Power Efficient Data Compression Hardware for Wearable and Wireless Biomedical Sensing Devices

Chengliang Dai

Doctor of Philosophy

University of York

Computer Science

March 2016

Abstract

This thesis aims to verify a possible benefit lossless data compression and reduction techniques can bring to a wearable and wireless biomedical device, which is anticipated to be system power saving. A wireless transceiver is one of the main contributors to the system power of a wireless biomedical sensing device, and reducing the data transmitted by the transceiver with a minimum hardware cost can therefore help to save the power. This thesis is going to investigate the impact of the data compression and reduction on the system power of a wearable and wireless biomedical device and trying to find a proper compression technique that can achieve power saving of the device.

The thesis first examines some widely used lossy and lossless data compression and reduction techniques for biomedical data, especially EEG data. Then it introduces a novel lossless biomedical data compression technique designed for this research called Log2 sub-band encoding. The thesis then moves on to the biomedical data compression evaluation of the Log2 sub-band encoding and an existing 2-stage technique consisting of the DPCM and the Huffman encoding. The next part of this thesis explores the signal classification potential of the Log2 sub-band encoding. It was found that some of the signal features extracted as a by-product during the Log2 sub-band encoding process could be used to detect certain signal events like epileptic seizures, with a proper method. The final section of the thesis focuses on the power analysis of the hardware implementation of two compression techniques referred to earlier, as well as the system power analysis. The results show that the Log2 sub-band is comparable and even superior to the 2-stage technique in terms of data compression and power performance. The system power requirement of an EEG signal recorder that has the Log2 sub-band implemented is significantly reduced.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to first thank my supervisors Dr Christopher Crispin-Bailey and Professor Jim Austin for their constant guidance and help throughout my doctorate. Their remarkable minds and inspiring ideas saved me many times from various difficulties I encountered in the past few years. The novel Log2 sub-band encoding technique introduced in this thesis would not exist without Chris.

I would also like to thank Cybula Ltd, the UK Technology Strategy Board and Mr Anthony R Moulds for providing the wonderful NAT-1 device for my research. NAT-1 offered me a perfect platform to verify my research hypothesis. My thanks go to every member of the Advanced Computer Architecture Group (ACAG), and specifically to Dr Amir Mansoor Kamali Sarvestani, my housemate Sun Hao and Dr Victoria J Hodge for providing me kind assistance. I am also grateful for access to the EEG data supplied by the Institute of Medical Sciences, University of Aberdeen.

I also want to thank my grandparents Haifeng and Xingsu, my aunts Luping, Yun and Jingping, and my uncles Lukang, Zhaochen and Yuanlin: your support means a lot to me.

Finally, a special thanks to my parents Jingquan and Lubin, for it is their love and steadfast support that give me the strength to achieve all I have today.

Declaration

I declare that the research described in this thesis is original work, which I undertook at the University of York during 2011 - 2016. Except where stated, all of the work contained within this thesis represents the original contribution of the author.

Some parts of the work described in chapters 4, 5 and 6 have been published in conference proceedings given below.

1. C. Dai and C. Bailey, "A time-domain based lossless data compression technique for wireless wearable biometric devices," in *SENSORCOMM 2013, The Seventh International Conference on Sensor Technologies and Applications*, pp. 104107, 2013.

2. C. Dai and C. Bailey, "Power analysis of a lossless data compression technique for wireless wearable biometric devices," in *Ph.D. Research in Microelectronics and Electronics (PRIME), 2015 11th Conference on*, pp. 97100, June 2015.

3. C. Dai and C. Bailey, "A lossless data reduction technique for wireless eeg recorders and its use in selective data filtering for seizure monitoring," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 61866189, Aug 2015.

Items that were published jointly with collaborators, the author of this thesis is responsible for the material presented here. For each published item the primary author is the first listed author.

The VLSI test chip fabrication was supported by the NOMAD project, Innovate-UK grant number: 131200. The VLSI design and layout were completed jointly in collaboration with Dr Christopher Crispin-Bailey in 2015.

Chapter 1

Introduction

1.1 Motivation

The human biological system generates various biosignals which can be observed as the electroencephalogram (EEG) electromyogram (EMG) and electrocardiogram (ECG). These are small electric currents caused by electric potential discrepancies across our body organs or nervous system. It has been a long time since we realized that these signals reflect some alterations inside our body, like our heart rate and body temperature. These alterations may be caused by diseases, external stimuli or other internal changes, and therefore the biomedical signal is one of the important keys for us to reveal any secrets of human physiological processes.

Since recording biomedical signals has been proved as reliable and usually non-invasive, these signals are now used widely in various applications, For instance, EEG and EMG signals are used in numerous areas like neurological disorder studies [1] [2], prosthetic devices design [3], and brain computer interfaces (BCI) [4].

However, the process of collecting the data often takes quite long time and causes inconvenience to the patients or research subjects as it largely restricts their mobility. For many years, long-term EEG recordings have had to be taken as an inpatient process because all the electrodes on a persons scalp have to be wired to a static system. In some sleep disorder studies, in order to monitor the EEG of a patients sleep stage, the subject needs to stay in hospital all night, which is impractical and inefficient. The monitoring time can be even longer in some epilepsy or other neurological disorder studies. As a result, the idea of a wearable wireless signal recording device that helps to free users from the inpatient environment and promises better real-time healthcare from caregivers to

patients with neurological or other diseases has drawn a lot of attention in the past few years.

Such a device has some obvious advantages compared with the old-fashion recording system.

a. Low cost

For instance, a standard EEG test costs around £120 to £450 or up to £1800 if extended monitoring is required [5]. A 24-hour ambulatory outpatient EEG monitoring is more than 50 per cent cheaper than inpatient monitoring [6]. Besides, maintaining a wearable set costs less.

b. Portable

To be integrated into a body area network (BAN) application [7], all wearable signal recorders are designed to be small and light, making them suitable for frequent use. Some commercialized products like Emotiv (for EEG) [8], Myo (for EMG) [9], Apple Watch [10] and NAT-1(for multiple purposes) [11] are very well designed and can hardly be seen when someone is wearing them.

c. Wireless

Wireless connection gives more flexibility and allows researchers to access the data remotely and in real-time. A wireless-connected signal recorder is clearly perfect for applications such as BCI and prosthetic devices.

Despite all the benefits a wearable wireless device can bring us, there are still a few challenges when designing such a device, and power consumption is clearly one of the critical issues.

To provide an acceptable user experience, miniaturisation and duration are two important factors to consider when designing a wearable wireless biomedical signal recorder, but these factors are difficult to achieve at the same time as smaller size inevitably leads to a smaller space for a battery. The truth is that many manufacturers choose size over battery life in order to make their products more appealing to the customers, and the operation time of these products is sometimes disappointing. Table 1.1 gives some information about the battery life of three typical wearable devices.

According to this table, these products hardly meet this criterion for any long-term signal recording mentioned earlier. One obvious way to solve the problem is to increase the battery capacity, but considering the limits on the size of a wearable device, making more room for a larger battery does not seem feasible, so we have to find some other means

Table 1.1: Battery information of wearable products

Product	**Battery type**	**Battery life**
Myo Gesture Control Armband [12]	lithium-ion	6h
Emotiv Insight [13]	lithium polymer	4h
Google Glass [14]	lithium polymer	30min

to extend battery life instead.

A typical wearable wireless biomedical signal recording system has at least two parts, as shown in Figure 1.3, and communication between the device and the workstation is via Bluetooth or other wireless transmission schemes. It is known that the transceiver is always one of the biggest contributors to the overall power consumption on a wearable signal recorder, so reducing the transmitted data size will be a significant answer to this power issue.

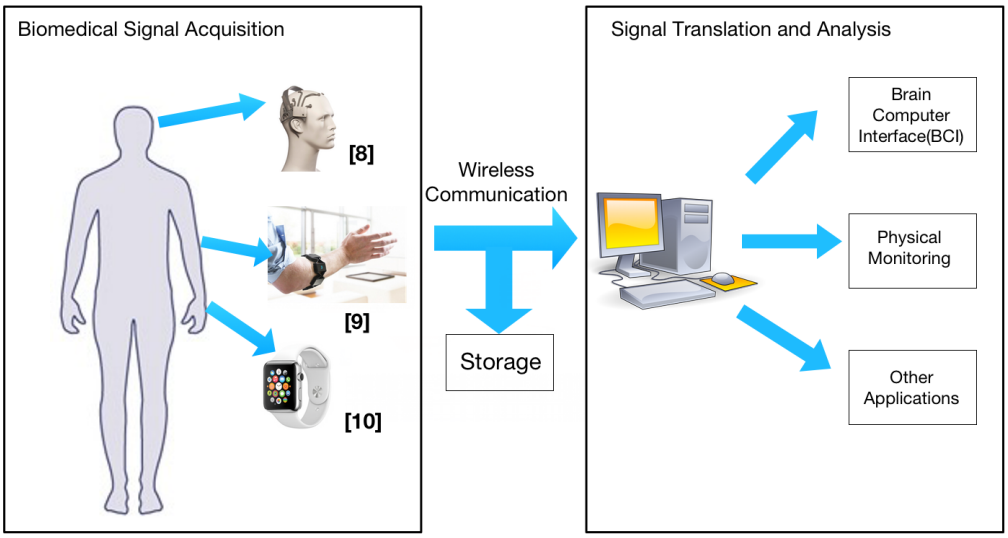

Figure 1.1: System structure

Data compression and reduction have several benefits.

a. Power saving

Power saving is the most obvious advantage. In addition to the power we can save from a transceiver, reducing the data size can further help to save power where an onboard storage medium, such as flash memory, is used, as high read/write rates may be reduced.

b. Time efficient

Transmitting compressed data between the device and the workstation takes less time than transmitting raw data. If the time for compressing and reconstructing the data is shorter than the time saved on transmission, the total time of a signal recording process will be shortened particularly at low bit-rates.

c. Low storage cost

Although the cost for data storage media decline over time, data compression can still lower the storage costs if a greater compression gain can be achieved. More space could be made for other circuit units if less storage is required.

d. Less work on signal analysis

As a spin-off of some data compression algorithms, some key features that are very helpful for diagnostic uses or other related applications are extracted from the original signals during the compression process, and this could save time for the scientists.

However, benefits come with costs, and introducing an extra unit that compresses the data will certainly add its own extra power to the device, and the compression process may remove some diagnostic information and impair the integrity of the original signal, which can be a serious problem for medical research. Needless to say, several other trade-offs, such as the complexity of the design and the time delay after having a compressor, also need to be taken into account. As a result, finding one of the best trade-offs in this matter is certainly a fruitful area for research.

This thesis first investigates some of the existing biomedical signals mainly EEG, compression and reduction methods and gives their pros and cons. Then a novel data compression technique that not only reduces the biomedical data size but also minimizes the power consumption of a wearable wireless device is introduced, followed by a biomedical data compression analysis of this novel compression technique and the 2-stage Huffman coding technique. Then an introduction to the signal classification potential of the proposed technique is given and analysis presented. Finally, a system power analysis with various use-case scenarios is given.

1.2 Main contribution

The main contributions of this thesis are summarized below.

1. Survey on biomedical data (mainly EEG) compression and reduction techniques.

2. Proposing a novel lossless compression method called Log2 Sub-band encoding that can be used for real-time biomedical data compression.

3. Providing a comprehensive evaluation of compressing the Human EEG data with the 2-stage Huffman coding and the Log2 sub-band encoding.

4. Analysis and comparison of the proposed data compression method and Huffman coding.

5. Suggesting a biomedical signal classification method (used for detecting seizure from the human EEG signal) based on the proposed compression technique.

6. Hardware implementation of the proposed method and Huffman coding in order to estimate the compressors power consumption accurately.

7. Proposing possible components for building up a simplified wearable and wireless biomedical device model.

8. Analysis of the data compressors impact on the power consumption of a simplified wearable and wireless biomedical signal recording system.

9. Examining the benefits of use of the compression method in various hypothetical use cases.

1.3 Thesis structure

The structure of the rest of this thesis is as follows.

Chapter 2: Background Information and Literature Review

This chapter first gives a background to the use of the biomedical signal and the wearable wireless signal recording system. It then examines some emblematic data biomedical data compression and reduction methods and their pros and cons.

Chapter 3: Hypothesis and Methodology

This chapter introduces the hypothesis of this research. The methodology of this research is given, and a set of performance metrics for evaluating the performance of the chosen data compression methods is illustrated at the end. A novel data compression technique called Log2 sun-band encoding as well as a 2-stage Huffman coding technique are introduced in this chapter, and a comprehensive introduction to the biomedical datasets used in experiments in this thesis is given.

Chapter 4: Data Compression Analysis

The data compression performance of both two techniques mentioned in chapter 3 is given in this chapter. Analysis and comparison of these two methods are discussed. Limitations of both methods are also referred to at the end of this chapter.

Chapter 5: Signal-Classification-Based Data Reduction with Log2 Sub-band

This chapter explores the signal classification potential of the Log2 Sub-band encoding. This compression method is used to detect seizure events from epilepsy patients EEG signals. Its result is analysed with the given performance metrics. Some possible improvements that can be made in future are given.

Chapter 6: Hardware Implementation

This chapter presents the hardware design of the compression methods mentioned in chapter 5. Analysis and comparison of the power consumption of both methods are presented.

Chapter 7: System Power Overview

The power information from the previous chapter is further discussed in this chapter. The electronic components such as amplifier and transceiver that are suitable for a signal recording system are examined first, and the overall power saving of the system built with the chosen component is then analysed. The impact on battery life is also discussed.

Chapter 8: Chip testing of the Log2 sub-band encoding

This chapter gives some testing results of the fabricated chip that the Log2 sub-band encoding is on. The layout of the hardware design and a picture of the fabricated chip are given.

Chapter 9: Conclusions

The final chapter reviews and concludes the work that has been carried out for this thesis, and outlines the main contributions based on the results given in this thesis. Some potential future work that can build upon this research is also outline at the end.

Chapter 2

Background information and literature review

2.1 Introduction

This chapter introduces the relevant background information and literature that motivate the research described in this thesis. A brief introduction of two typical biomedical signals is given first, including their characteristics and diagnostic uses. Then, this chapter reviews work of other researchers on biomedical data compression and reduction techniques, and the hardware implementation of some of these techniques. Finally, it gives a fundamental hardware model that can be applied to most biomedical signal recorders, and a power model of it is presented as well.

2.2 Biomedical signals

Biomedical signals are usually collected by electrodes that are either implanted under human skin, or more often, on the skin as a non-invasive process. Using these signals to monitor the human biological system was first discovered more than a hundred years ago [15], when people found that electrodes placed on the skin could record small electric currents and those recordings could then be used to describe graphically the activity or change of internal organs like the heart or brain. Since then studies on these signals have never stopped especially in medical fields, and more solid links between various diseases and related biomedical signals have been found. Two typical biomedical signals that are used for experiments in this research are discussed in this section.

2.2.1 EEG

In late nineteenth century, Hans Berger first recorded electrical brain waves from a human. These waves were collected by the electrodes attached to the subjects scalp, and they showed a time-varying, oscillating behaviour that varied in shape from location to location on the scalp [16].

Berger′s discovery formed the foundation of electroencephalography, and now it is an important non-invasive tool for studying the human brain and diagnosing neurological disorders. Some other potential uses of EEG have also been explored whilst applications such as BCI and prosthetic limb control are developed.

The electrical field that we can record on the scalp is the consequence of the joint activity of countless cortical neurons, and in general, these electrical signals have amplitudes ranging from a few microvolts to about 150μV, and with a frequency spectrum ranging from 0.5 to 60Hz [17]. Based on the frequency spectrum, electroencephalographic rhythms are usually classified into five different frequency bands, as indicated below [18].

Delta rhythm (<4 Hz) can be mostly observed during deep sleep and has a large amplitude. Although it is not commonly discovered while awake, it is quite indicative of cerebral damage.

Theta rhythm (4-7 Hz) is encountered during drowsiness and some certain stages of sleep.

Alpha rhythm (8-13 Hz) is observed when subjects are awake, especially when they are relaxed with their eyes closed, and the rhythm gets less prominent when the eyes are open.

Beta rhythm (14-30 Hz) is a rapidly changing rhythm with relatively low amplitude, and it can be recorded mainly from the frontal and central regions of the scalp.

Gamma rhythm (>30 Hz) is found to be more active when the cortex is processing certain information, e.g. during finger movement [19].

The display of these rhythms often changes with the subjects age and state of mind, and some neurological diseases also have a significant impact on these frequency bands. As the rhythms reflect all these alternations, they are particularly useful in clinical studies.

EEG signals are normally considered as generated by a nonlinear dynamic system, and whether they are more deterministic or more stochastic has still not yet been settled as signals show different characteristics under different conditions. Since neural activities are often organized in groups with some substantial internal dependence, EEG signals are

hardly Gaussian, but in some cases, Gaussian probability density function(PDF) is still valid when processing the signals. The ambiguity of the characteristics of EEG signals clearly brings some difficulties for researchers.

When collecting EEG signals for clinical use, a standardized International 10/20 system, which is a guideline for electrode placement [20] is often used. This system requires 21 electrodes to be attached to the surface of the scalp at certain locations. However, the number of electrodes with this system is inadequate sometimes, which causes the electrical activity to be inaccurately represented. Some studies have pointed out that 64, as shown in Figure 2.1, or even more electrodes should be used in brain mapping applications in order to acquire the detail required [21].

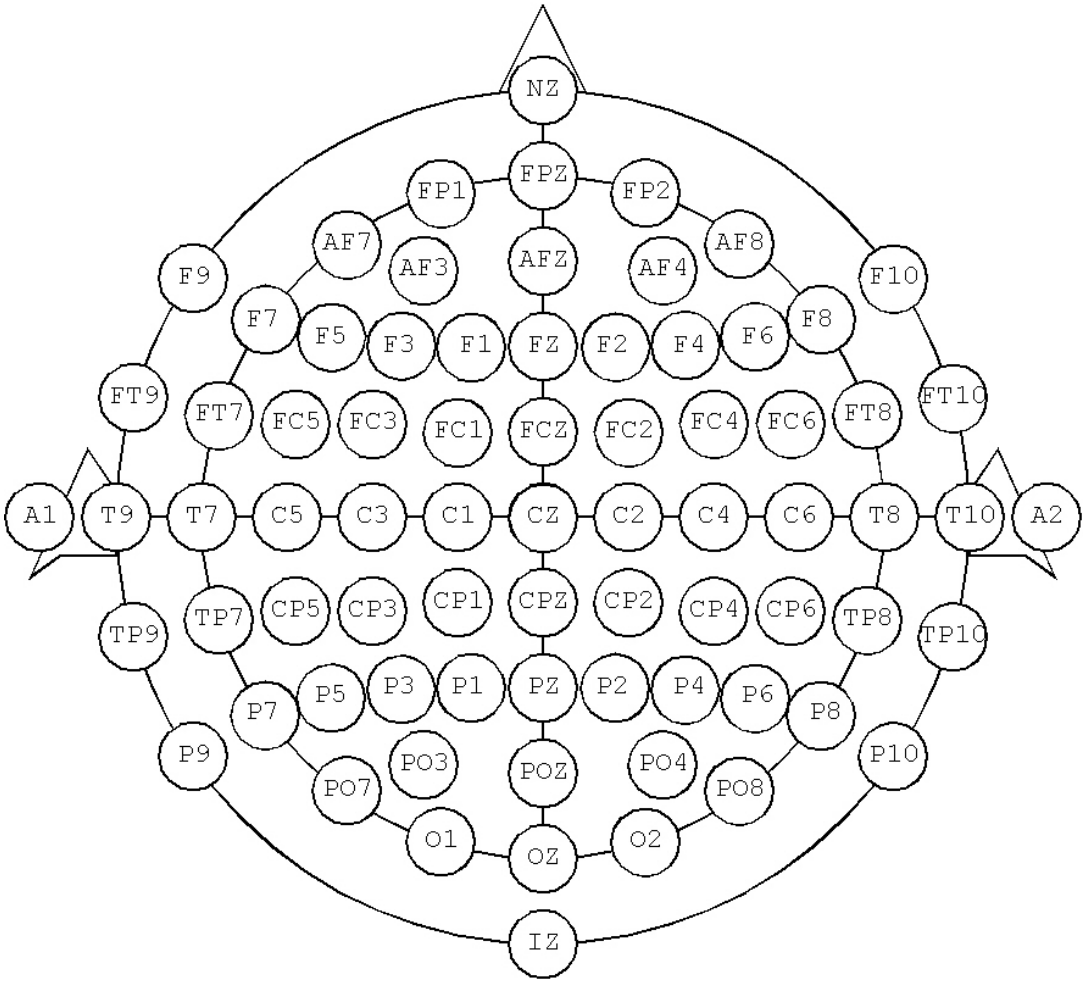

Figure 2.1: 64-channel system [21]

During EEG acquisition, artifacts and noise are inevitable, and they are either of physiological origin or of technical origin. The former is the result of activities of any

other electrical source in human body, which includes electro-oculogram(EOG) from eye movement and blinks, EMG from muscle activity and ECG from cardiac activity. The latter is mainly caused by electrodes and recording equipment, and the artifact might occur when electrodes move or the recording process is conducted in a place with complex electromagnetic fields.

2.2.2 EMG

Similar to the discovery of the EEG, it has been known for a long time that the electrical field detected on human skin comes from the muscle, but it was not until the 1960s that the EMG signal was found to be clinically useful. The EMG signal helps to reveal the secret of muscles that make our body move as it usually contains information about controller function of the central and peripheral nervous systems on the muscles [22].

Myoelectric activity is mostly measured in two ways: the invasive way of using a needle electrode and the non-invasive way of attaching electrodes on the skin overlying the muscle. The peak-to-peak amplitude of the EMG signal is normally from 0.25 to 5 mV, and the surface EMG has most of its frequency below 500 Hz, suggesting a minimum 1 kHz sampling rate is required [23], while the needle EMG requires a sampling rate up to 50 kHz.

It has been proved that the EMG signal is stochastic given the fact that it can be well represented by a Gaussian distribution function.

As with the problem encountered when recording other biomedical signals, some artifacts and noise are also collected in the process of EMG recording, and removing artifacts from these signals becomes more important now.

2.3 Biomedical data compression techniques

Considering biomedical signals share a high similarity in their characteristics, and the EEG signal, as probably the most complex one, has both stochastic and deterministic characteristics, this part will be mostly reviewing EEG data compression and reduction techniques. Some techniques are developed based on the stochastic feature of the EEG signal, and they are also suitable for the EMG signal.

Every message contains redundancies, and most data compression algorithms are designed to remove them. Redundancy is the difference between the message being trans-

mitted and the actual information in that message. In information theory, entropy is used to quantify the expected value of information in a signal, which is defined as

$$H(X) = -\sum_{i=1}^{N} P(x_i) * \log P(x_i) \tag{2.1}$$

where $P(x_i)$ is the probability of the ith symbol of source. The unit of H(X) in this case is in bits. If symbols are all independent and identical distributed(i.i.d), the shortest average length of code to represent every symbol can be obtained from 2.1 which is the best result a lossless compression technique can achieve. A good compression algorithm is able to make symbols very close to i. i. d. For EEG signals, there are two types of redundancy: temporal and spatial. Temporal redundancy is caused by repeated symbols and correlations between samples in signals while spatial redundancy is induced by a high degree of similarity of neighbouring channel structures, which is quite common as many are placed electrodes are close to each other.

To measure the performance of a data compression technique for this thesis, the compression ratio(CR) is introduced as

$$CR = Size_{Original}/Size_{Compressed} \tag{2.2}$$

where $Size_{Original}$ and $Size_{Compressed}$ are the size of data before and after the compression.

As some of the most complex biomedical signals, EEG signals are highly non-Gaussian, non-stationary and non-linear as mentioned earlier, and since they are commonly used for diagnostic purposes, the integrity of data is crucial. Any discrepancy between the original and recovered signals may have a significant impact on the diagnosis of a disease. For those reasons, methods that can be applied in compressing EEG data are quite restricted and are preferable to be lossless. A comprehensive survey on lossless data compression was given in [24], and G. Antoniol tested some well-known lossless compression algorithms based on the peculiarities of EEG, and an average CR of 2.3 was achieved which is quite extraordinary. However, lossless algorithms always have their constraints, such as low bit rates, complex computations, etc. And with more applications of EEG signals being discovered, lossy or near-lossless compression methods sometimes are also acceptable as long as reconstructed data serve all the research purposes.

Percentage Root-Mean-Square Difference (PRD) is given as

$$PRD = \sqrt{\frac{\sum_{i=1}^{l} (x_{original}(i) - x_{reconstructed}(i))^2}{\sum_{i=1}^{l} (x_{original}(i))^2}} * 100 \tag{2.3}$$

where l is the length of signals, and $x_{Original}$ and $x_{Reconstructed}$ are the values of original and reconstructed signals which are used for evaluating the distortion between original and recovered signals when using lossy compression algorithms. A larger PRD indicated a greater loss of original information after reconstruction. A table that contains the results of all the mentioned techniques is given in table at the end of this part.

2.3.1 Time-domain based compression techniques

2.3.1.1 Entropy coding

a. Huffman coding

In [24] and [25], as the most commonly used method, Huffman coding was tested on EEG data. In Huffman coding, every different symbol is assigned with a variable-length code based on the probability of that symbols occurrence, and symbols that occur more frequently in the dataset are given shorter codes, so the average bit string length of encoded signals is shorter than the original ones. As it is shown in Figure 2.2, if there are four symbols (A, B, C, D) with different probabilities of occurrence (0.5, 0.3, 0.1, 0.1), and it requires at least 2 bits to encode with the blocking encode method, but with Huffman coding, only 1.7 bits are needed.

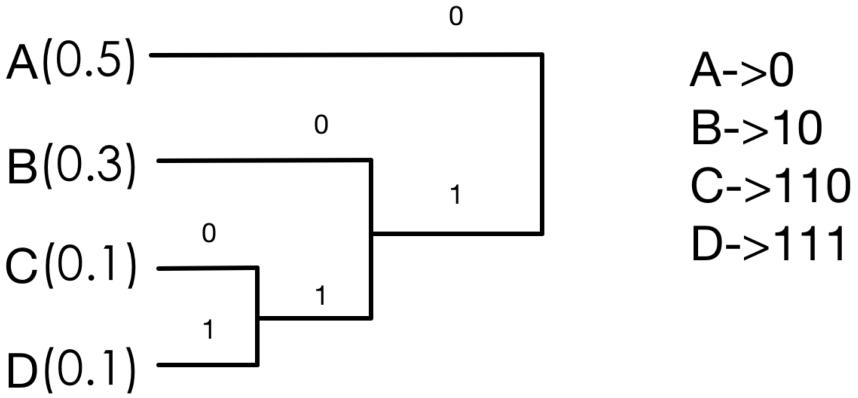

Figure 2.2: Huffman coding

Huffman coding gives a CR of 1.65 in [24]. Moreover, G. Antoniol generated a file-independent Huffman tree with a train dataset to simplify the encoder and decoder op-

erations under the hypothesis that all EEG data are highly alike, and then all the data were encoded with this pre-generated tree, and a CR of 1.69 was attained. It is important to note that this experiment was done in 1990s, and the data used in [24] only had 20 channels, with 8 bits accuracy, at a sampling rate of 128 Hz, which cannot satisfy todays requirements. A similar experiment was conducted in [25], and by using signals with a resolution of 16 bits sampled at 1.024 kHz, a CR of 1.33 was achieved.

b. Arithmetic coding

Arithmetic coding was explored as another famous entropy encoding technique in [26]. Compared to Huffman coding, it encodes all the symbols into one fraction n where $0<n<1$. In arithmetic coding, an initial range will be assigned based on the frequency of occurrence of every symbol, and as the input data are encoded, this range decreases. In Figure 2.3, there are two symbols(A, B) with probabilities of occurrence of 2/3 and 1/3. The shortest binary code that falls into the range of the received data will be picked as a representation, and in this case, data AA can be encoded as .01. Compression is achieved as the more probable it is that a symbol occurs the wider range it needs, and this leads to fewer bits to present this range. Apparently, arithmetic coding has a higher complexity in computation, but a CR of 1.46 was reached in [26], which is better than the performance of Huffman coding when compressing the same EEG data [25].

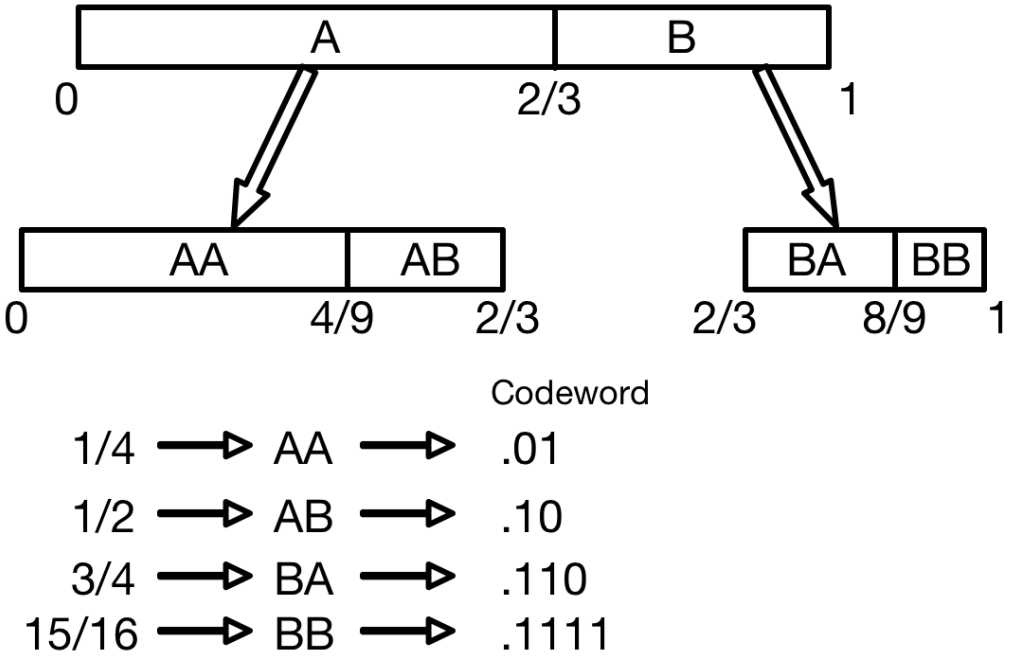

Figure 2.3: Arithmetic coding

c. Dictionary-based encoding

Dictionary-based encoding techniques such as run-length encoding (RLE) and the Lempel-Ziv family (LZ77, LZW, etc.) are designed to exploit the frequent reoccurrence of certain patterns in the data, but such patterns are rarely found in EEG data because of the nondeterministic nature of EEG. Therefore, dictionary-based encoding performs are rather disappointing on EEG data.

Taking Lempel-Ziv 77 (LZ77) for instance, this scheme replaces parts of the input sequence with a length-distance pair representation of an identical sample that appears earlier in the input stream. This can be achieved by using a non-overlapping sliding window to scan the input stream and build up a dictionary of code words based on the input seen so far, and this dictionary can be used to encode the repetitions of those identical samples that come later in the stream. A length-distance pair is used to represent these repetitions of identical samples, which point to the dictionary location of the same sample that appeared earlier in this data stream. So according to this mechanism, no prior knowledge of data stream characteristics is needed, and the scheme is quite adaptive.

The limitations of LZ77 when used on EEG or other biomedical signals are obvious.

LZ77 needs a search buffer and a dictionary which can be indefinitely increased and extended, and for signals like EEG, the resource LZ77 may consume is tremendous, and it certainly has a negative effect on CR. LZ77 also needs enough input to build up a dictionary of sufficient length before it really starts compressing any data, so this leads to more bits being used to represent the length-distance pair than the original sample at the beginning, and the rest of the data must have enough redundancy to be removed to compensate for these extra bits.

The unsatisfying performance of LZ77 has been proved in [24], when an application called Gzip, a combination of LZ77 and Huffman coding, was used in EEG data compression and a CR of 1.63 was achieved which is even lower than the CR of using Huffman coding only which is 1.65.

2.3.1.2 Predictive coding

Predictive encoding techniques are usually used in waveform compression, especially in early speech coding. During the predictive coding process, the coder (predictor) predicts the value of the current sample x_k based on the previous samples, so if the prediction value is $\hat{x}_k$, then only the difference (prediction error)

$$e_k = x_k - \hat{x}_k \tag{2.4}$$

needs to be transmitted. And when the decoder receives the data, by using the same predictor, the original sample can be fully reconstructed through

$$x_k = e_k + \hat{x}_k. \tag{2.5}$$

Entropy encoding such as Huffman coding or arithmetic coding is used to further compress residues of prediction errors before transmission. According to information theory, predictive encoding compresses the data because it efficiently removes some inter-sample correlations by eliminating a redundancy of

$$\sum\nolimits_{(x_1,x_2,...x_k)} P(x_1, x_2, ...x_k) \log_2 \left[P(x_k|x_{k-1}, ...x_1)\right] \tag{2.6}$$

Several predictive compression techniques were tested in [24], and G. Antoniol reached a CR of 2.37 from the Markov predictor, 2.32 from the digital filtering predictor, 2.42 from the linear predictive coding (LPC), 2.48 from the adaptive linear predictive coding, and 2.22 from the artificial neural network predictor. The differences between these predictors

come from the training techniques they use, and since some of these predictors are not widely applied, most of these encoding methods will not be introduced in this chapter.

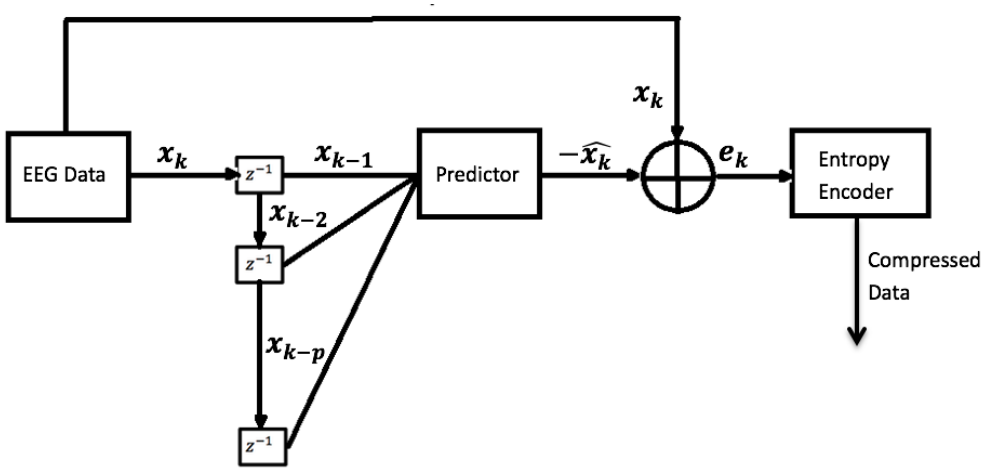

Figure 2.4: Predictive encoding process

a. Differential pulse-code modulation(DPCM)

DPCM is a simple but effective predictive coding technique, and it assumes all signal samples are equal. The encoder transmit only the difference between adjacent signal samples, which can be described as:

$$y_i[n] = x_i[n] - x_i[n-1] \tag{2.7}$$

where $y_i[n]$ is the output of the encoder. It takes advantage of the high correlation of the adjacent time domain symbols. These correlated values are generally very similar, hence the differences between them are small, and this similarity is largely observed in biomedical signals. DPCM has been applied in many EEG compression experiments as the first stage of signal processing because it is a simple reversible process which can also eliminate short-term redundancies efficiently. By introducing this encoder before entropy encoding(e.g. Huffman coding), a higher CR of 2.04 was achieved in [25], and 1.98 in [26], which are all significantly better than the results from experiments without DPCM.

b. context-based linear predictive encoding

Further, in [27], an enhanced near-lossless context-based predictive encoding using the autoregressive (AR) model was introduced (Figure 2.4). The AR model is a linear predictor, and its encoding process can be defined as

$$x_k = \sum_{i=1}^{P} a_i x_{k-1} + e_k = \hat{x}_k + e_k, \tag{2.8}$$

where P is the order of the AR process, $\hat{x}_k$ is the prediction value, e_k is the prediction error for x_k, and a_i is the parameter of the AR model which can be derived by the LevinsonDurbins method or other algorithms. Instead of sending e_k directly like in other predictive encoding processes, the error is sent to a correction encoder. Inside the encoder, by assigning each EEG data sample to a context C, the conditional expectations of prediction error E(e|C) can be calculated along with the conditional mean $\bar{e}(C)$, and since $\bar{e}(C)$ is the most possible prediction error within C, a prediction result of $\hat{x}_k + \bar{e}(C)$ is more accurate than $\hat{x}_k$, and the information to be transmitted, which is

$$e_k = x_k - \hat{x}_k - \bar{e}(C) \tag{2.9}$$

is reduced (Figure 2.5). The compression contexts were usually formed by taking the differences between adjacent EEG data samples.

