We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Data Reduction Techniques in Neural Recording Microsystems

Mohsen Judy, Alireza Akhavian and Farzad Asgarian

Additional information is available at the end of the chapter

1. Introduction

Nowadays, implantable devices developed for electrically interfacing to the brain are of great interest. Such devices, also known as *brain-machine interfaces (BMI)*, are expected to revolutionalize so many aspects of the human life, such as the way we interface with the external world, and how we cure deseases and disabilities such as the Parkinson's desease, paralysis, and blindness. General concept of intra-cortical neural recording using implantable microsystems along with an example of such systems is illustrated in Figure 1. In a wide variety of applications for such systems, there is a need for recording neural activities from a certain region of the brain with enough spatial resolution. To be able to come up with meaningful information from the region of interest in the brain, implantable neural recording devices are typically designed to record from tens to hundreds of recording sites [1-3].

1.1. General building blocks

A neural recording system, in general, comprises two parts: a neural recording implant, and an external setup. Implantable cortical neural recoding microsystems (the implant) typically consist of three main parts:

- **neural recording front-end;** This module is in charge of sensing extracellular neural activities, and consists of a recording microelectrode array followed by analog signal preconditioning circuitry. A 4-site silicon probe fabricated based on the Michigan approach [4] is shown in Figure 2 [5].

- **neural signal processing module;** This is where most of the signal handling and signal processing tasks take place, e.g., analog signal processing, analog/digital conversion, and digital signal processing.
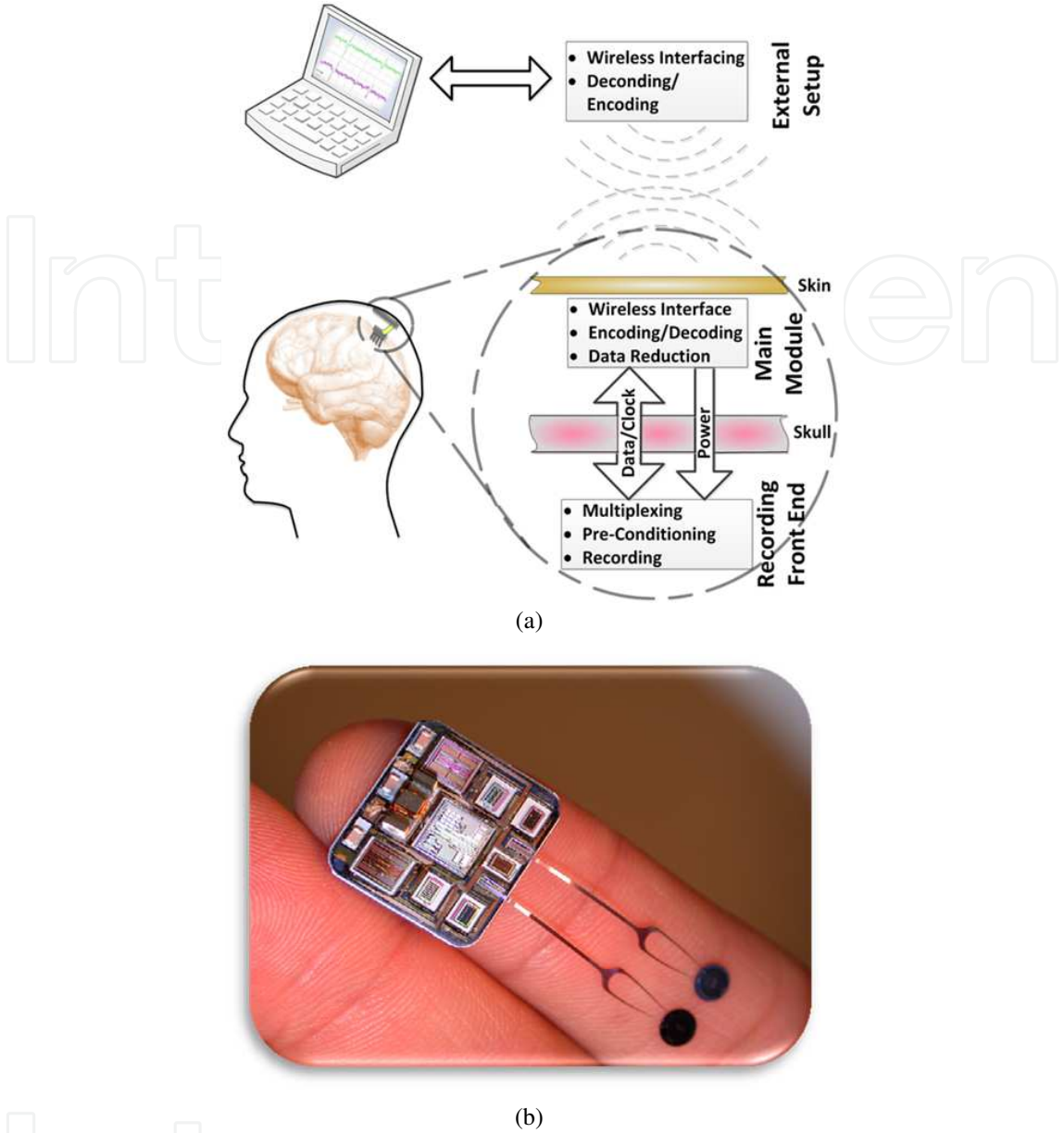
(a)

(b)

Figure 1. (a) Intra-cortical neural recording using an implantable microsystem [6] (b) a 64-channel neural recording microsystem developed at the University of Michigan [1]

- **wireless interface module;** This module is used for data exchange with the external setup and in some cases for supporting power telemetry from the outside to the implant.

1.2. Challenges in the development of high-density neural recording microsystems

As the number of recording channels for a wireless neural recording microsystem increases, many aspects of the design of the system will be challenging. From among the more important design challenges, one can point to the low power consumption and small physical dimensions of the system. For tens to hundreds of recording channels, transferring a huge amount of neural
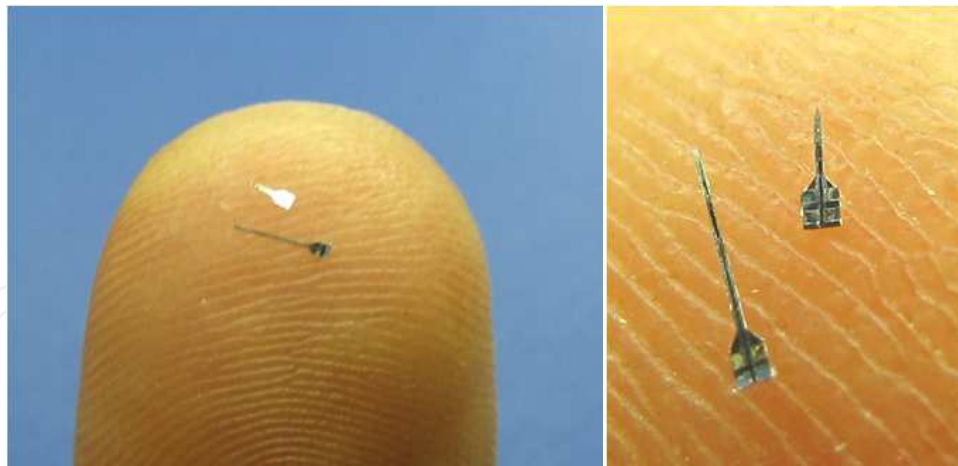
Figure 2. Passive silicon probes with four 20μmx20μm recording sites [5]

data through a wireless link is also a design bottleneck. This is simply due to the fact that an implantable neural recording device needs to transmit the recorded neural information to the external world through wireless connection, and the frequency band used for wireless communication is not unlimited. One of the efficient ways to overcome this problem is to either compress the data being telemetered or at least to extract and transmit only the useful information needed for the target application.

2. Spike reporting

An intra-cortically-recorded neural signal, in general, comprises three major components: *action potentials* (also known as *neural spikes* or simply *spikes*), *local field potential (LFP)*, and *background noise*. It is believed that most of the important information in neural signals is reflected in the occurrence rate of neural spikes. As a result, in some applications (e.g., prosthetic applications) only the occurrence of spikes is detected and reported to the external world. In some other applications (e.g., neuroscientific research), however, researchers and scientists need more information on how or where the neural activities occur.

Recording the entire neural signal (action potentials superposed with background noise) is the maximum function expected from a general neural recording system, which allows for studying different components of a neural signal including the background noise. For multi-channel wireless neural recording implants, because of the limited bandwidth available for transmitting the neural data, the number of recording channels will be limited if the entire signal is intended to be telemetered. In many applications, the rate of spike occurrence is the most important information that is expected from a neural recording system to provide. Hence, it will be much more bandwidth-efficient if the spikes are detected by the implanted recording system and only the occurrences of the spikes are reported to the external host rather than transmitting the entire neural signal.

2.1. Spike detection

In addition to the small action potentials with the amplitude of around 100~500μV, a neural signal contains background noise and probably low-frequency baseline variations. To prepare the neural signal for spike detection, it is amplified with a gain of around 40~60dB and also its low-frequency (below 1~10Hz) and high-frequency (above 7~10kHz) contents are filtered out [7], [2]. Then, this preconditioned signal is delivered to a spike detector.

There are various spike detection approaches that can be classified into two major categories: feature-based spike detection methods, and spike detection by hard thresholding. In the former, a preprocessor searches the input neural signal for certain features of a spike to occur, while in the latter, a threshold level is defined and a spike is detected when the neural signal goes beyond the threshold.

