTS

The IC was fabricated in a 0.35 μm 4 M2 P CMOS process and the chip size is 8.8 × 7.2 mm² (Fig. 10). The 8 front-end blocks consume 50% of the total power. The UWB transmitter
circuitry consumes 1.6 mW with PPM. The power consumption of the DSP block is only 0.1 mW and the power consumption caused by leakage current is negligible in 0.35 μm CMOS process. Test bench was first performed with multiple signal generators connected to the inputs of the front-end blocks to evaluate and characterize the performance of each circuit blocks and is summarized in Table I.

Extracellular recordings were made from a dissected snail brain to verify the functionality of the entire system in a real recording environment. An intact circumesophageal ring was extracted from *H. aspersa* and placed in a glass dish filled with Ringer’s solution. A concentric bipolar electrode with 25 μm platinum tip was placed close to circumesophageal ring by micromanipulators. The biological experimental setup for *ex vivo* test is shown in Fig. 11. The measured impedance of the electrode in solution was 190 kΩ at 1 kHz. The gain of the amplifier was set to its maximum value because the amplitude of the extracellular action potential from the dissected snail brain was less than 200 μV due to a small value of the seal resistance, which determines how closely the neuron is attached to the measuring electrode. Because most of the signal energy of the extracellular action potential is concentrated between 100 Hz and 10 kHz, the high-frequency roll-off was set to 10 kHz. The low-frequency roll-off is set to 100 Hz eliminating the 60 Hz interference from the power line. Fig. 11 also shows the obtained waveforms at the output of amplifier, spike sorting DSP, UWB transmitter, and UWB receiver. The physical dimension of the UWB antenna used in the experiment is 10 × 10 × 0.8 mm³. The power spectral density (PSD) of the transmitted UWB pulses was measured with spectrum analyzer and the result shows that the PSD is within the FCC mask regulations.

### Table I

<table>
<thead>
<tr>
<th>Number of channels</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal gain of the preamp</td>
<td>40dB</td>
</tr>
<tr>
<td>Input impedance (at 1kHz)</td>
<td>16MΩ</td>
</tr>
<tr>
<td>Input referred noise</td>
<td>4.9μVrms</td>
</tr>
<tr>
<td>CMRR or the preamp</td>
<td>90dB</td>
</tr>
<tr>
<td>PSRR of the preamp</td>
<td>80dB</td>
</tr>
<tr>
<td>LF roll-off of the preamp</td>
<td>0.1Hz ~ 200kHz</td>
</tr>
<tr>
<td>HF roll-off of the preamp</td>
<td>2kHz ~ 20kHz</td>
</tr>
<tr>
<td>Signal gain of the 2nd amp</td>
<td>17dB ~ 20dB</td>
</tr>
<tr>
<td>ADC resolution</td>
<td>6 ~ 9 bits</td>
</tr>
<tr>
<td>ADC sampling rate</td>
<td>640ksample/sec</td>
</tr>
<tr>
<td>DSP’s functionality</td>
<td>On-the-fly spike feature extraction for one selected channel</td>
</tr>
<tr>
<td>Power dissipated by DSP</td>
<td>0.1mW</td>
</tr>
<tr>
<td>Maximum UWB data rate</td>
<td>90Mbps</td>
</tr>
<tr>
<td>Power dissipated by UWB</td>
<td>1.6mW</td>
</tr>
<tr>
<td>Power supply level</td>
<td>±1.65V</td>
</tr>
<tr>
<td>Total chip power dissipation</td>
<td>6.0mW</td>
</tr>
<tr>
<td>Technology</td>
<td>0.35μm 4M2P CMOS</td>
</tr>
<tr>
<td>Total chip area</td>
<td>8.8mm × 7.2mm</td>
</tr>
</tbody>
</table>

### VII. Conclusion

A low-power fully integrated single chip neural recording system is presented in this paper for wireless recording of 128 channels simultaneously. The on-the-fly detection and feature extraction for one selected channel is implemented on this chip. Simultaneous processing for 128 channels with more functionalities of neural classification and decoding will be integrated in the future. The sequential turn-on method is used to minimize the power consumption of front-end blocks, which are the most power consuming circuitries. The wireless transmission of high data rate was achieved through a low-power implantable UWB wireless transmitter. The total power consumption is only 6 mW even when transmitting raw data from all 128 channels. The front-end has been designed to have amplifiers with programmable gain and bandwidth to offer flexibility in recording various bio-potentials. The bench test was carried out to evaluate the performance of the chip, and *ex vivo* extracellular recording was done to verify the proper operation of the system in a real environment.
Fig. 11. Photograph of experimental setup ex vivo and measurement results of the IC with neural signal from a snail brain.

ACKNOWLEDGMENT

The authors thank E. Basham for his help in snail dissection for testing setup of the chip ex vivo and, T. Chen for layout synthesis, J. Kim for help in circuit design, and B. Liang for help in bench top test.

REFERENCES


Moo Sung Chae received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1998 and 2000, respectively. He is currently working toward the Ph.D. degree at the University of California, Santa Cruz. In 2000, he joined Samsung Electronics, Korea, where he had been working on circuit design of high-speed DRAMs. His research interests are integrated neural recording systems, neural prosthetic devices, and retinal prostheses.