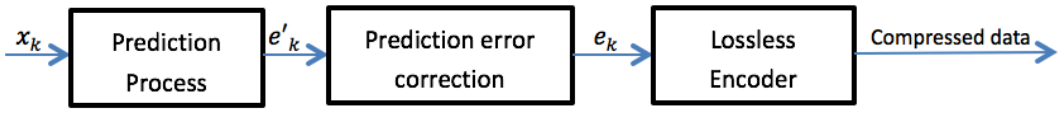

Figure 2.5: Context-based bias cancellation

Later in [27], a near-lossless compression method that allows small discrepancies between original data and reconstructed data to minimize entropy of predictive error sequence was investigated. an error bound θ was introduced in [27] as

$$\theta \geq |x_k - \tilde{x}_k|, \tag{2.10}$$

where x_k and $\tilde{x}_k$ are the values of original and reconstructed signals respectively. With a $\theta = 0$, the lossless compression is guaranteed, and a $\theta > 0$ leads to lossy reconstruction. So a new prediction error within a range

$$\bar{e}_k \in [e_k + \theta, e_k - \theta] \tag{2.11}$$

is given which allows an error sequence with the lowest entropy to be chosen. The e_k is uniformly quantized with

$$Q_k = \lfloor \frac{e_k + \theta}{2\theta + 1} \rfloor \tag{2.12}$$

where $\lfloor . \rfloor$ denotes the integer part of the argument, and is then transmitted. The $\bar{e_k}$ is decoded with

$$\bar{e}_k = Q_k * (2\theta + 1) \tag{2.13}$$

When $\theta = 1$, it gave a better CR at the sacrifice of less than 1% error in [27] when compressing the EEG signal collected from the piglet. In order to satisfy the near-lossless criteria, EEG signals recorded under different physiological conditions have diverse requirements of θ values ranging from 1 to 10, and a good survey on this was given in [28].

A more detailed evaluation of the near-lossless linear predictive encoding was conducted in [29], and a CR of 2.44 with a PRD of 1.65 was achieved with near-lossless encoding when $\theta = 3$. Both of them applied the Huffman coder as the entropy encoder.

According to the result of the near-lossless linear predictor in [29], even with the sacrifice of some original data and the simplicity of the compressor, it seems that the CR does not improve a lot compare to the lossless techniques, so near-lossless predictive encoding may not be an efficient data compression method.

In general, a predictors structure is sophisticated, and it takes a long time to train especially in the case of biomedical signals, so further hardware implementation may be improbable.

2.3.2 Transform-based compression techniques

In general, transform-based compression projects a signal onto a suitable basis, and the projected coefficients are encoded, usually with entropy coding techniques, and are compressed to send, so the compression is achieved if data are expressed more concisely in projection form than in their original form. The Fourier family, including Fast Fourier Transform(FFT) and Discrete Fourier Transform(DFT), and Discrete Cosine Transform (DCT) are broadly used in speech coding and image coding which yield very good results for speech and image signals. These algorithms employ stationary waveforms as functions of basis, so only the frequency information of a signal will be captured, therefore a data reduction can be achieved. As mentioned earlier, biomedical signals like EEG are highly non-stationary, and Fourier family and DCT are not very often chosen to compress these signals for this reason. Other lossless and lossy techniques such as the Karhunen-Loeve transform (KLT), Wavelet family are, on the other hand, more commonly used to compress biomedical data as they reveal the time-frequency information of a signal.

2.3.2.1 KLT

KLT is claimed to be the optimum lossless transform-based compression method because of its good performance in decorrelating non-stationary data like the EEG signal. When applied to EEG data, the KLT processes a sequence of N EEG signal samples as a point X in an N-dimensional space. A more concise representation of X can be obtained by introducing an orthogonal transformation

$$Y = TX \tag{2.14}$$

where Y denotes the transformed vector and T denotes the transformation matrix [30]. After KL transformation, a subsequence of Y which contains M components can be derived. Since T is generated from the covariance matrix of X, Y always adapts to the transformation process. The compression is achieved because M is smaller than N, and (N-M) components are discarded. A CR of 2.36 is achieved on EEG in [25]. The KLT is a complex process and there is no fast algorithm for it, so probably it is computationally inefficient for wearable devices.

2.3.2.2 DWT

Discrete wavelet transform (DWT) employs non-stationary waveforms as functions of basis so that both the time and frequency information can be extracted from a signal [31]. The DWT expresses a signal as a weighted sum of basis functions, and these basis functions are derived from dilated and translated versions of a function which is called mother wavelet [31] [32]. So the original signal can be defined with coefficients of the set of basis functions. The mother wavelet is converted and varied in scale to extract both time and frequency information of a signal. The basis functions with large scales are used to extract low frequency information from the signal while small scales are used to extract high frequency components. The DWT coefficients $c_{m,k}$ are defined as the inner product of the original signal x(n) and the selected basis functions $\psi_{m,k}$ [31],

$$c_{m,k} = \langle x_n, \psi_{m,k} \rangle \tag{2.15}$$

where m decides the wavelets scale and k controls the wavelets translation.

With only these coefficients a signal can be represented. The ability of wavelet analysis to localise the signals energy components from both time and frequency perspectives makes it suitable for compressing non-stationary signals like EEG.

A typical implementation of DWT consists of recursive decomposition of the original signal using quadrature mirror low-pass and high-pass filters [30].

JPEG 2000 is a well-known application of DWT, and it has been applied to lossy EEG compression in [33], and the arithmetic coder was used as the entropy encoder. The EEG data used here is the same as it in [25], and a CR of 8 was reached with a PRD of 30 which was claimed to be acceptable for most clinical uses.

2.3.2.3 DCT/FFT

One of the major limitations of Fourier-based analysis is its inability to capture the information about the time that different frequencies of a signal occur. As mentioned earlier, Fourier-based techniques are sometimes inadequate for non-stationary signals like EEG whose frequencies vary with the time. However, due to DCT/FFTs good performance in compressing ECG [34], their potential for decorrelating and reducing multichannel EEG signals′ inter-channel and inter-sample redundancy has been mentioned in [35]. According to the experiment in [35], CRs of at least 16.7 and 17.6 with PRDs of 0.26 and 0.78 can be achieved from DCT and FFT respectively. Since the original EEG data used to get this result have never been used in other studies, the Fourier familys performance on the EEG signal is still questionable, but its ability to compress other biomedical signals like EMG and ECG has been proved.

2.3.3 Compressive sensing

Unlike the above mentioned techniques that compress the data after the original signal is sampled, compressive sensing allows compression while sampling. In signal processing, the signal must be sampled at or above Nyquist rate, which is twice the maximum frequency component (bandwidth) in that signal, and compressive sensing is a lossy compression technique that lowers the effective sampling rate of the signal. Due to a signal′s sparsity, its "information rate" sometimes is much smaller than the bandwidth to represent it, and compressive sensing assumes that a proper basis can be a concise representation of a natural signal that is often sparse and compressive [36]. In the case of EEG acquisition, 100 Hz or higher sampling rates are normally used, and with compressive sensing, the sampling rate can be reduced, and fewer data are collected as a result.

To further illustrate this technique, consider a single-channel EEG signal, and after the analog-to-digital converter(ADC) we get a signal x, which in this case is an N X 1

vector. We assume x can be expanded in an orthonormal basis(assume a Fourier basis):

$$x = \sum_{i=1}^{N} s_i \psi_i or X = \Psi s \tag{2.16}$$

where s is an N by 1 vector that contains the coefficient sequence of x, and $s_i = \langle x, i \rangle$, and Ψ is an N by N basis matrix. If vector s is sparse, then s can be well approximated by a s_m which is s_i with all but m largest values set to zero, and with compressive sensing, x can be recovered by s_m which is more compressive than s_i.

To conduct the compression is actually to find another basis Φ used for sensing the x and then get:

$$y = \Phi x \tag{2.17}$$

where y is the x after compressive sensing and Φ is an M by N basis matrix. As long as M < N, and the coherence between M and N is very low, e.g. N could be the canonical or spike basis in this case, the compression can be achieved.

In [37], a CR of at least 2.3 can be achieved with different PRDs from 20.39 to 184.77 depending on which basis is chosen.

2.4 Biomedical data reduction

There are several other methods to reduce the size of biomedical data, and they are mainly designed for removing noise, artifacts in the signal. These noise and artifacts degrade the quality of the signal and bring redundancies. Sometimes parts that researchers are not interested in can also be removed to reduce the size of collected data, for instance, if some neurologists only interested in EEG signal during the epileptic seizures, then other background EEG signals can be removed as they are irrelevant to the study. Unlike the lossless data compression techniques mentioned earlier, these filtering-based data reduction processes are usually irreversible, therefore the original signal is unable to be reconstructed after the data reduction process is finished, which brings some risks of losing important information in the original signal. On the other hand, the automated removal of these "unwanted" parts not only reduces the data collected, but also saves time for researchers as the job is mostly done by visual analysis now.

2.4.1 Artifacts and noise

Artifacts and noise often come from the source given below, and there are techniques to detect or reduce the interference originated from the artifacts and the noise.

2.4.1.1 DC offset

As mentioned earlier, artifacts and noise are either from a physiological origin or a technical origin. DC offset comes from the latter, and is caused by the movement of electrodes during signal acquisition. DCPM is usually used to remove DC offset. If the baseline of the original signal moves up or down due to the movement of electrodes, DPCM can easily remove the offset.

2.4.1.2 Artifacts from other electrical source in human body

Given the fact that many biomedical signals are alike, removing artifacts from other physiological sources has always been difficult. The current solution is to record the signals from the source of artifacts simultaneously, and as for EEG acquisition, several channels of reference EOG data are necessary.

With the reference data, some techniques like the least mean-square(LMS) algorithm could be used [38] [39].

Note that another widely used artifacts cancellation technique is the independent component analysis(ICA) [40], which doesn not require any pre-knowledge of the signal or reference channel of the artifact source. A more comprehensive introduction can be found in [41].

2.4.2 Data reduction techniques

If the "unwanted" parts of the biomedical signal are found, several techniques can be applied to reduce the data size. Finding the "unwanted" parts requires very accurate signal classification techniques to make sure all the important information is kept. The classification techniques will not be fully reviewed here.

2.4.2.1 Adaptive sampling

In many experiments, some parts of biomedical signals are taken as reference data being used to distinguish those important signals, thus a lower accuracy of the reference signal

is still acceptable. For example, different stages of sleep can be analysed by extracting relevant EEG features [42], and these parameters can help to diagnose sleep disorders. According to [42], patients with sleep apnoea syndrome show some different spectral features in their EEG signals during sleep stages, so neurologists are more interested in these sleep stages than in the $'$awake$'$ stages. As a result, a wearable EEG recorder may not keep to the same sampling frequency all the time, and the sampling frequency during the $'$awake$'$ stages can be tuned lower, which will definitely reduce the data size without compromising the diagnosis. If a subject stays awake for 40% of the experiment time, then decreasing the sampling rate by 80% during the awake stages can reduce the data size by 32%.

Obviously, the adaptive sampling requires a prior knowledge of the data, and it is presumably achievable in this case because the fact that dominant frequency changes in the different stages of sleep, and due to these changes, an ADC with an adaptive sampling rate controlled by these frequency changes can be built.

2.4.2.2 Discontinuous recording

Discontinuous recording is similar to the adaptive sampling, but instead of decreasing the sampling rate, it simply stops recording when the data is thought to be unrelated. And here is an example of its application: according to many reports, neurologists find that epileptic EEG traces can be separated into two phases − ictal, which contains seizure activities, and interictal, which contains spikes and waves between seizures [43]. Data with seizures are more important to neurologists because they are considered to be the most prominent feature of neurological dysfunction. For a long time, neurologists have had to manually identify these two phases from other background EEG signals which are not diagnostically useful. However, if a preliminary classification can be done on the wearable device, and only relevant data are transmitted, the data size along with peoples workload will be largely reduced. Detecting and eliminating the background signals is not a new concept, and it has been applied to inpatient epilepsy monitoring for years [44], but it can possibly be implemented in wearable recorders in the future.

It is also worth mentioning that abnormal signals caused by epilepsy can be detected with wavelet transforms (DWT, CWT, etc.) [45] [46], and, as presented earlier, wavelet transforms are also used in data compression, so a better result may be achievable if we can combine ictal and interictal detection and transform-based data compression together.

The prerequisite of using the adaptive sampling rate or discontinuous recording is the

related knowledge about the signals for picking out useful ones from those less important ones, and this requires massive experiments and field research.

A technology called Real-Time EEG Analysis for Event Detection (REACT) has been proposed in recent years [47]. REACT contains several steps including feature extracting, classification and post-processing stages, and it can be summarized as a sliding window scheme which gives out the probability of seizures for each window of EEG data. With the help of REACT, the discontinuous recording as well as the adaptive sampling rate could be more powerful and accurate when applied to clinical EEG data reduction.

2.5 Compression results summary

Table 2.1 summarises the performance of EEG compression techniques mentioned above, and irreversible data reduction techniques are not included in this table.

From the results shown in Table 2.1, it is clear that lossy compression techniques can usually achieve higher CRs than lossless techniques, and that is because more information is discarded during the compression process. The CRs that lossless techniques can achieve are close to each others even though the datasets used to evaluate the compression techniques are different.

Table 2.1: Compression results summary

Reference	[24]	[25]	[26]	[24]	[24]	[25]
Compression method	Huffman	Huffman	Arithmetic	L77+Huffman	Predictive encoding with various predictors	DPCM+Huffman
Compression type	Lossless	Lossless	Lossless	Lossless	Lossy	Lossless
Data sampling rate	128Hz	1.024kHz	1.024kHz	128Hz	128Hz	256Hz
Data bit-width	8-bit	16-bit	16-bit	8-bit	8-bit	16-bit
CR	1.69	1.33	1.46	1.63	2.22-2.48	1.98
PRD	N/A	N/A	N/A	N/A	Not mentioned	N/A
Reference	[27]	[25]	[33]	[35]	[35]	[37]
Compression method	Context-based predictive	KLT	JPEG2000	DCT	FFT	Compressive Sensing
Compression type	Near-lossless	lossless	Lossy	Lossy	Lossy	Lossy
Data sampling rate	173.6Hz	1.024kHz	256Hz	128Hz	128Hz	200Hz
Data bit-width	12-bit	16-bit	16-bit	14-bit	14-bit	12-bit
CR	2.08	2.36	8	16.7	17.6	2.3
PRD	1	N/A	30	0.26	0.78	20.39

2.6 Wearable wireless EEG signal recorder

The portable EEG signal recorder is an good example of how wearable wireless devices will bring convenience to the users. In this part, two typical applications of wearable wireless EEG recorders are given, followed by an introduction to the simplified hardware model of an EEG signal recorder. The power consumption analysis based on the given hardware model is presented at the end of this part.

2.6.1 Applications

2.6.1.1 Epilepsy

Epilepsy is a neurological disease caused by pathological conditions such as brain injury, stroke or genetic factors, or more often is of unknown aetiology.

The EEG is used as the principle test for diagnosing epilepsy as well as gathering information of seizures. As seizures are not frequently observed in an epilepsy patient, a long-term EEG recording is always required, and it sometimes takes days and can be expensive if it is done on an inpatient basis. A small portable device that can be worn during everyday activities is no doubt a better alternative, and a wireless wearable device clearly has a very promising future in this area. An EEG recording device that can be attached to a belt around the subjects waist is described in [48], while other products may be released in the future.

2.6.1.2 BCI

A BCI builds a direct link between the subject and the outside world without using other parts of the human body as we usually do. Commands from the brain will be conveyed and interpreted by the BCI, and used to control any peripheral apparatus. The BCI is a technique that can significantly increase the quality of life of patients with severe neuromuscular disorders or those who are paralysed.

The concept of BCI was first proposed nearly 50 years ago [49], but it was not until the 1990s that people were able to build reliable BCIs after obtaining more knowledge about the EEG signal with the help of computers [50] [51].

As the accuracy of the interpretation of the EEG signal increases, smaller BCI devices make the user experiences even better. Devices like Emotiv [8] and iFocusBand [52] further broaden the area that wearable BCI devices can used in including drone control and driver

vigilance monitoring.

2.6.2 Simplified hardware model

A simplified hardware model of a wearable wireless EEG recorder is mentioned in [53] to help to estimate the system power of such device, and it is showed in Figure 2.6. It contains the most essential components for recording the EEG signal including amplifier(s), ADC(s) and a radio transmitter. The extra reference (REF) channels showed in Figure 2.6 are for recording EOG signals that are often needed for eliminating the artifacts from the acquired EEG signal.

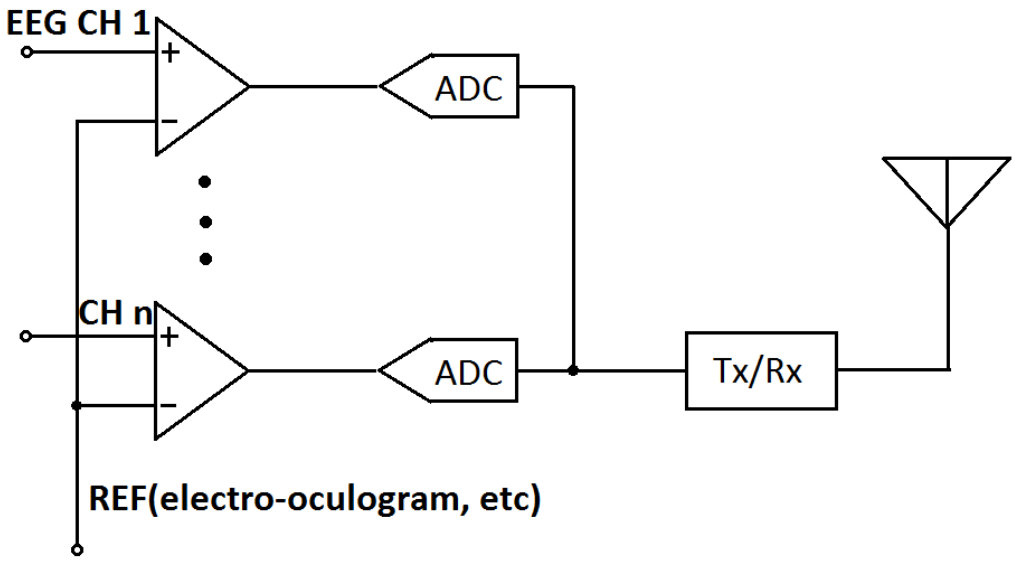

Figure 2.6: Simplified hardware model of wearable wireless device

The storage medium is optional to this simplified hardware model as it consumes relatively lower power than the listed components, but similar to the transceiver, its power consumption is also susceptible to the amount of data sending in and out, so a brief power analysis of the storage medium will be presented separately.

2.6.3 Power analysis

2.6.3.1 System power analysis

Based on the model given earlier, an overall power consumption of the system can be easily derived as:

$$P_{sys} = P_{amp} + P_{ADC} + P_{Tx} \tag{2.18}$$

where P_{Amp} and P_{ADC} are the power consumption of amplifier(s) and ADC(s) respectively, and P_{Tx} is the power consumption of the transceiver, which is further defined as

$$P_{Tx} = D * P_{Trans} \tag{2.19}$$

where P_{Trans} is the power of the transceiver when operating at the highest transmission rate, and D is the duty cycle rate which is defined such that:

$$D = R_{Orig}/R_{Tx}. \tag{2.20}$$

where R_{Orig} is the required data rate, and R_{Tx} is the maximum data rate of the transceiver.

The power of the transceiver when operating at the highest transmission rate is further defined as

$$P_{Trans} = N_{bit} * E_{tx/bit} \tag{2.21}$$

where N_{bit} is the number of bits the transceiver can send at its maximum data rate, and $E_{tx/bit}$ is the energy needed for transmitting one bit data.

The amount of power consumed by the ADCs and amplifiers depends on the signal collecting system chosen, and more power will be consumed if the system requires more channels of data to be recorded. Nevertheless the wireless transceiver is always the one of the biggest contributors to the power consumption of an EEG acquisition system compare to the ADCs and amplifiers.

There are several off-the-shelf electronic devices that can help to estimate the power consumption during the signal acquisition process.

Assuming an EEG acquisition system that collects 24 channels of data, and therefore 24*1 μW amplifiers [54], 24*1 μW ADCs [55], and a 22mW off-the-shelf Bluetooth transceiver nRF8001 [56] are chosen to estimate the system power. The maximum data rate of nRF8001 is 1Mbps, so to transmit 1 bit of data, it consumes around 22nJ power. If the system operates at a sampling rate of 1 kHz, and the signal is digitised into 12-bit, a data rate of 288kbps is needed from the transceiver.

The overall power consumption can be estimated with (2.18), where $P_{Amp} = 24\mu W$ and $P_{ADC} = 24\mu W$, and $P_{Trans} = 22mW$. The duty cycle rate can be calculated with (2.20), and D = 0.288 when R_{Orig} and R_{Tx} are 1Mbps and 288kbps respectively. Therefore, $P_{Tx} = 6.36mW$ and $P_{sys} = 6.41mW$ in this case. If using a standard large coin cell battery with a capacity of 160mAh [57] to power this device, a lifetime of approximately 25 hours is expected which is inadequate for many applications.

2.6.3.2 Power analysis of the storage medium

The flash memory is most likely to be used on a wearable and wireless device as it consumes less power and the flash memory is non-volatile [58–62]. In general, the power of the storage medium can be measured with [63]:

$$P_{read} = N_{rb} * E_{read/bit} \tag{2.22}$$

$$P_{write} = N_{wb} * E_{write/bit} \tag{2.23}$$

where P_{read} and P_{write} is the power of memory when read and write, and N_{rb} and N_{wb} is the number of bits read and write per second. $E_{read/bit}$ and $E_{write/bit}$ is read and write energy per bit which are given as:

$$E_{read/bit} = \frac{P_{active,read} * t_{cycle,read}}{N_{I/O}} \tag{2.24}$$

$$E_{write/bit} = \frac{P_{active,write} * t_{cycle,write}}{N_{I/O}} \tag{2.25}$$

where $P_{active,read}$ and $P_{active,write}$ are the active read and write power respectively, and $t_{cycle,read}$ and $t_{cycle,write}$ are the minimum read and write cycle time respectively. $N_{I/O}$ is the number of Input/Outputs.

From the equation (2.22) and (2.23), it is quite clear that a lower data rate will result in a lower power consumption of the storage medium.

2.7 Hardware implementation of the biomedical data compression technique

There are not much literature mentioning the hardware implementation of the biomedical data compression techniques and how these techniques help to reduce the system power of a wearable biomedical device.

The hardware implementation for a 2-stage lossless biomedical data compression technique that combines predictive and entropy coding for brain-heart monitoring systems was introduced in [64]. The average CR achieved from EEG/ECG/diffuse optical tomography (DOT) was 2.05. The compressor was designed with 65 nm complementary metal-oxide semiconductor (CMOS) technology and claimed to save about 43% energy from a monitoring system with Bluetooth.

In [65–67], the hardware implementations of three different compression technologies for ECG sensing applications and their estimated power consumptions were given. These

techniques are similar to the one described in [64], which were consisted of the predictive coding and entropy coding. A CR of 2.53 was reported in [65], and 2.25 in [66], and 2.43 in [67]. The impacts on the system power these techniques can bring were not mentioned.

2.8 Summary

The literature outlined earlier has presented many promising data compression/reduction methods for EEG or other biomedical data, and a CR of 17 can be reached from EEG data which is extraordinary, but very little information can be found on the aspect of the implementation of these methods, let alone a full power analysis of these compressors. Many of these techniques are designed to save the power of wireless transceivers or memory units on the wearable wireless devices as mentioned in [53, 64, 68–71] because transmitting or storing less data clearly reduces the power consumed by these components, so CR and PRD as the only measurements for evaluating the performance of the data compression/reduction technique are inadequate. To achieve system power saving, the power consumption of a data compressor should also be taken into account. As a result, the rest of this thesis explores a lossless data compression technique that not only gives good CR but also has been proved to be able to save power.

Chapter 3

Hypothesis and methodology

In this chapter, the hypothesis of this research is given, and the methodology that will be used to prove the given hypothesis is also outlined.

3.1 Hypothesis

A review of different techniques that can be used to reduce the size of biomedical data especially EEG data, is given in chapter 2. As briefly mentioned earlier, these techniques are mainly developed to reduce the overall power consumption of the wearable wireless biomedical signal recorder, but despite the high CR they usually achieve, very little work has been reported to evaluate the hardware cost of introducing an extra data compressor to a signal recording device, and whether this data compressor will consume more power than it can save is not yet very clear.

This research assumes that a good data compression/reduction technique can reduce the size of recorded biomedical data without consuming too much power itself, and therefore having a data compressor should help to reduce the system power consumption of the device. To achieve that, the data compression/reduction technique this research aims to find should deliver at least the following benefits:

3.1.1 Power saving from the transceiver

The technique this research provides should reduce the data transmission rate with a minimum cost. A simplified hardware model of the wireless EEG signal recorder is given in Figure 2.6 to evaluate the power of such devices, and in order to estimate the compressor′s impact on the overall system power consumption, a hardware model was proposed in [53],

and it is given in Figure 3.1, which includes the data compression/reduction unit. Based on this new model, the new system power P'_{sys} after data compression/reduction can then be calculated with

$$P'_{sys} = P_{amp} + P_{ADC} + P_{compressor} + P'_{Tx} \tag{3.1}$$

where P_{Amp} and P_{ADC} are the power consumption of amplifier(s) and ADC(s) respectively, and P'_{Tx} is the power consumption of the transceiver for transmitting compressed data, which is defined as

$$P'_{Tx} = D * P'_{Trans} \tag{3.2}$$

where D is the duty cycle rate defined in (2.20), and P'_{Trans} is the power of the transceiver after introducing the data compression, which can be further defined as

$$P'_{Trans} = N_{bit} * E_{tx/bit}/CR \tag{3.3}$$

where CR is the compression ratio defined in (2.2), and N_{bit} and $E_{tx/bit}$ are the number of bits the transceiver can send at its maximum data rate and the energy needed for transmitting one bit data, which are defined in (2.21).

In theory, by consuming the same amount of energy for sending uncompressed data, more data can be transmitted after the compression, therefore the equivalent energy E_{eqtx}, which is defined as

$$E_{eqtx} = E_{tx/bit}/CR \tag{3.4}$$

for sending one bit is lower than before.

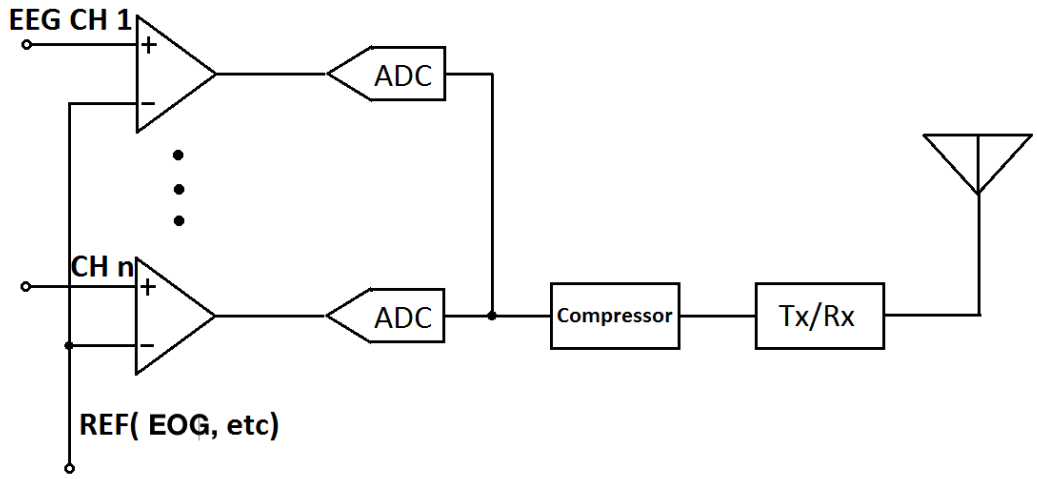

Figure 3.1: Simplified wearable wireless device model with data compression unit implemented

From equations (3.1) and (3.3), it is quite clear that a higher CR will help to reduce the data rate, and therefore the power that the transceiver needs. According to the review of the biomedical data compression techniques given in the chapter 2, complex compression methods like DWT often yield better CR results compared with simple methods like entropy encoding techniques, and as the CR increases, so $P_{compressor}$ is very likely to increase as well, since more computational resources will be needed to fulfil the task like wavelet transform. There is a trade-off between the CR that a compression method can achieve and the hardware resource it may need. A simple compression technique that gives an acceptable CR, but consumes a minimum resource, is often a good alternative to those hardware-hungry methods.

3.1.2 Power saving from the memory unit

As mentioned in chapter 2, the data storage unit does not consume much power compared with a transceiver on a wearable wireless device, but it is worth mentioning that the data compression/reduction technique this research looks for should also be able to reduce the data read/write rate of the memory unit at a low cost, which will help to further reduce the power consumption.

A method of estimating the power consumption of the memory unit during read and write is given in (2.22) and (2.23), and if the data size is reduced, the new number of bits read and written per second $N_{rb}^{'}$ and $N_{wb}^{'}$ can be calculated as

$$N_{rb}^{'} = N_{rb}/CR \tag{3.5}$$

$$N_{wb}^{'} = N_{wb}/CR \tag{3.6}$$

where CR is defined in (2.2),so the read and write power consumption of the memory unit $P_{read}^{'}$ and $P_{write}^{'}$ will be

$$P_{read}^{'} = N_{rb}^{'} * E_{read/bit} = N_{rb} * E_{read/bit}/CR \tag{3.7}$$

$$P_{write}^{'} = N_{rb}^{'} * E_{write/bit} = N_{wb} * E_{write/bit}/CR \tag{3.8}$$

where $E_{read/bit}$ and $E_{write/bit}$ is given in (2.24) and (2.25) respectively.

Similar to the transceiver power, the equivalent energy for reading and writing the memory can be calculated with

$$E_{eqr} = E_{read/bit}/CR \tag{3.9}$$

and

$$E_{eqw} = E_{write/bit}/CR \tag{3.10}$$

The CR is decisive to the power saving of the memory unit, and a data compressor that gives a higher CR with minimum power consumption is always preferable.

3.1.3 System power saving

A wearable wireless signal recorder with the data compression/reduction unit should operates longer than a recorder without such unit, and to guarantee this, the unit should consume less power than the power saved from the transceiver or the memory unit. The whole research rests on this assumption.

3.1.4 Storage saving

Storage saving is another benefit that the data compressor can bring to the signal recorder. The memory can store more data if they are compressed, and if the recording operates longer, extra memory is less demanded if data is compressed.

3.2 Methodology

Based on the hypothesis, the research methodology is outlined here to illustrate how this research will identify a good set of trade-offs between the compression techniques and the system power saving.

3.2.1 Data compression technique candidates

Many signal acquisition devices are used in clinical research, and therefore only the lossless data compression technique that can fully preserve the original information of the signal will be considered in this research, as it satisfies the strict requirements of clinical uses, and techniques like transform based algorithms will not be discussed. A performance metric that contains the CR and the power consumption of the compressor is introduced for evaluating the existing compression technique, and developing the new technique as the ultimate purpose of introducing a data compressor is to reduce the system power.

3.2.1.1 DPCM + Huffman coding

A data compression process that combines DPCM and the Huffman coding is selected for further analysis for the following reasons.

a. Computational simplicity

There are several lossless data compression techniques that can be considered for further analysis based on the review given in chapter 2, and they are Huffman coding, Arithmetic coding, KLT, and 2-stage encoding techniques including LZ77+Huffman coding and DPCM+Huffman coding. Most of these techniques are time-domain based except for the KLT, which is transform based. Given the fact that the KLT certainly requires more computational resource whilst unable to deliver a significantly higher CR when compressing the EEG signal than the time-domain based techniques [25], it will not be taken into the next stage of the research. As a result, only the DPCM+Huffman coding with the other mentioned time-domain based techniques fit the purpose regarding the aspect of computational simplicity.

b. Hardware efficiency

The arithmetic coding and the Huffman coding are two well-known time-domain based entropy coding techniques, and as indicated earlier, both of them require comparatively low computational resource. As introduced earlier in chapter 2, they have similar work schemes, and both algorithms can achieve the data compression via assigning a shorter code to symbol(s) based on the PDF of the signal. The Huffman coding assigns a unique binary code to each possible symbol, and the arithmetic coder encodes multiple symbols into one single and unique binary fraction, and both assigned codes are shorter than the original symbols.

In [26] and [25], it is shown that the arithmetic coding yields a slightly better CR than the Huffman coding does, but [26] also indicates a power consumption of 41mW from the arithmetic encoder when compressing the 16-bit EEG signal, which is even higher than the power that a wireless transceiver needs to send uncompressed data, and according to equations (2.19) and (2.20), the power would be 8.65mW if transmitting 24 channels of EEG signal with nRF8001 [56]. Therefore, the arithmetic coding is not suitable for this research in respect of the hardware efficiency.

c. Promising data compression performance

The 2-stage technique LZ77+Huffman coding is ruled out since this process performs even worse than the Huffman coding alone according to [24], and obviously it also consumes

more computational resource.

The Huffman coding itself gives an acceptable performance when compressing the EEG signal as it is shown in Table 2.1. Moreover, with a DPCM encoder added before the Huffman encoder to remove short-term redundancies, the Huffman coding reached the highest CR in [25] among all the other time-domain based techniques. The performance of the DPCM+Huffman coding is even close to the KLT.

Due to above reasons, the DPCM+Huffman coding technique is chosen as a possible data compressor that can be used on the simplified hardware model of the signal recording system, and since little work has been conducted on the power analysis of the DPCM+Huffman coding, this technique will be selected for a further test.