Feature-Based Approaches. Only a few years after artificial neural networks (ANN's) were introduced as an efficient tool to implement artificial intelligence, due to the feature extraction capability that certain types of ANN's had, they showed to be attractive candidates for automatic spike detection either by themselves or in conjunction with preprocessors [8-9]. Kohonen and Grossberg networks with unsupervised learning, and Multi-Layered Perceptron network with "error back-propagation" as a supervised learning algorithm have been used to perform spike detection. Although there are cases where the raw neural signal is fed to the ANN for spike detection [9-10], it is mostly preferred to use a preprocessor for extracting certain spike features first, and then use an ANN for processing them and detecting the spikes, as illustrated in Figure 3. Because of their relatively large area- and power-consuming electronic implementations, this class of spike processors has never been used in implantable neural recording microsystems.
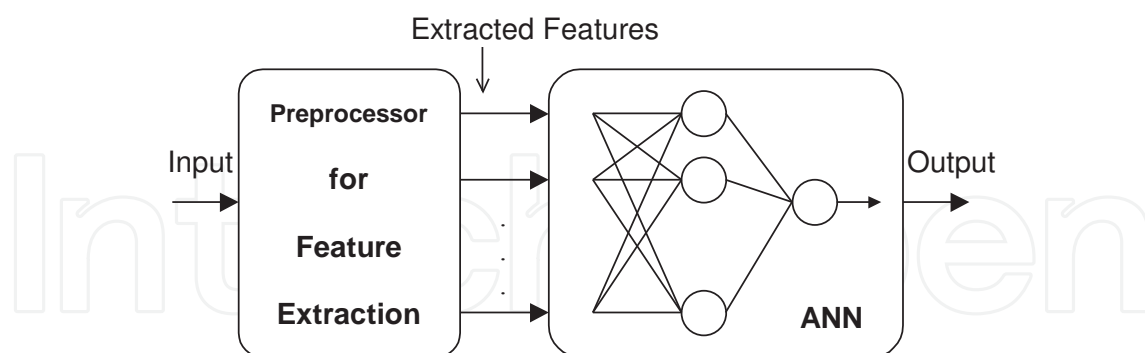

Figure 3. Employing neural networks for spike detection

Spike Detection Based on Nonlinear Energy Operator (NEO). Traditional spike detectors (explained above) usually need prior information about action potentials, which is usually not available before recording the neural signals in real applications. In contrast with these methods, which are mostly based on the amplitude or time-domain features of the neural signal, detection of action potentials, i.e., spikes, can also be performed based on the energy content of the signal. Direct square of the signal, absolute value operator, and variance

estimator are the energy-based operators commonly used for the detection of bio-potentials. The nonlinear energy operator (NEO), also called the Teager energy operator (TEO) is one of the unsupervised action potential detectors exhibiting satisfactory performance in the case of low signal-to-noise ratio (SNR) for the neural signal and also convincing speed for the alignment of spikes in the real time. In its original form, continuous-time NEO is defined as:

$$\psi(x(t)) = \left(\frac{dx(t)}{dt}\right)^2 - x(t)\left(\frac{d^2 x(t)}{dt^2}\right). \tag{1}$$

which is sensitive to signal with short time interval and at a high-frequency band [11]. An efficient hardware implementation for an NEO neural signal processor employing custom OTA[1]-C analog circuits was reported in [12].

2.2. Hard thresholding

Four possible ways of spike detection by hard thresholding are illustrated in Figure 4. Spike detection is mostly performed to detect either positive (Figure 4 (a)) or negative (Figure 4 (b)) spikes [7], [2], [13]. Having such a fixed pre-assumption for the polarity of the spikes limits the operation of the system. Recognition of both positive and negative spikes, i.e., bi-phasic spike detection, can be realized in two major ways, illustrated in Figure 4 (c) and (d). In Figure 4 (c), the spike detector returns a logical "1" on the Spike Occurrence (S.O.) output upon the detection of a spike, no matter if it is positive or negative. This is a bandwidth-efficient way of bi-phasic spike detection, which requires almost the same bandwidth as the uni-phasic methods, but pays the price by losing the spike polarity. The simplest realization of this idea is to filter out the DC component of the input signal, find its absolute value, and then detect the spikes using one comparator and one threshold [14] as shown in Figure 5(a). Aside from the need for a precise full-wave rectifier in this realization, the fact that both the positive and the negative spikes are compared with the same threshold level might be considered as a drawback. Figure 5 (b) shows another realization of bi-phasic spike detection with no polarity, which uses two comparators and an OR gate and also allows for comparing positive and negative spikes with separate threshold levels. This approach is used in [7],[13] with positive and negative thresholds, $V_{TH,P}$ and $V_{TH,N}$, defined by a threshold value (THR) and a threshold offset (ThrOS), as shown in Figure 6.

The circuit shown in Figure 5 (b) can also be used with minor modifications to realize the complete bi-phasic spike detection method illustrated in Figure 4 (d), which returns two bits per detected spike. These two bits can be either Spike Occurrence (S.O.) and Spike Polarity (S.P.) as shown, or one bit assigned to detected positive spikes and the other to detected negative spikes.
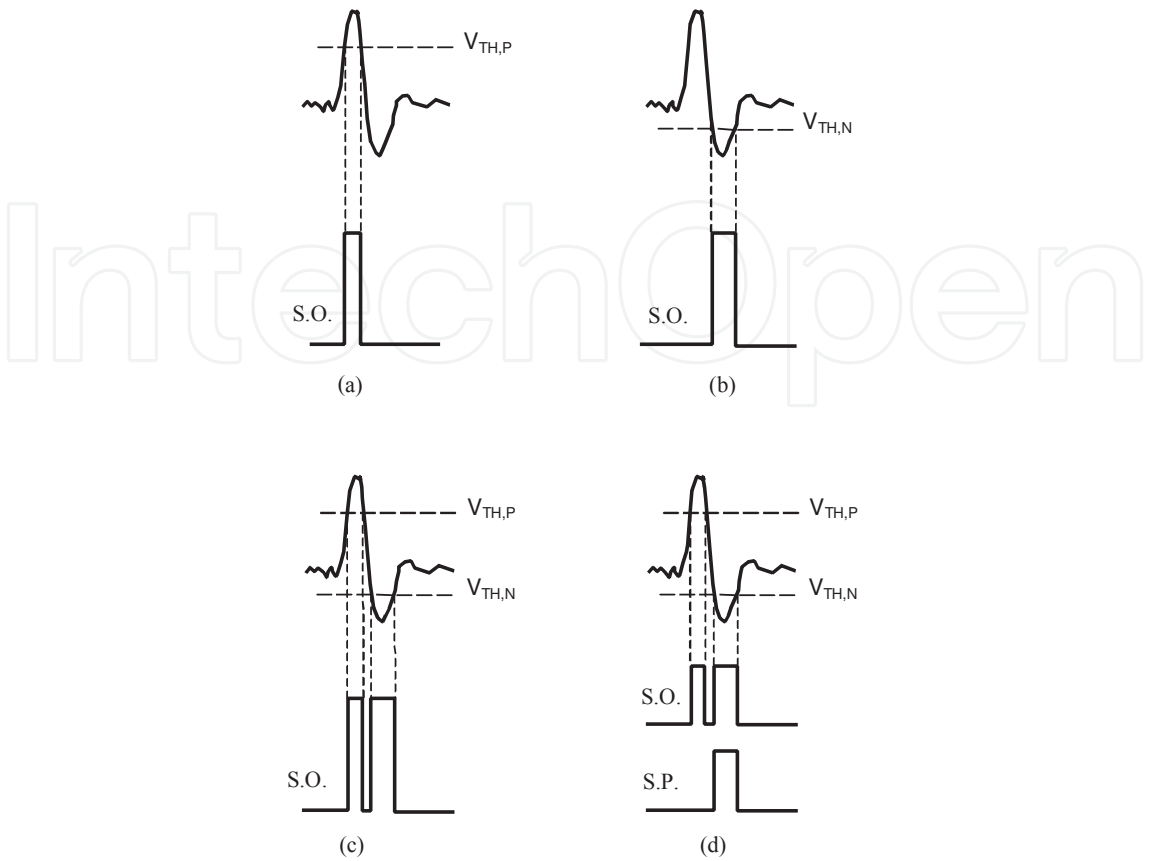
---

1 Operational Transconductor Amplifier

Figure 4. Spike detection approaches (a) Positive (b) Negative (c) Simple bi-phasic (d) Bi-phasic with spike polarity [13]
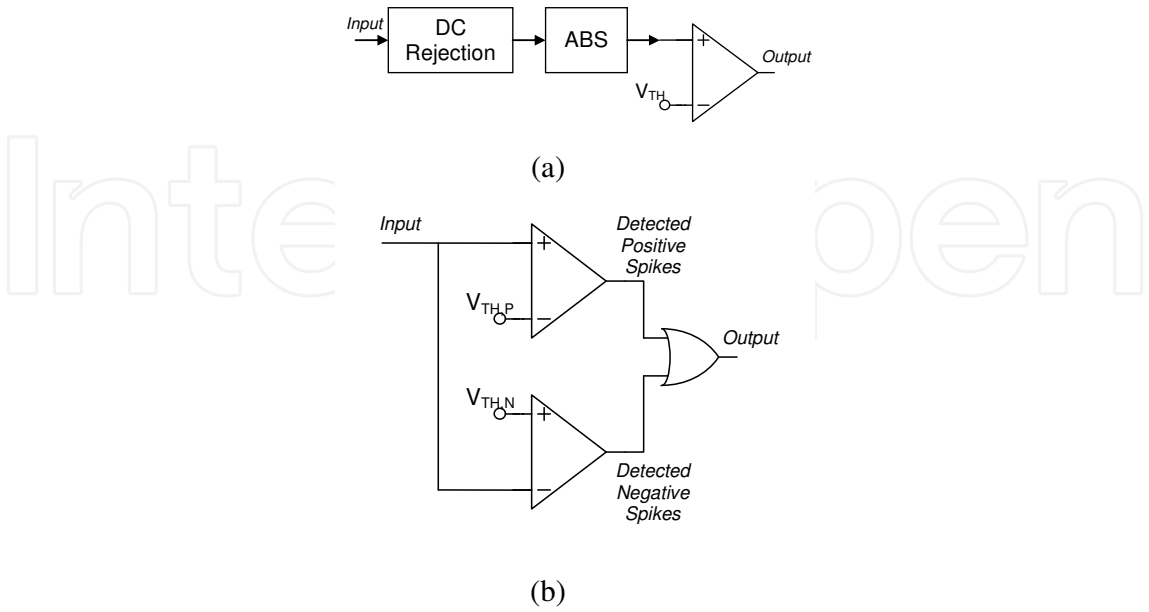

Figure 5. Implementations of biphasic spike detection with no polarity

There is a variety of methods for generating the thresholds required for spike detection. The threshold can be either statically defined by the user [2], [13] or automatically set by the internal circuitry.