3.2.1.2 Log2 sub-band encoding - a new algorithm

All entropy coding techniques can only conduct compression with the prior knowledge of the PDF of the original signal, and the Huffman coding inevitably suffers from this limitation. The Huffman coding requires a pre-generated code book which contains all the possible symbols it may encounter and all the corresponding Huffman codes of these symbols. The code book can be easily generated with a training dataset, however when compressing biomedical signals, some unexpected symbols sometimes occur, and they are usually caused by sudden changes within the physiological system such as seizure spikes detected in EEG signals or by artifacts from other sources.

The code book is always a compromise-relying on typical cases to achieve gains, and it contains all occurred symbols included in the training dataset, but it not always covers all the symbols in the dataset to be compressed.

To overcome some of these limitations, a novel technique called Log2 sub-band is developed for this research, based loosely on the principles of the DPCM+Huffman coding but with considerable simplification.

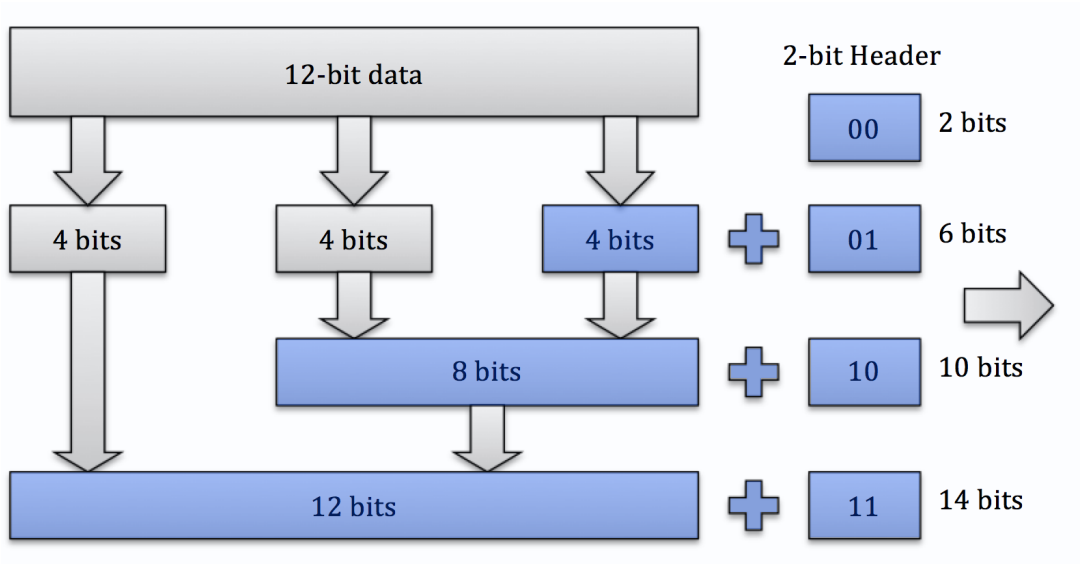

Figure 3.2: Work scheme of the Log2 sub-band compression technique

In general, the Log2 sub-band divides each binary data sample into several parts, and each part may be a few bits long. The technique then compares each part of the current sample with the part in the previous sample. The comparison process goes on part by part until one part in the current data sample is found different from the part in previous one. In the end, only the part that is different and parts after that different part will be transmitted or stored. To reconstruct the data, a header is needed to indicate how many parts of the current data sample are transmitted or stored.

A more specific example of the Log2 sub-band is illustrated in Figure 3.2, and in this case assumes the original biomedical data is digitized into 12-bit, and the Log2 sub-band first divides each signal sample into three 4-bit segments(nibbles). In this case, each nibble then is compared with the same part of the previous data sample, and instead of transmitting or storing the whole current sample, only the nibbles that are different from the previous sample′s are transmitted or stored. A header will be added to each compressed sample to indicate the number of nibbles transmitted or stored, and it could be 3 nibbles, 2 nibbles, 1 nibble, or, as in this case none, so a 2-bit header is added at the end of the comparison process. Schemes can be devised for any bit width, and each band can have arbitrary bit width, for instance Figure 3.2 presents a {4, 4, 4} scheme, and it could also be {3, 4, 5} with a 2-bit header or {3, 3, 3, 3} with a 3-bit header for 12-bit data.

The Log2 sub-band is a lossless data compression technique that combines the sim-

ilarities of the DPCM and the Huffman coding into a single stage process. Unlike the Huffman coding, the Log2 sub-band does not require the PDF of the original signal to generate a code book before compressing, so it will not be interrupted if any new symbols show up. Compared to the DPCM+Huffman coding, the Log2 sub-band is expected to be more adaptive to biomedical signals. A detailed compression result comparison will be given in the next chapter.

3.2.2 Biomedical data for testing

The EEG, EMG and some other biomedical data are chosen for evaluating the compression performance of the DPCM+ Huffman coding technique and the Log2 sub-band.

3.2.2.1 EEG data

a. Human EEG data

One of the EEG datasets used in this thesis is recorded from subjects with different physiological conditions at the University of Bonn on non-linear deterministic patterns of brain electrical signals [72]. These data are used by other literature in [73–78] to test their EEG data compression techniques.

The EEG signals are categorized into five groups based on subjects′ conditions. They are signals recorded from healthy people with eyes open and closed, and from within and outside epilepsy diagnosed patients′ epileptogenic zone (seizure generating area) during the seizure free interval and during the seizure. The signals are digitized into 12-bit at a sampling of 173.6Hz as shown in Figure 3.3. Each group has 100 data segments of 23.6 sec duration of signals. The EEG data of healthy people with eyes open, within epileptogenic zones and seizures are chosen for this research. These three types of data can very well reflect the impact of a neurological disease on human EEG signals and help to demonstrate the performance of a compression technique when compressing different EEG signals.

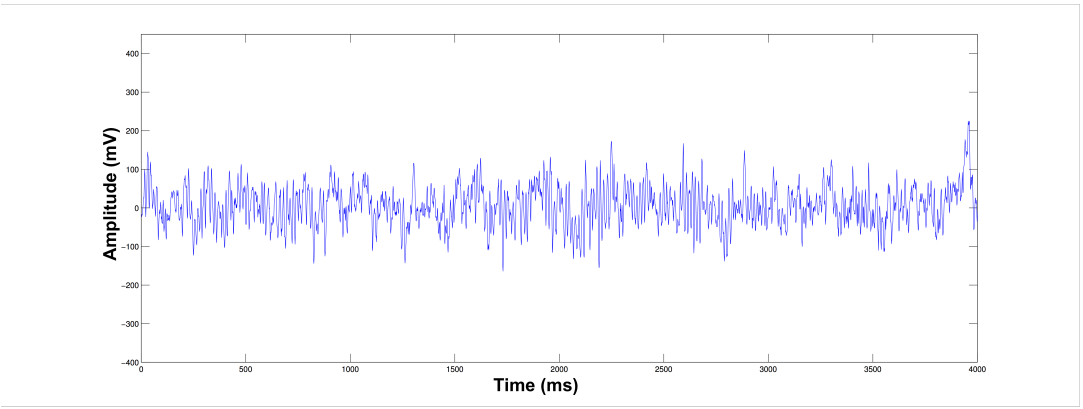

Figure 3.3: A segment of healthy human EEG data

b. Mice EEG data

The mice EEG data are recorded for an Alzheimer′s disease study by a research team in Aberdeen University [79]. The signals as shown in Figure 3.4 are collected and digitized into 12-bit at 1000Hz and 500Hz respectively by the University of York/Cybula Neural Acquisition Tracker (NAT) device [11]. The data used in this research are randomly picked 10-minute segments, and the same selections are used for all compression methods.

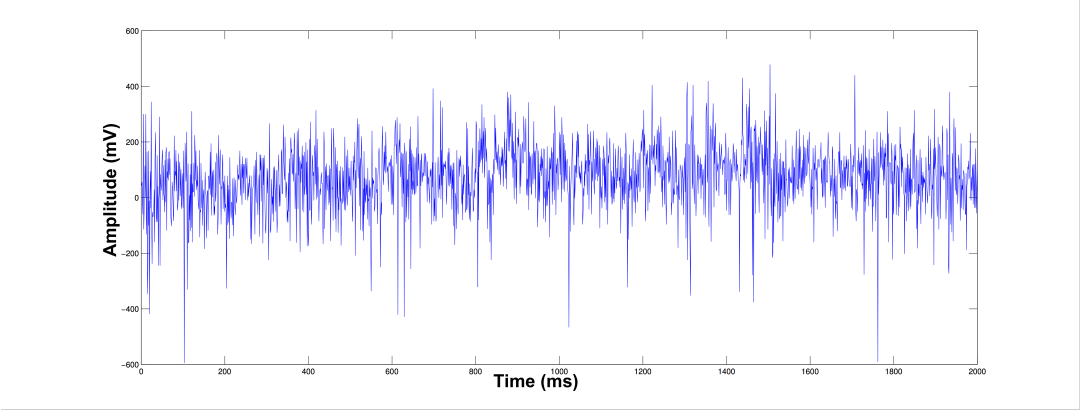

Figure 3.4: A segment of mice EEG data

3.2.2.2 EMG data

The EMG signal is from Physionet [80]. It is collected from the *tibialis anterior* muscle of a 44-year-old man with no history of neuromuscular disease, and the subject was asked to dorsiflex the foot gently against resistance . The signal is digitized into 12-bit, and first sampled at 50kHz, and then downsampled to 4kHz.

3.2.2.3 Others

To illustrate the adaptivity of the Log2 sub-band encoding, some other bio-signals were chosen for testing the compression performance of the Log2 sub-band, which include Electrohysterogram (EHG) [81], ECG [82] and Event-related potentials(ERP) [83].

3.2.3 Data compression result analysis

The DPCM+Huffman coding and the Log2 sub-band are tested with the selected datasets via MATLAB simulations. Since these data have different characteristics depending on their sources and the subjects′ physiological conditions, the performance of these two data compression techniques when compressing these data is expected to change with the conditions.

The DPCM+Huffman coding and the Log2 sub-band encoding are both lossless compression techniques, so only the CR will be used to measure their compression performance, and a higher CR indicates that the technique is more capable of compressing the biomedical data.

Some experiments were also conducted to test the limitations of the DPCM+Huffman coding when using a code book with data from non-training sets.

3.2.4 System power analysis

At this stage, the chosen compression techniques will be transferred into hardware design. Accurate power consumption results of the data compressor can be acquired, and therefore the system power can be estimated.

With the compression results acquired from MATLAB simulations, it is possible to estimate the power saving from the transceiver/storage with equations (3.1), (3.7) and (3.8). For example, assuming an EEG recording system using the transceiver nRF8001 mentioned in chapter 2, a data compression unit that brings a CR of 2 will reduce the power consumption of the transceiver by 50%, which is $3.18mW$ in this case. This amount of power saving is the power budget for the data compressor, and any data compression unit that consumes more power than $3.18mW$ would be considered as sub-optimal.

The power budget for the data compressor changes with the different transceivers or storage media chosen, and the overall system power will be analysed under these different scenarios.

Ideally, the best combination of the data compressor and the transceiver/storage can maximize the battery lifetime on a wearable wireless device.

3.3 Summary

In this chapter, the hypothesis of this research is illustrated, and the methodology of the study is also outlined. The research framework can then be summarized as it is shown in Figure 3.5.

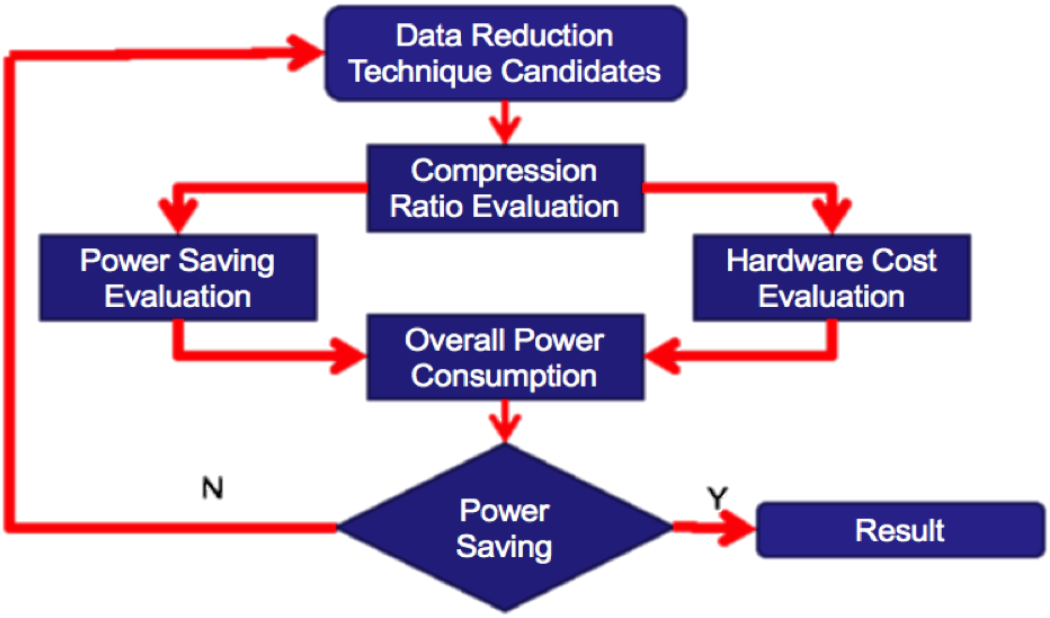

Figure 3.5: Research flow

A data compression/reduction technique that can help to save the system power of a wearable wireless biomedical signal recorder is expected to be demonstrated.

Chapter 4

Compression results analysis

This chapter gives the data compression results of the DPCM+Huffman and the Log2 sub-band encoding, and then their performance will be analysed. A comparison of the performance of the two techniques will be discussed at the end. Some of the work mentioned in this chapter was published in [84–86].

4.1 DPCM+Huffman coding performance analysis

4.1.1 Code book generation

As mentioned earlier, the Huffman coding requires some prior knowledge of the original data, which is the PDF of the signal symbols. A code book based on this information must be generated before the compression process. Therefore, part of the testing data will be used as a training dataset for generating the code book, and then all the data will be compressed with the acquired code book. The whole process is shown in Figure 4.1.

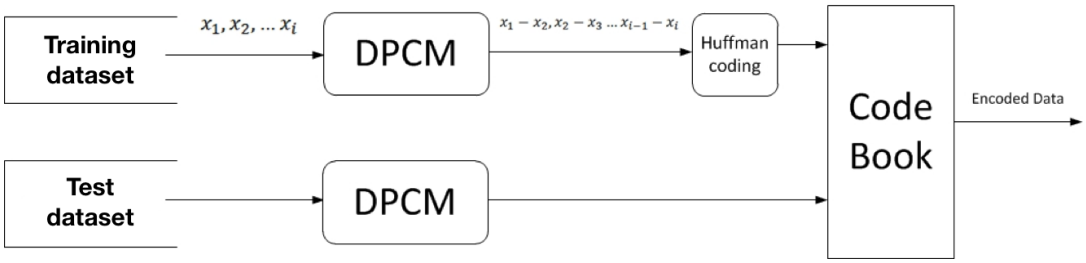

Figure 4.1: DPCM+Huffman coding process

Table 4.1: The compression ratio results of the DPCM + Huffman coding

Data type	**Compression ratio**
EEG from the healthy people	1.82
Seizure-free EEG from the epilepsy patients	2.2
Seizure	1.51
Mice EEG data@1000Hz	1.47
Mice EEG data@200Hz	1.45
EMG data	2.59

Table 4.2: Number of entries of the Huffman code book

Data type	**Number of entries**
EEG from the healthy people	324
Seizure-free EEG from the epilepsy patients	604
Seizure	1910
Mice EEG data@1000Hz	925
Mice EEG data@200Hz	839
EMG data	473

4.1.2 Compression results

The compression results of the DPCM+Huffman coding are given in Table 4.1. The performance of the DPCM+Huffman coding decreased when compressing the seizure signal. That is due to the rapid changes that occurred in the signal, and each occurred symbol usually has a low frequency of occurrence. The Huffman code book generated based on such signals will have more entries, and the average code length to represent one symbol becomes longer. The number of entries of the code book is given in Table 4.2. However, a larger number of entries does not always indicate lower compression results. The code book of the EMG data has more entries than that of the EEG signal from the healthy people, but if some symbols rarely show up, the compression result will still be better, as it is shown in Table 4.1. Overall, this technique performs better when compressing relatively stable signals.

The CR result of the mice EEG signal recorded at different frequencies has proved that the sampling frequency has no significant impact on the compression performance of the

Table 4.3: The compression ratio results of the Log2 sub-band

Data type	**Compression ratio**
EEG from the healthy people	1.94
Seizure-free EEG from the epilepsy patients	2.58
Seizure	1.66
Mice EEG data@1000Hz	1.55
Mice EEG data@200Hz	1.44
EMG data	2.6

2-stage technique, and although the signal with a higher resolution should be containing more information, the number of entries of its code book is not significantly larger.

4.2 Log2 sub-band encoding performance analysis

The Log2 sub-band is more adaptive to the signals compared to the DPCM+Huffman coding. There is no need to pre-process the data before the compression process, therefore all the data were used for evaluating the compression performance. As all the original data are 12-bit, the compression process is exactly the same as is shown in Figure 3.2. The MATLAB code for testing Log2 sub-band is given in Appendix C1.

The compression ratios are given in Table 4.3. Similar to the performance of the DPCM+Huffman coding, the Log2 sub-band yields higher CR for compressing stable signals, and lower for the rapidly changed signals.

The probability distributions of CRs of each one of the 100 EEG data segments [72] from three groups of data are shown in Figures 4.2 to 4.4. The Y-axis in these three figures indicates the number of data segments that achieve the CR given on the X-axis. The CRs of these data shown in Table 4.3 are the averages of the CR of each data segment.

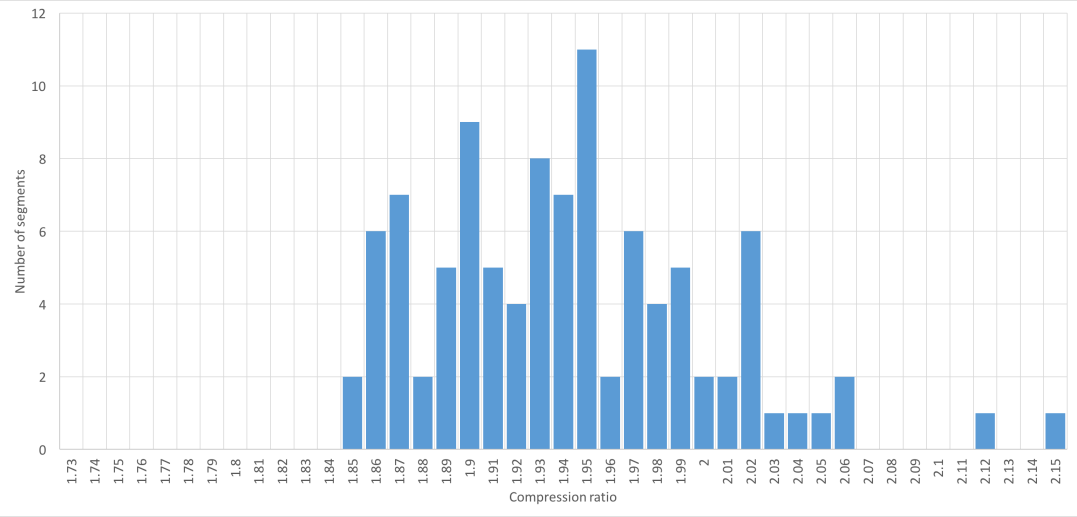

Figure 4.2: The compression ratio of 100 healthy EEG signal segments

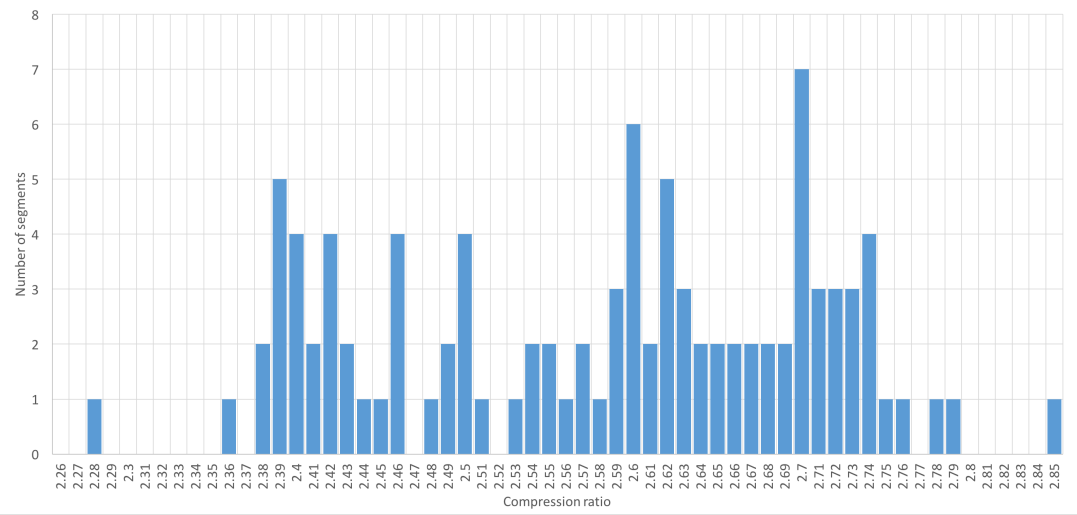

Figure 4.3: The compression ratio of 100 seizure-free EEG signal segments

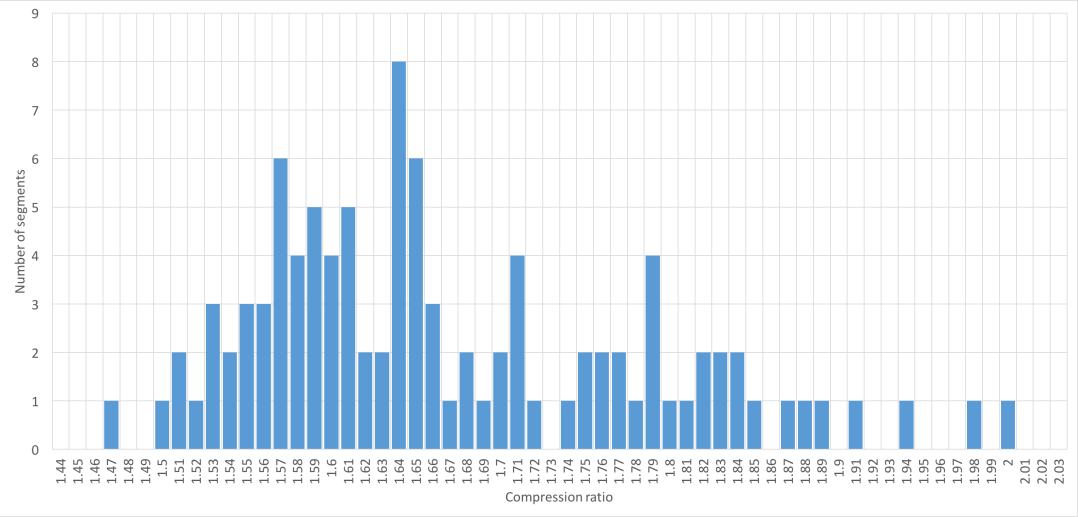

Figure 4.4: The compression ratio of 100 seizure EEG signal segments

Table 4.4: Performance of Log2 sub-band in compressing some other bio-signals

Data type	CR of the Log2 sub-band
EHG [81]	3.02
ECG [82]	2.14
ERP [83]	1.38

From the compression results, it can be inferred that the biomedical signals like the seizure signal require more bands (nibbles) to transmit, and signals with fewer fluctuations requires fewer bands. Moreover, the distribution of the number of bands required to transmit every single signal symbol reflects some of the characteristics of the biomedical signals. One way to utilise this distribution information will be raised in the next chapter.

As a novel lossless compression technique, the Log2 sub-band can compress not only the EEG and the EMG data, but also some other kinds of bio-signals. Its performance in compressing other types of bio-signals [80] is given in Table 4.4, which proves the versatility of this technique.

4.3 Comparison of the techniques

A comparison of the performance of the presented techniques is given in Table 4.5. The compression results indicate that the Log2 sub-band technique performs better than the 2-stage technique in most cases, yet the pre-knowledge of the original signal is not required.

Table 4.5: Comparison of the compression performance

Data type	**CR of 2-stage technique (partially covered code book)**	**CR of the the Log2 sub-band**	**Log2 sub-band′s improvement (%)**
EEG from the healthy people	1.82	1.94	6.6
Seizure-free EEG from the epilepsy patients	2.2	2.58	17.3
Seizure	1.51	1.66	9.9
Mice EEG data@1000Hz	1.47	1.55	5.4
Mice EEG data@200Hz	1.45	1.44	-0.7
EMG data	2.59	2.6	0.4

4.4 Limitations

4.4.1 DPCM+Huffman coding

4.4.1.1 Pre-knowledge of the signal required

The Huffman coding requires the generation of a code book based on the PDF of the original signal. To achieve the best compression result, a large training dataset that can mostly reflect the characteristics of the original signal is required when generating the code book, which makes the compression process inefficient sometimes.

4.4.1.2 Biased code book

A code book generated with a limited number of training datasets will usually not cover all the possible symbols. For instance, the human seizure signal used in this work has a bit width of 12, and after the DPCM, the bit width should become 13-bit, so in theory there will be 2^{13} possible values for the input signal. However, as is shown in Table 4.2, the code book of the seizure signal only has 1910 entries, which means only 1910 different symbols were found in the training dataset, and if any value that is not included in this code book occurs afterwards, the compression process might fail. To overcome this problem, all the theoretical symbols of the target signal have to be included into the training dataset to ensure the code book covers all the symbols later in a real signal. The PDF of the signal

Table 4.6: Comparison of the partially covered and fully covered code books

Data type	**CR of using the partially covered code book**	**CR of using the fully covered code book**
EEG from the healthy people	1.82	1.71
Seizure-free EEG from the epilepsy patients	2.2	2.04
Seizure	1.51	1.37
Mice EEG data@1000Hz	1.47	1.36
Mice EEG data@200Hz	1.45	1.34
EMG data	2.59	1.9

symbols is altered as a result, and the compression ratio will inevitably deteriorate as the probability distribution is stretched, and the average length of code for representing one symbol is longer than before.

Table 4.6 provides a comparison of the code book generated with and without adding extra symbols in the training datasets. It is quite obvious that the compression ratio decreases when using a code book that covers all the possible signal values. The compression performance of the Huffman coding deteriorates more significantly if the original code book has fewer entries, since adding all the missing symbols brings a bigger change to the PDF of the original signal.

A solution to mitigate this problem is to duplicate the original training datasets several times before filling them with the extra symbols. This method reduces the overall percentage of the extra symbols, and strengths the PDF of the original training dataset in the new training dataset.

The CRs of using enhanced code books are slightly improved, as shown in Figures 4.5 to 4.10 and Table 4.7.

Table 4.7: Comparison of the code books with and without duplicating the original training datasets

Data type	**CR of using the fully covered code book without duplicating**	**CR of using the fully covered code book with the original training datasets duplicated 8 times**
EEG from the healthy people	1.71	1.77
Seizure-free EEG from the epilepsy patients	2.04	2.11
Seizure	1.37	1.39
Mice EEG data@1000Hz	1.36	1.46
Mice EEG data@200Hz	1.34	1.43
EMG data	1.9	2.44

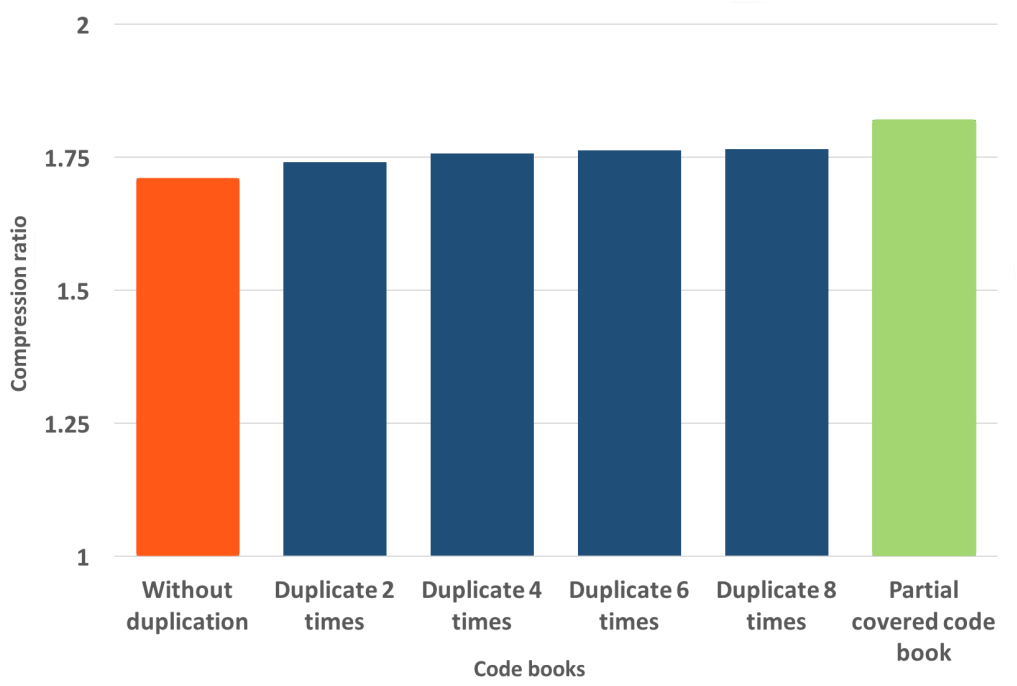

Figure 4.5: Performance improvement of code book generated with the healthy people's EEG signal

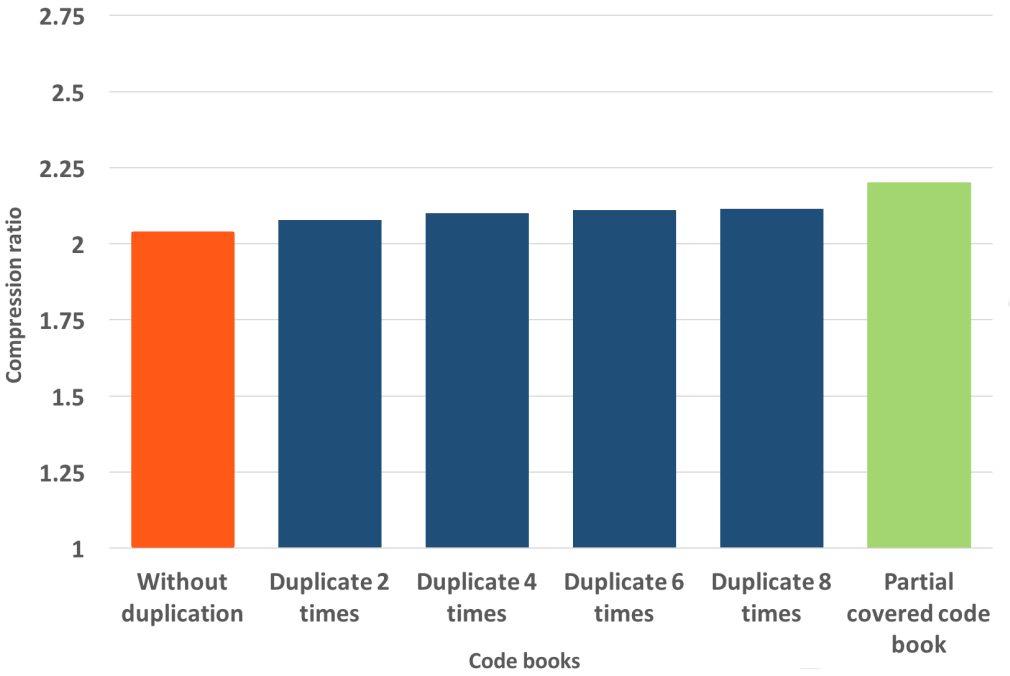

Figure 4.6: Performance improvement of code book generated with the seizure-free EEG signal

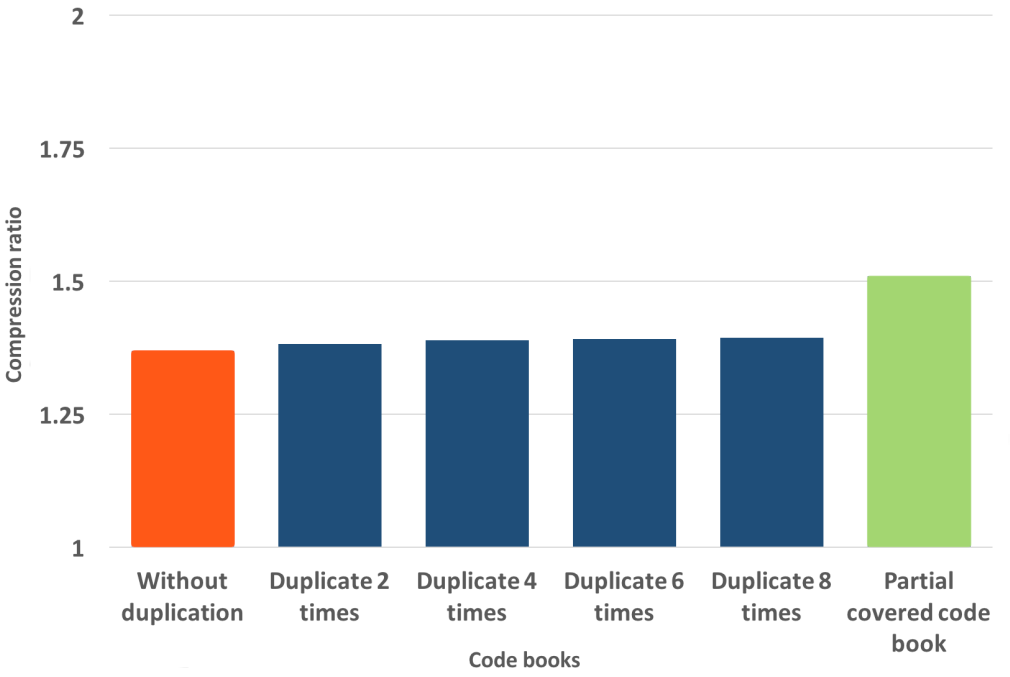

Figure 4.7: Performance improvement of code book generated with the seizure EEG signal

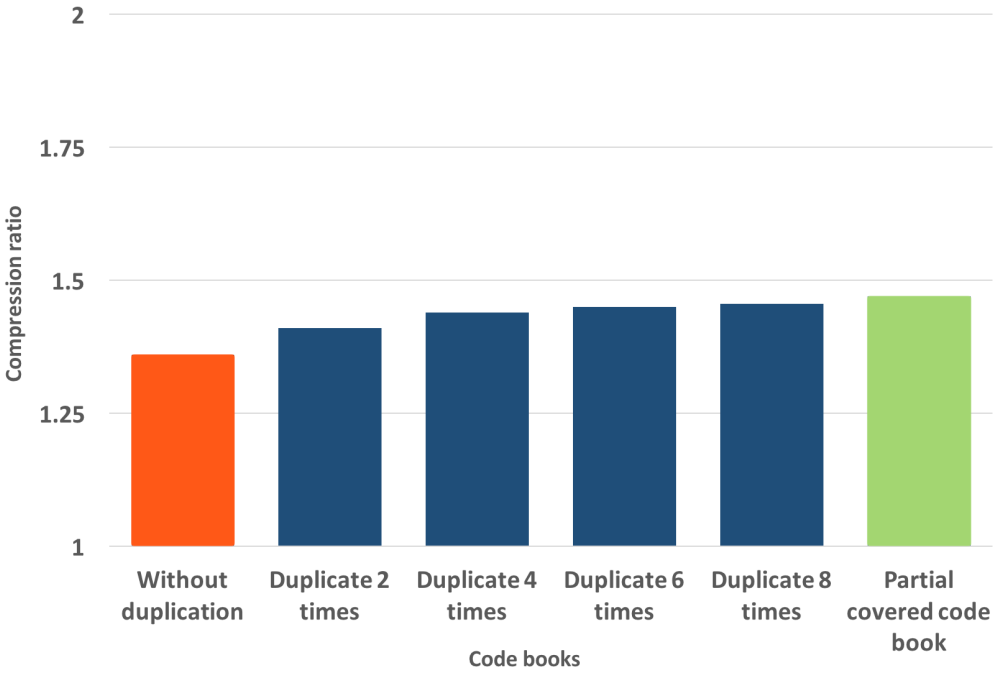

Figure 4.8: Performance improvement of code book generated with the mice EEG signal sampled at 1000 Hz

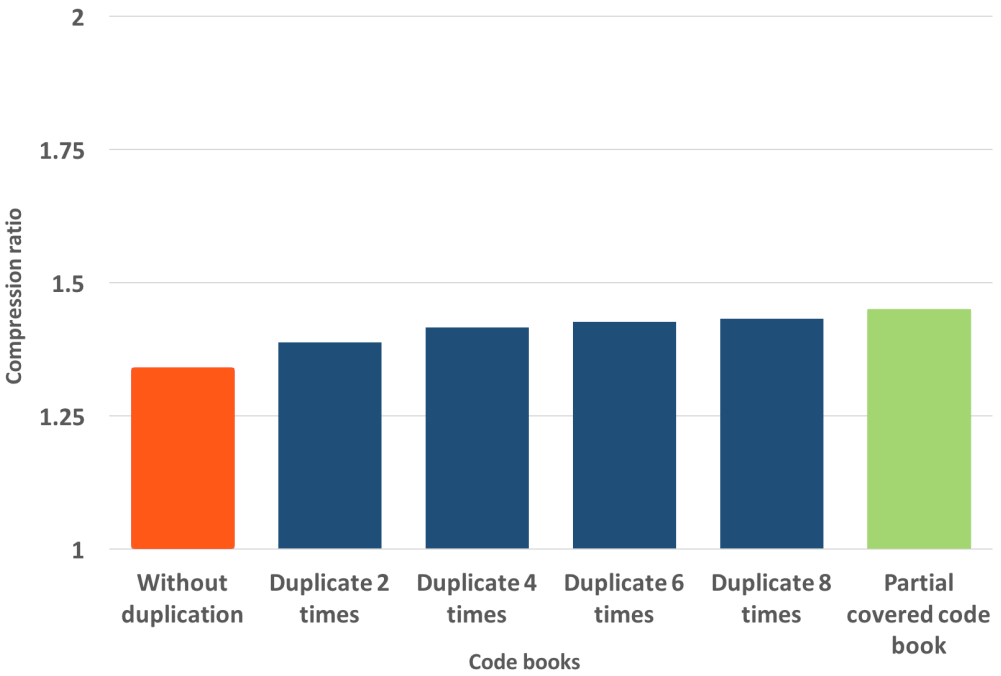

Figure 4.9: Performance improvement of code book generated with the mice EEG signal sampled at 200 Hz

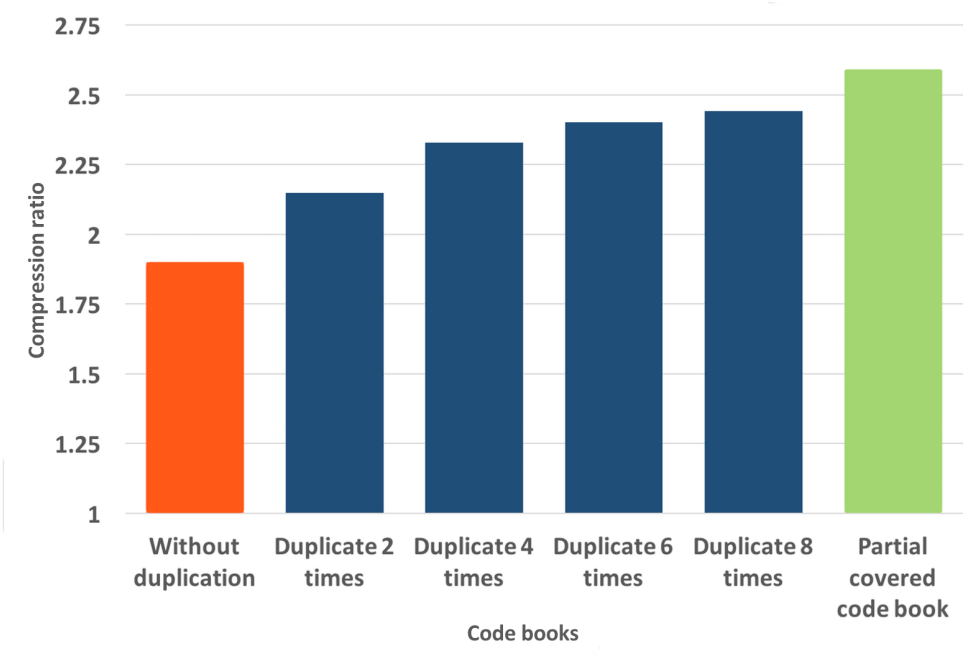

Figure 4.10: Performance improvement of code book generated with the EMG signal

4.4.1.3 Specified code book

As referred to earlier, there are three different code books trained with human EEG data based on the signals of three different neurological conditions including the signal from the healthy people and the seizure and the seizure-free signals from the epilepsy patients, and each one of the code books is trained by a specific dataset to get the best compression result from the given signal. In reality, the wearable wireless device that has this Huffman compressor implemented could be used by anyone regardless of their neurological conditions. The compression ratio can be much worse if compressing the signal with a code book that is unable to fully reflect the PDF of the input signal.

Table 4.8 shows the compression ratio achieved when using different code books to compress the given data. All the code books listed cover all the possible input symbols like the ones shown in Table 4.6, but they were generated without duplicating the original training dataset.

For instance, when compressing the EEG of healthy people with the code book trained by the seizure signal, the performance of this 2-stage technique is weakened.

There is clearly a set of trade-offs when choosing the code book for implementation as it will be used to encode mixed signals like the signal from an epilepsy patient, which

contains both seizure-free and seizure signals. A code book trained by the EEG data that contains both a seizure-free and seizure signals will inevitably give a lower CR for both types of signal. A typical seizure monitoring was described in [87] and [88]. One subject with epilepsy disease was recorded with 82 second seizure signals during a 70-minute monitoring process, and therefore all the seizure episodes takes around 2% of the whole monitoring time. If using the code books SF-S, S-H and SF-H as given in Table 4.8, for estimation, the overall compression ratios can be estimated by

$$CR_{ms} = (CR_{sf} * t_{sf} + CR_s * t_s)/t_{all} \tag{4.1}$$

, where CR_{ms}, CR_{sf} and CR_s are compression ratios of mixed signal, seizure-free signal and seizure signal respectively, t_{all} is the overall observation time, and t_{sf} and t_s are the total time with and without the observing of the seizure. Therefore, CRs achieved from these code books are 1.99, 1.86 and 2.05 respectively, and the code book SF-H yields better CR than the other two code books. However, when seizure happens more frequently, the SF-S or S-H might outperform the others.

This set of trade-offs will not be further discussed in this thesis, but it could be an interesting area to look into in the future.

4.4.2 Log2 sub-band encoding

As one of the time-domain based compression techniques, the Log2 sub-band is very sensitive to the fluctuation of the signal, as mentioned earlier in this chapter. According to the simulation results, it gives a higher compression ratio when compressing more stable signals.

Unlike the Huffman coding and many other compression techniques, the Log2 sub-band might increase the size of the original data in the worst case where no same band can be found between adjacent data samples due to its adding 'header' scheme.

Unfortunately, there is no good solution to these limitations, but as long as it serves the purpose given earlier in this thesis, these drawbacks are acceptable.

4.5 Summary

The data compression performances of both techniques are given in this chapter. The performance of the Log2 sub-band technique that this research proposed is clearly superior

Table 4.8: Compression performance with different code books

H: code book trained with the healthy people's EEG signals

SF: code book trained with seizure-free signals from seizure-free EEG from the epilepsy patients

S: code book trained with seizure signals

SF-S: code book trained with seizure-free and seizure signals

S-H: code book trained with seizure and healthy signals

SF-H: code book trained with seizure-free and healthy signals

SF-S-H: code book trained with all three types of signals

		Code book chosen for 2-stage technique							
		H	**SF**	**S**	**SF-S**	**S-H**	**SF-H**	**SF-S-H**	**Log2 sub-band**
Test data	EEG from the healthy people	1.71	1.62	1.62	1.68	1.71	1.74	1.72	1.94
	Seizure-free EEG from the epilepsy patients	1.94	2.12	1.71	2.01	1.89	2.08	2	2.58
	Seizure	1.19	1.1	1.37	1.35	1.37	1.16	1.34	1.66

to the existing DPCM+Huffman coding despite some limitations it has. Whether or not the compression results would help to reduce the system power consumption will be discussed later in this thesis.

Chapter 5

Signal-classification-based data reduction with the Log2 sub-band encoding

As mentioned in chapter 4, the Log2 sub-band compression process provides the distribution of the number of bands needed to represent a symbol, which carries some characteristics of the original signals. This chapter will illustrate a method of using this distribution information to filter out non-seizure events from the EEG signals. The impact of this data filtering process on the data reduction will also be noted. Some of the work described in this chapter was published in [86].

5.1 Seizure detection

During the compression process, the information on band distribution can be gathered by the Log2 sub-band encoder. As for the EEG signals used in this research, the distributions of healthy human signals and epilepsy patients′ seizure free and seizure signals are presented in Figures 5.1 to 5.3. A comparison of these three distributions can be found in Figure 5.4. '0 band' on the X-axis in these figures means the current symbol is identical to the previous one, therefore no band and only a header is needed to represent the symbol.

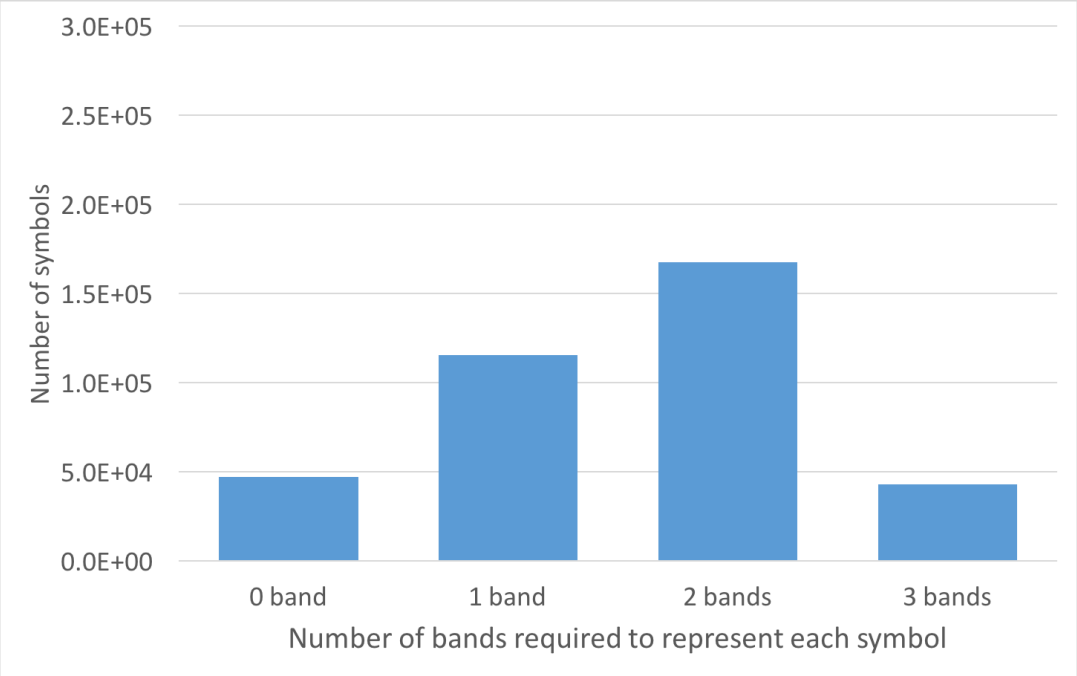

Figure 5.1: Band distribution of the healthy EEG signal

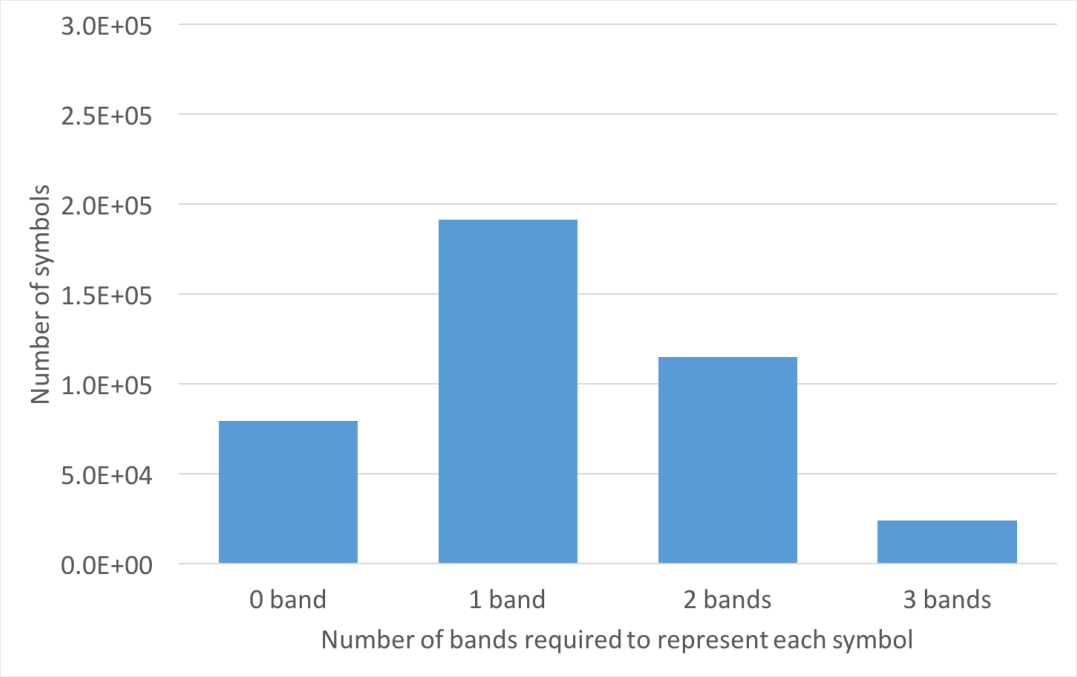

Figure 5.2: Band distribution of the epilepsy patients′ seizure-free EEG signal

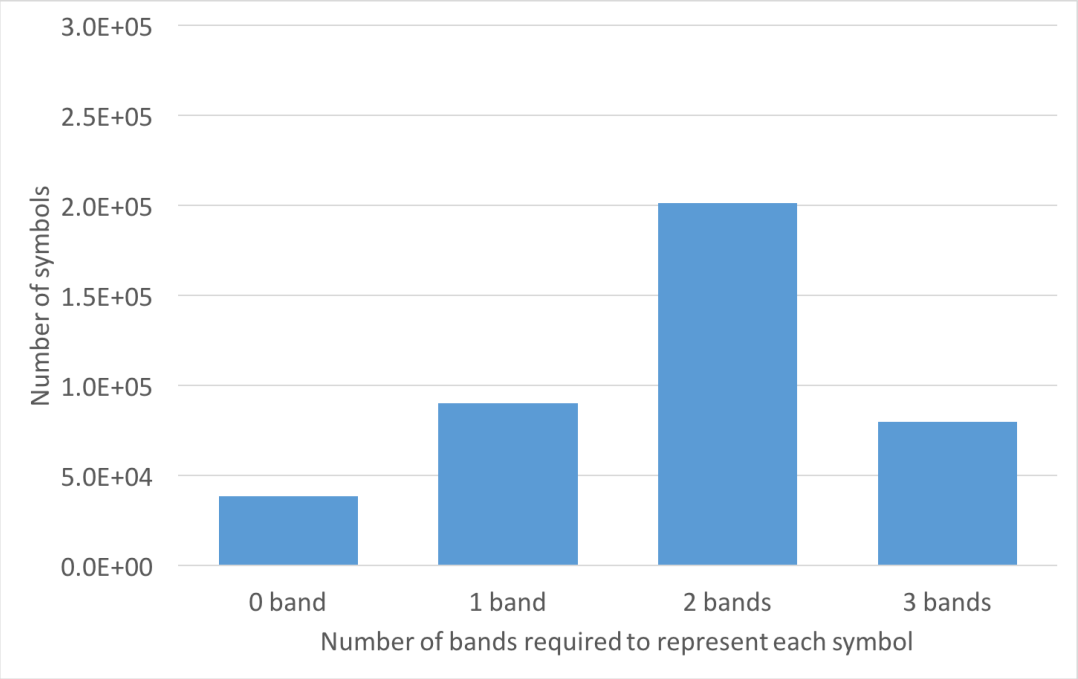

Figure 5.3: Band distribution of seizure EEG signal

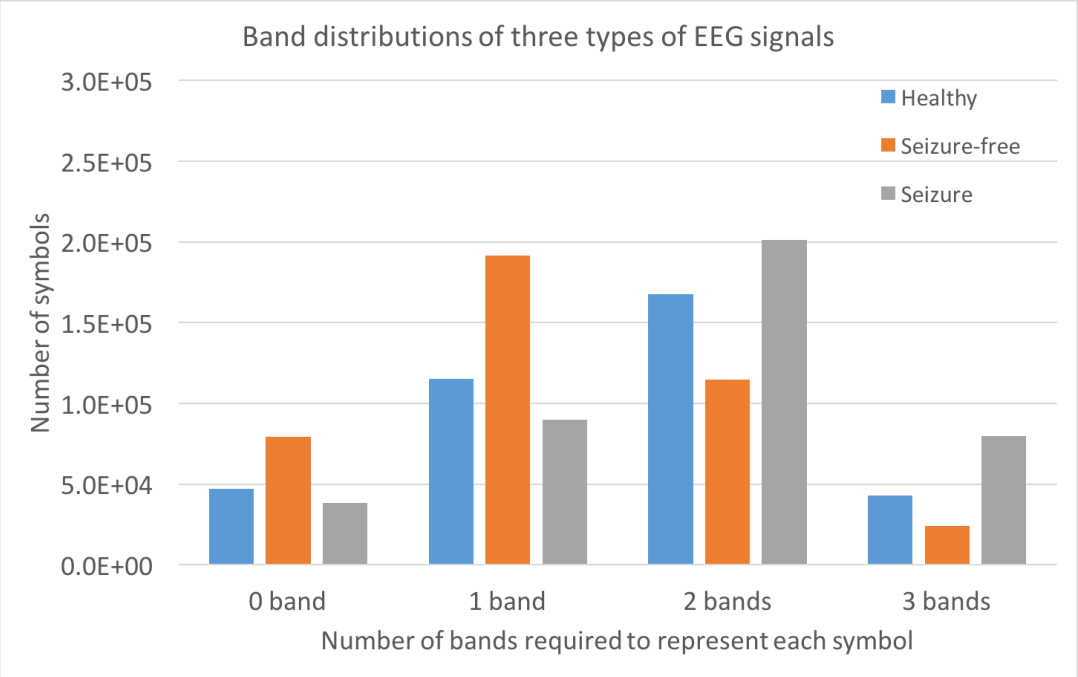

Figure 5.4: Band distribution of 3 kinds of EEG signals

A pattern can be easily found in that more symbols from the rapid changing signals like seizure signals will be transmitted with 2 or 3 bands whilst symbols from the healthy people′s EEG signals and the epilepsy patients′ seizure-free EEG signals need fewer bands

to transmit in general.

The proportions of symbols that are transmitted with 0 and 1 band of three different EEG signals are given in Table 5.1.

It is quite clear that the symbols of the seizure signals that were represented with 0 and 1 band are only half of those of the seizure-free EEG signal′s symbols, and a method to identify the seizure event in real-time can be attempted based on this significant discrepancy.

Table 5.1: Proportions of symbols that are transmitted with 0 and 1 band

Data type	**Proportion (%)**
Healthy people with eyes closed	48.64
Patients′ epileptogenic zone	66.11
Seizure	32.38

A sliding window scheme was designed, as shown in Figure 5.5. This scheme keeps analysing the band distribution of the symbols within the sliding window. A threshold of the proportion of the signal symbols that are represented with 0 and 1 band is set to identify the start of a seizure event based on the assumption that the seizure signals have less proportion of symbols that are transmitted with 0 and 1 band. If the proportion of symbols within a sliding window is found to be lower than the threshold, then the position of this sliding window will be seen as the start of a *possible* seizure episode. The primary aim of this method is to find a 'possible' seizure for further analysis and not to verify it clinically.The result of this selective data filtering technique will be given later.

5.2 Result analysis

To evaluate the performance of this sliding window scheme, a simulated EEG signal that contains both seizure-free and seizure data is used. As the original EEG signals from [72] are pre-categorised into different groups including seizure-free and seizure signals, so 20 segments of data were randomly selected from the seizure group and then inserted into a period of seizure-free data to simulate a real epilepsy patient′s EEG signal.

The performance is measured in terms of sensitivity (SEN), false positive rate (FPR) and speed of detection (SoD), which are the percentage of successfully detected seizure, the percentage of false positive detection, and the average time cost to identify the seizure.

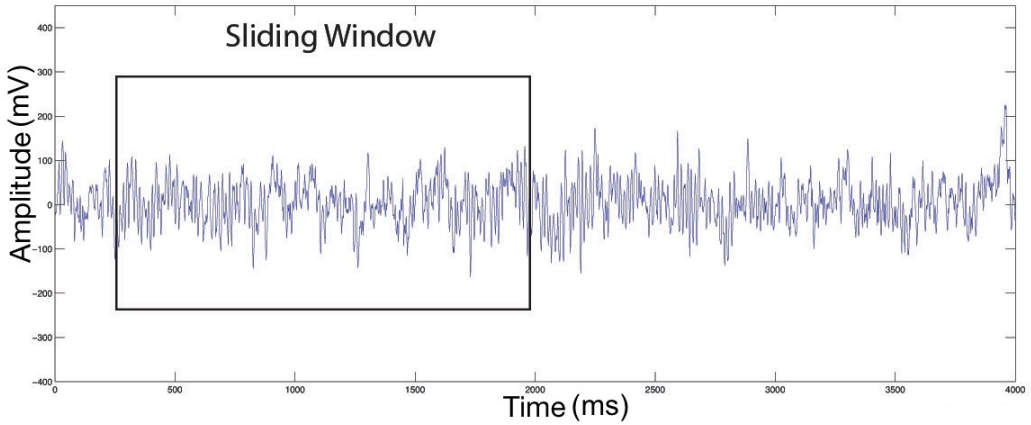

Figure 5.5: Sliding window scheme

The window size is set to 100 data symbols, which is about 580 ms data. The window size was chosen rather arbitrarily in this case, but some principles need to be followed. The epileptic seizure does not last very long, and as it is mentioned in [89], the average duration of seizure of 123 patients is 71 seconds, therefore the window size has to be small to achieve a early detection. In this case, a window size of 100 is quite reasonable.

The detection process is repeated with different threshold settings from 30% to 45%, and as mentioned earlier, if the proportion of the symbols that were represented with 0 and 1 band is smaller than the threshold, the window′s location will possibly be the start of a seizure event. The result is shown in Table 5.2.

Table 5.2: Seizure detection result

Threshold (%)	**30**	**35**	**40**	**45**
SEN	95%	100%	100%	100%
FPR	0	9.1%	31.0%	37.5%
SoD(sec)	2.9	2.5	1.4	0.6

The result indicates that if the threshold is set to a lower value, part of the seizure signal might be missed, and a longer delay in identifying the seizure event will be expected, but the false positive rate will decrease significantly. On the other hand, with a higher threshold value, some parts of the no-seizure signal will be identified as the seizure, but a higher speed of detection can be achieved.

This seizure extraction method is quite simple without adding too much computational burden to a biomedical signal recorder, and it brings a set of trade-offs between SEN, FPR

and SoD that we need to consider. For studies that require higher precision in the data, a higher threshold value is recommended, but it will also inevitably sacrifice the data reduction performance as more irrelevant data will be extracted as well.

For EEG signal recording, there usually is more than one channel of EEG data being recorded. More channels of data may offer a much more accurate detection result because more channels provide more detection results, and a judging system can be built upon these results to give a more accurate result.

The band distribution data is generated as a by-product of the compression process, so it is almost 'free' in terms of hardware cost. There are some more accurate techniques, but considering the hardware cost they may add to the device, it is better to implement them locally on a station.

5.3 Data reduction

Epileptic seizures are caused by sudden burst of uncontrolled electrical activity occurring in a group of neurons in the cerebral cortex, and therefore the EEG signal is a practical way of studying the seizure as it reflects the brain's electrical activities [90–94]. However, a seizure often occurs quite randomly, and most signals recorded and transmitted for epileptic seizure research are actually seizure-free. The sliding window scheme offers a crude method of extracting and sending the signal events that interest researchers only, and it helps to reduce the power consumed by the transceiver. The original data can be stored on the device as a backup if a further examination is required.

One of the subjects mentioned in [87] and [88] was observed having two seizure episodes with a total 163-second duration in in an EEG recording process lasting seven hours. If a high threshold value is chosen for the sliding window to better preserve the seizure data, the technique will give an FPR of 37.5% as is shown in Table 5.2. Therefore, 260-second of data will be extracted, which gives approximately a 99% data saving. This estimation is based on the assumption that the EEG signals recorded in [87] have the same characteristics as the signals from [72]. In practice, the FPR might be larger or smaller depending on the source signal, but this technique clearly has a positive impact on data reduction.

Besides the discontinuous data transmission method mentioned above, adaptive sampling may also be introduced based on the seizure detection result. If the 'possible' seizure is detected, the system uses a higher sampling rate, and a much lower sampling rate can

be used for the rest of the signal recording process. Therefore, the data can be reduced significantly.

5.4 Summary

In this chapter, a sliding window scheme is introduced to be used along with the Log2 sub-band encoding to extract the seizure signal from the epilepsy patient′s EEG signal. The result proves that this data filtering method clearly helps to detect the seizure with just a small extra hardware cost. If only the seizure signals are needed by the researcher, a significant reduction of the data as well as the transceiver′s power can be expected if a suitable threshold of the sliding window is selected.

As this technique is simple and efficient, a more complex method can be applied off-line to further increase the accuracy of the seizure detection.

The possibilities of using this sliding window scheme to detect and extract other event-related biomedical signals are worth further investigation. In particular, when used in systems with many channels the ability to detect seizures with good sensitivity and FPR may improve significantly, so it seems a worthwhile area for further investigation.

An important point here is that the Log2 sub-band encoding provides seizure information as a useful by-product of compression with minimal effort, neither the Huffman or the arithmetic coding can easily offer such a capability.

Chapter 6

Hardware simulation and power analysis of the data compressor

In this chapter, the hardware design of both the Huffman coding and the Log2 sub-band encoding mentioned earlier will be given. The power consumption of these two compressors will be estimated with the given design, and the power analysis will be provided based on the estimated result. The biomedical data used in this chapter are human EEG data only. These data have similar features, but they were recorded from subjects with different physiological conditions, they can best represent an application case in reality.

6.1 Hardware design

6.1.1 DPCM+Huffman coding

The hardware model of the DPCM+Huffman coding is straightforward, as shown in Figure 6.1. In theory, the entire 2-stage compressor should contain a DPCM coder that differentiates the input biomedical signal, and a quantizer to convert each signal symbol from the DPCM into an address that can be looked up in the memory, as shown in Figure 6.2. A data shifter that shifts out the right number of bits is also required. The power analysis of the shifter has been done for this research. Based on the preliminary result, the shifter power is so small compared to the RAM power that it can be neglected in the power analysis of the 2-stage technique. For the same reason, the power analysis excludes the power of the DPCM and quantizer, and only the power of the RAM was evaluated as it consumes much more power than other parts.

The Huffman code book is generated based on the pre-acquired and differentiated EEG

signals, and the code book contains the entries of all the possible signal symbols, and their corresponding code words. The code book is stored in the memory, as shown in Figure 6.2, and during the compression process, every original signal symbol will be assigned with a code word based on the code book stored in the memory. As the code words in the code book have different lengths, they are padded with 1 to get the same length before storing in the memory. The code words in the memory will be shifted out by the shifter to get a correct length.

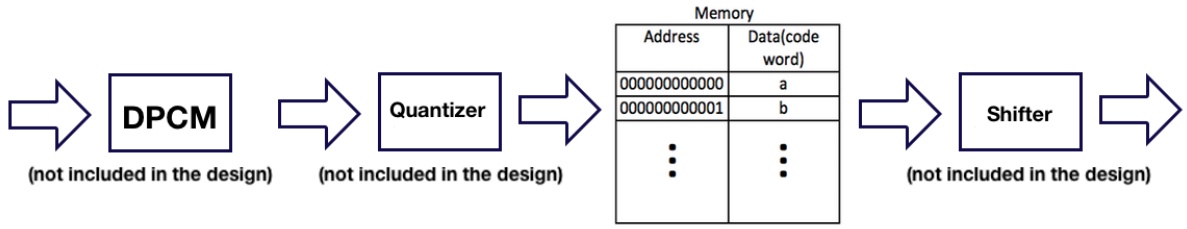

Figure 6.1: 2-stage technique implementation

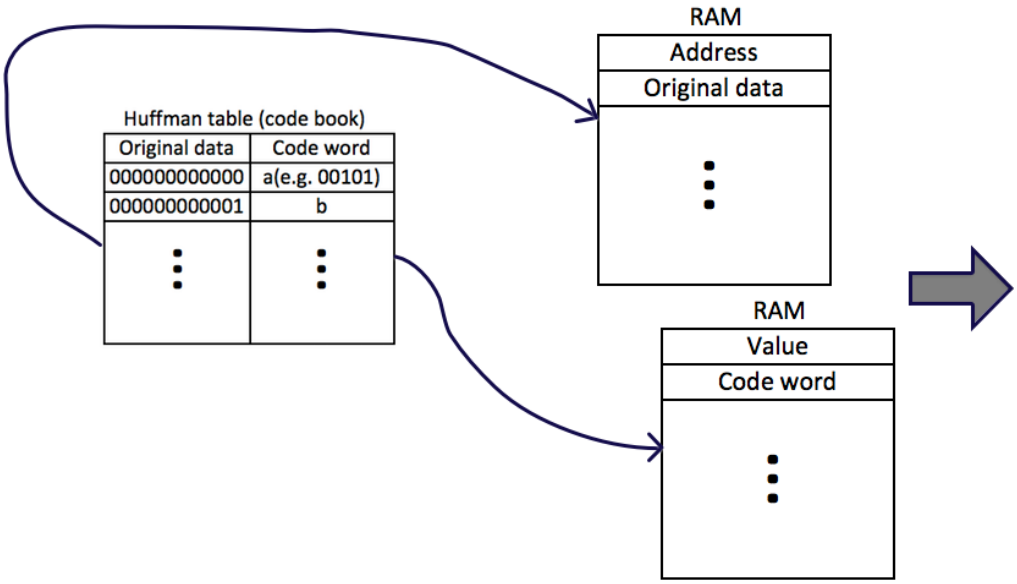

Figure 6.2: Huffman code book implementation

For the hardware implementation of the 2-stage compression technique, the core is the RAM that the code book is stored in. To measure the power of the RAM, a tool called CACTI [95–97] is used to estimate the memory access time, cycle time, area and static and dynamic power of the static random-access memory(SRAM) that stores the Huffman code book. The static power of a circuit is the power consumed when the circuit is in quiescent mode, and the dynamic power of a circuit is the power consumed by circuit switching. With the memory size, number of read/write ports and size of technology node set, CACTI can provide an accurate power estimation of the required memory.

6.1.2 Log2 sub-band

The hardware design of the Log2 sub-band is given in Figure 6.3, and in this case, the original signal is 12-bit, and each compressed signal symbol will be 14, 10, 6 or 2 bits long after adding a header. A control unit will decide how many bits to shift out. Two different designs were used for power analysis. One is designed to process one data symbol per clock cycle, which in this case means feeding in 12 bits in parallel every clock cycle. The other design takes the data in serial, which takes 12 cycles to compress one data symbol. These two designs can be described as parallel-in parallel-out (PIPO) and serial-in serial-out (SISO).