Automatic Threshold Setting. The 32-channel spike detector ASIC reported by [7] uses a straightforward approach for automatic threshold generation. In this method, the average (AVG) and the standard deviation (SD) of the neural signal is calculated, and then the two thresholds required for bi-phasic spike detection are set above and below the average value as:

$$THR = AVG \pm k.SD \tag{2}$$

where k is a constant. Typical value for k varies from 3 to 7 depending on the signal-to-noise ratio (SNR) of the recorded neural signal. Functional block diagram of this spike processor implementing the above method in digital domain is shown in Figure. Thirty two channels of preconditioned neural signals, which are already time-division multiplexed on four lines in analog domain by a recording front-end (not shown), are delivered to the spike detector ASIC. The four multiplexed inputs each carrying 8 channels of neural signals are first converted to digital by four A/D converters simultaneously. The *Sample Distributor*, which is synchronized with the time-division multiplexer on the recording front-end, demultiplexes the amplitude samples into 32 digital neural channels. Then, the digital spike detector computes the averages and standard deviations of the 32 channels separately and accordingly calculates their threshold values. After the *Threshold Calculation* program is executed, the *Spike Detection* program is run. The amplitude sample of each channel is compared to the associated threshold, and if it is beyond the threshold level, it is considered as a detected spike. As long as a channel is active, its amplitude samples are tagged with the associated channel address, and put in a buffer to be sent to a wireless interface.
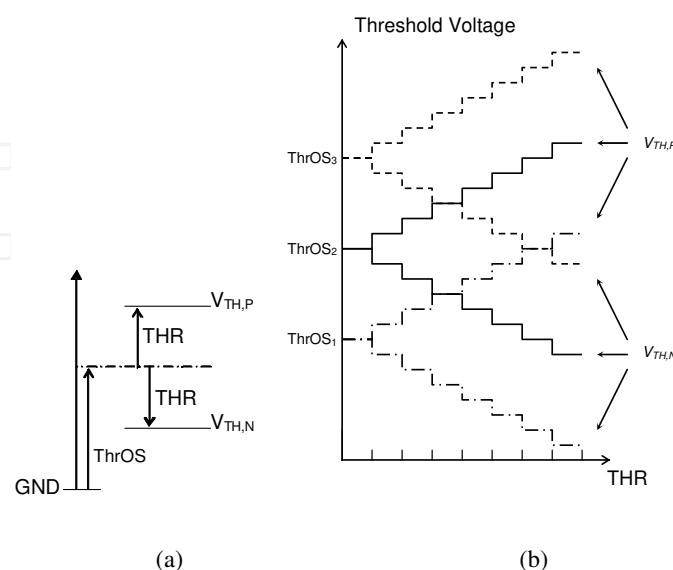

(a)   (b)

Figure 6. Defining positive and negative thresholds using a threshold value (THR) and an offset (ThrOS) [13]
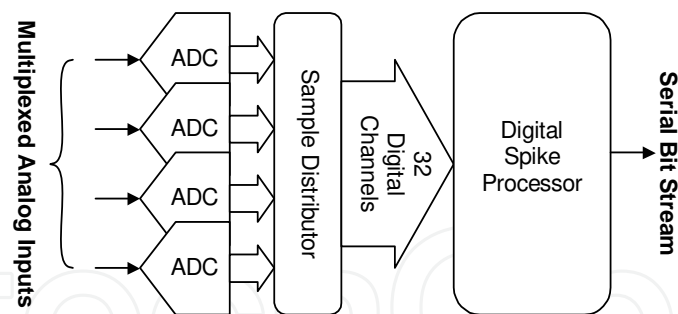
Figure 7. Functional block diagram of the 32-channel digital spike detector ASIC reported in [7]

In [15], a spike detector circuit is reported that computes the detection threshold in analog domain. Functional diagram of this circuit is shown in Figure 8. One of the advantages of this circuit is that unlike the spike detector in [7], the threshold is computed in real time. The circuit assumes that the input signal has already been amplified and band-pass filtered, and the background noise has a Gaussian distribution. Since the input signal is assumed to have no DC component, the noise can be described by its RMS value, $V_{1\sigma}$, which is equivalent to its standard deviation, σ. In order to be well above the noise level, the threshold voltage is set to $V_{K\sigma} = K.V_{1\sigma}$, with K set to 5 in [15]. Although it is assumed that the low-frequency baseline variations of the input signal have been already filtered out, further analysis of this method shows that the detection threshold can adaptively follow the baseline variations. There is, however, an upper bound on the frequency-amplitude product of the baseline variations that the adaptive threshold is capable of following [16]. Power dissipation of the implementation of this approach is very low, in the range of microwatt and occupies very small silicon area. The drawbacks of this method include the circuit's sensitivity to the absolute value of some of the circuit elements, which are usually subject to relatively large fabrication tolerances, and the difficulty in implementing the low-frequency low-pass filter required in the RMS block.
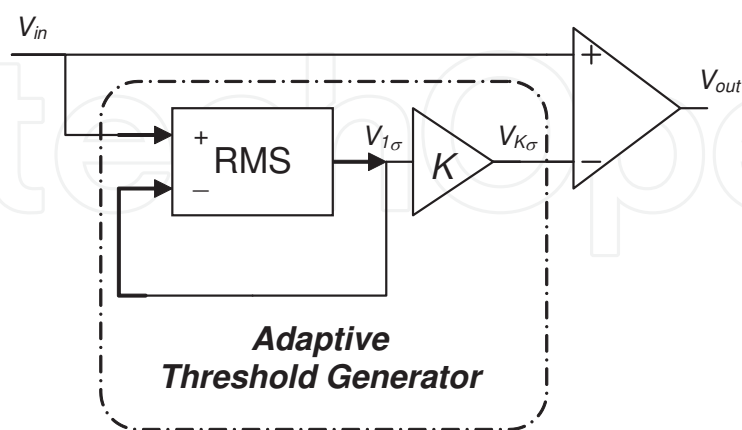

Figure 8. Functional diagram of the analog spike detector of [15]

Figure 9. shows another idea in analog spike detection, in which two OTA-based low-pass filters with different cut-off frequencies play the key role [17]. One filter has a higher cut-off

frequency to remove high frequency noise, and the other has a lower cut-off frequency to make a local average. The difference between the high-pass filtered signal and its local average is provided by an OTA, and is recognized by another OTA as a detected spike when exceeds a certain reference value (V_{ref}). This method is robust against changes in both noise level and the input signal's DC offsets, both of which are likely to happen in long-term neural recording. The OTAs operate in subthreshold region to reduce power. The τ bias voltages are set off chip to enable adjustment of the cut-off frequencies after fabrication. In this circuit, the spike detection threshold level is set by V_{bias}, which along with the other bias and reference voltages should be properly set, and probably fine-tuned for long-term recordings.
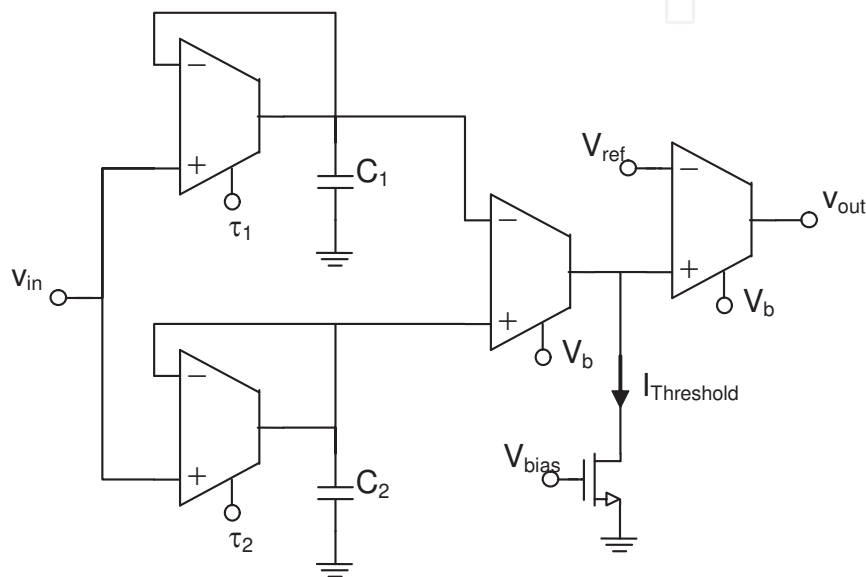

Figure 9. Schematic diagram of the analog spike detector of [16]

In both of the above analog approaches, there are device and circuit parameters that should be set by the user for proper operation, which make these circuits inappropriate for implantable applications.

3. Mathematical approaches

Mathematical transforms are among the most common methods of data compression. Recently, the Discrete Wavelet Transform (DWT) has been successfully employed in neural recording microsystems to compress the neural information, while preserving the wave shape of action potentials [18-26]. The DWT transforms discrete signals from the time domain into the time-frequency domain. One-level DWT for a given signal is achieved by convolving the signal samples through low-pass and high-pass decomposition filters [19,21]. The filtering is then followed by sub sampling to obtain the *approximation* and *detail* coefficients. For multi-level DWT, approximation coefficients should be fed to the same decomposition filters recursively [27]. Characteristics of the filters are determined by the wavelet basis. For neural

signal compression, the optimal choice is a wavelet function that can approximate the action potentials waveform with minimum DWT coefficients and error. It has been shown that by proper selection of the wavelet basis, most of the spike energy is concentrated in a few large coefficients, while many small coefficients carry insignificant information and are mainly attributed to noise [19]. Therefore, in order to achieve higher data reduction rates, the DWT coefficients are passed through a thresholding stage. In this block, data coefficients smaller than desired certain threshold level are set to zero, while others are left unchanged. It is obvious that the threshold value plays an important role in the overall data compression rate, and also in the quality of the reconstructed signal. Hence, the threshold level should be set carefully based on the requirements of the target application.