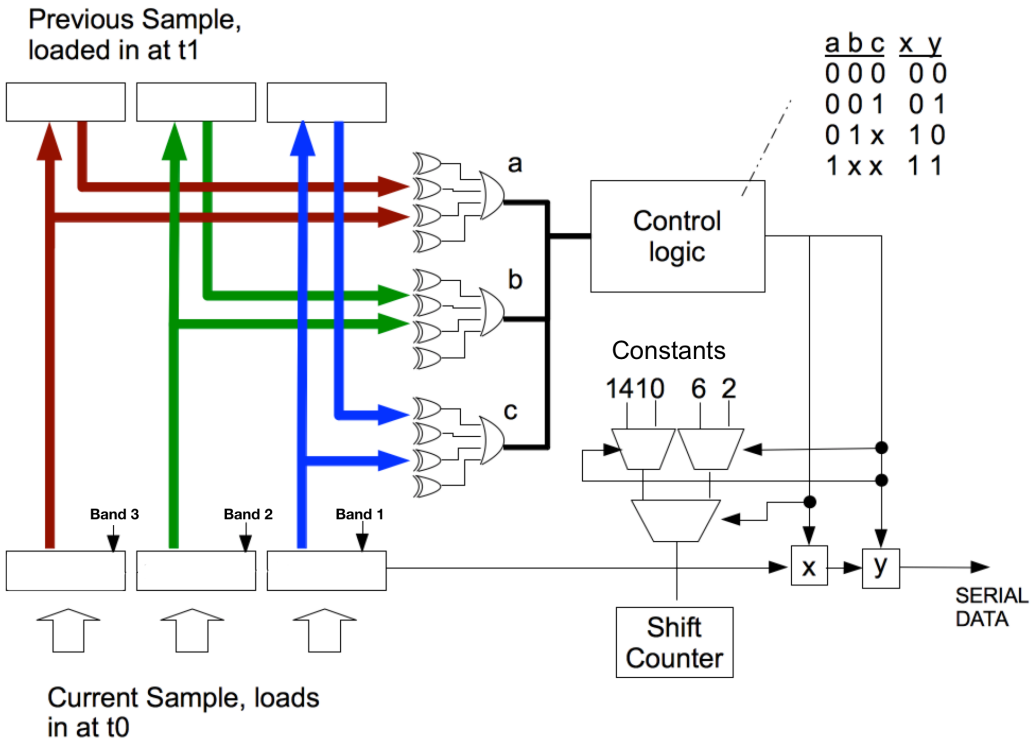

Figure 6.3: Log2 sub-band design

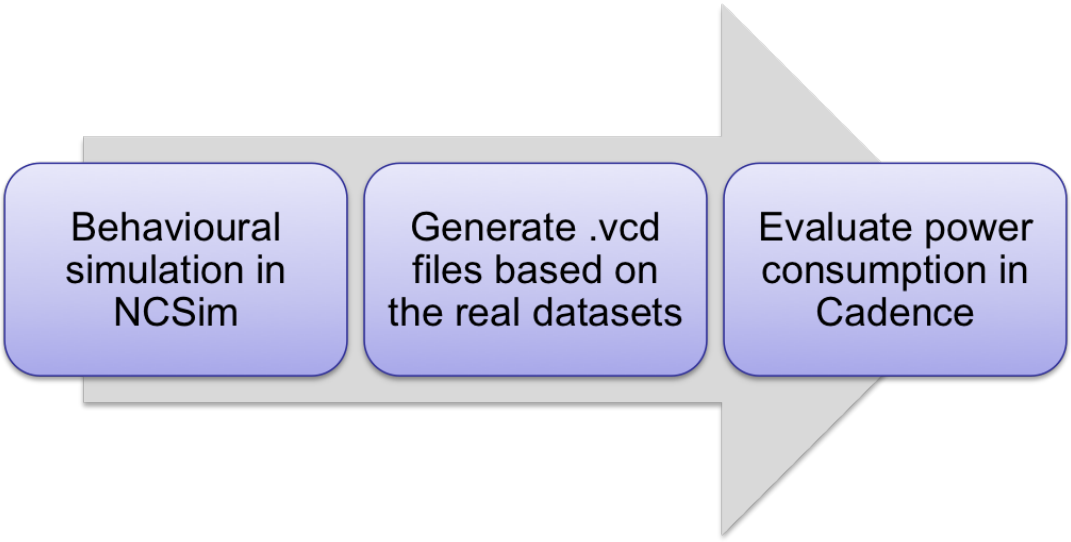

Figure 6.4: Power analysis process of the data compressors

The process of power analysis is given in Figure 6.4. The human EEG signals [72] used earlier in the compression performance analysis were fed into the Log2 sub-band encoding unit, and value change dump(vcd) files were generated to record the value changes of the signals in the data compressor. The vcd files are for estimating the power of the data

compressors while compressing the given biomedical signals. The hardware design and analysis were conducted within the Cadence Virtuoso environment using 65 nm UMC CMOS process node. The Verilog and VHDL code for Log2 sub-band and its testbench are given in Appendix C2 and C3.

6.2 Power analysis

In this part, the power consumption of the 2-stage compression technique and the Log2 sub-band will be provided based on the hardware models given earlier in this chapter.

6.2.1 DPCM+Huffman coding

6.2.1.1 Power result

The memory is initiated by CACTI based on the size of the EEG code book implemented. As discussed earlier in chapter 4, the code books generated based only on the training datasets are smaller, and they have fewer entries than the numbers of possible signal symbols. Although storing a smaller code book certainly consumes less memory power, the compression process may fail if an unknown symbol occurs. To prevent this from happening, the code books that cover all the possible symbols are used here to keep the system stable, regardless of the deterioration of the power and the compression performance.

The bit width of every input symbol is 13 after the DPCM, and the longest code word in these code books is 23-bit, so the book has 2^{13} entries in total. Due to the limitation of CACTI, the number of read-out bits can only be set to a multiple of 8, so it was set to 24 to take into account the length of the longest code word in the Huffman code books. As a result, the RAM size can be calculated with

$$size = (Nr.of BitsReadOut) * (Nr.of Entries)/8 \tag{6.1}$$

, in this case 24576 bytes.

The numbers of read and write ports are both set to 1, and the technology node is set to 65 nm which will be sufficient for this application.

CACTI provides other setting options that are about the power consumption of the RAM.

(a) RAM cell/transistor type in data array

It can be set to high performance (HP) which means using fast transistors with short gate lengths, low static power (LS) which means using slower transistors with longer gate length, and low operating power (LO) which means using slowest transistors therefore consuming the lowest dynamic power.

(b) Peripheral and global circuitry transistor type in data array

Similar to (a), it can be set to high performance (HP)/low static power (LS)/low operating power (LO).

(c) Interconnect projection type

It can be set to aggressive (A) which assumes aggressive use of low-k dielectric causing insignificant resistance degradation, and conservative (C) which assumes limited use of low-k dielectric causing significant resistance degradation.

(d) Type of wire outside mat

It can be set to semi-global (S) which indicates a pitch of 4F(F = Feature size), and global (G) which indicates a pitch of 8F.

The power results with different configurations are given in Table 6.1 and Table 6.2, and a comparison of different settings is given in Figure 6.5 and Figure 6.6. The dynamic power results were acquired based on a sampling rate of 1 kHz, which is enough for processing most of the biomedical signals.

The labels of the x axis are the settings of (a), (b), (c) and (d) mentioned above. For instance, if (a) and (b) are set to high performance, (c) is set to aggressive and (d) is set to semi-global, then the corresponding label would be HP-HP-A-S.

Table 6.1: Static power of SRAM

Power type/Settings	LO-LO-C-G	LS-LO-C-G	HP-LO-C-G	LO-LS-C-G	LS-LS-C-G	HP-LS-C-G
Static power(μW)	300.11	69.58	12795.29	231.62	0.83	12729.49
Power type/Settings	LO-HP-C-G	LS-HP-C-G	HP-HP-C-G	LO-LO-A-G	LS-LO-A-G	HP-LO-A-G
Static power(μW)	5700.05	5272.39	16673.95	301.43	70.9	12793.29
Power type/Settings	LO-LS-A-G	LS-LS-A-G	HP-LS-A-G	LO-HP-A-G	LS-HP-A-G	HP-HP-A-G
Static power(μW)	231.63	0.83	12729.49	5328.37	5472.39	16778.96
Power type/Settings	LO-LO-C-S	LS-LO-C-S	HP-LO-C-S	LO-LS-C-S	LS-LS-C-S	HP-LS-C-S
Static power(μW)	298.31	67.79	12795.27	231.62	0.83	12729.49
Power type/Settings	LO-HP-C-S	LS-HP-C-S	HP-HP-C-S	LO-LO-A-S	LS-LO-A-S	HP-LO-A-S
Static power(μW)	5698.83	5271.17	16673.06	301.42	70.89	12793.29
Power type/Settings	LO-LS-A-S	LS-LS-A-S	HP-LS-A-S	LO-HP-A-S	LS-HP-A-S	HP-HP-A-S
Static power(μW)	231.63	0.83	12729.49	5327.9	5471.92	16680.2

Table 6.2: Dynamic power of SRAM at 1 kHz

Power type/Settings	LO-LO-C-G	LS-LO-C-G	HP-LO-C-G	LO-LS-C-G	LS-LS-C-G	HP-LS-C-G
Dynamic power(nW)	15.4	13.43	18.17	25.19	34.29	26.48
Power type/Settings	LO-HP-C-G	LS-HP-C-G	HP-HP-C-G	LO-LO-A-G	LS-LO-A-G	HP-LO-A-G
Dynamic power(nW)	26.27	28.56	26.37	15.92	18.58	16.48
Power type/Settings	LO-LS-A-G	LS-LS-A-G	HP-LS-A-G	LO-HP-A-G	LS-HP-A-G	HP-HP-A-G
Dynamic power(nW)	24.56	27.18	26.34	21.6	28.78	26.71
Power type/Settings	LO-LO-C-S	LS-LO-C-S	HP-LO-C-S	LO-LS-C-S	LS-LS-C-S	HP-LS-C-S
Dynamic power(nW)	15.16	17.65	17.35	24.29	26.79	26.8
Power type/Settings	LO-HP-C-S	LS-HP-C-S	HP-HP-C-S	LO-LO-A-S	LS-LO-A-S	HP-LO-A-S
Dynamic power(nW)	24.74	26.95	24.66	15.59	18.2	16.1
Power type/Settings	LO-LS-A-S	LS-LS-A-S	HP-LS-A-S	LO-HP-A-S	LS-HP-A-S	HP-HP-A-S
Dynamic power(nW)	25.39	26.48	25.88	20.96	27.96	26.57

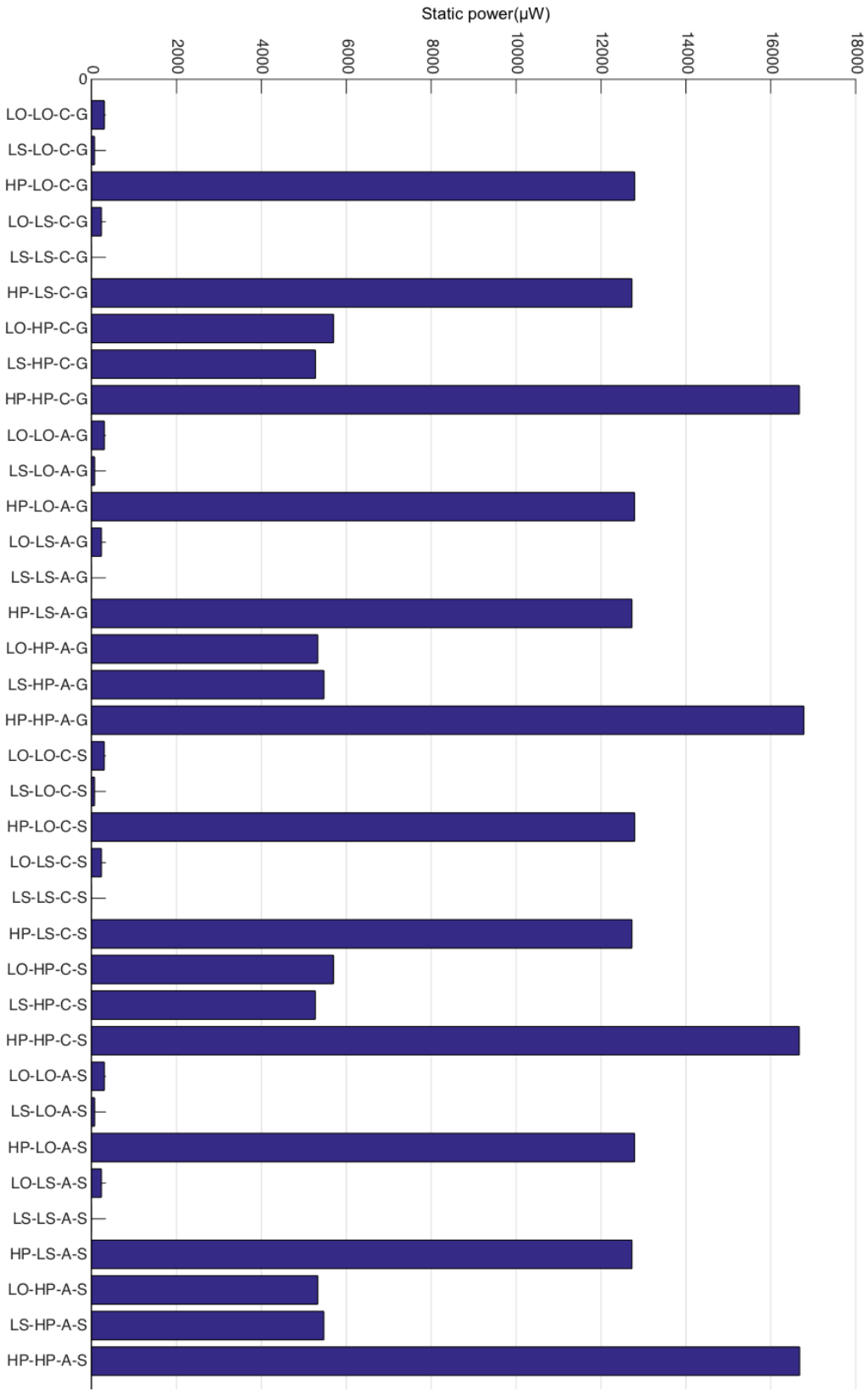

Figure 6.5: Static power of the SRAM with the code book stored in

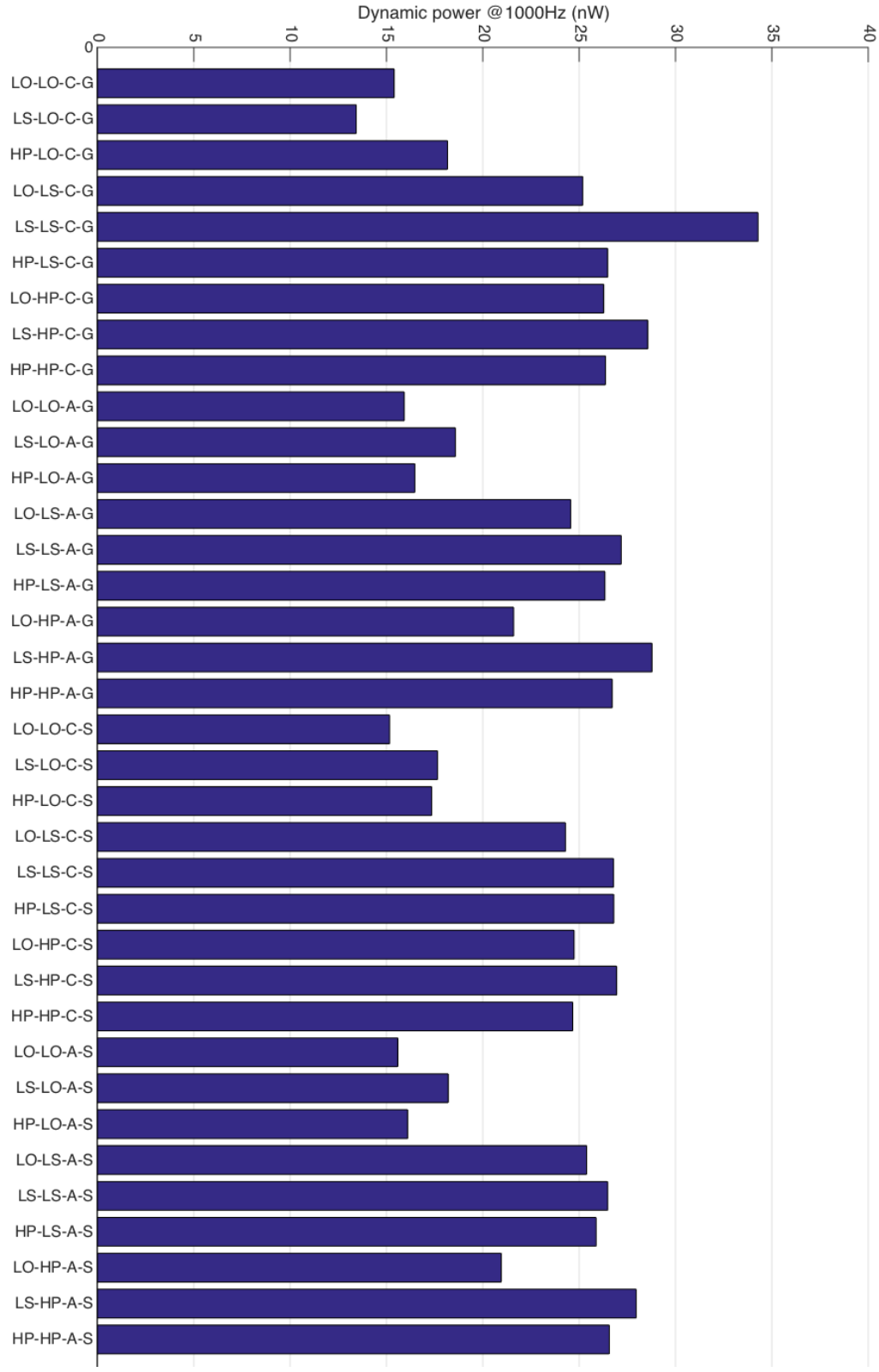

Figure 6.6: Dynamic power of the SRAM with the code book stored in

6.2.1.2 Result analysis

From the result shown in Table 6.1, when the RAM cell/transistor type and the peripheral and the global circuitry transistor type in data array are set to low static power, the RAM has the lowest static power, which is around 0.83 μW.

According to Table 6.2, the RAM consumes the lowest dynamic power when the RAM cell/transistor type and the peripheral and the global circuitry transistor type in data array are set to low operating power, which is around 15 nW. However, the static power of the RAM is around 300 μW with these settings, which makes the overall power larger than the case when both types are set to low static power.

After comparing all the possible settings, it is quite clear that the RAM consumes the lowest overall power when the four types are set to low static, low static, aggressive, and semi-global respectively, where the static power and dynamic power are 0.83 μW and 26.48 nW respectively.

In this case, the 2-stage technique will consume 856.48 nW to compress one channel of biomedical data. Since the RAM usually operates at a very high frequency, only one RAM unit is needed even for compressing more channels of signals. Compressing more channels of data will only increase the dynamic power of the RAM. In this case, the static power of the RAM will remain the same when compressing four channel of EEG data, but the dynamic power will increase from 26.48 to 105.92 which is four times as much as the original dynamic power.

6.2.2 Log2 sub-band

6.2.2.1 Power result

With the process shown in Figure 6.4, the power consumptions of the Log2 sub-band circuits with PIPO and SISO when compressing one channel of EEG signal are given in Table 6.3 and Table 6.4.

Only the human EEG signals [72] were fed into the Log2 sub-band compressors, and they can help to reveal the relationship between the CR and the power performance of the Log2 sub-band. To better demonstrate the relationship between the achieved CRs and the power consumed to achieve those CRs, each type of the human EEG data was further divided into ten subsets, and therefore each type contains ten segments of 23.6 seconds of EEG signal. A power regression analysis between CRs and power consumptions will be

given later in this chapter.

Table 6.3 and Table 6.4 give both the static and the dynamic power of the Log2 sub-band compressor, and they are the average power consumption for compressing the ten data subsets.

Both the SISO and PIPO circuits are designed to run at the clock rate of 50 MHz. The 50 MHz was chosen here to reduce the time cost of the simulation process since lower frequency will require longer time to simulate the design. Furthermore, as static power is frequency independent, and dynamic power scales linearly with frequency, knowing the dynamic power at 50 MHz allows dynamic power to be calculated at any chosen frequency even when compressing the EMG signal, which requires a much higher frequency than EEG signal does.

The dynamic power of the SISO design was estimated by

$$P^{'}_{dynamic} = P_{dynamic} * f_{sample} * W_{in}/f_{clock} \tag{6.2}$$

, where f_{clock} and f_{sample} are the clock rate and the required sampling rate respectively, and W is the bit width of input signal, and $P_{dynamic}$ and $P^{'}_{dynamic}$ are the dynamic power at the original clock rate f_{clock} and at a given sampling rate f_{sample} respectively. In this case, it takes 12 clock cycles to compress a 12-bit data sample, and f_{clock} and f_{sample} are 50 MHz and 1 kHz respectively.

The dynamic power of PIPO was estimated with

$$P^{'}_{dynamic} = P_{dynamic} * f_{sample}/f_{clock} \tag{6.3}$$

, which is similar to (6.2), but it only takes one clock cycle for PIPO to compress a data sample. Both static and dynamic power were estimated with Cadence design tools mentioned earlier.

6.2.2.2 Results analysis

From the results shown in Table 6.3 and Table 6.4, the negative correlation between the dynamic power and the CR is quite clear, which indicates that the dynamic power decreases when CR increases whilst the static power remains stable. The reason behind this is that usually fewer circuit state changes are needed when a signal can be largely compressed by the Log2 sub-band, and if the CR was high, most data symbols would be represented by fewer bands after the compression process.

Table 6.3: Power consumption of Log2 sub-band with PIPO

Data Type	**CR**	**Static Power(nW)**	**Dynamic Power(nW)**	**Total**
EEG from the healthy people	1.94	Avg. 230.97	Avg. 3.14	234.11
Seizure-free EEG from the epilepsy patients	2.58	Avg. 230.81	Avg. 2.72	233.53
Seizure	1.66	Avg. 232.54	Avg. 3.14	235.69

Table 6.4: Power Consumption of Log2 sub-band with SISO

Data Type	**CR**	**Static Power(nW)**	**Dynamic Power(nW)**	**Total**
EEG from the healthy people	1.94	Avg. 147.23	Avg. 12.92	160.15
Seizure-free EEG from the epilepsy patients	2.58	Avg. 147.29	Avg. 12.68	159.97
Seizure	1.66	Avg. 147.29	Avg. 13.37	160.66

The power regression analysis of the CR and the dynamic power of the parallel and serial input circuits are given in Figure 6.7 and Figure 6.8. According to the trend lines, as the CR is infinitely approaching 1, which indicates that no data size reduction achieved, the power consumed by the Log2 sub-band will become higher. It will gradually become inefficient to apply such a compression process as the compressor would consume more power than it could save. On the other hand, the power consumption will also be infinitely approaching a minimum value as the CR goes higher.

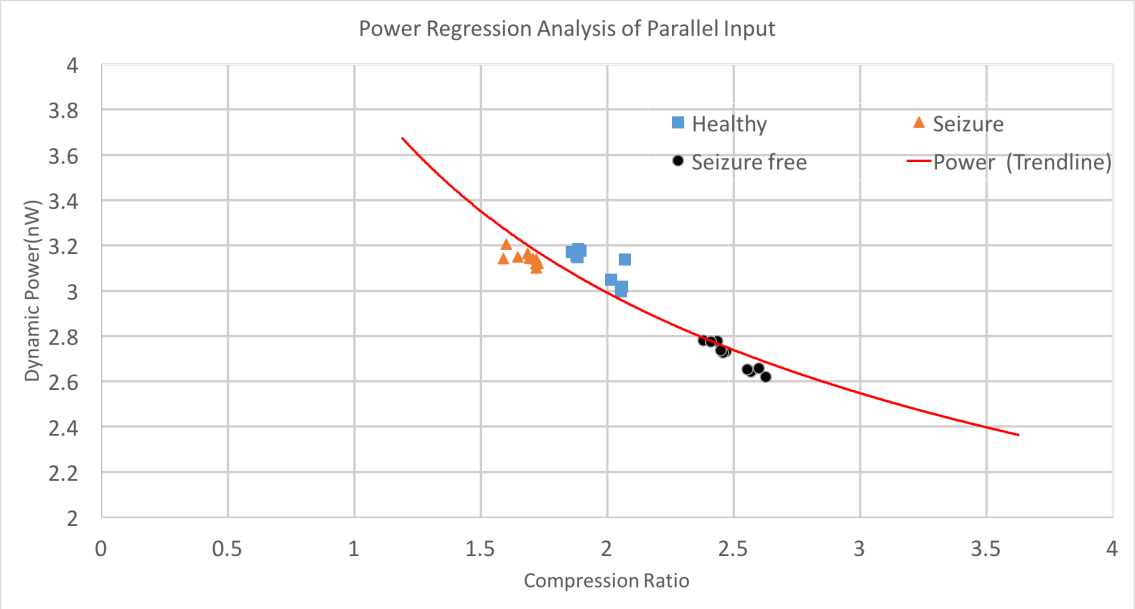

Figure 6.7: Power regression analysis of Log2 sub-band with parallel input

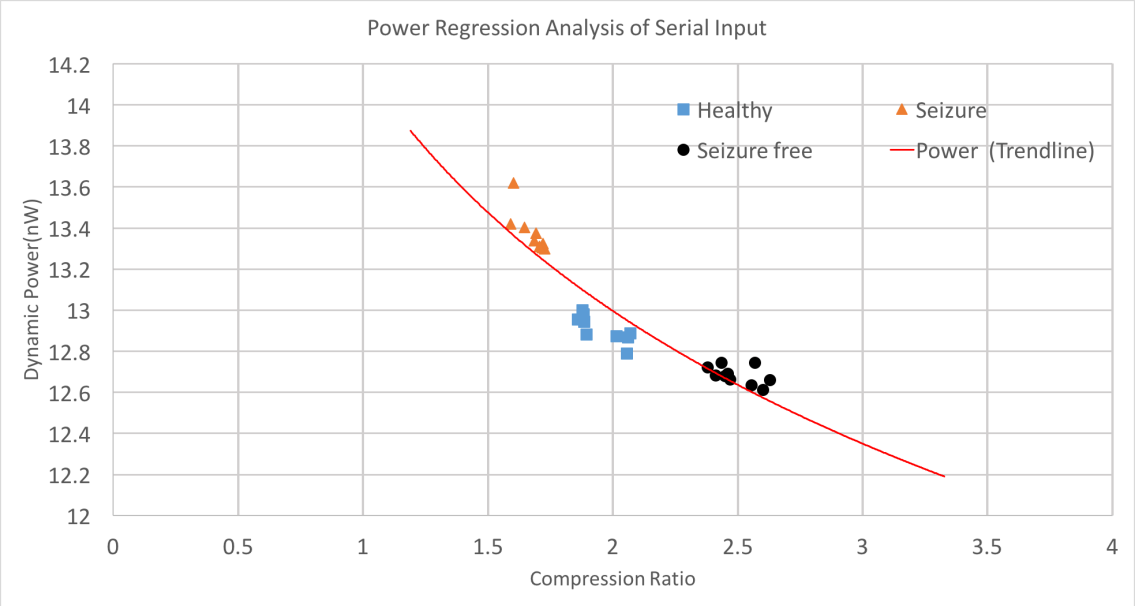

Figure 6.8: Power regression analysis of Log2 sub-band with serial input

In general, the circuit with parallel input consumes more static power but less dynamic power than the one with serial input, and the former consumes more overall power than the latter.

A Log2 sub-band compressor is designed to compress only one channel of biomedical data, and if there are more than one channel of data is required, more duplicated compressors are needed. As the number of channels increases, the static power and the dynamic

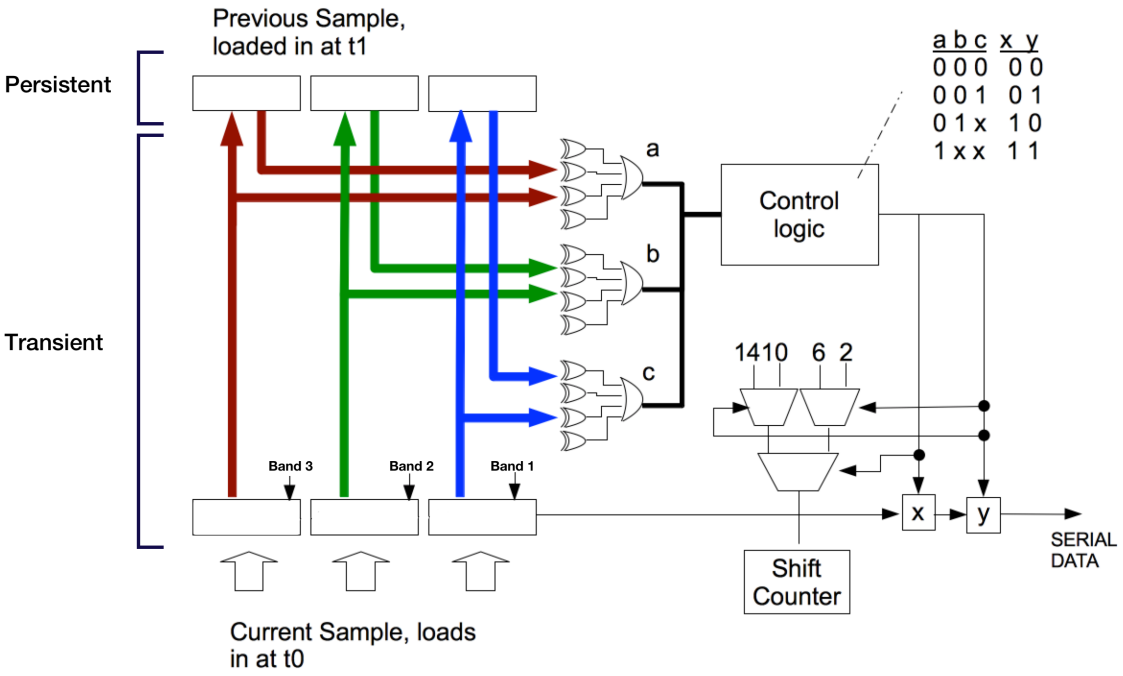

Figure 6.9: Persistent and transient parts of the Log2 sub-band design

power of the compression unit accumulate, and it will become less efficient to use data compression.

6.2.2.3 Power optimisation

The hardware design shown in Figure 6.3 can be further optimised given the low sampling rate normally required for processing the biomedical signals. For the EEG signals mentioned earlier in this chapter, the sampling rate needed is 1 kHz. As shown in Figure 6.9, only the registers that store the previous sample need to stay powered all the time, which is the persistence part of this design, and the rest of this design is transient in that it can be switched off at certain intervals to save static power. The power-gating needs to be implemented to control the transient circuit, therefore a small amount of extra circuitry would be required , and that is neglected in the estimated power saving.

For instance, when compressing one channel of the EEG signal from the healthy people with a SISO Log2 sub-band compressor, the static power consumed by the compressor at 50 MHz is 147.23 nW if all components are powered at all times, but the logic comparing unit of the compressor is only needed for 1000 clock cycles per second to achieve a sampling rate of 1 kHz. The load-in and shift-out units will need a few more cycles every second. Comparing to the clock rate, the total operating time of the transient part of

the compressor is very short, so the static power consumed by this part can be neglected if it is switched off when not needed. Therefore, the static power of the Log2 sub-band compressor approximately equals to the power consumed by the persistent part shown in Figure 6.9 after optimisation, and it is around one quarter of the original static power, which is 37 nW in this case.

6.2.3 Comparison of two techniques

As briefly mentioned earlier, the RAM that stores the Huffman code book can be used to compress multiple channels of input data simultaneously after increasing the clock rate. One Log2 sub-band unit is designed to compress one channel of data each time, therefore more units are required when compressing multiple channels. In the case of recording and processing one channel of data, the Log2 sub-band consumes less power than the 2-stage technique does. As the number of channels increases , the RAM used by 2-stage technique only needs an extra dynamic power to increase its clock rate to process the extra channels, but the Log2 sub-band encoding needs to consume extra static and dynamic power because each Log2 sub-band unit can only process one channel of input, and it is impossible for one unit to process multiple channels of input due to the hardware design of the Log2 sub-band.

Table 6.5 shows the total power that the 2-stage technique and the SISO Log2 sub-band need to compress the given number of channels of healthy human EEG signals, and Table 6.6 gives the power consumptions of the DPCM+Huffman coding and the optimised Log2 sub-band.

Figure 6.10 illustrates the average power consumed by the 2-stage technique and the Log2 sub-band with and without the power optimisation for compressing each channel of EEG data.

From Table 6.5 and 6.6 and Figure 6.10, it is quite clear that if the device has more than six channels of input data, the 2-stage compression technique will be more efficient than the Log2 sub-band. However, the optimised Log2 sub-band compressor is superior to the DPCM+Huffman coding in terms of power when the number of input channels is less than thirty-five.

The data shifter was not included in the previous analysis, nevertheless the power estimation of the shifter has been done for this research. Based on the design given in Appendix D, the dynamic power consumed by the shifter is around 32 nW, and the static

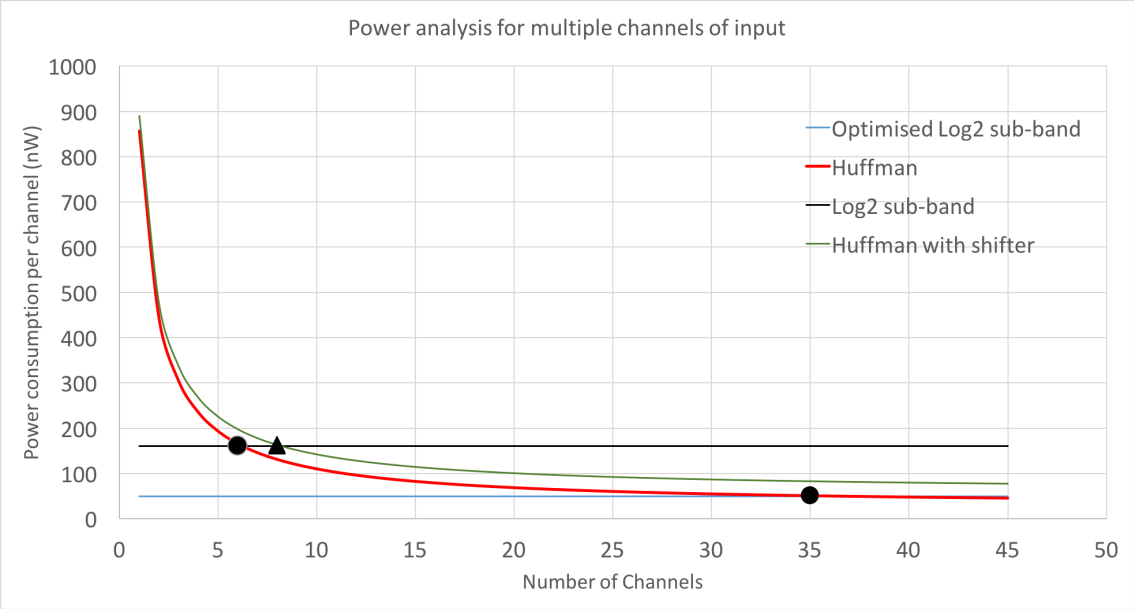

Figure 6.10: Power consumed by compressing one channel

power is around 1 nW. If take the shifter power into account, the Log2 sub-band encoding will not consume more power than the 2-stage technique if the number of input channels is less than eight. For the optimised Log2 sub-band design, it will always consume less power than the 2-stage technique.

It is still worth noting that even though the Log2 sub-band loses its advantage over the 2-stage compression technique in terms of power efficiency at some point, it still has other benefits such as its high adaptivity compared to the DPCM+Huffman coding and many other compression techniques. Because every channel of data needs an independent Log2 sub-band unit, every unit can be tailored to give the best compression result for each channel. For instance, a {4,4,4} setting can be changed to other patterns to get a higher CR. Moreover, the hardware design of the 2-stage technique used in this chapter does not include the DPCM part, so this technique will definitely consume more power in reality.

6.3 Summary

This chapter evaluated the power performance of the simplified DPCM+Huffman coding (only has the RAM) and the Log2 sub-band encoding techniques. In compressing one channel of healthy human EEG signal the Log2 sub-band consumes less power than the DPCM+Huffman coding does, but as the number of channels increases, the 2-stage technique would become more efficient since one RAM can be used by multiple channels.