Due to power and size constrains in biomedical implants, VLSI implementation of DWT is of great importance. In [20], it has been shown that from a signal compression standpoint, *symmlet4* wavelet basis is advantageous over other wavelet functions for neural recording applications. It is believed that this is mainly because of the similarity of this function to the general wave shape of action potentials. For hardware implementation of *symmlet4* function, lifting method is proposed in [21]. Furthermore, two different circuit designs, pipelined and sequential, are presented and compared for the lifting scheme. It is demonstrated that for single-level single-channel integer DWT, the pipelined approach consumes lower power, but occupies more silicon area compared with the sequential implementation. On the other hand, 4-level multi-channel implementation of the two designs indicates that the sequential approach requires significantly smaller chip area, while the power consumption of both is almost the same. As a result, the sequential execution architecture is employed in [23] to design a complete 32-channel compression system based on the 4-level *symmlet4* DWT. The chip consumes 3 mW of power and occupies only 5.75 mm² in a 0.5-μm CMOS technology. Also, with a sampling rate of 25 KSample/Sec per channel and 10-bits data samples, the system provides data compression of more than 20 times, resulting in a total output bit rate of less than 370 kbps.

In [26], a neural signal compression method based on the Discrete *Haar* Wavelet Transforms (DHWT) is proposed. From the standpoint of data compression, *Haar* basis function may not perform as efficient as complex functions such as high-order *Daubechies* and *symmlet*, but due to its simple hardware implementation, it can be easily used for large number of neural channels with less concern about power and area. As discussed in [26] for two-point DHWT, data coefficients can be calculated by only a buffer, an adder, and a subtractor. Moreover, in order to compare *Haar* and *symmlet4* basis functions, they have been both designed for processing a single channel with 8-bits data samples. Results indicate that before the thresholding stage, relative error (between the original signal and the corresponding reconstructed signal) for the DHWT is only 0.01% larger than the *symmlet4* case, which is obviously negligible. On the other hand, hardware implementation of the DHWT shows around 83% saving in number of transistors, and more than 90% in the occupied silicon area, when physically laid out in a 0.13-μm CMOS technology. The complete 64-channel DHWT-based neural compressor achieves a compression rate of 112 with an error of 2.22%. Additionally, the compressor circuit consumes as low as 0.12 mW @1.2 V supply voltage, and occupies less than 0.1 mm² in a 0.13-

µm CMOS technology. Therefore, it can be said that with this architecture, the gain in circuit simplicity and the bit-rate improvement are much more significant than the penalty paid by the noise added to the signal. However, it is worth mentioning that, in general, the appropriate architecture should be selected based on the application.

4. Hardware approaches

To avoid adding extra power- and area-hungry signal processing blocks for data reduction, and at the same time preserving important information of the neural signals, there is a different category of data reduction techniques, known as *hardware approaches*. These are the approaches focused on modifying the hardware of the recording system in such a way that considerable bit-rate reduction can be achieved. Obvious advantages of these approaches are smaller silicon area and power consumption as compared with the mathematical approaches explained in the previous section.

To benefit from the advantages associated with digital signal processing and also digital data communication (as opposed to their analog counterparts), neural recording devices are commonly designed to convert neural signals into digital as the first step. As a result, analog-to-digital converters (ADCs) are known as one of the key building blocks in such systems. Recently some efforts are put on designing application-specific ADCs to efficiently utilize the bandwidth allocated for wireless data telemetry. In this section, an efficient method for analog-to-digital (A/D) conversion of neural signals is discussed. This method results in significant reduction of data-rate for multi-channel cortical neural recording microsystems.

4.1. Anti-logarithmc quantization

Although linear ADCs are typically used to digitize neural signals in neural recording microsystems, it is beneficial to design a nonlinear ADC for such specific signals. Choosing the best-suited nonlinearity function for a specific signal requires recognition of the concentration of information along the signal amplitude range. As illustrated in the left side of Figure 10, in general, signals can be categorized into three types according to how the information they carry is distributed along the amplitude range.

Type-I signals are named "Signals with *Non-Concentrated Information* (*NCI*)" due to their almost uniform distribution of information concentration. Important information for *Type-II* signals is concentrated at the lower side of the amplitude range. Audio signals are of this type, referred to as "signals with *Information Concentration at Low Amplitudes* (*ICLA*)". Conversely, for "signals with *Information Concentration at High Amplitudes* (*ICHA*)", i.e., *Type-III* signals, more information is present at higher side of the amplitude range, with neural signals as examples.

Figure 10 provides intuitive illustration of the choice of different quantization functions for the three signal types discussed. With a constant slope (i.e., linear) quantization function, an NCI signal is better digitized. Decreasing slope quantization functions such as logarithmic function are recommended for ICLA signals. These functions put more emphasis on lower

amplitudes, where more information is concentrated. For example, compressing/expanding (companding) of audio signals in communications systems is based on logarithmic quantization, which increases the dynamic range and improves the SNR [28]. Whereas quantization functions with increasing slope along the input amplitude range, such as the exponential function, put more resolution in the quantization of the larger amplitudes and are preferred for ICHA signals.
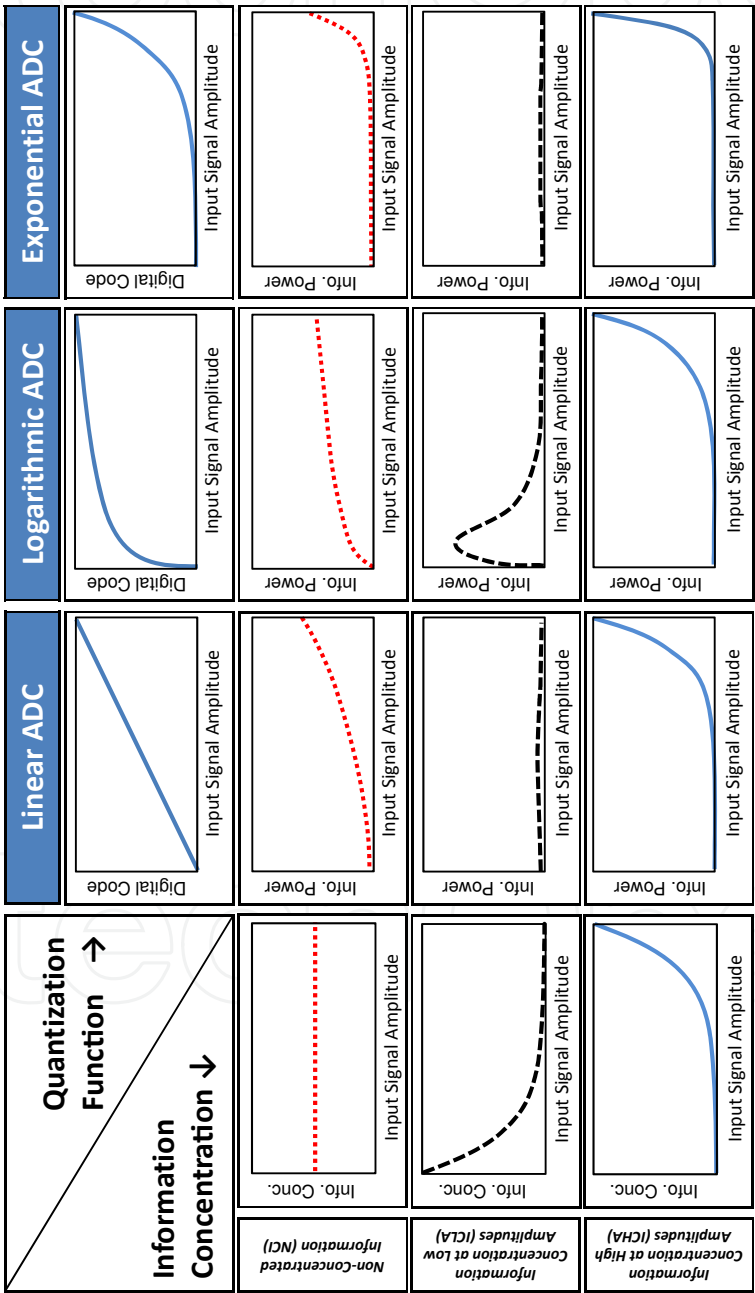

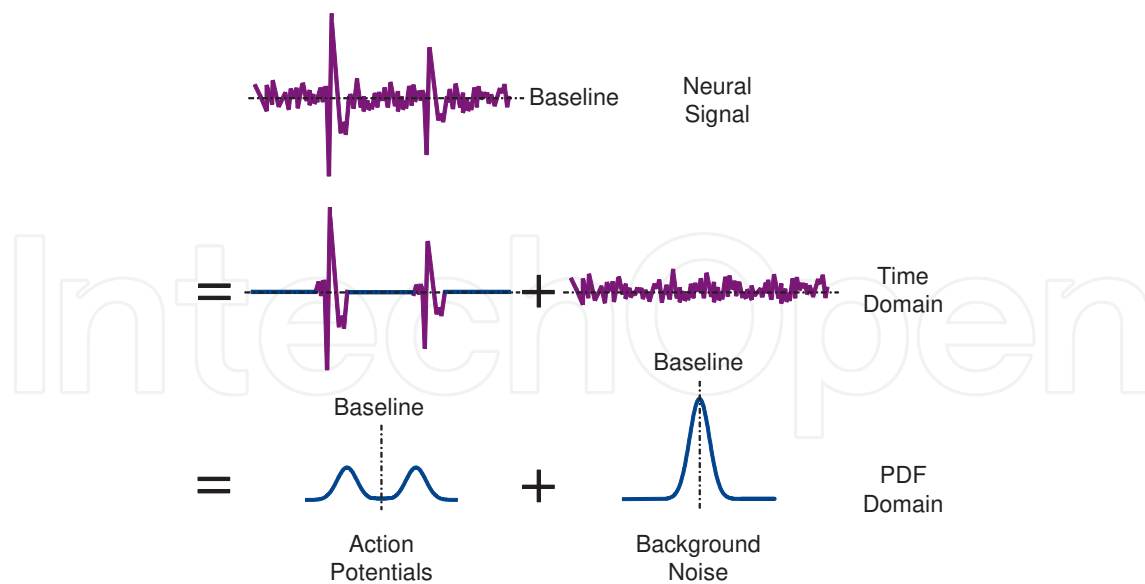**Figure 10.** Selecting quantization function for NCI, ICLA, and ICHA signals