Table 6.5: Power comparison between the Huffman and the Log2 sub-band

Nr of channels	**Static power (Huffman)**	**Static power (Log2 sub-band)**	**Dynamic power (Huffman)**	**Dynamic power (Log2 sub-band)**	**Total power (Huffman)**	**Total power (Log2 sub-band)**
1	830	147.23	26.48	12.92	856.48	160.15
2	830	294.46	52.96	25.84	882.96	320.3
3	830	441.69	79.44	38.76	909.44	480.45
4	830	588.92	105.92	51.68	935.92	640.6
5	830	736.15	132.4	64.6	962.4	800.75
6	830	883.38	158.88	77.52	988.88	960.9
7	830	1030.61	185.36	90.44	1015.36	1121.05
8	830	1177.84	211.84	103.36	1041.84	1281.2
9	830	1325.07	238.32	116.28	1068.32	1441.35
10	830	1472.3	264.8	129.2	1094.8	1601.5
11	830	1619.53	291.28	142.12	1121.28	1761.65
12	830	1766.76	317.76	155.04	1147.76	1921.8

Table 6.6: Power comparison between the Huffman and the optimised Log2 sub-band

Nr of channels	**Static power (Huffman)**	**Static power (Log2 sub-band)**	**Dynamic power (Huffman)**	**Dynamic power (Log2 sub-band)**	**Total power (Huffman)**	**Total power (Log2 sub-band)**
1	830	37	26.48	12.92	856.48	49.92
5	830	185	132.4	64.6	962.4	249.6
10	830	370	264.8	129.2	1094.8	499.2
15	830	555	397.2	193.8	1227.2	748.8
20	830	740	529.6	258.4	1359.6	998.4
25	830	925	662	323	1492	1248
30	830	1110	794.4	387.6	1624.4	1497.6
35	830	1295	926.8	452.2	1756.8	1747.2
40	830	1480	1059.2	516.8	1889.2	1996.8
45	830	1665	1191.6	581.4	2021.6	2246.4

According to the simulation, the Log2 sub-band without any optimisation consumes more power than the simplified 2-stage technique if there are more than six channels of input signal. An optimised Log2 sub-band design, it consumes less power than the 2-stage technique for a total number of channels of less than thirty-five.

If including the shifter in to the power analysis of the 2-stage technique, the original Log2 sub-band design will consume less power than the 2-stage technique if the number of input channels is less than eight. The optimised Log2 sub-band design will always consume less power than the 2-stage technique. If including the DPCM, the advantage of using the Log2 sub-band encoding will be more obvious.

The impact on system power that these two techniques may bring to a wearable and wireless biomedical signal recorder will be introduced in the next chapter.

Chapter 7

System power overview

In this chapter, the system power analysis of a wireless and wearable biomedical signal recorder with and without the DPCM+Huffman coding and the Log2 sub-band encoding implemented is given. Some of the possible components that can be used to build the device are listed at the beginning, and the possible power saving from the transceiver and the memory is estimated based on the chosen components. The system power and the expected battery lifetime of the signal recorder is demonstrated at the end of this chapter. The power analysis of the DPCM+Huffman coding and the Log2 sub-band encoding techniques was evaluated with the human EEG data in the previous chapter. In order to further analyse the system power in this chapter, the same data are used here. Some of the work described in this chapter was published in [85] and [86].

7.1 Components selected for the simplified system

Based on the hardware model given in Figure 2.6, some of the off-the-shelf and experimental ADCs, amplifiers, transceivers and memory units are listed here. For each component, a most suitable candidate is chosen from all the choices. The system power will be estimated based on these chosen components.

7.1.1 Amplifier

Some possible amplifiers for biomedical devices are given in Table 7.1, and amplifier [54] is the best option in terms of power performance, but, as introduced earlier, the biomedical signals such as EEG and EMG have a frequency spectrum ranging from 0.5 to 500Hz, therefore the amplifier [98] is the best option if everything is taken into account.

Table 7.1: Amplifiers

Amplifier	[99]	[100]	[98]	[101]	[102]	[54]
Technology (nm)	180	130	90	90	350	1500
Power Supply (V)	1.8	1.2	1.2	1.2	±1.5	±2.5
Bandwidth (Hz)	10-7.2k	192-7.4k	0.07-20k	0.2-20k	13-8.9k	0.025-30
Area (mm^2)	0.0625	0.053	not mentioned	0.058	0.022	0.22
Power consumption (μW)	7.92	1.92	11	23.5	6	1

7.1.2 ADC

Some ADCs that are specifically designed for converting biomedical signals are given in Table 7.2. ADC [103] consumes the least power compared to other candidates, and it gives a sample rate up to 40 kHz, which is enough for most of the biomedical signals. As a result, [103] is the best option for the simplified hardware model.

Table 7.2: ADCs

ADCs	[103]	[104]	[105]	[106]	[107]	[108]
Technology (nm)	65	110	180	350	180	180
Area (mm^2)	0.076	0.092	0.63	1	0.7	0.43
Supply Voltage (V)	0.6	0.9	1	1.15	1	1.8
Sample rate (S/s)	40k	1000k	100k	1k	100k	2k
Resolution (bit)	12	12	12	12	12	12
Power (μW)	0.097	16.5	25	0.233	3.8	0.455
Energy/Sample (nJ/S)	0.002	0.017	0.25	0.233	0.038	0.228

7.1.3 Transceiver

The transceiver is the most power hungry component in the simplified hardware model of the signal recorder given earlier, and some widely used wireless transmission solutions are given in Table 7.3. These transceivers offer various achievable ranges, starting from 1 m to 20 m. To reach a larger range, more power is needed, as shown in Table 7.3.

Based on the different applications a wearable and wireless device may be used for,

different achievable ranges will be needed. Therefore, transceivers [109], [110] and [56] that allow different mobilities are chosen for the simplified system to simulate various situations a signal recorder may be used for, and [110] is selected because it consumes the least power compared to other transceivers that have the same achievable range.

7.1.4 Memory

As mentioned in chapter 2, the flash memory is likely to be used to store recorded biomedical data on a wearable and wireless device. Table 7.4 gives the power performances of two typical flash memory chips. The performances were measured by power consumed for reading and writing one page of data. The page is the basic unit in a flash memory chip, and a flash memory contains multiple blocks, and blocks are composed of pages.

The memory read/write operation consumes very little power, as shown in Table 7.4. Even though memory writing consumes more power than memory reading, which is 2.32 nJ for writing one bit data [115], it is still less than 1/10 of the power for sending one bit data with the most efficient transceiver, which is 22 nJ [56] for sending 1 bit data.

7.2 Power saving

The power saving is either from the transceiver or the memory, and the following part gives some details of these two components.

7.2.1 Transceiver power

The main purpose of introducing the data compression technique is to reduce the power consumed by the wireless transceiver. By applying one of two techniques listed below, the equivalent energy required for sending one bit data will be reduced. The equivalent energy can be derived from Equation (3.4).

7.2.1.1 DPCM+Huffman coding

A wireless biomedical device, in this case an EEG signal recorder, is usually designed to record the signal collected from a subject with any physiological conditions. For instance, a wireless EEG signal recorder maybe used by both healthy people and epilepsy patients, and a tailored code book may be optimal to a specific signal, but it will not give the best result when compressing a different signal, as has been discussed in chapter 4. The SF-H

Table 7.3: Transceivers

Transceiver	[109]	[111]	[112]	[113]	[114]	[110]	[56]
Type	FSK	ZigBee	Zigbee	FSK	Bluetooth	Bluetooth Low Energy	Bluetooth
Frequency	862-960MHz	2.4GHz	2.4GHz	300-1000MHz	2.4GHz	2.4GHz	2.4GHz
Tx data rate (kbps)	250	250	250	76.8	2178	4000	1000
Supply voltage (V)	2.1	2.1	1.8	2.1	2.8	3.6	1.9
Tx power consumption (mW)	52.5	36.5	25.2	24.8	140	138.6	22
Achievable range (m)	20	10	8	10	10	10	1
Energy/bit (nJ/b)	210	146	101	322	64	34.7	22

Table 7.4: Memory chips

Memory chip	[115]	[116]
Pages/block	64	128
Page size (KB)	2	4
Read energy (μJ/Page)	4.72	1.63
Read energy (nJ/bit)	0.29	0.05
Write energy (μJ/Page)	38.04	29.75
Write energy (nJ/bit)	2.32	0.91

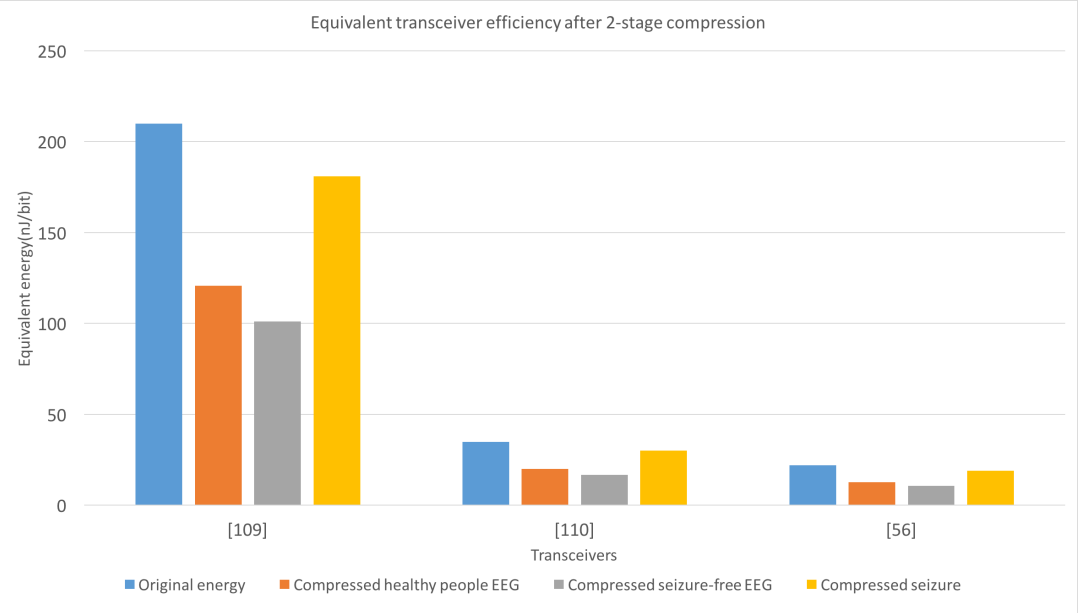

Figure 7.1: Equivalent data transmission energy after 2-stage compression

code book mentioned in chapter 4 will be used here due to its good performance when compressing the EEG signals from both healthy people and epilepsy patients .

The power saving is reflected in the equivalent energy for sending one bit data, as shown in Figure 7.1, and power can be saved from all chosen transceivers. Compressing the seizure-free signal brings the most significant power saving.

7.2.1.2 Log2 sub-band

Similar to the impact that the 2-stage compression technique brings , the Log2 sub-band also reduces the equivalent energy of all transceivers, but more significantly than the former technique. Compressing the seizure-free signal will also bring the biggest power

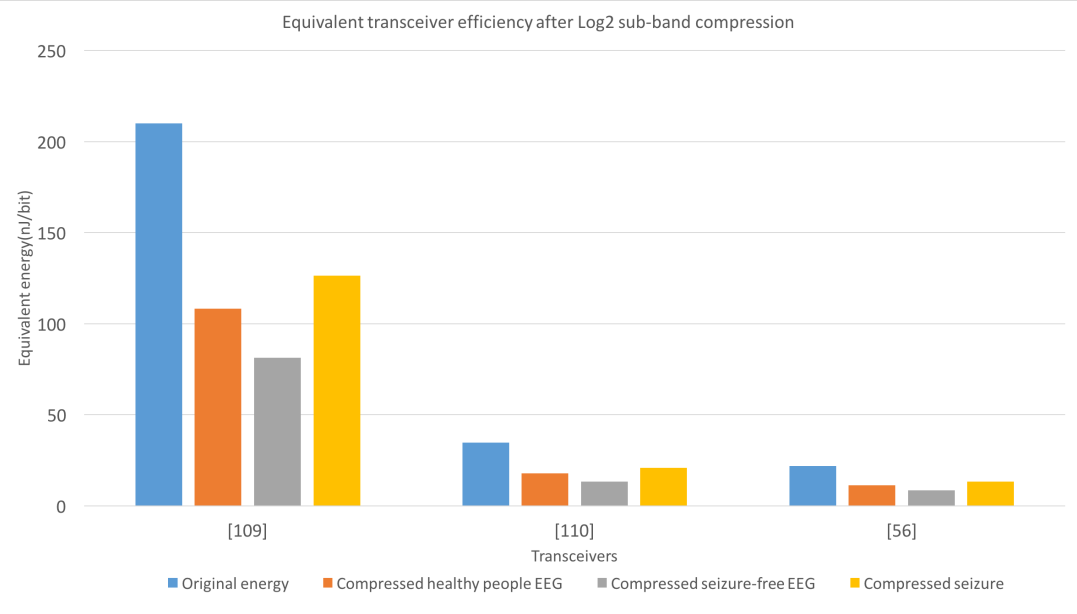

Figure 7.2: Equivalent data transmission energy after Log2 sub-band compression

reduction. The result is shown in Figure 7.2.

It is worth mentioning that if a proper signal-classification-based data reduction technique is applied, as discussed in chapter 5, the data transmitted by the wireless transceiver can be even less, and more power can be saved compared to simply compressing the data. However, this saving is based on the assumption that only certain signal events such as a seizure signal are needed by the user, and therefore this case will not be included here.

7.2.2 Memory power

Compared to the transceiver, the memory only consumes a little power, as shown in Table 7.4. The equivalent energy for reading and writing memory after data compression can be estimated by Equations (3.9) and (3.10). Considering that the original memory power is very low, the data compression technique will not bring a significant power saving from the memory unit, but the possible memory space saving could be more beneficial.

7.2.2.1 DPCM+Huffman coding

The impact of the DPCM+Huffman coding on memory power can be found in Figure 7.3, assuming that the code book used here is the same as the one in the transceiver power analysis. Given the fact that the highest CR achieved is from compressing the seizure-free signal, more power can be saved when writing/reading these data into/from the memory.

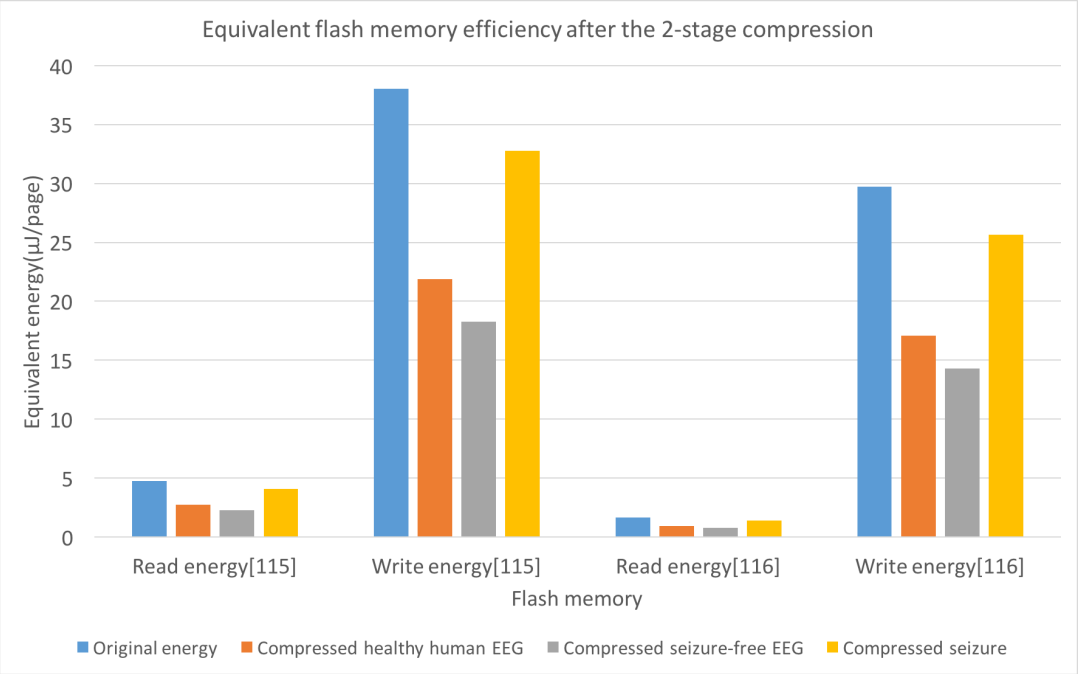

Figure 7.3: Equivalent memory energy after 2-stage compression

7.2.2.2 Log2 sub-band

Since the Log2 sub-band achieves higher CRs compared to the DPCM+Huffman coding technique, it saves more power than the 2-stage technique does. The result is shown in Figure 7.4.

7.3 System power overview

The system power of the simplified hardware model is derived from all the power information acquired so far, including the power of the chosen components as well as the power needed by the two compression techniques given in the previous chapter. Assuming the system only records one channel of EEG data, and the power of a system without the data compression technique is given in Table 7.5.

The flash memory often consumes much less power than the transceiver, so data compression and reduction techniques usually have very limited impact on saving memory power. Therefore, the system power in this part will only be based on the simplified hardware model given in Figure 3.1, in which the memory unit is not included.

As previously mentioned, the EEG data from epilepsy patients can never be pure seizure-free or seizure, and during the signal recording process conducted in [87] [88], the

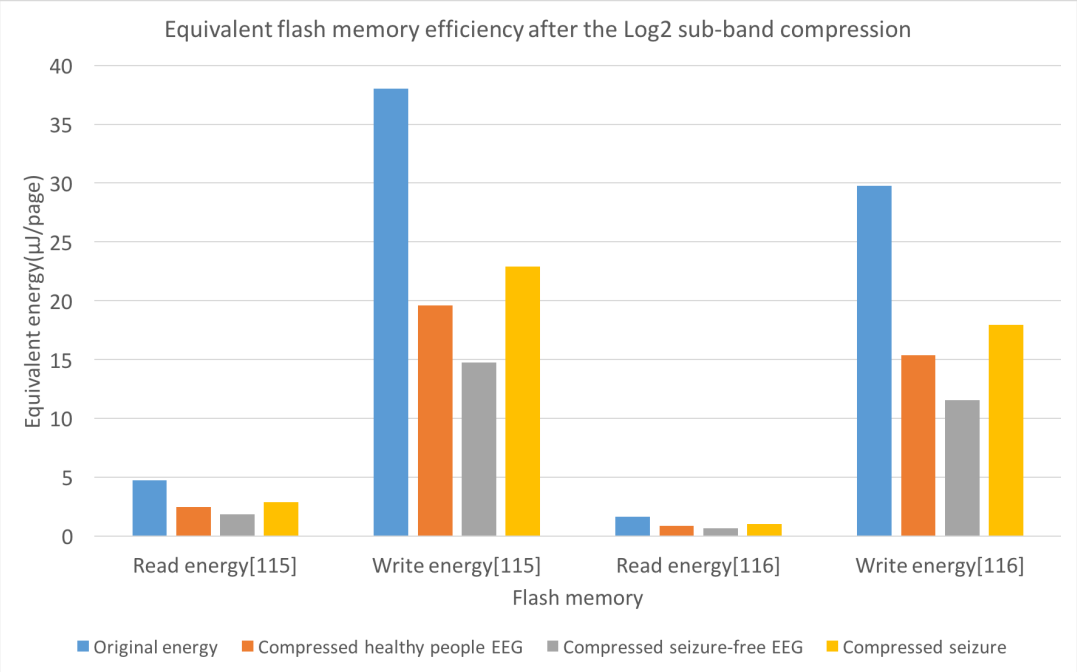

Figure 7.4: Equivalent memory energy after Log2 sub-band compression

seizure only happened for less than 2% of all observation time in most cases. In order to estimate the power saving that the 2-stage and the Log2 sub-band techniques can bring to the system, it is assumed that the EEG data to be compressed here are from a hypothetical patient and consist of 2% of seizure data and 98% of seizure-free data. Therefore, the CR both techniques can achieve from the simulated epileptic EEG data can be calculated by Equation (4.1).

For the 2-stage technique, the code book SF-H proved to be most efficient for the hypothetical epilepsy patient, as outlined in chapter 4, and it gives a CR of 2.05 based on Equation (4.1). For the Log2 sub-band encoding, the CR it can achieve from the simulated epileptic EEG data can be estimated similarly, and is 2.56.

7.3.1 DPCM+Huffman coding

For a signal recorder that has the 2-stage technique implemented, the power consumption of the transceiver can be estimated with Equation (3.3), and the system power can be estimated with Equation (3.1).

Table 7.6 shows all the power information of a system with and without the 2-stage data compression technique when recording one channel of the EEG signal. The compression technique achieved a higher CR from the epileptic data, therefore the system consumes

Table 7.5: System power without data compression overview

Sampling rate	1 kHz		
Number of channels	1		
Date rate	12 kbit/s		
Amplifier power (μW)	11 [98]		
ADC power (nW)	2 [103]		
Transceiver	FSK [109]	Bluetooth Low Energy [110]	Bluetooth [56]
Range (m)	20	10	1
Tx power (mW)	2.52	0.42	0.26
Total (mW)	2.53	0.43	0.27

less power if it is used by a epilepsy patient than it is used by a healthy user.

The Figure 7.5 and 7.6 show the power consumption of the multi-channel system. As the required data rate of the transceiver rises with the number of channels, the system power increases significantly. If the number of channels doubles, based on Equations (3.2) and (3.3), the transceiver power will double accordingly, and more ADCs and amplifiers will be needed. Although for the DPCM+Huffman technique the power of the data compressor will not double, as shown in Table 6.6, the system power growth will still be close to linear if more channels of data are recorded, because the transceiver power dominates the overall system power.

7.3.2 Log2 sub-band

The power information of a system with and without the Log2 sub-band implemented when compressing one channel of EEG data is given in Table 7.7. The system power of the multiple channels system is given in Figure 7.7 and Figure 7.8. For the Log2 sub-band encoding, each channel of input signal requires a Log2 sub-band unit for compression, and the system power growth is linear.

7.3.3 Comparison of two systems

The simulation described in chapter 4 indicates that the Log2 sub-band achieved better CRs from all the EEG data than the DPCM+Huffman coding technique did, therefore more power can be saved from the transceiver on a system with the Log2 sub-band im-

Table 7.6: System power with the implementation of the 2-stage technique

Data	Healthy	Epilepsy	Healthy	Epilepsy	Healthy	Epilepsy
Transceiver	[109]		[110]		[56]	
Tx power without compression(mW)	2.52		0.42		0.26	
Total power without compression(mW)	2.53		0.43		0.27	
CR	1.74	2.05	1.74	2.05	1.74	2.05
Tx power after compression(mW)	1.45	1.23	0.24	0.21	0.15	0.13
Compressor power(nW)	856.48					
Total power after compression(mW)	1.46	1.24	0.25	0.22	0.16	0.14

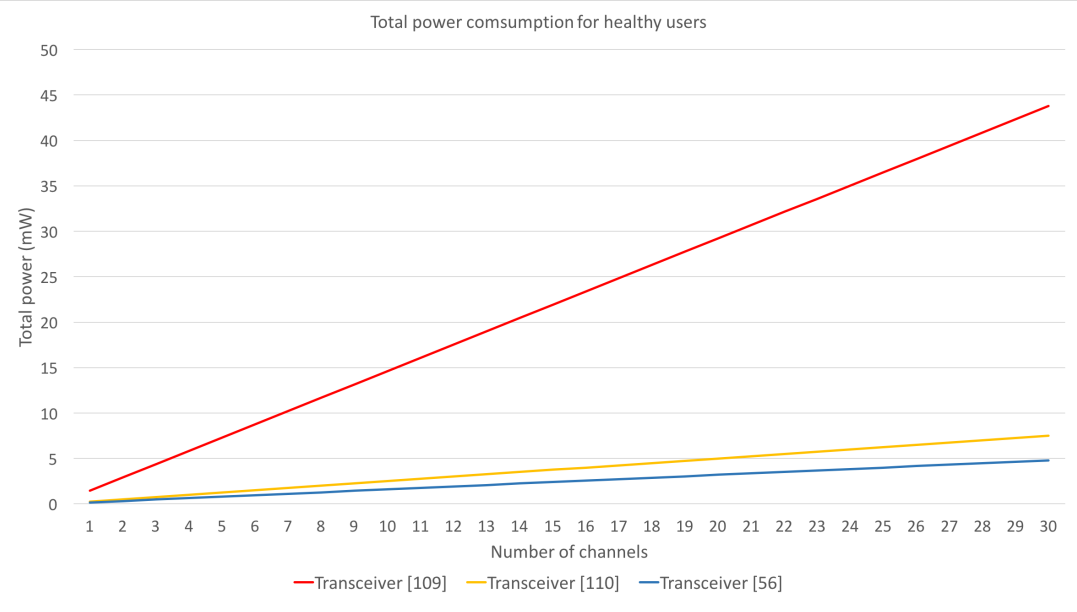

Figure 7.5: Total power consumption for healthy user with 2-stage compressor implemented

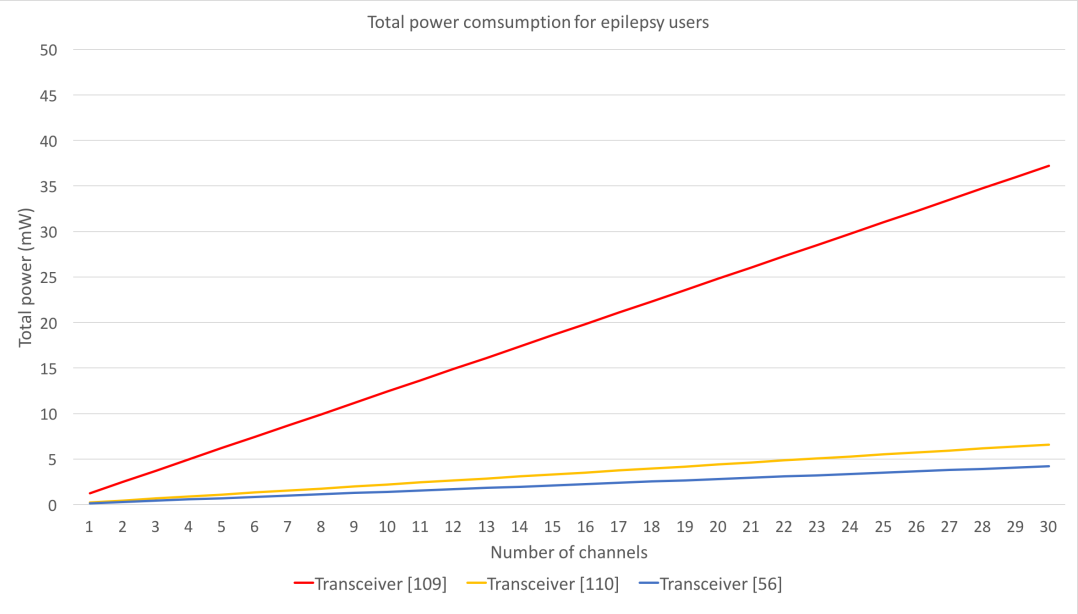

Figure 7.6: Total power consumption for epilepsy user with 2-stage compressor implemented

Table 7.7: System power with the implementation of the Log2 sub-band

Data	Healthy	Epilepsy	Healthy	Epilepsy	Healthy	Epilepsy
Transceiver	[109]		[110]		[56]	
Tx power without compression(mW)	2.52		0.42		0.26	
Total power without compression(mW)	2.53		0.43		0.27	
CR	1.94	2.56	1.94	2.56	1.94	2.56
Tx power after compression(mW)	1.3	0.98	0.22	0.16	0.13	0.1
Compressor power(nW)	50					
Total power after compression(mW)	1.31	0.99	0.23	0.17	0.14	0.11
Power reduction comparing to 2-stage technique(%)	11.4	25.2	8.7	29	14.3	27.3

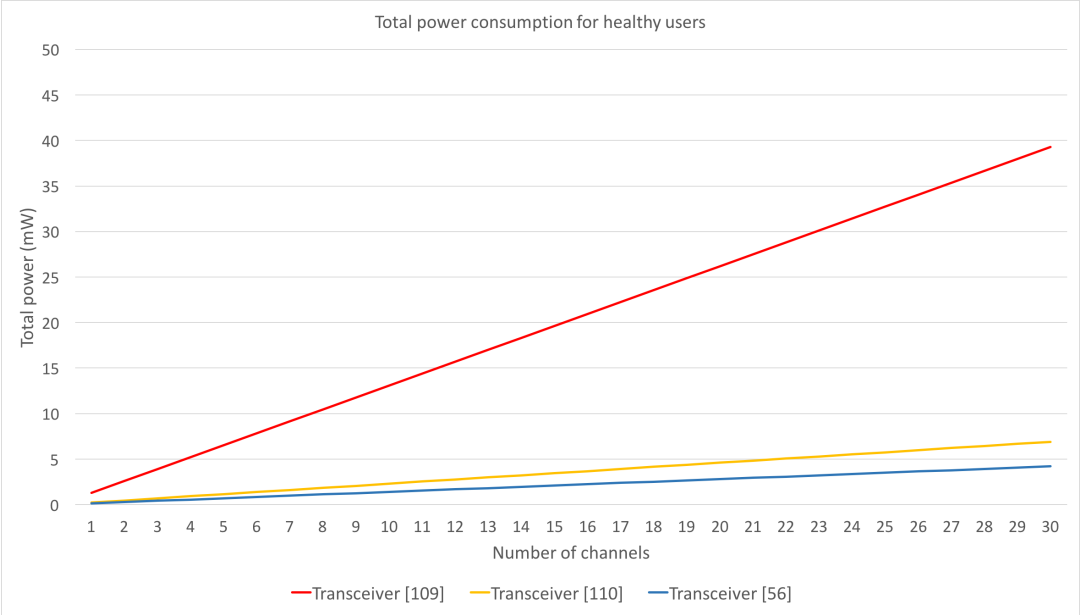

Figure 7.7: Total power consumption for healthy user with the Log2 sub-band compressor implemented

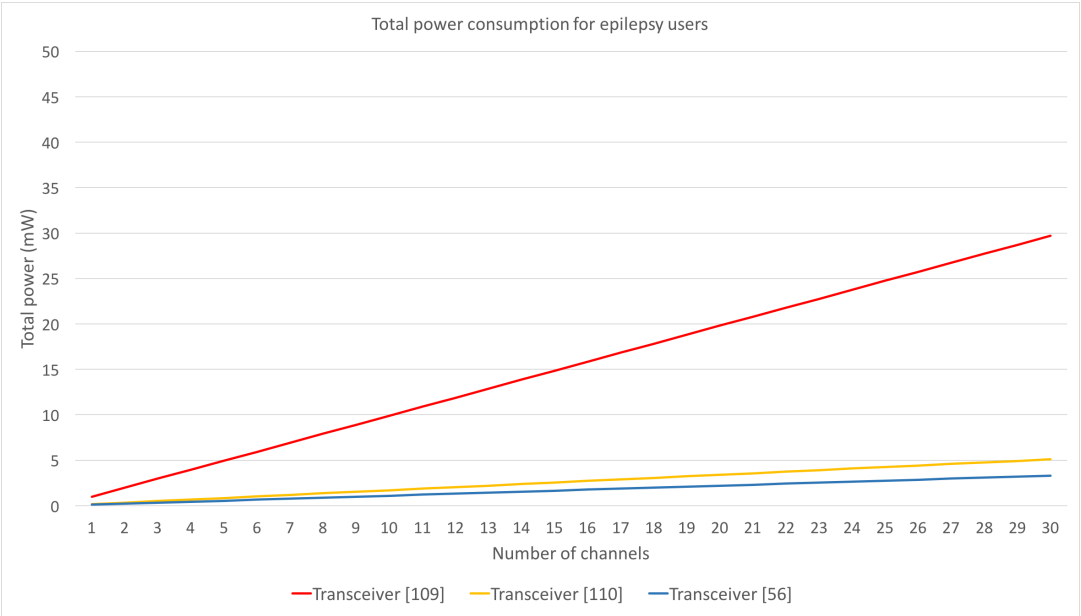

Figure 7.8: Total power consumption for epilepsy user with the Log2 sub-band compressor implemented

plemented. The results are shown in Table 7.6 and Table 7.7. If there are more than 35 channels of input, the Log2 sub-band encoding consumes more power than the 2-stage technique, as has been discussed in chapter 6, but the extra power saving the Log2 sub-band can bring cancels this disadvantage.

If there are more channels of inputs, the system with the Log2 sub-band can have a tailored setting of the data compressor for each channel, hence a higher CR and power saving might be possible. On the other hand, the system with the 2-stage technique only implements one pre-generated code book, and the power performance might be worse with multiple channels of input.

7.4 Impact on operating time

A lower system power consumption enables a longer operating time, and therefore both data compression techniques listed above can help to extend the battery lifetime. In the following section, the operating time improvements of the system will be analysed.

7.4.1 Battery overview

Some battery candidates that can be used in a wearable and wireless biomedical signal recorder are listed in Table 7.8. Considering the size of such a device, the large coin cell (LCC) and the small coin cell (SCC) batteries are two common options. The NAT-1 mentioned earlier uses SCC, and some other commercialised equipments may use their own customised power units, which will not be discussed here. The AA and AAA batteries may be used by devices that have more space for the battery, and since these batteries are larger, they often have a much higher capacity and longer lifetime compared to LCC or SCC batteries.

Table 7.8: Batteries

Make	[117]	[118]	[119]	[120]
Type	AAA	AA	LCC	SCC
Voltage (V)	1.5	1.5	3	1.55
Size(diam.*height in mm)	10.5*44.5	14.5*50	20*2.5	6.8*2.1
Weight (g)	11.5	24	2.5	0.4
Capacity (mAh)	1000	2850	170	20

7.4.2 Battery lifetime analysis

Because the CRs the two compression techniques can achieve are different, their impacts on the battery lifetime are different. A higher CR brings a more significant battery lifetime extension.

The lifetime time of a battery on a given EEG recorder can be estimated by

$$lifetime = battery\ capacity * voltage/device's\ power\ consumption. \tag{7.1}$$

7.4.2.1 DPCM+Huffman

Tables 7.9 to 7.11 show the operating time of the EEG signal recorder with and without the 2-stage technique implemented. The system may be powered by the four different batteries previously referred to, assuming it has one input channel or 21 channels if a standardised International 10/20 system [20] is required. The epileptic EEG data is easier to compress, therefore the system operates longer for epilepsy users. The result proves the 2-stage technique helps to extend the battery lifetime.

7.4.2.2 Log2 sub-band

Table 7.12 to 7.14 give the operating time before and after having a Log2 sub-band data compressor assuming the system has one or 21 input channels.

Similarly to the system with the DPCM+Huffman coding, the operating time of an EEG signal recorder with the Log2 sub-band is longer for epilepsy users than for healthy users.

7.4.2.3 Comparison of two techniques

Table 7.15 shows the battery lifetime improvements that both techniques can bring to the wearable and wireless biomedical signal recorder. As both compression techniques have better performances on the epileptic EEG data, the recorder can run longer on epilepsy users no matter which technique is implemented.

The Log2 sub-band technique has achieved higher CRs from all the EEG data, and it can save more transceiver power, as mentioned earlier, so the system with the Log2 sub-band brings a longer extension of the system operating time.

Overall, if the transceiver power takes a greater proportion from the system power, implementing a data compression technique will bring more power saving to the system.