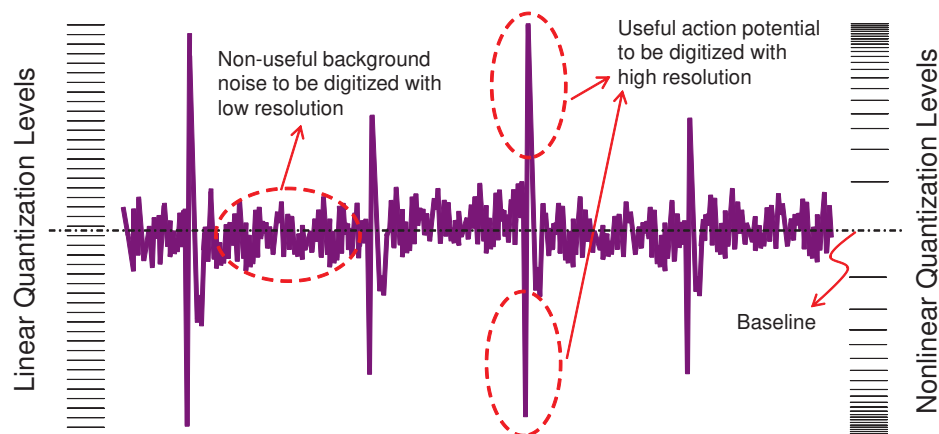**Figure 11.** Neural signal composed of APs and B-Noise

Figure 12. Quantization of a neural signal with linear and nonlinear quantization steps

Basic Idea. As shown in Figure 11, in time domain a typical intracortically-recorded extracel-lular neural signal can be divided into two parts: action potentials (APs) and background noise (B-Noise). In probability density function (PDF) domain, APs are concentrated at large amplitudes while B-Noise is concentrated at small amplitudes. In a wide variety of neuro-scientific and neurophysiological studies, as well as in many neuroprosthetic applications, it is the APs that carry the useful information embedded in neural signals. As illustrated in Figure 12, in implantable neural recording microdevices, neural signals are usually digitized using linear ADCs, i.e., ADCs with linear quantization characteristics. This means that the non-useful B-Noise is digitized with the same resolution as the useful APs are. In other words, when telemetering a digitized neural signal, part of the outgoing bit-rate is wasted to carry the noise content present in the neural signal. In [29] the idea of digitizing neural signals using an ADC

with non-uniform quantization steps has been proposed. According to the classification presented in the previous section, neural signals are categorized under Type-III (i.e., ICHA). Hence, the best type of nonlinearity function for the quantization of neural signals is signals with increasing slopes such as parabolic and exponential functions.

Digitizing neural signals using exponential ADC (exp-ADC) helps saving the bandwidth in wireless data telemetry between the implanted device and the external host. Data reduction for an 8-bit exp-ADC is 24% as compared with its linear counterpart. Along with data reduction, anti-logarithmic quantization of neural signals significantly reduces the power consumption of the ADC, comparing with a standard linear ADC. This is due to less number of digital code transitions for the exp-ADC. Moreover, anti-logarithmic quantization increases the SNR of neural signal by reducing its noise content.

Converting Back to Analog. The transfer characteristics for conventional linear analog-to-digital-to-analog (A/D/A) conversion process is a linear function, i.e., analog input signal is digitized by a linear ADC and then is converted back to the analog domain using a linear DAC. On the other hand, in a nonlinear A/D/A conversion process, analog input signal is digitized by a nonlinear quantization function. As shown in Figure 13 to convert back to the analog domain, digital signal should be passed through a nonlinear DAC with exact inverse characteristic. The resulted A/D/A conversion transfer characteristic is similar to that of a linear A/D/A conversion process, except that the quantization steps are non-uniform. In the case of anti-logarithmic A/D/A conversion, quantization steps are decreasing in length along the input amplitude range.
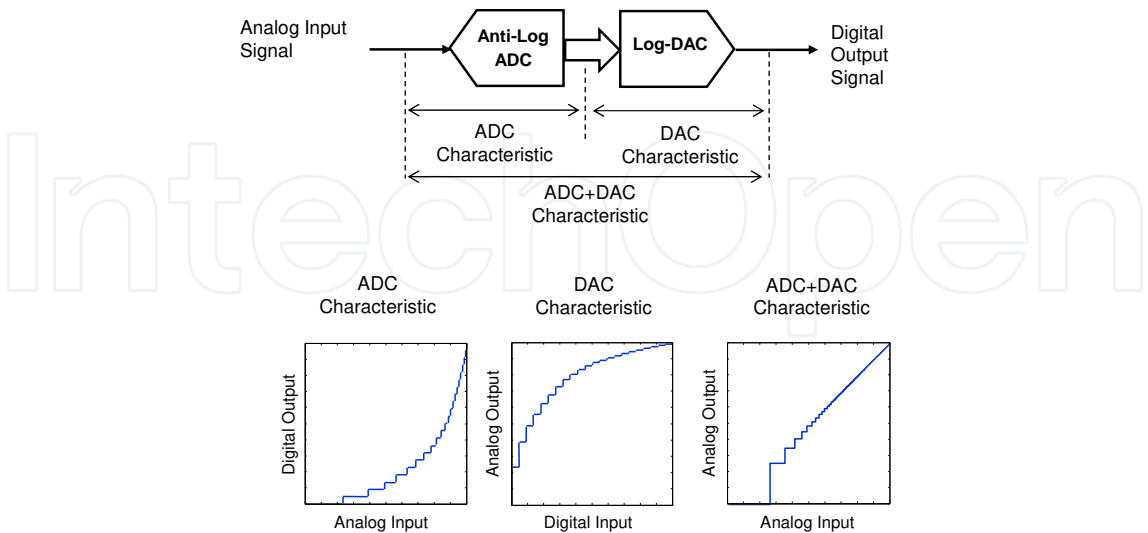

Figure 13. Transfer characteristics of an anti-logarithmic A/D/A conversion
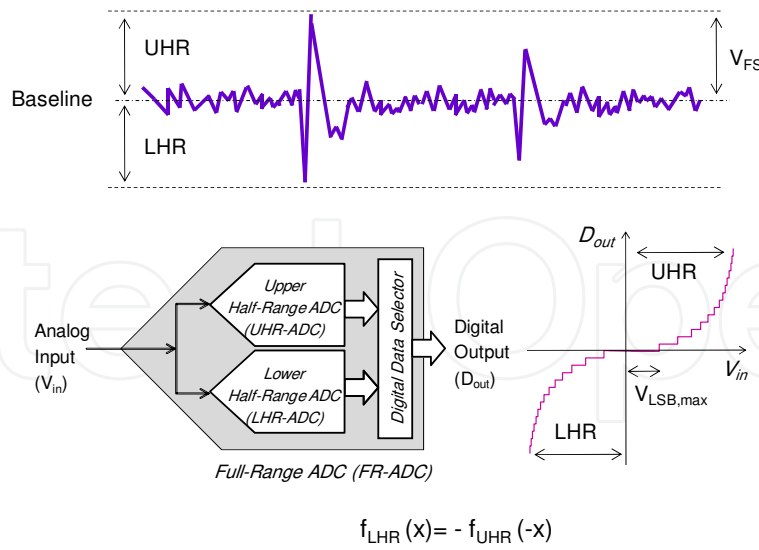
$$f_{LHR}(x) = -f_{UHR}(-x)$$

Figure 14. An FR-ADC designed to digitize neural signals using two HR-ADCs with exponential quantization function

Covering the Full Range. Assuming that the neural signal is preamplified and positioned around a certain baseline level, as illustrated in Figure 14, the nonlinear quantization function needs to be defined in an odd symmetric form around the baseline of the signal. Therefore, the nonlinear ADC needed to cover the entire input signal range, full-range ADC (FR-ADC), is realized using two complimentary half-range ADCs (HR-ADCs) each covering half of the input signal range. Hence, assuming that the basic nonlinear quantization function is used for the upper half-range ADC (UHR-ADC), $f_{UHR}(x)$, the quantization function used for the lower half-range ADC (LHR-ADC) will be:

$$f_{LHR}(x) = -f_{UHR}(-x). \tag{3}$$

The nonlinear ADC discussed hereafter is assumed to be the ADC that covers the upper half of the input signal range, i.e., the UHR-ADC, unless otherwise stated.

Half-Range Characteristic Function. The input-output relationship for an N-bit HR-ADC with exponential quantization function is:

$$\frac{b_{N-1}2^{N-1} + ... + b_0}{2^N} = a.[\exp(\frac{v_{in}}{V_{FS}} \times k) - 1], \tag{4}$$

where $(b_{N-1}...b_1b_0)$ is digital representation of analog input, v_{in}, and V_{FS} is the full-scale input range for the UHR-ADC. To satisfy the boundary conditions for minimum and maximum values of v_{in} it can be shown that:

$$k = \ln\left(\frac{1}{a} + 1\right). \tag{5}$$

Parameter a sets the curvature of the characteristic function. The smaller this parameter is, the more rapid the exponential input-output relationship will be. Quantization steps along the input range are known, in general, as least significant bits (LSB), and are calculated as:

$$V_{LSB,i} = \frac{V_{FS}}{k} \ln\left(\frac{1}{i + a2^N} + 1\right) \tag{6}$$

for $i = 0, 1, \dots, 2^N-1$. The largest and the smallest quantization steps, LSB_{max} and LSB_{min}, are calculated using eq.(6) for $i=0$ and 2^N-1, respectively, as:

$$V_{LSB,max} = \frac{V_{FS}}{k} \ln\left(\frac{1}{a2^N} + 1\right) \tag{7}$$

and

$$V_{LSB,min} = \frac{V_{FS}}{k} \ln\left(\frac{1}{(a+1)2^N - 1} + 1\right). \tag{8}$$

In general, dynamic range (DR) of a nonlinear ADC is defined to be the ratio of the full-scale input voltage to the smallest resolvable signal, $V_{LSB,min}$ [32]. DR in the case of exp-ADC is achieved as:

$$DR = \frac{V_{FS}}{V_{LSB,min}} = \frac{k}{\ln\left(\frac{1}{(a+1)2^N - 1} + 1\right)}. \tag{9}$$

The choice of the largest quantization step, $V_{LSB,max}$, is perhaps the most critical decision in forming the quantization function for the exp-ADC. This is because of its key role in the reduction of the noise content of the neural signal. The largest LSB is responsible for the largest quantization error. It is along the $V_{LSB,max}$ that variations of the input signal are intentionally not seen and replaced with 0 (the baseline level in our design). By using this method, not only the quantization error is not a disturbing phenomenon for the signal, but it also plays a denoising role as it replaces the B-Noise around the baseline with 0. To achieve significant reduction in the B-Noise power, $V_{LSB,max}$ is set to 3σ, where σ is the standard deviation of the B-Noise PDF. This way, most of the B-Noise will be intentionally removed from the neural signal during the digitization process.
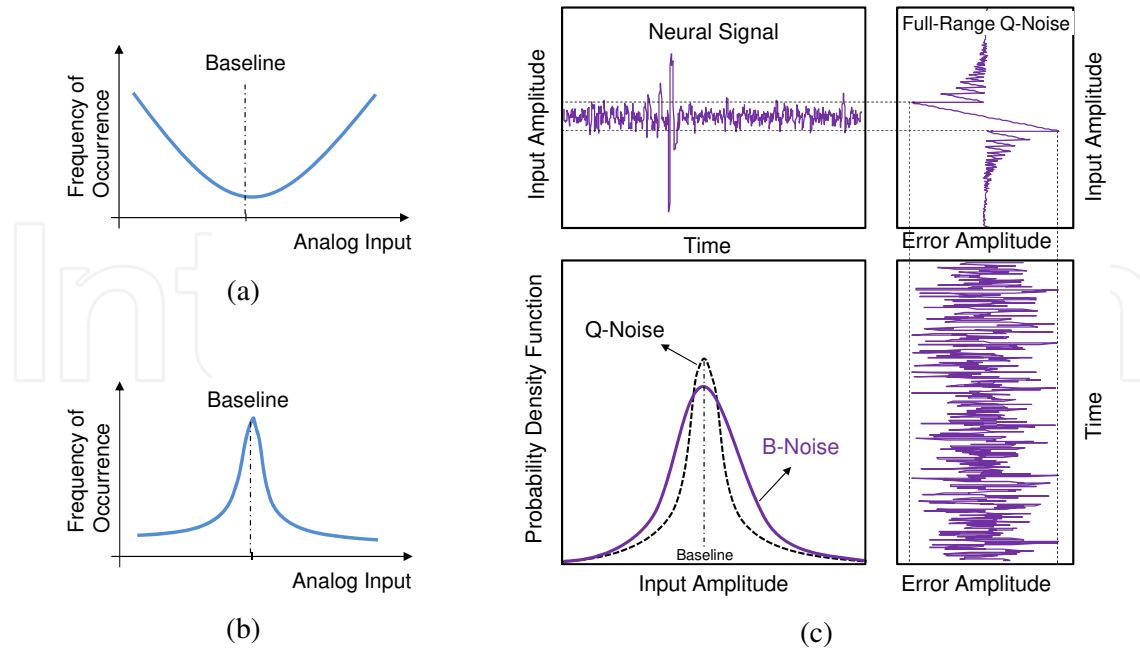
Figure 15. Q-Noise PDF for (a) logarithmic ADC and (b) exp-ADC; (c) Q-Noise for exp-ADC vs. B-Noise of a neural signal

Noise Analysis. PDF of quantization noise (Q-Noise) for linear ADC is uniform along the input amplitude range. This can be shown by eq. (10) which formulates PDF of the Q-Noise associated with code n in an N-bit linear HR-ADC [30]:

$$P_{Lin,HR}(n) = \frac{1}{2}\left[\left(\frac{n}{2^N}\right) - \left(\frac{n-1}{2^N}\right)\right] = \frac{1}{2^{N+1}}. \tag{10}$$

Uniform distribution of Q-Noise along the input amplitude makes linear ADCs suitable for digitizing NCI signals. For specific signals an NLADC might be useful in terms of SNR improvement. This advantage comes from the fact that NLADCs exhibit non-uniform Q-Noise distribution. In a Logarithmic ADC, Q-Noise energy is shifted to large amplitudes. As a result, logarithmic ADC has widely been used in digitizing of ICLA signals, such as audio. PDF of Q-noise associated with code n for an N-bit logarithmic HR-ADC is formulated in eq. (11) and is illustrated in Figure 15(a):

$$P_{Log,HR}(n) = \frac{a}{2}\left[\exp\left(\frac{kn}{2^N}\right) - \exp\left(\frac{k(n-1)}{2^N}\right)\right]. \tag{11}$$

In the case of the anti-logarithmic N-bit HR-ADC, the Q-Noise PDF associated with code n is derived as:

$$P_{Exp,HR}(n) = \frac{1}{2k}\left[\ln\left(\frac{n}{2^N a}+1\right) - \ln\left(\frac{n-1}{2^N a}+1\right)\right].\tag{12}$$

As shown in Figure 15(b), exp-ADC shapes the Q-Noise in such a way that most of its energy concentrates at small amplitudes, making it suitable for digitizing ICHA signals. As an example, in neural signals, APs with large amplitudes are quantized with higher SQNR as opposed to the B-Noise with rather small amplitudes. Figure 15(c) illustrates that most of the noise content of the signal (B-Noise) lies within the very first LSBs. The interesting point here is that since it is some of the noise content of the neural signal that is lost during the quantization process, not only the associated quantization error is not undesirable, but it is also welcomed as it leads to noise-content reduction and consequently to significant improvement in the SNR of the neural signal being digitized.

Noise-Content-Reduction Ratio, *NCRR*, is a measure of capability of an ADC in reducing the noise content of the neural signal being digitized, and can be defined as [29]:

$$\text{NCRR (dB)} = 10\log_{10}\left(\frac{\text{average noise power @ input}}{\text{average noise power @ output}}\right)\tag{13}$$

where the average noise power at the input of the ADC is calculated as:

$$\overline{n_i^2(t)} = \int_{-\infty}^{+\infty} n_i^2(t)P_{n_i}(n_i)dn_i\tag{14}$$

In this equation, $P_{ni}(n_i)$ is the probability density function for noise content of the neural signal, $n_i(t)$. Similarly, the average noise power of the signal at the output of the ADC is derived to be:

$$\overline{n_o^2(t)} = \int_{-\infty}^{+\infty} n_o^2(t)P_{n_i}(n_i)dn_o\tag{15}$$

where $n_o(t)$ is the noise content of the neural signal after passing through the ADC:

$$n_o(t) = n_i(t) \times \left[ADC\ Characteristic\ Curve\right].\tag{16}$$

Figure 16 illustrates how the B-Noise is reduced as it passes through the exp-ADC, already formulated by equations (14)-(16).
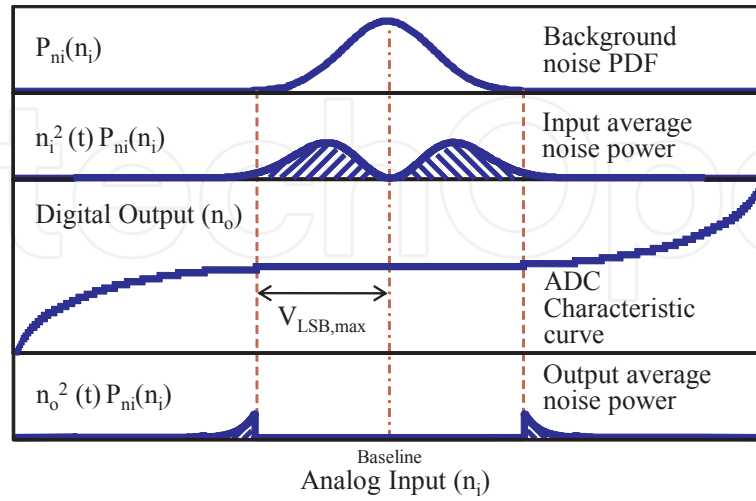

Figure 16. The process of B-Noise reduction as the signal passes through the NLADC, formulated by eqs. (14)-(16)
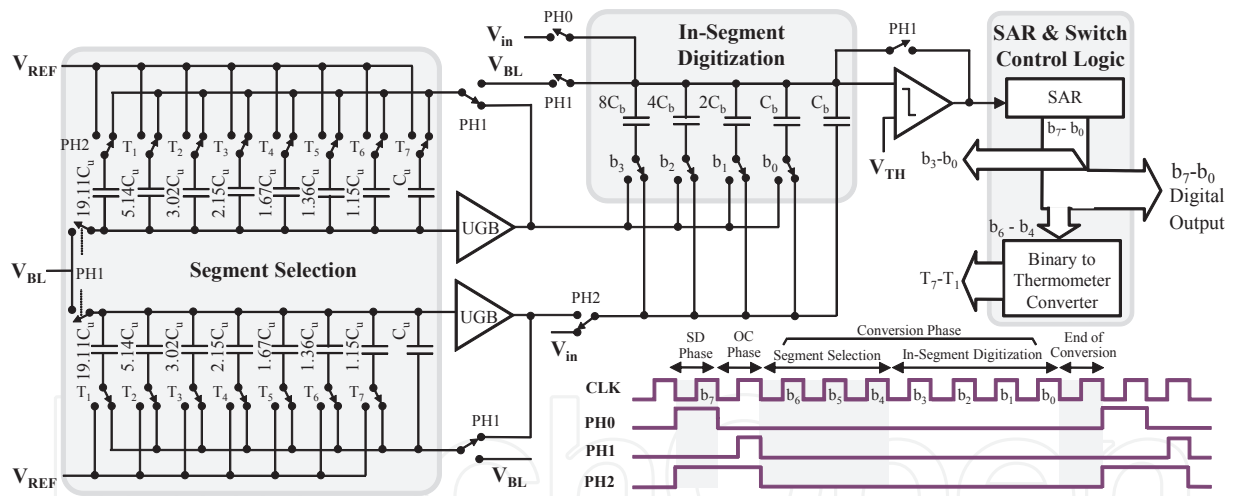