Table 7.9: Battery lifetime for device using transceiver [109] with and without the 2-stage technique implemented

Battery	AAA				AA				LCC				SCC			
Number of channels	1		21		1		21		1		21		1		21	
Data type (H: healthy E:epilepsy)	H	E	H	E	H	E	H	E	H	E	H	E	H	E	H	E
Total power without compression (mW)	2.53		53.13		2.53		53.13		2.53		53.13		2.53		53.13	
Total power after compression (mW)	1.46	1.24	30.66	26.04	1.46	1.24	30.66	26.04	1.46	1.24	30.66	26.04	1.46	1.24	30.66	26.04
Battery lifetime without compression (hours)	592.89		28.23		1689.72		80.46		201.58		9.60		12.25		0.58	
Battery lifetime after compression (hours)	1027.40	1209.68	48.92	57.60	2928.08	3447.58	139.43	164.17	349.32	411.29	16.63	19.59	21.23	25.00	1.01	1.19

Table 7.10: Battery lifetime for device using transceiver [110] with and without the 2-stage technique implemented

Battery	AAA				AA				LCC				SCC			
Number of channels	1		21		1		21		1		21		1		21	
Data type (H: healthy E: epilepsy)	H	E	H	E	H	E	H	E	H	E	H	E	H	E	H	E
Total power without compression (mW)	0.43		9.03		0.43		9.03		0.43		9.03		0.43		9.03	
Total power after compression (mW)	0.25	0.22	5.25	4.62	0.25	0.22	5.25	4.62	0.25	0.22	5.25	4.62	0.25	0.22	5.25	4.62
Battery lifetime without compression (hours)	3488.37		166.11		9941.86		473.42		1186.05		56.48		72.09		3.43	
Battery lifetime after compression (hours)	6000	6818.18	285.71	324.68	17100	19431.82	814.29	925.32	2040	2318.18	97.14	110.39	124	140.91	5.9	6.71

Table 7.11: Battery lifetime for device using transceiver [56] with and without the 2-stage technique implemented

Battery	AAA				AA				LCC				SCC			
Number of channels	1		21		1		21		1		21		1		21	
Data type (H: healthy E: epilepsy)	H	E	H	E	H	E	H	E	H	E	H	E	H	E	H	E
Total power without compression (mW)	0.27		5.67		0.27		5.67		0.27		5.67		0.27		5.67	
Total power after compression (mW)	0.16	0.14	3.36	2.94	0.16	0.14	3.36	2.94	0.16	0.14	3.36	2.94	0.16	0.14	3.36	2.94
Battery lifetime without compression (hours)	5555.56		264.55		15833.33		753.97		1888.89		89.95		114.81		5.47	
Battery lifetime after compression (hours)	9375	10714.29	446.43	510.2	26718.75	30535.71	1272.32	1454.08	3187.5	3642.86	151.79	173.47	193.75	221.43	9.23	10.54

Table 7.12: Battery lifetime for device using transceiver [109] with and without the Log2 sub-band implemented

Battery	AAA				AA				LCC				SCC			
Number of channels	1		21		1		21		1		21		1		21	
Data type (H: healthy E: epilepsy)	H	E	H	E	H	E	H	E	H	E	H	E	H	E	H	E
Total power without compression (mW)	2.53		53.13		2.53		53.13		2.53		53.13		2.53		53.13	
Total power after compression (mW)	1.31	0.98	27.51	20.79	1.31	0.98	27.51	20.79	1.31	0.98	27.51	20.79	1.31	0.98	27.51	20.79
Battery lifetime without compression (hours)	592.89		28.23		1689.72		80.46		201.58		9.60		12.25		0.58	
Battery lifetime after compression (hours)	1145.04	1530.61	54.53	72.15	3263.36	4362.24	155.4	205.63	389.31	520.41	18.54	24.53	23.66	31.63	1.13	1.49

Table 7.13: Battery lifetime for device using transceiver [110] with and without the Log2 sub-band implemented

Battery	AAA				AA				LCC				SCC			
Number of channels	1		21		1		21		1		21		1		21	
Data type (H: healthy E: epilepsy)	H	E	H	E	H	E	H	E	H	E	H	E	H	E	H	E
Total power without compression (mW)	0.43		9.03		0.43		9.03		0.43		9.03		0.43		9.03	
Total power after compression (mW)	0.23	0.17	4.83	3.57	0.23	0.17	4.83	3.57	0.23	0.17	4.83	3.57	0.23	0.17	4.83	3.57
Battery lifetime without compression (hours)	3488.37		166.11		9941.86		473.42		1186.05		56.48		72.09		3.43	
Battery lifetime after compression (hours)	6521.74	8823.53	310.56	420.17	18586.96	25147.06	885.09	1197.48	2217.39	3000	105.59	142.86	134.78	182.35	6.42	8.68

Table 7.14: Battery lifetime for device using transceiver [56] with and without the Log2 sub-band implemented

Battery	AAA				AA				LCC				SCC	
Number of channels	1		21		1		21		1		21		1	
Data type (H: healthy E: epilepsy)	H	E	H	E	H	E	H	E	H	E	H	E	H	E
Total power without compression (mW)	0.27		5.67		0.27		5.67		0.27		5.67		0.27	
Total power after compression (mW)	0.14	0.11	2.94	2.31	0.14	0.11	2.94	2.31	0.14	0.11	2.94	2.31	0.14	0.11	2.94	2.31
Battery lifetime without compression (hours)	5555.56		264.55		15833.33		753.97		1888.89		89.95		114.81		5.47	
Battery lifetime after compression (hours)	10714.29	13636.36	510.2	649.35	30535.71	38863.64	1454.08	1850.65	3642.86	4636.36	173.47	220.78	221.43	281.82	10.54	13.42

As shown in Table 7.15, [109] is the most power hungry transceiver, and by implementing data compression on a system that has a this transceiver, the battery lifetime will increase the most.

7.5 Summary

In this chapter, a comprehensive system power analysis is given. The benefits of having one of two data compression techniques on the simplified system are illustrated. It is clear that both techniques can help to reduce the system power. Given the high compression and adaptivity performance the Log2 sub-band has, it is superior to the DPCM+Huffman coding technique, and it still has more potential to explore in the future.

Table 7.15: Battery lifetime improvement after implementing the data compression unit

User	Healthy					
Transceiver	[109]		[110]		[56]	
Compression technique	DPCM+Huffman	Log2 sub-band	DPCM+Huffman	Log2 sub-band	DPCM+Huffman	Log2 sub-band
Battery lifetime improvement	73.29%	93.13%	72%	86.96%	68.75%	92.86%
User	Epilepsy					
Transceiver	[109]		[110]		[56]	
Compression technique	DPCM+Huffman	Log2 sub-band	DPCM+Huffman	Log2 sub-band	DPCM+Huffman	Log2 sub-band
Battery lifetime improvement	104.03%	158.16%	95%	152.94%	92.86%	145.45%

Chapter 8

Hardware fabrication of the Log2 sub-band encoding

This chapter gives the very-large-scale integration(VLSI) layout of the Log2 sub-band encoding technique, and some test results from the fabricated chip of the technique.

8.1 VLSI layout

The circuit is modelled in VHDL, and test benches are prepared in Verilog. The design and test benches are given in Appendix C2 and C3. Synthesis is performed by the Cadence RTL compiler using 65nm UMC technology files and faraday standard cell libraries. The circuit is optimised for timing and power using appropriate synthesis options. The resulting netlist is then translated to a VLSI layout and Graphic Database System II(GDSII) output using automated layout tool flows.

The circuit is initial tested as a behavioural model using real input data sequences and known results, with a check mechanism to validate circuit functional correctness. The same test-bench is then used for the post-synthesis and post-layout circuits.

The final layout is incorportaed into a 65nm IC design GDSII layout, which was fabricated via IMEC EUROPRACTICE in 2015.

The layout of the Log2 sub-band encoding unit with a {4,4,4} band setting on the fabricated chip is given in Figure 8.1. The utilisation rate, which is the area cost of the cells that performs the data compression over the circuit area, of this design is around 50%. The total area cost of compressing one channel signal can be calculated as

$$area\,cost = layout\,length * layout\,width * utilisation\,rate, \tag{8.1}$$

given the dimension of the circuit, which is 115 μm by 20 μm, the area cost is 1150 μm^2. According to the result from CACTI, the area cost of the RAM that stores the code book for the 2-stage compression technique is at least 2.1 mm^2, so the Log2 sub-band design takes much less space than the 2-stage technique even if there are more channels of input.

In the VLSI implementation, the Log2 sub-band circuits layout was placed and routed using Cadence software, and the utilisation rate was the default setting for the layout tools. However, higher utilisations are possible, at the cost of using higher metal layers. In this application, where high speed use (e.g. 100KHz , 1MHZ, etc) is not envisaged, that is not likely to affect performance. An improved layout generated by Cadence automated layout tools is shown in Figure 8.2. The area cost of this improved design is 182 μm^2, which is only around 1/11538 of the area cost of the 2-stage technique for compressing one channel of data.

8.2 Data compression test result

The fabricated chip that the Log2 sub-band encoding is on is shown in Figure 8.3, and the design was part of a multi project integrated circuit (IC). The Figure 8.4 shows the close-up view of the chip.

Some test data were applied to the data compression circuit via a computer and an Input/Output interface board, and the compression result of the circuit is generated by the Cadence and shown in Figure 8.5. The column of input in Figure 8.5 shows the previous(PREV) and the current(NEW) data samples. Each output sample is 14-bit long since the longest code word for a {4,4,4} band setting is 14, and each sample contains the compressed data, and if the compressed data is less than 14 bits, the rest of bits in the output sample would be set to 1. The column of compressed data are compressed data samples that will be shifted out, and HD in this column indicates the 2-bit header of each compressed data sample. For instance, if previous and current data samples are both 000 h, the output would be 11111111111100, because only a header 00 is needed to represent the current sample in this case, and the compressed data sample is 00 as highlighted in 8.5. It is clear that the chip is fully functional for compressing the given data with the

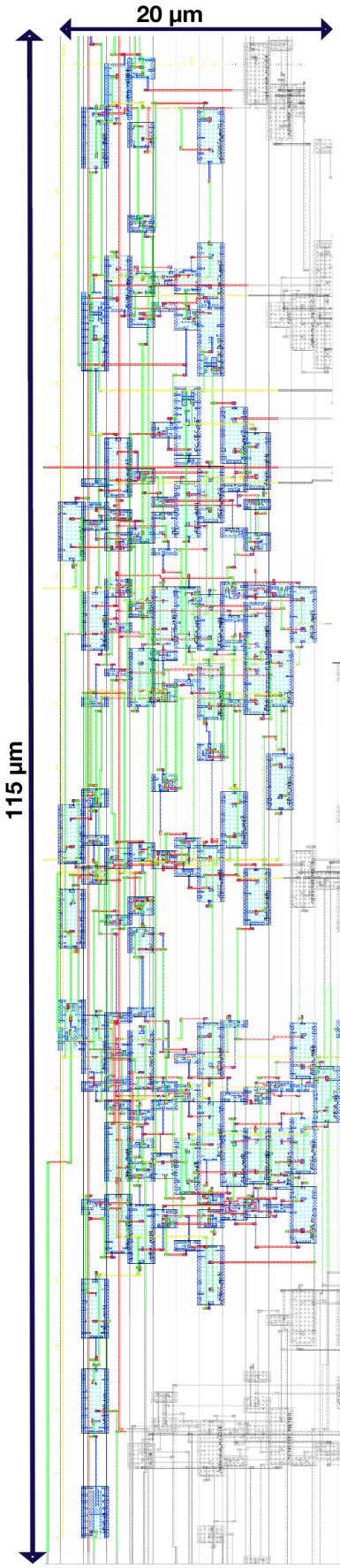

Figure 8.1: VLSI layout of the Log2 sub-band

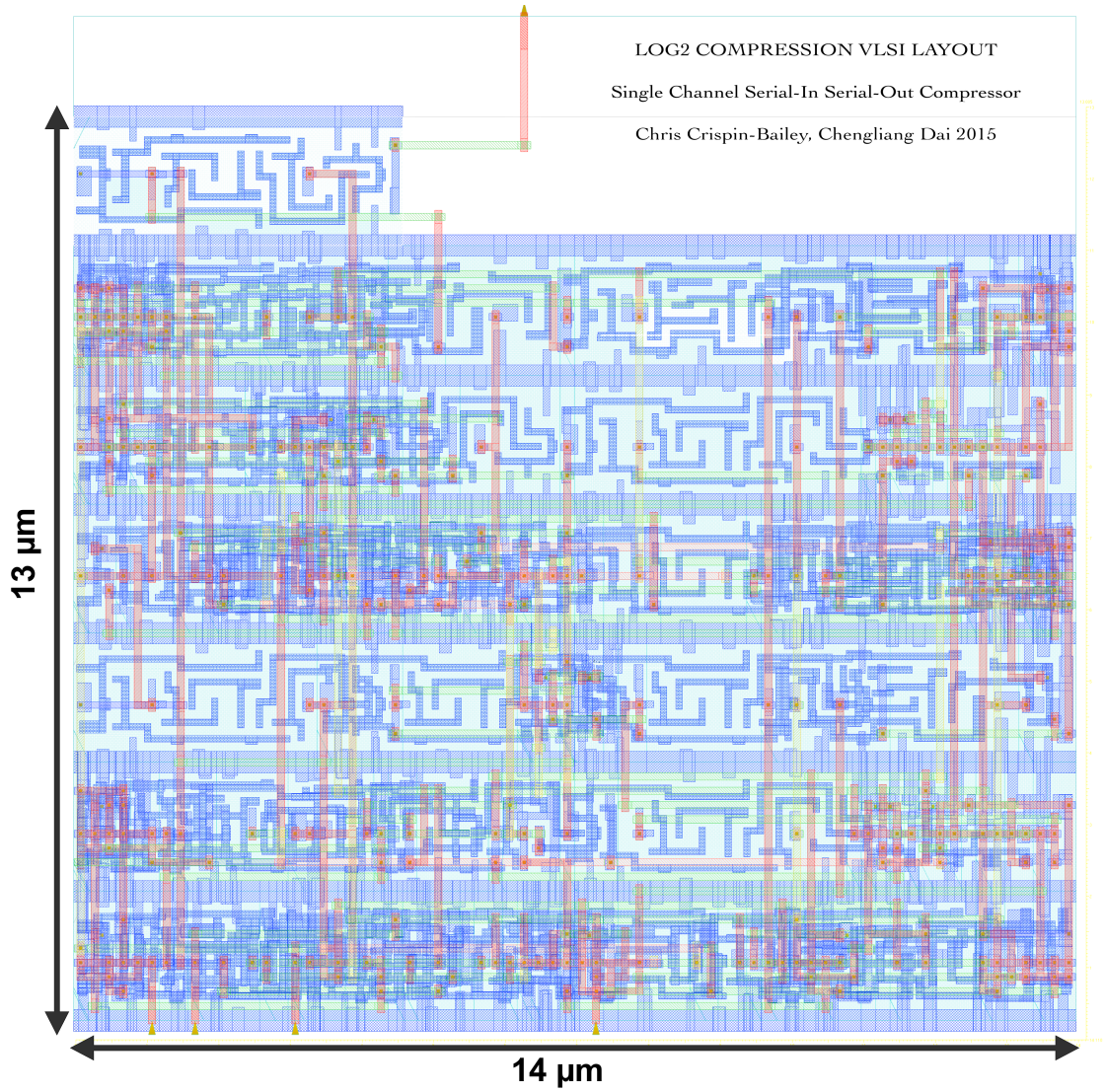

Figure 8.2: Optimised VLSI layout of the Log2 sub-band

scheme of Log2 sub-band encoding.

Figure 8.6 is a snapshot from the oscilloscope, and it shows the waveform of the 12-bit input and compressed output. The waveform has proved that the Log2 sub-band unit can shorten the input data when possible. For instance, in the test case 1, the input symbol changes between 000h and 007h, and based on the scheme of Log2 sub-band encoding showed in Figure 3.2, only one nibble plus a 2-bit header are needed to represent each symbol, so the output of the chip is a header 10 followed by either 1110 or 0000 in this case.

8.3 Summary

Some details of the Log2 sub-band unit that is on a fabricated chip are given in this chapter, including the area cost and a possible improved design of it. The preliminary test result of the unit is also introduced, and the result has shown that the design can work properly for compressing the given data.

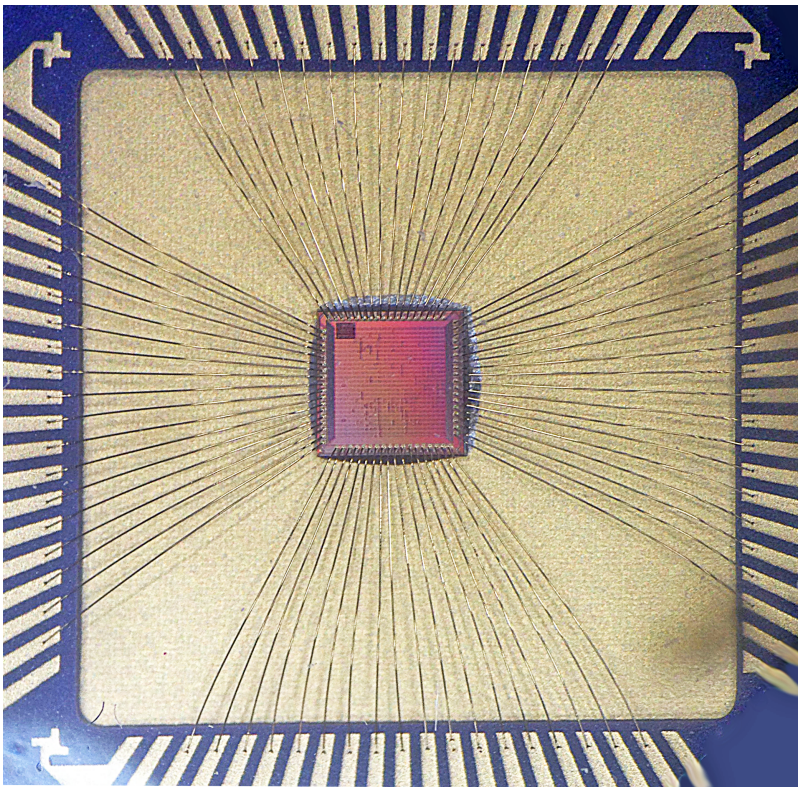

Figure 8.3: Fabricated chip of the Log2 sub-band

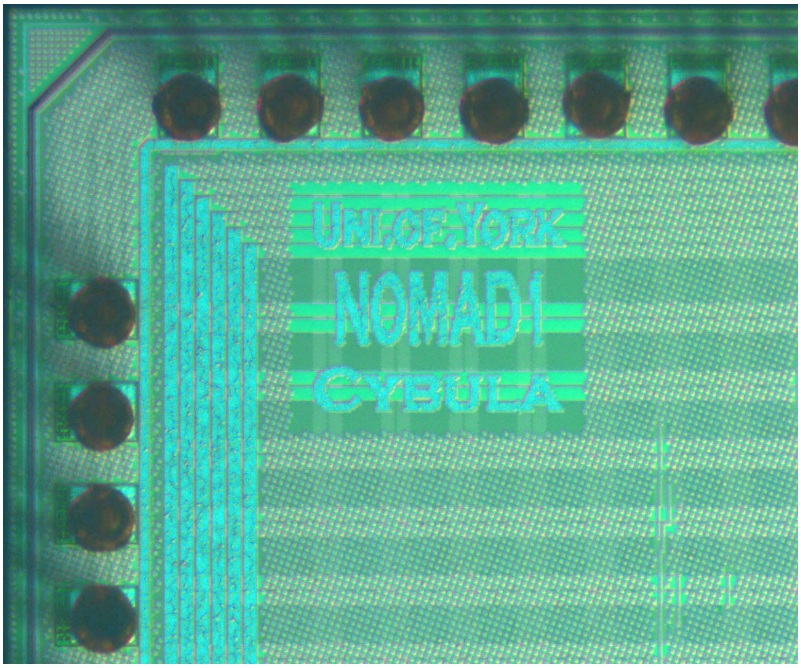

Figure 8.4: Close-up view of the chip top-left corner

Input	Output		Compressed data
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[005]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[005] NEW[050]	1 1 1 1 1 1 1 1 0 1 0 1 0 1 x0x1	--	HD[01] [0101]
PREV[050] NEW[500]	1 1 1 1 0 1 0 1 0 0 0 0 1 0 x0x1	--	HD[10] [0101] [0000]
PREV[500] NEW[530]	0 1 0 1 0 0 0 0 0 0 0 0 1 1 x0x1	--	HD[11] [0101] [0000] [0000]
PREV[530] NEW[531]	1 1 1 1 0 0 1 1 0 0 0 0 1 0 x0x1	--	HD[10] [0011] [0000]
PREV[531] NEW[531]	1 1 1 1 1 1 1 1 0 0 0 1 0 1 x0x1	--	HD[01] [0001]
PREV[531] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	0 0 0 0 0 0 0 0 0 0 0 0 1 1 x0x1	--	HD[11] [0000] [0000] [0000]
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[001]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[001] NEW[002]	1 1 1 1 1 1 1 1 0 0 0 1 0 1 x0x1	--	HD[01] [0001]
PREV[002] NEW[004]	1 1 1 1 1 1 1 1 0 0 1 0 0 1 x0x1	--	HD[01] [0010]
PREV[004] NEW[008]	1 1 1 1 1 1 1 1 0 1 0 0 0 1 x0x1	--	HD[01] [0100]
PREV[008] NEW[010]	1 1 1 1 1 1 1 1 1 0 0 0 0 1 x0x1	--	HD[01] [1000]
PREV[010] NEW[020]	1 1 1 1 0 0 0 1 0 0 0 0 1 0 x0x1	--	HD[10] [0001] [0000]
PREV[020] NEW[040]	1 1 1 1 0 0 1 0 0 0 0 0 1 0 x0x1	--	HD[10] [0010] [0000]
PREV[040] NEW[080]	1 1 1 1 0 1 0 0 0 0 0 0 1 0 x0x1	--	HD[10] [0100] [0000]
PREV[080] NEW[000]	1 1 1 1 1 0 0 0 0 0 0 0 1 0 x0x1	--	HD[10] [1000] [0000]
PREV[000] NEW[000]	1 1 1 1 0 0 0 0 0 0 0 0 1 0 x0x1	--	HD[10] [0000] [0000]
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----
PREV[000] NEW[000]	1 1 1 1 1 1 1 1 1 1 1 1 0 0 x0x1	--	HD[00] ----

Figure 8.5: The compression test result from the fabricated chip

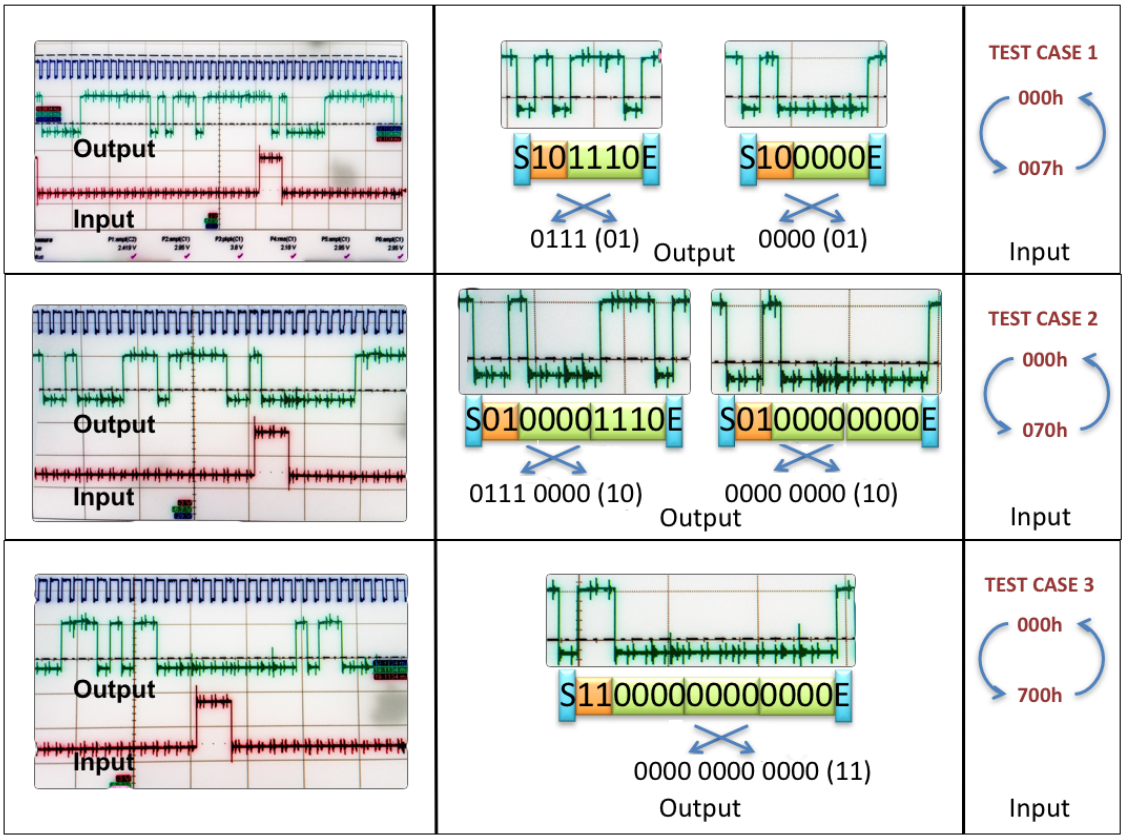

Figure 8.6: The waveform of the Log2 sub-band circuit

Chapter 9

Conclusions and future work

The final chapter summarises the work described in this thesis and recaps the main contributions of the research. All the observations and results will be presented here, and some possible future work that can build upon this research will be given at the end of this chapter.

9.1 Thesis overview

The key goal of this research is to find out if biomedical data compression and reduction technique can help to reduce the system power consumed by a wearable and wireless device beyond state of the art. The thesis has investigated the current techniques for reducing the size of the biomedical data, and has proposed and introduced a new time-domain-based lossless data compression technique called the Log2 sub-band encoding.

The thesis further tested the compression performance of the presented techniques on different bio-signals such as EEG and EMG, and the results were compared with a lossless 2-stage compression technique consisting of the DPCM and the Huffman coding.

The next part of the research discussed the possibility of using the features that the Log2 sub-band extracted from the EEG signals to achieve the seizure detection, and then presented a crude method to detect the seizure.

The final part of the research investigated the power consumption of the Log2 sub-band and the 2-stage techniques for compressing the EEG signal, and estimated the possible power saving of the transceiver after having a data compression unit implemented. All the results were combined at the end to analyse the benefits that data compression could bring to the whole system, and to the battery lifetime. A test chip has been fabricated

and validated in bench-tests.

9.2 Main conclusions

The conclusions of each chapter from chapters 4 to 7 are given below

1. Chapter 4

Both the DPCM+Huffman coding and the Log2 sub-band encoding lossless compression techniques can reduce the size of the EEG and EMG data. Under the best-case scenario for the 2-stage technique, where the pre-generated code book contains only the symbols that occurred in the training dataset, and the code book is trained and used to compress a specific kind of data, the Log2 sub-band encoding still achieved a CR at least 0.4% higher than the 2-stage technique for most of the datasets (Table 4.5).

If a code book covers all possible symbols for the 2-stage technique, the compression performance will get worse. This problem can be partly solved by enhancing the original PDF of the data in the new code book, but a decrease of the CR is inevitable (Table 4.6 and Table 4.7).

The code book used for the 2-stage technique is trained for a specific kind of data, and for instance the code book can only achieve a high CR from the EEG data of the seizure. To have a code book that is optimal for all kinds of signals is unlikely (Table 4.8).

The Log2 sub-band is more adaptive and efficient than the 2-stage technique even though their work schemes are close. The Log2 sub-band encoding can be used for compressing other biomedical signals such as EHG, ECG and ERP (Table 4.4).

Both compression techniques perform better on signals with less rapid changes like the EEG signal of the healthy people (Table 4.1 and Table 4.3).

2. Chapter 5

The Log2 sub-band encoding can provide the band distribution information during the compression process. The information can be used as the signal feature to extract certain signal events. For instance, the seizure and the seizure-free data from an epilepsy patient have different band distributions (Figure 5.1 to Figure 5.4), and by using the right technique, it is possible to detect the seizure episode with a high accuracy. The band distribution is the by-product of the data compression process, and it is a 'free' benefit of the Log2 sub-band encoding.

3. Chapter 6

The DPCM+Huffman coding needs a block of memory to store the code book, and

the same memory block can be used for compressing multiple channels of signals. On the other hand, each channel of signal requires a Log2 sub-band unit for the compression.

An optimised SISO Log2 sub-band compressor consumes around 50 nW for compressing one channel of healthy human EEG data at 1kHz, and the data compressor consumes a bit more power for compressing signals with more rapid changes (Table 6.5).

The 2-stage technique consumes around 856 nW for compressing one channel of EEG data if only considering the power consumed by the RAM, and only the dynamic power increases if the system has more input channels (Table 6.6).

If a system has more than thirty-five channels of input, the 2-stage technique will consume less power than the optimised Log2 sub-band design does for compressing the data (Figure 6.10). If including the shifter power, the 2-stage technique will always consume more power than the optimised Log2 sub-band design.

4. Chapter 7

Both data compression techniques can reduce the transceiver and the memory power, and the Log2 sub-band can achieve a more significant power saving because of the higher CRs it can achieve (Table 7.6 and Table 7.7).

Because the transceiver is one of the biggest contributors to the system power consumed by a wearable and wireless biomedical device, both data compression techniques can help to reduce the system power.

The battery lifetime of a simplified wearable and wireless EEG signal recorder can be extended by both compression techniques, but the Log2 sub-band encoding brought more significant improvements to the battery lifetime, which was at least 87% longer for healthy users and 145% for epilepsy users (Table 7.9 to 7.15).

Some of the results given in this chapter have been published in papers listed in Appendix A and posters given in Appendix B.

5. Chapter 8

The preliminary results have shown that the Log2 sub-band unit on the fabricated test chip can help to compress the input signals (Figure 8.5 and Figure 8.6). The area cost of the Log2 sub-band encoding is much smaller than the 2-stage technique that requires a RAM to store the code book.

9.3 Future work

Some future work that can build upon this research is described below.

1. Lossy compression technique analysis

This research focused on a lossless data compression technique that can be implemented to a wearable and wireless biomedical device, and the same methodology can be applied to explore lossy compression techniques. Lossy techniques are suitable for applications that allow some discrepancies between original and reconstructed data. Even though some lossy compression techniques such as DWT are complex, and they might consume more power to run, it is still worth paying more attention to this area considering the remarkable CRs these techniques usually achieve.

2. Log2 sub-band optimisation

In this thesis, the Log2 sub-band was set to divide a 12-bit data sample into three nibbles, which is a form of {4,4,4}. The CR might get better when compressing different signals with different lengths of bands, for instance, {3,3,3,3} or {3, 4, 5} for 12-bit data. For a multi channels system, each channel can have its tailored design to optimise the overall CR. Finding possible measures to improve compression performance will be an interesting area to explore in the future.

3. Log2 sub-band used as a second stage for data compression

Some preliminary work has been done to explore the possibility of using the Log2 sub-band encoding as the second stage in a compression process. Using DPCM as the first stage technique has been tested, but the results are not included in this thesis as the CRs this process achieved were inferior to the results of using the Log2 sub-band only. The DPCM and the Log2 sub-band have similar work schemes, so combining these two techniques together will not bring extra gains to the overall CR. However, there are other compression algorithms that can be used as the first stage technique, and it is worth exploring the possibility of combining these techniques with the Log2 sub-band. As discussed in this thesis, the possible techniques for the first stage processing should not consume more system power than they can save.

4. DPCM+Huffman code book improvement

As mentioned earlier, a Huffman code book has to be generated with a training dataset before compression starts, and to get the best result, the PDF of the uncompressed data should be very similar to that of the training dataset. However, a biomedical device is often used to collect multiple channels of data simultaneously, and it may be used by people with different physiological conditions. Finding a code book that gives good compression results under all conditions is very important for the hardware implementation of the 2-

stage technique, and thorough research into all the trade-offs of generating a code book can be done in the future.

5. Signal classification with the Log2 sub-band

In chapter 5 a primitive method was presented to show seizure detection with the signal features extracted by the Log2 sub-band. There are many other more sophisticated techniques, like an artificial neural network, that make the signal classification more accurate. For multi channels system, a more reliable decision can be made based on the signal classification result of each channel. These techniques will help to exploit more potential from the Log2 sub-band encoding, and it will be another interesting area to look at in the future.

Appendix A

Published papers based on this thesis

The following published conference papers are based on the research work described in this thesis, and I am the primary author of the papers given below.

C. Dai and C. Bailey, "A time-domain based lossless data compression technique for wireless wearable biometric devices," in *SENSORCOMM 2013, The Seventh International Conference on Sensor Technologies and Applications*, pp. 104107, 2013.

C. Dai and C. Bailey, "Power analysis of a lossless data compression technique for wireless wearable biometric devices," in *Ph.D. Research in Microelectronics and Electronics (PRIME), 2015 11th Conference on*, pp. 97100, June 2015.

C. Dai and C. Bailey, "A lossless data reduction technique for wireless eeg recorders and its use in selective data filtering for seizure monitoring," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 61866189, Aug 2015.

A Time-Domain Based Lossless Data Compression Technique for Wireless Wearable Biometric Devices

Chengliang Dai, Christopher Bailey
Department of Computer Science
University of York
York, United Kingdom
Email: cd633@york.ac.uk, christopher.crispin-bailey@york.ac.uk

***Abstract*—This paper presents a promising lossless compression technique called Log2 Sub-band encoding, which is suitable for implementing on wireless wearable biometric devices. Data compression promises a power saving from the transceiver during the data transmission and further extends battery lifetime on a device. The performance of this technique is measured in term of compression ratio (CR). Our simulations suggest a CR that is comparable, and indeed superior to the combination of Differential pulse-code modulation (DPCM) and Huffman coding, whilst using minimal hardware. The simulations primarily use electroencephalogram (EEG) data, and a estimated power saving is given whilst the implementation issues and possible influence of different biomedical data on technique′s performance will also be considered.**

***Index Terms*—Wearable device; Bioelectric data; Lossless compression technique; Power consumption**

I. Introduction

It has been a long time since people realized that bioelectric signals could reflect alternations inside our biological system, and such signals now can help researchers to study many diseases or other external and internal stimuli that cause these alternations. As a result, wireless wearable biometric devices have drawn much attention in recent years, and they promise to make collecting bioelectric data more flexible and convenient. Applications and prototypes have been widely used in medical monitoring, brain-computer interfaces (BCI), and other relevant areas [1] [2]. Especially in medical fields, wearable devices significantly improved patients quality of living whilst allowing caregivers and doctors to provide a better healthcare for patients with various neurological and physical diseases like epilepsy, Alzheimers, and insuring early detection of emergency conditions for high risk patients.

With all the advantages of using wireless wearable devices, power consumption now becomes a design obstacle to the prevailing of these devices. A wearable device is usually battery powered, and the limited size of the device restricts the size and capacity of batteries, so power efficiency is extremely crucial for a device to extend the operational lifetime and further improve user experience. There are several ways to reduce device power consumption, and since the transceiver is one of the significant contributors to the power consumption, reducing the size of transmitted data might be a fruitful objective.

Nevertheless, compressing the data increases the complexity of the system and itself consumes power, such that it is necessary to strike a balance between the power that a compression unit is able to save and the power it consumes. A compression technique that demands less hardware resource but delivers high compression ratio (CR) will be a desirable solution, particularly where large numbers of channels become more common, and hence multiply the hardware and power costs of compression.

Bioelectric signals are highly non-Gaussian, non-stationary and non-linear which make data compression a difficult task. Moreover, signals for clinical uses require a high consistency of original data and reconstructed data that rules out several lossy data compression techniques [3]. Various compression methods have been tested on bioelectric signals especially on EEG signals including entropy encoding [4] [5], predictive encoding [4] [6], transform-based encoding [3] etc. Entropy encoding techniques such as Huffman coding are versatile and widely used on many kinds of data, but they require the pre-knowledge of the signals, the probability density of samples for instance, before conducting any compression. This is impractical for a wireless biometric device where data are collected, processed and transmitted in real-time. Most predictive encoding techniques encounter the same problem. Transform-based encoding techniques can achieve impressive compression results, but they are either lossy or far too complex in system level. Therefore, a new method called Log2 Sub-band encoding is developed to overcome these limitations.