Figure 17. Circuit schematic for the two-step SAR ADC designed to implement exp-ADC and its simplified timing diagram

Circuit Design. To realize the proposed exp-ADC with reasonable power and silicon area, successive-approximation register (SAR) architecture was chosen [31]. In order to facilitate the realization of the exponential quantization function needed for the ADC, a piecewise-linear (PWL) approximation of the required function was implemented. As shown in the timing diagram of Figure 17, the proposed ADC operates in three phases: sign detection (SD), offset cancelation (OC), and conversion. In the SD phase, the analog input voltage, v_{in}, is compared with a certain threshold voltage, V_{TH}, which is temporarily set to the baseline voltage, V_{BL}. The result of this comparison determines the half range in which the input voltage is located. In

the OC phase, an ordinary offset cancellation technique is applied on the comparator and buffers. In the conversion phase, successive approximation algorithm first finds the segment of interest, encoded by $b_6b_5b_4$, in 3 clock cycles. These 3 bits are converted to a 7-bit thermometer code, T_7~T_1, which will be used in the *Segment Selection* block to generate two analog voltages associated with the endpoints of the segment of interest. An in-segment linear A/D conversion process is then performed to determine the remained 4 LSBs, $b_3b_2b_1b_0$, in 4 clock cycles. Finally, an end of conversion signal is generated to reset the ADC and prepare for the next conversion cycle. A low-voltage band-gap reference (BGR) was designed to generate the required baseline, reference and threshold voltages for the exp-ADC.
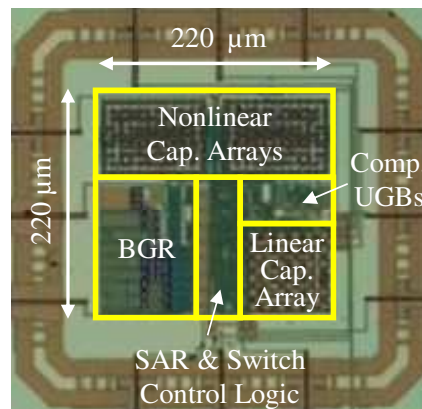

Figure 18. Chip photograph of the proposed NLADC

Experimental Results. The presented exp-ADC was fabricated in a standard 0.18-μm CMOS process. A chip photograph is shown in Figure 18, in which the chip occupies a total size of 220 ×230 μm². The measured worst-case differential non-linearity (DNL) and integral non-linearity (INL) are +0.8/-0.9 LSB and +4.3/-2.1 LSB, respectively. Table 1 summarizes the specifications of the NLADC and compares it with some of the nonlinear and linear ADCs reported in the literature.

A proof-of-concept prototype of a 4-channel neural recording system based on anti-logarithmic quantization was reported in [6]. As shown in the block diagram of Figure 19, a time-domain multiplexer (TDM) shares an anti-logarithmic ADC (AL-ADC) between 4 channels (each sampled at 20kSps). The output digital codes are then packed by the data packaging block to be transmitted to the outside world via a wireless link. At the external host, received signal is first recovered and then converted back to analog via a PC software. This inverse conversion is performed using a logarithmic DAC. Quantization characteristic function for the DAC is an exact inverse for that of the NLADC used in the recording system. To evaluate the operation of the system, neural signals recorded from the auditory cortex of a Guinea pig are used for in-vitro tests. Figure 20 shows the input signal to one of the neural channels before entering the nonlinear quantization process on the implant side along with the associated signal on the external setup after reconstruction. To verify the concept of the noise reduction caused by

overlapping PDFs for B-Noise and Q-Noise, distributions of the measured quantization error along the input amplitude are depicted in Figure 21.
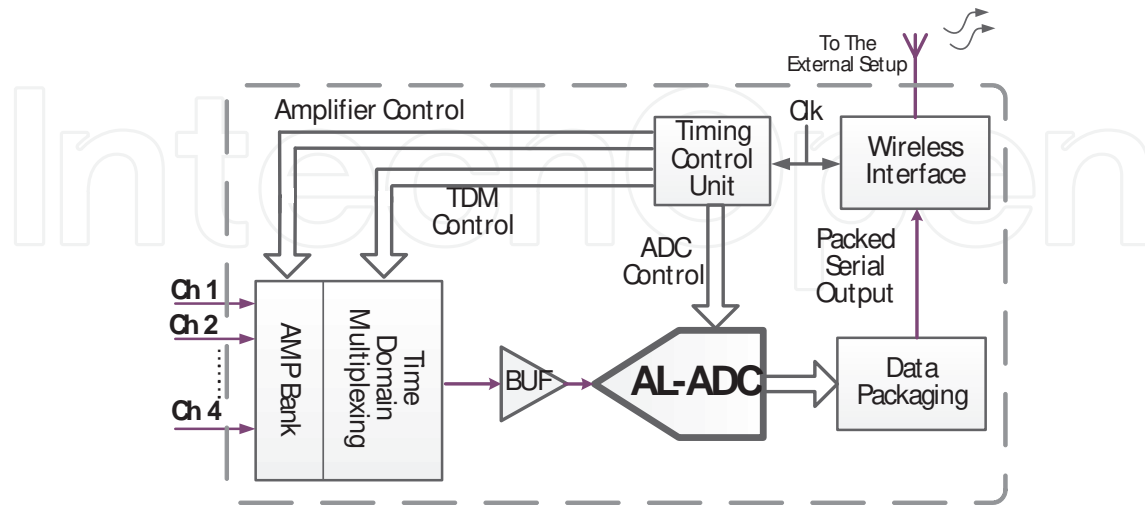

Figure 19. Block diagram of the prototyped neural recording system

Specification	This Work	[34]	[35]	[36]	[37]
Process (µm)	0.18	0.18	0.18	0.5	0.8
ADC Type	Exp.	Log.	Linear	Linear	Linear
No. of Bits	8	7	8	10	10
P_{diss} (µW)	87.2	9.7	7.4	100	17.9
Area (mm²)	0.036	0.003	0.04	0.15	0.8
Supply (V)	1.8	1.8	1.8	3.3	2.8
Input Range (V)	1	N/A	0.6	2.5	0.8
DR (dB)	64.1	80	39.1	N/A	50.5
fs (kS/s)	25	5	30	15	2.9
INL (LSB)	+4.3/ -2.1	±0.86	±1	0.6	0.98
DNL (LSB)	+0.8/ -0.9	±0.44	N/A	0.6	0.67

Table 1. Performance Comparison

5. Conclusion

To overcome bandwidth limitation in the wireless telemetry of recorded neural data, a wide variety of data reduction techniques has been reported. These techniques range from spike reporting approaches such as spike detection and spike sorting techniques to mathematical approaches such as the discrete wavelet transform. Although it is proven that spike reporting

approaches contain enough information to actuate prosthetic devices [33], in some other applications, such as neuroscientific studies, they are not satisfactory due to considerable loss of important information, e.g., spike wave shapes. Mathematical approaches, on the other hand, have been successful from the standpoint of data compression, while preserving wave shape of the spikes. Nonetheless, increasing the number of recording channels can result in the potential problems of these approaches: large silicon area and high power consumption. In contrast with all of the mentioned techniques, hardware approaches are capable of data reduction without adding any extra block to the microsystem. This is achieved by modifying the present hardware of the implant. An implementation of one of these approaches focusing on the ADC circuit of the system was presented and discussed in detail. The proposed method results in considerable reduction of bit-rate in multi-channel neural recording microsystems. Thus, efficient design of application-specific circuits for building blocks of neural implants should be taken into account as an appropriate method of data reduction.
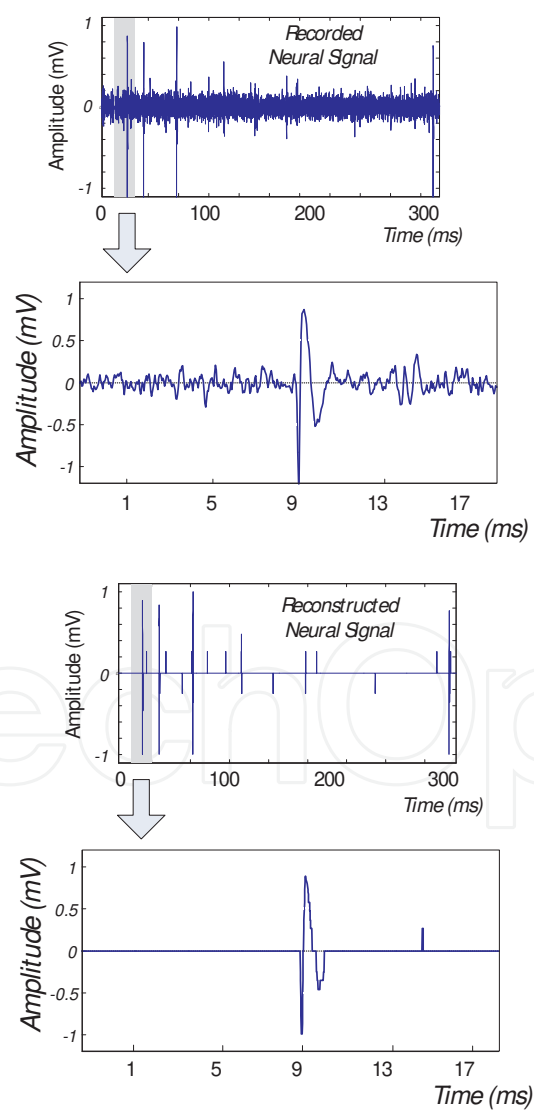

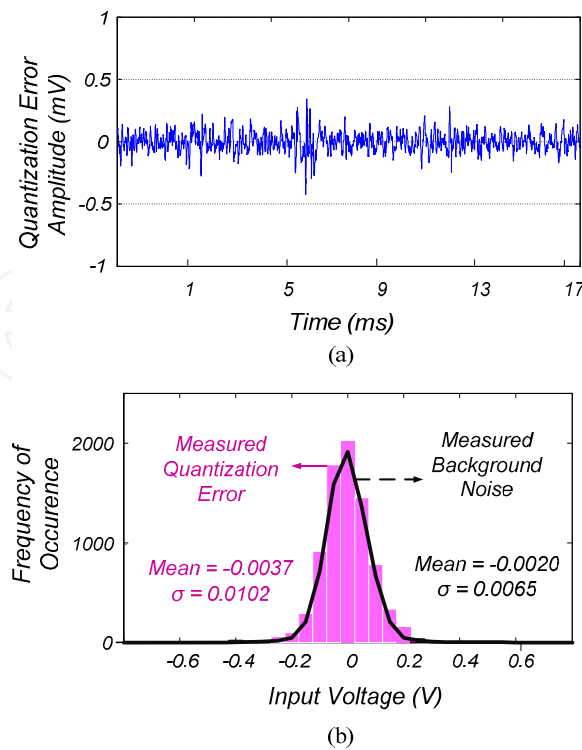**Figure 20.** Recorded neural signal and reconstructed neural signal

Figure 21. (a) Measured quantization error in the time domain (b) Distributions of the measured quantization error and the measured B-Noise

Author details

Mohsen Judy[1], Alireza Akhavian[2] and Farzad Asgarian[3]

*Address all correspondence to: mjudy@ieee.org

1 EECS Dept., University of Tennessee, Knoxville, USA

2 ECE Dept., George Mason University, Fairfax,VA, USA

3 EECS Dept., University of Michigan, Ann Arbor, Michigan, USA

References

[1] Sodagar A.M., Perlin G.E., Yao Y., Najafi K., Wise K.D. An Implantable 64-Channel Wireless Microsystem for Single-Unit Neural Recording. IEEE Journal of Solid-State Circuits 2009; 44(9) 2591-2604.

[2] Harrison R., Watkins P., Kier R., R. Lovejoy, Black D., Normann D., Solzbacher, F. A Low-Power Integrated Circuit for a Wireless 100-Electrode Neural Recording System. IEEE International Solid-State Circuits Conference 2006; 38. 958-965.

[3] Chae M.S., Yang Z., Yuce M.R. , Hoang L., Liu W. A 128-Channel 6 mW Wireless Neural Recording IC with Spike Feature Extraction and UWB Transmitter. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2009; 17, 312-21.

[4] Wise K.D., Angell J.B., Starr A. An Integrated Circuit Approach to Extracellular Microelectrodes. IEEE Transactions on Biomedical Engineering 1970; 17, 238-247.

[5] Sodagar A.M. Integrated Circuit and System (ICAS) Lab Internal Report. K. N. Toosi University of Technology 2011; Tehran.

[6] Akhavian A., Judy M., Sodagar A.M. Anti-Logarithmic Quantization for Data Reduction in Multi-Channel Intra-Cortical Neural Recording Implants. Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Jan. 2012, Hong Kong, 416-419.

[7] Olsson III R., Wise K.D. A Three-Dimensional Neural Recording Microsystem with Implantable Data Compression Circuitry. IEEE Journal of Solid-State Circuits 2005; 40(12).

[8] Eberhart R.C, Dobbins R.W., Webber W.R.S. EEG Waveform Analysis Using Case-Net. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society 1989; 2046-2047.

[9] Vaz F., Principe J.C. Neural Networks for EEG Signal Decomposition and Classification. Proceedings of the IEEE 17th Annual Conference on Engineering in Medicine and Biology Society1995; 1, 793-794.

[10] Eberhart, R.C., Dobbins, R.W. and Webber, W.R.S. Neural Network Design Considerations for EEG Spike Detection. Proceedings of the 1989 Fifteenth Annual Northeast Bioengineering Conference 1989; 97 – 98.

[11] Choi J.H., Jung H.K., Kim T. A New Action Potential Detector Using the MTEO and Its Effects on Spike Sorting Systems at Low Signal-to-Noise Ratios. IEEE Transactions on Biomedical Engineering 2006; 53, 738-746.

[12] B. Gosselin and M. Sawan. An Ultra Low-Power CMOS Automatic Action Potential Detector IEEE Transactions on Neural Systems and Rehabilitation Engineering 2009; 17, 346-353.

[13] Sodagar A.M, Wise K.D, Najafi K. A Fully Integrated Mixed-Signal Neural Processor for Implantable Multichannel Cortical Recording. IEEE Transactions on Biomedical Engineering 2007; 54(6), 1075-1088.

[14] Iyad Obeid, Patrick D Wolf. Evaluation of spike-detection algorithms fora brain-machine interface application. IEEE Transactions on Biomedical Engineering 2004; 51, 905-911.

[15] Harrison R.R. A Low-Power Integrated Circuit for Adaptive Detection of Action Potentials in Noisy Signals. Proceedings of the 2003 International Conference of the IEEE Engineering in Medicine and Biology Society 2003; 3325-3329.

[16] Sodagar A.M. Analysis and Design of Analog Spike Detectors with Adaptive Threshold. Internal Report, Center for Wireless Integrated Microsystems (WIMS) 2005, University of Michigan; Ann Arbor, MI.

[17] Rogers C.L., Harris J.G. A Low-Power Analog Spike Detector for Extracellular Neural Recordings. Proceedings of the 2004 11th IEEE International Conference on Electronics Circuits and Systems 2004; 290-293.

[18] Roy J.F., Sawan M. A Fully Reconfigurable Controller Dedicated to Implantable Recording Devices. The 3rd International IEEE-NEWCAS Conference 2005; 303-306.

[19] Dumortier C., Gosselin B., Sawan M. Wavelet Transforms Dedicated to Compress Recorded ENGs from Multichannel Implants: Comparative Architectural Study. Proceedings of the 2006 IEEE International Symposium on Circuits and Systems 2006; 2129-2132.

[20] Oweiss K.G. A Systems Approach for Data Compression and Latency Reduction in Cortically Controlled Brain Machine Interfaces. IEEE Transactions on Biomedical Engineering 2006, 53(7), 1364-1377.

[21] Oweiss K.G., Mason A., Suhail Y., Kamboh A.M., Thomson K.E. A Scalable Wavelet Transform VLSI Architecture for Real-Time Signal Processing in High-Density Intra-Cortical Implants. IEEE Transactions on Circuits and Systems-Part I: Regular Papers 2007; 54(6), 1266-1278.

[22] Kamboh A.M., Raetz M., Oweiss K.G., Mason A. Area-Power Efficient VLSI Implementation of Multichannel DWT for Data Compression in Implantable Neuroprosthetics. IEEE Transactions on Biomedical Engineering 2007; 1 (2), 128-135.

[23] Kamboh A.M., Oweiss K.G., Mason A.J. Resource Constrained VLSI Architecture for Implantable Neural Data compression Systems. Proceedings of the 2009 IEEE International Symposium on Circuits and Systems 2009; 1481-1484.

[24] Kamboh A.M., Yang Y., Oweiss K.G., Mason A.J. Design of a Configurable Neural Data Compression System for Intra-Cortical Implants. Proceedings of the 2010 IEEE International Symposium on Circuits and Systems 2010; 3473-3476.

[25] Zhang, F., Aghagolzadeh, M. and Oweiss, K.G. An Implantable Neuroprocessor for Multichannel Compressive Neural Recording and On-the-fly Spike Sorting with Wireless Telemetry. Proceedings of the 2010 IEEE Biomedical Circuits and Systems Conference 2010; 3-6.

[26] Shaeri M.A., Sodagar A.M., Abrishami-Moghaddam H. A 64-Channel Neural Signal Processor/ Compressor Based on Haar Wavelet Transform. Proceedings of the 2011 Annual IEEE Engineering in Medicine and Biology Society Conference 2011; 6409-6412.

[27] Mallat S. G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 1989; 2(7).

[28] Li B., Tenhunen H. A Second Order Sigma Delta Modulator Using Semi-uniform Quantizer with 81dB Dynamic Range at 32x OSR. Proc. of the 28th European Solid-State Circuits Conference, Florence, Sept. 2002, 579-582

[29] Judy M., Sodagar A.M., lotfi R. A Nonlinear Signal-Specific ADC for Efficient Neural Recording. Proceedings of the 2010 IEEE Conference on Biomedical Circuits and Systems, Paphos, Cyprus Nov. 2010, 17-20.

[30] Burns M., Roberts GW. An Introduction to Mixed-Signal IC Test and Measurement, Oxford University Press, 2000.

[31] Judy M., Sodagar A.M., Lotfi R., Sawan M. Nonlinear Signal-Specific ADC for Efficient Neural Recording in Brain-Machine Interfaces, IEEE Transactions on Biomedical Circuits and Systems 2014; 8(3), 371-381.

[32] Lee J., Rhew H.G., Kipke D., Flynn M. A 64 Channel Programmable Closed-loop Deep Brain Stimulator with 8 Channel Neural Amplifier and Logarithmic ADC. Symposium on VLSI Circuits Digest of Technical Papers, Hawaii, June 2008. 76-77.

[33] Nurmikko A.V., Donoghue J.P., Hochberg L.R., Patterson W.R., Song Y., Bull C.W., Borton D.A., Laiwalla F., Park S., Ming Y., Aceros J. Listening to Brain Microcircuits for Interfacing with External World, Progress in Wireless Implantable Microelectronic Neuroengineering Devices. Proceedings of the IEEE 2010; 98 (3), 375-388.

[34] Guo J., Sonkusale S. An Area-Efficient and Low-Power Logarithmic A/D Converter for Current-Mode Sensor Array, IEEE Sensors Journal 2009; 9(12), 2042-2043.

[35] Robert P., Gosselin B., Ayoub A.E., Sawan M. An Ultra-Low-Power Successive-Approximation-Based ADC for Implantable Sensing Devices, in Proceedings of the 49th IEEE Int. Midwest Symposium on Circuits and Systems Aug. 2006, 7-11.

[36] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Sozbacher, "A Low-Power Integrated Circuit for a Wireless 100-Electrode Neural Recording System," IEEE J. Solid-State Circuits 2007; 42(1), 123-133.

[37] Bonfini G., Brogna A.S., Garbossa C., Colombini L., Bacci M., Chicca S., Bigongiari F., Guerrini N.C., Ferri G. An Ultralow-Power Switched Opamp-Based 10-b Integrated ADC for Implantable Biomedical Applications, IEEE Transactions on Circuits and Systems I: Regular Papers 2004; 51(1), 174-177.