II. Log2 sub-band compression

The fundamental design of this algorithm is simple: after analog-to-digital conversion, every data sample is divided into several bit-fields (chunks), and each chunk is compared with the same part of the previous sample. Chunks of the current sample will not be sent if they are identical to the parts of previous sample. An extra header is sent with every sample to indicate number of chunks transmitted. For instance, if a bioelectric signal is digitalized into 12 bits, the whole procedure can be illustrated with Fig. 1.

As it is shown in Fig. 1, each data sample is chopped into 4-bit chunks, making 3 nibbles, and there will be four scenarios including transmitting whole sample, 2 nibbles of the sample, 1 nibble or only the header after each comparison.

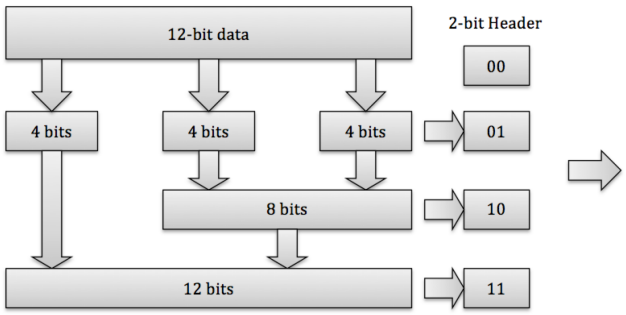

Fig. 1. Log2 Sub-bang encoding

There will be a 2-bit header added before transmission to guide the receiver to reconstruct compressed data. The procedure is similar to the simplest predictive encoding technique, Differential Pulse-Code Modulation (DPCM) [7], which eliminates short-term redundancies by taking the difference of adjacent samples. However, the bit width of each sample after applying DPCM depends on the adjacent samples that have the biggest difference, and as for bioelectric signals in which sudden spikes may happen because of any stimuli or pathological reasons, the performance of DPCM is often disappointing. Even though such problem can be partly solved by using entropy coding unit as a second processing stage [7], it is still not compatible enough with complex bioelectric signals. The Log2 Sub-band algorithm, due to its encoding scheme, is adaptive to any sudden changes of the signal with a minimum cost of hardware, and it simply takes more bits to represent the parts with severe fluctuations and fewer bits when the signals become stable.

Meanwhile, the algorithm only processes the signal within the time-domain, and no information is dropped due to the compression technique, so the raw data will be fully reconstructed after reception, and are lossless.

III. Introduction to the data for testing

The simulations of Log2 Sub-band encoding are conducted with EEG signals, and they come from two sources: humans EEG data are from a research on nonlinear deterministic patterns of brain electrical signals at the University of Bonn [8], and mouse EEG data from an on-going Alzheimer disease study in Aberdeen University [9].

A. Human EEG data

All EEG signals were recorded with a 128-channel amplifier system, and digitized into 12 bit samples. The final signals were acquired at a sampling rate of 173.61 Hz, and the band-pass filter was set to 0.53-40 Hz [8].

Data were categorized into 5 groups, which are recordings from healthy people with eyes open and closed, signals originated from within and outside epilepsy diagnosed patients epileptogenic zone (seizure generatingarea) during the seizure free interval, and recordings of seizures. Signals from healthy people were recorded extra-cranially with severe eye movement artifacts (EOG), but EOG interference was removed manually afterwards by the research group in Bonn. Recordings of epilepsy patients were recorded intra-cranially. Each group has 100 data segments of 23.6- sec duration of signals, and 4096 samples in every segment.

B. Mice EEG data

The EEG signals of mice with novel knock-in Alzheimer were recorded intra-cranially from three areas, and then digitized into 12 bit/sample at a sampling rate of 200 Hz and 1000 Hz respectively. Each data segment contains 5 minutes of gathered EEG signals, captured using the University of York/Cybula Neural Acquisition Tracker (NAT-1) device.

IV. Result analysis

To evaluate the performance of Log2 Sub-band encoding, the results from Huffman coding are used as a benchmark baseline. The compression ratio (CR) in the following analysis is defined as the size of original data vs compressed data such that:

$$CR = Size_{Original}/Size_{Compressed} \quad (1)$$

A. Huffman Coding

Optimal Huffman coding requires the probability distribution of signal samples to be known before compressing the data, and to achieve real time compression, a file-independent codebook was trained with 40 EEG data fragments from both healthy people and epilepsy patients in advance using the described data sets. To degrade the inter-channel correlation and minimize the size of codebook, DPCM was applied before conducting the Huffman coding, and it also renders better results than solely using Huffman coding [7]. Twenty EEG data fragments from healthy people with eyes closed and twenty more from patients epileptogenic zone were tested, and a CR of 1.82 was achieved from the former and 2.2 from the latter. The lowest CR of 1.55 was observed with unstable seizure signals. Twenty recording fragments from healthy people with eyes closed, patients epileptogenic zone, seizure, and mice with novel knock-in Alzheimer. And the results are showed in Fig. 2 and Table I.

B. Log2 Sub-band encoding

As it is shown in Fig. 2, the type of data and sampling frequency have significant impacts on the performance of this technique, but in all cases the Log-2 Sub-band method is superior to Huffman by between roughly an order of 5% to 15% better. Apparently, both Log2 Sub-band encoding and Huffman coding perform better when processing the signals with less fluctuations and higher similarity of adjacent samples. Huffman coding is less capable to process some bioelectric signals with a large value range since the more values they contain the longer the codes required for less frequent samples. However, Log2 Sub-band encoding is adaptive to all these signals without the need to modify the basic circuit design for each case.

Considering different results the technique delivered, it is worth looking into the changes of source signals that can be

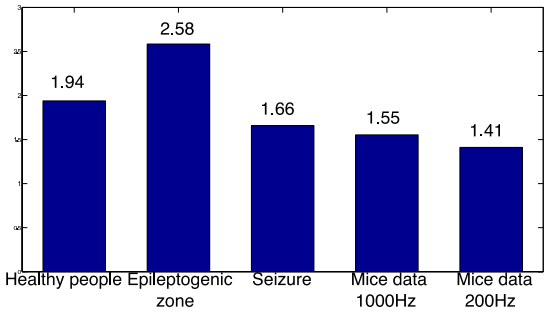

Fig. 2. Compression ratios

TABLE I
SIMULATION RESULTS

Data Type	CR of Log2 Sub-band Encoding	CR of Huffman and DPCM
Healthy people with eyes closed	1.94	1.82
Patients' epileptogenic zone	2.58	2.20
Seizure	1.66	1.51
Mice data@1000Hz	1.55	N/A
Mice data@200Hz	1.41	N/A

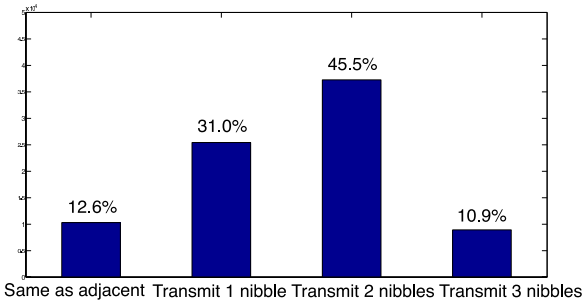

Fig. 3. The distribution of the length of transmitted healthy peoples data samples

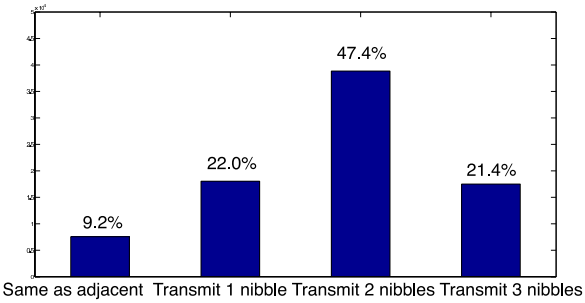

Fig. 4. The distribution of the length of transmitted seizures data samples

reflected by this compression technique. For instance, Fig. 3 and Fig. 4 are the distributions of length of transmitted data from healthy people and seizure periods. Healthy peoples EEG signals are relatively smooth so that most data samples were transmitted with 1 or 2 nibbles (76.5%), but there is an apparent drop to the number of 1 nibble data samples while compressing the seizure signals (31.0% to 22.0%). This noticeable discrepancy could be used to warn the caregivers that patients physical conditions might have changed. And this preliminary pattern recognition will not add on more workload to the system.

To reduce the power consumption is the primary purpose of introducing this encoding method. An off-the- shelf Bluetooth transceiver nRF8001 [10] is chosen to evaluate its effect. The power of this transceiver is 22mW, and it takes 22nJ to transmit 1 bit data. Considering the lowest CR (1.41) reached earlier, which is around 30% of size reduction, and if uncompressed data were supposed to be transmitted at 64Kb/s, this encoding method would bring at least 200μW′s power saving before introducing the power overhead of the compression itself. Hence, as long as the compression circuit consumes less than 200μW a significant saving might be obtained for each recorded channel.

V. Hardware Implementation

A simplified hardware design of this encoding method for processing the EEG signals above mentioned is shown in Fig. 5. Current sample and previous sample are loaded respectively, and different sub-bands (chunks) are sent to logic gates to compare, with the result used to determine the header and to control a shifter that shifts right number of bits to transmit.

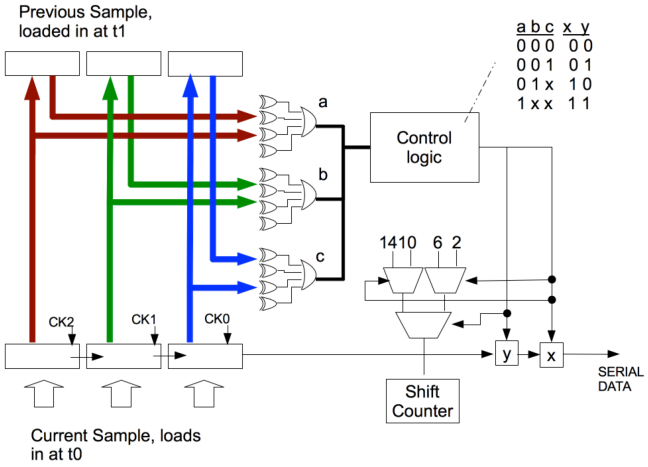

Fig. 5. Simplified Hardware Design

The number of bits represents the data payload of 0,4,8, or 12 bits, and the header of 2 bits in each case. In this case, it would therefore be a total symbol length of 2 bits, 6 bits, 10 bits, and 14 bits.

VI. Conclusion

As mentioned earlier, Log2 Sub-band encoding has a better performance on EEG data and simpler hardware design compare to other current techniques, based upon our initial evaluations. Since bioelectric signals share many common

behaviors, we can anticipate that the Log2 Sub-Band method can be applied widely to other signals besides EEG, and a more diverse analysis would be very desirable. Improvements to the system may be possible. We have already considered the possibility of variable-sub-band encoding, such that the width of bands can be varied in such a way as to ensure that the distribution of changes in each band is optimized, for instance $\{3,3,4,6\}$ might be better for some signal types with a resolution of 16-bit. Also the possibility of switching between two encodings depending upon the relative activity or idleness of signal behavior might permit higher overall compression rates with relatively minimal circuit changes.

ACKNOWLEDGMENT

We are grateful for access to the mouse EEG data supplied by the Institute of Medical Sciences, Univerysity of Aberdeen.

Mouse data was recorded using the NAT-1 (Neural Activity Tracker) device, developed by University of York and Cybula Ltd. This device was developed with support from the UK Technology Strategy Board.

REFERENCES

[1] L. Boquete, J. M. R. Ascariz, J. Cantos, R. Barea, J. M. Miguel, S. Ortega, and N. Peixoto, "A portable wireless biometric multi-channel system," *Measurement*, vol. 45, no. 6, pp. 1587–1598, Jul. 2012.

[2] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects." *Journal of neuroscience methods*, vol. 167, no. 1, pp. 115–25, Jan. 2008.

[3] W. Optimization, L. Brechet, M.-f. Lucas, C. Doncarli, and D. Farina, "Compression of Biomedical Signals With Mother Wavelet Packet Selection," vol. 54, no. 12, pp. 2186–2192, 2007.

[4] G. Antoniol and P. Tonella, "EEG data compression techniques." *IEEE transactions on bio-medical engineering*, vol. 44, no. 2, pp. 105–14, Feb. 1997.

[5] D. O'Shea and R. McSweeney, "Efficient Implementation of Arithmetic Compression for EEG," *IET Proceedings of Irish*, 2011.

[6] N. Sriraam, "A high-performance lossless compression scheme for EEG signals using wavelet transform and neural network predictors." *International journal of telemedicine and applications*, vol. 2012, p. 302581, Jan. 2012.

[7] Y. Wongsawat, S. Oraintara, T. Tanaka, and K. Rao, "Lossless Multi-channel EEG Compression," *2006 IEEE International Symposium on Circuits and Systems*, vol. 6, pp. 1611–1614, 2006.

[8] R. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, Nov. 2001.

[9] B. Platt, B. Drever, D. Koss, S. Stoppelkamp, A. Jyoti, A. Plano, A. Utan, G. Merrick, D. Ryan, V. Melis, H. Wan, M. Mingarelli, E. Porcu, L. Scrocchi, A. Welch, and G. Riedel, "Abnormal cognition, sleep, EEG and brain metabolism in a novel knock-in Alzheimer mouse, PLB1." *PloS one*, vol. 6, no. 11, p. e27068, Jan. 2011.

[10] "nRF8001 Bluetooth® low energy Connectivity IC," http://www.nordicsemi.com/eng/Products/Bluetooth-R-low-energy/nRF8001, 2012, [Online; accessed 09-April-2013].

Power Analysis of a Lossless Data Compression Technique for Wireless Wearable Biometric Devices

Chengliang Dai, Christopher Bailey
Department of Computer Science
University of York
York, United Kingdom
Email: cd633@york.ac.uk, christopher.crispin-bailey@york.ac.uk

***Abstract*—This paper presents a promising time-domain based lossless compression technique called Log2 Sub-band encoding, which is designed for using on wireless biomedical devices. Data compression can help to save power from the wireless transceiver during data transmission, and from the storage medium during reading and writing, ultimately leading to a longer battery life of the device. The performance of Log2 Sub-band is measured in terms of its compression ratio (CR) on EEG data and its power consumption. Our simulation results indicate a CR that is comparable and even superior to the well-known Huffman coding, whilst consuming minimal hardware resource. The simulations primarily use electroencephalogram (EEG) data, and the power consumption during compressing process is given to evaluate the system′s improvement on its power performance. The possible influence of different biomedical data on technique′s performance will also be considered and the signal classification potential of Log2 Sub-band will be noted.**

Keywords*—Wearable device; Bioelectric data; Lossless compression technique; Power consumption

I. Introduction

Wearable technology has drawn a lot of attention in recent years, especially after people have discovered more potential of our bioelectric signals. The wireless wearable biomedical devices usually promise a more flexible and convenient way of collecting bioelectric data, and therefore a growing number of devices have been now used in medical monitoring, brain-computer interfaces and other relevant areas [1] [2]. When it comes to medical uses, wearable devices allow caregivers to provide a better healthcare for patients with neurological and cardiovascular diseases such as epilepsy and cardiopathy, and insuring early detection of emergency conditions for high-risk patients.

Miniaturization is always preferable and sometimes critical to a wearable device, and the available battery size is usually limited as a result. To overcome this design obstacle, and to achieve a longer battery life and better user experience, one straight answer is to reduce device power consumption. There are several means to achieve this purpose, and since the transceiver, as we know, is one of the most significant contributors to a wireless wearable device power consumption, transmitting minimum data might be a promising way of addressing the problem.

On the other hand, introducing a module that compresses data increases the complexity of the system, and the module itself consumes power. The situation might get even worse where more channels of signals become more common, and hence multiply the hardware and power costs of data compression, such that there are several sets of trade-offs we need to take into account.

A simplified wearable EEG recorder model only needs amplifier(s), ADC(s), a transceiver and a data compression unit (Fig. 1) [3], so the system power consumption P_{sys} can be roughly modeled as:

$$P_{sys} = P_{amp} + P_{ADC} + P_{comp} + (D * P_{Tx})/CR \qquad (1)$$

where P_{Amp} and P_{ADC} are the power consumption of amplifier(s) and ADC(s), and P_{Tx} and P_{Comp} are the power consumption of the transceiver and the data compressor. CR is the compression ratio which is given as:

$$CR = Size_{Original}/Size_{Compressed} \qquad (2)$$

And D is duty cycle rate which is defined such that:

$$D = R_{Orig}/R_{Tx}. \qquad (3)$$

where R_{Orig} is the required data rate before compression, and R_{Tx} is the maximum data rate of transceiver. According to equation(1), if we can find a technique that yields promising CR results without a high P_{Comp}, a significant power saving will be achieved.

Biomedical signals are highly non-Gaussian, non-stationary and non-linear, and clinical uses require a high consistency of original and reconstructed signals. The characteristics of bioelectric signals and the strict requirement of clinical uses limit the number of compression techniques that can be applied to compressing biomedical data, and lossless algorithms are preferable. Some of the well-known compression algorithms have been tested on biomedical signals especially on EEG signals including entropy encoding [4] [5] , predictive coding [4] [6] , transform-based encoding [7] , etc. Huffman coding is one of the most widely used algorithms, and it gives good compression result via a simple scheme, but like other entropy encoding methods, it requires the pre-knowledge of the signals, the probability density of different signal samples for instance, before compressing the data, and therefore it is difficult to be used on wireless wearable devices where data are collected, processed and transmitted in real-time. As to predictive coding algorithms and transform-based algorithms, they are either lossy or too complex to implement on the system level. As a result, Log2 Sub-band, a time-domain based compression technique, is designed to tackle the problems mentioned above.

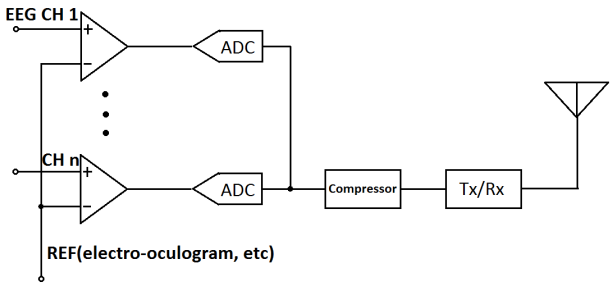

Fig. 1. A simplified EEG recorder system

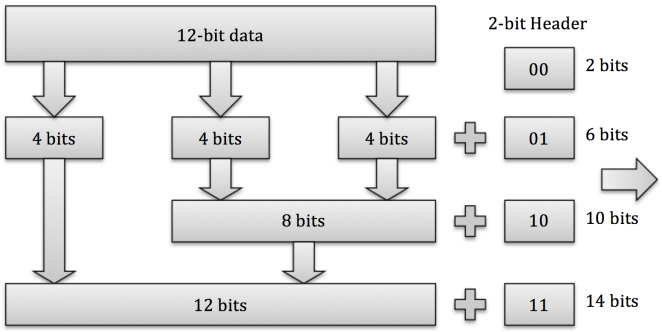

Fig. 2. Log2 Sub-band encoding

II. Log2 Sub-band Compression

Log2 Sub-band is a simple time-domain based compression algorithms [8], and its scheme is shown in Fig. 2. For instance, an EEG signal is digitalized into 12 bits after analog-to-digital conversion on a signal recorder, and every data sample is then divided into several bit-fields, and in this case, into 4-bit nibbles. Each nibble is compared with the same part of the previous sample, and instead of transmitting or storing the whole sample, only the nibbles that are different from previous sample′s are transmitted or stored. A header will be added to each compressed sample to indicate the number of nibbles transmitted or stored, and it could be 3 nibbles, 2 nibbles, 1 nibble, and none in this case, so a 2-bit header is added in the end of comparison process.

The scheme of Log2 Sub-band is similar to Differential Pulse-Code Modulation (DPCM), a simple predictive coding algorithm, which eliminates short-term redundancies by taking the difference of adjacent signal samples. However, the performance of DPCM when compressing signals with many sudden fluctuations such as seizure signals is always disappointing, and the compression result can be slightly better by combining the Huffman coding as a second processing stage [9], but the alogrithm is still very sensitive to the nature of the input signals, and to effects of scale and level offsets. The way that Log2 Sub-band works makes it more adaptive to these biomedical signals, and it takes more bits to represent samples with sudden changes, and less bits when the signal becomes stable. The simulation results will be presented in next part. .

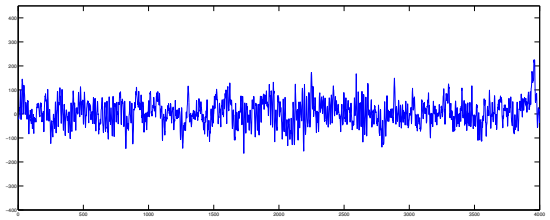

Fig. 3. Healthy Human EEG

III. Compression Ratio Results

The data used in simulations are from two sources: human EEG data (Fig. 3) are from a research on nonlinear deterministic patterns of brain electrical signals at the University of Bonn [10] that were digitized into 12-bit at a sampling rate of 173.61Hz, and mice EEG data are from an Alzheimer disease study in Aberdeen University [11] that have the same bit width as human′s and a sampling rate of 200 Hz and 1000 Hz respectively captured by the University of York/Cybula Neural Acquisition Tracker (NAT-1) device.

Human EEG data are categorized into 5 groups, which are data recorded from healthy people with eyes open and closed, signals originated from within and outside epilepsy diagnosed patients′ epileptogenic zone (seizure generating area) during the seizure free interval, and recordings of seizures, and each group has 100 data segments of 23.6 sec duration of signals. The data of healthy people with eyes open, within epileptogenic zones and seizures are chosen for the simulations, and each group is further divided into 10 sub-groups and each contains 10 data segments.

To evaluate the performance of Log2 Sub-band, the results of DPCM and Huffman coding are used as a benchmark.

A. DPCM and Huffman coding

As mentioned earlier, DPCM removes short-term redundancies, and it gives fair results on all sorts of data. Huffman coding requires a codebook, in which every possible signal sample is assigned with a code based on its frequency of occurrence, and by applying DPCM before Huffman coding, the codebook is smaller and therefore Huffman coding yields better results than DPCM or Huffman alone. As it is shown in Table I, DPCM achieves a CR of 1.38 to 1.86 on human EEG, and the combination of DPCM and Huffman gives a CR of 1.51 to 2.20 on human EEG, and around 1.46 on mice EEG. Obviously, the characteristics of the EEG data have significant impacts on the performance of algorithms, and algorithms perform better on signals with less fluctuations and higher similarity of adjacent samples.

B. Log2 Sub-band

Similar to DPCM and Huffman coding, Log2 Sub-band gives lower CRs when compressing seizure signals, and according to its results on mice EEG, this technique is more susceptible to the sampling rate of biomedical signals, but in all cases the Log2 Sub-band method is superior to Huffman and DPCM by achieving 5% to 39% higher CRs. Moreover,

TABLE I
COMPRESSION RATIO RESULTS

Data Type	CR of DPCM	CR of DPCM +Huffman	CR of Log2 Sub-band
Healthy people with eyes closed	1.73-1.84 (Avg. 1.77)	1.76-1.91 (Avg.1.82)	1.86-2.07 (Avg.1.94)
Patients′ epileptogenic zone	1.75-2.00 (Avg. 1.86)	1.92-2.27 (Avg. 2.20)	2.38-2.63 (Avg.2.58)
Seizure	1.34-1.42 (Avg. 1.38)	1.45-1.58 (Avg. 1.51)	1.59-1.73 (Avg. 1.66)
Mice data@1000Hz	N/A	1.47	1.55
Mice data@200Hz	N/A	1.45	1.44

as Log2 Sub-band doesnt require any pre-knowledge of these signals, it is more adaptive to all circumstances without the need to modify the basic circuit design for each case.

IV. HARDWARE EXPERIMENTAL RESULTS

A. Hardware Implementation

The hardware design of Log2 Sub-band is shown in Fig. 4. In this design the data is converted to a serial bitstream during compression, although there are also variations of the design which produce a parallel word output form. The function is extremely simple considering the gain in compression achieved. The preceding 12-bit sample is retained in the upper three 4-bit register banks, and compared to the new incoming 12 bit sample value. Each nibble pairing is compared to generate a match/no-match state 'a', 'b' and 'c' respectively. Generation of a header is achieved simply by translating the three inputs abc into a two bit header code, representing one of four cases representing the highest order nibble where a change is detected. The identified nibble, and all nibbles subordinate to it, are transmitted as an encoded word, which in this case means 0,1,2 or 3 nibbles are sent with a 2-bit header, resulting in 2,6,10 or 14 bits being transmitted per sample. The maximum compression in this configuration therefore reduces a 12-bit input to a 2-bit header, giving a maximum CR of 6.0. The worst case sees a 12 bit input encoded as a 14-bit output, with a minimum CR of 0.86 (which is actually more bits than the original sample). However due to the probability of each of the four possible cases a positive overall CR is achieved with signals exhibiting a largely contiguous variation behaviour, as we will show.

Note that the scheme has multiple variants. A sample width of any length is possible and the number of sub-bands and their size can vary. One could for example have a 20 bit sample with sub-bands of 8 ,6, and 6 bits respectively. The number and size of the sub-bands is an optimisation problem that relates to the nature of the data behaviour and elements such as noisiness.

Our evaluation work for this case of a serial-output three-band '4,4,4' encoder is conducted in 65-nm CMOS technology. We have modelled this and produced trial layouts, and estimated power with various input test cases. This is described in the next section.

B. Power Consumption and Power Saving

The human EEG data used earlier are used again to evaluate the power consumption of the design, and the result whilst

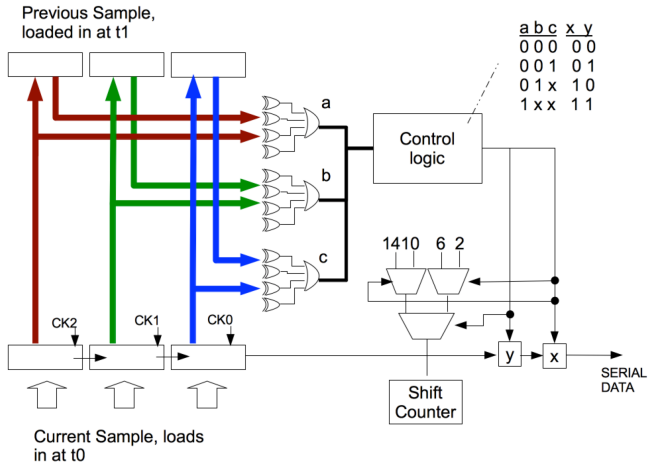

Fig. 4. Simplified Hardware Design

TABLE II
POWER CONSUMPTION OF LOG2 SUB-BAND

Data Type	CR	Static Power(nW)	Dynamic Power(nW)	Total
Healthy people with eyes closed	Avg. 1.94	Avg. 147.23	Avg. 12.92	160.15
Patients′ epileptogenic zone	Avg. 2.58	Avg. 147.29	Avg. 12.68	159.97
Seizure	Avg. 1.66	Avg. 147.29	Avg. 13.37	160.66

compressing one channel of different human biomedical signals is given in Table II. The total power consumption consists of both static and dynamic power, where the sample rate is set to 1 KHz in this case. Static power dominates the total power consumption with typical biomedical sample rates, but when more channels of signals are required, dynamic power will increase. According to our results, the circuit consumes more dynamic power when compressing seizure signals, as compared to non-seizure signals. Dynamic power is influenced by compression ratio for given sequence of data. The linear regression between CR and dynamic power consumption is given in Fig. 5.

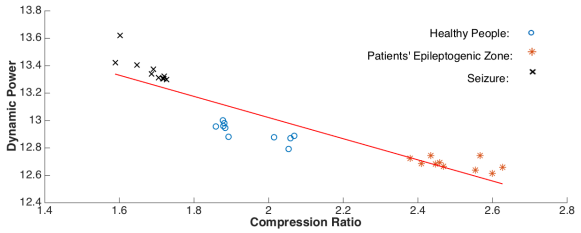

Fig. 5. regression analysis

To evaluate a possible power saving on the transceiver, a typical 22mW off-the-shelf Bluetooth transceiver nRF8001 is chosen [12], and the maximum data rate is 1Mbps, and therefore we assume that to transimit 1 bit of data, it consumes 22nJ′s power.

The results are shown in Table III. We see for instance, if we transmit one channel of healthy people′s EEG signals,

TABLE III
POWER SAVING

Data Type	Power Saving(μW)	Extra Circuit Cost(nW)	Total Saving(μW)
Healthy people with eyes closed	127.91	160.15	127.75
Patients′ epileptogenic zone	161.67	159.97	161.51
Seizure	104.96	160.66	104.80

TABLE IV
BATTERY LIFETIME

Data Type	Battery Lifetime With/Without a Compression Unit
Healthy people with eyes closed	145.5 / 75.6
Patients′ epileptogenic zone	192.5 / 75.6
Seizure	124.8 / 75.6

a system with the proposed data compression unit could save 127.91μW′s power from the transceiver with only an extra cost of 160.15 nW.

C. Battery Lifetime

The battery lifetime is calculated based on the simplified EEG recorder model mentioned earlier, and we assume 1μW amplifers [13] and 1μW ADCs [14] are used in this case, and 24 channels of data are transmitted to meet the minimum clinical requirement for certain applications [15]. The battery chosen for this simplified system is a large coin cell CR2025 [16], and its capacity is 160mAh, which keeps a device without a data compression unit running for 75.6 hours. The results of this type of battery whilst working on a system with a compression unit are given in Table IV. The results clearly indicate a significant extension of battery lifetime can be expected from a wireless EEG recorder with a compressor, and the lifetime is almost doubled when when the recorder is used on healthy people.

V. CONCLUSION

Based on all the results presented, Log2 Sub-band encoding is clearly a promising compression candidate for wearable EEG recordering devices, and it could be applied to other biomedical signals such as electromyogram(EMG) and electrocardiogram(ECG) after making some slight changes on its scheme. Our research team have recently prototyped the described compression scheme in 65nm test chip which is currently being fabricated for Apr, 2015 delivery.

As we already noticed, power consumption and compression ratios change considerably between normal and seizure signal behaviours. This suggests some potential to exploit this for dectecting such events in real-time. The possibility of varing the width of sub bands either statically or dynamically to take into account signal behaviour is also a interesting area for further research.

ACKNOWLEDGMENT

We are grateful for access to the mouse EEG data supplied by the Institute of Medical Sciences, Univerysity of Aberdeen.

Mouse data was recorded using the NAT-1 device, developed by University of York and Cybula Ltd. This device was developed with support from the UK Technology Strategy Board.

REFERENCES

[1] L. Boquete et al., "A portable wireless biometric multi-channel system," Measurement, vol. 45, no. 6, pp. 1587–1598, Jul. 2012.

[2] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects." Journal of neuroscience methods, vol. 167, no. 1, pp. 115–25, Jan. 2008.

[3] D. C. Yates and E. Rodriguez-Villegas, "A Key Power Trade-off in Wireless EEG Headset Design," 2007 3rd International IEEE/EMBS Conference on Neural Engineering, pp. 453–456, May 2007.

[4] G. Antoniol and P. Tonella, "EEG data compression techniques." IEEE transactions on bio-medical engineering, vol. 44, no. 2, pp. 105–14, Feb. 1997.

[5] D. O'Shea and R. McSweeney, "Efficient Implementation of Arithmetic Compression for EEG," IET Proceedings of Irish, 2011.

[6] N. Sriraam, "A high-performance lossless compression scheme for EEG signals using wavelet transform and neural network predictors." International journal of telemedicine and applications, vol. 2012, p. 302581, Jan. 2012.

[7] W. Optimization, L. Brechet, M.-f. Lucas, C. Doncarli, and D. Farina, "Compression of Biomedical Signals With Mother Wavelet Packet Selection," IEEE transactions on bio-medical engineering, vol. 54, no. 12, pp. 2186–2192, 2007.

[8] C. Dai and C. Bailey, "A time-domain based lossless data compression technique for wireless wearable biometric devices," in SENSORCOMM 2013, The Seventh International Conference on Sensor Technologies and Applications, 2013, pp. 104–107.

[9] Y. Wongsawat, S. Oraintara, T. Tanaka, and K. Rao, "Lossless Multi-channel EEG Compression," 2006 IEEE International Symposium on Circuits and Systems, vol. 6, pp. 1611–1614, 2006.

[10] R. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Physical Review E, vol. 64, no. 6, p. 061907, Nov. 2001.

[11] B. Platt et al., "Abnormal cognition, sleep, EEG and brain metabolism in a novel knock-in Alzheimer mouse, PLB1." PloS one, vol. 6, no. 11, p. e27068, Jan. 2011.

[12] "nRF8001 Bluetooth® low energy Connectivity IC," http://www.nordicsemi.com/eng/Products/Bluetooth-Smart-Bluetooth-low-energy/nRF8001, 2014, [Online; accessed 09-Dec-2014].

[13] R. Harrison and C. Charles, "A low-power low-noise cmos amplifier for neural recording applications," Solid-State Circuits, IEEE Journal of, vol. 38, no. 6, pp. 958–965, June 2003.

[14] J. Sauerbrey, D. Schmitt-Landsiedel, and R. Thewes, "A 0.5-v 1- mu;w successive approximation adc," Solid-State Circuits, IEEE Journal of, vol. 38, no. 7, pp. 1261–1265, July 2003.

[15] G. Deuschl and A. Eisen, Recommendations for the Practice of Clinical Neurophysiology: Guidelines of the International Federation of Clinical Neurophysiology, ser. Electroencephalography and clinical neurophysiology: Supplement. Elsevier, 1999. [Online]. Available: http://books.google.co.uk/books?id=HV2qE7BPEkgC

[16] "Duracell 2025 Battery," http://www.duracell.com/en-us/products/button-batteries/duracell-2025-battery, 2014, [Online; accessed 29-Dec-2014].

A lossless data reduction technique for wireless EEG recorders and its use in selective data filtering for seizure monitoring

Chengliang Dai, Christopher Bailey

Abstract— **This paper presents a time-domain based lossless data reduction technique called Log2 Sub-band encoding, which is designed for reducing the size of data recorded on a wireless electroencephalogram (EEG) recorder. A data reduction unit can help to save power from the wireless transceiver and from the storage medium since it allows lower data transmission and read/write rates, and then extends the life time of the battery on the device. Our compression ratio(CR) results show that Log2 Sub-band encoding is comparable and even superior to Huffman coding, a well known entropy encoding method, whilst requiring minimal hardware resource, and it can also be used to extract features from EEG to achieve seizure detection during the compression process. The power consumption when compressing the EEG data is presented to evaluate the system′s overall improvement on its power performance, and our results indicate that a noticeable power saving can be achieved with our technique. The possibility of applying this method to other biomedical signals will also be noted.**

I. Introduction

Electroencephalography (EEG) signals are known to have great potential in multiple applications such as Brain-Computer Interface, Prosthetic Devices, Physical Activity Monitors, etc. Since these signals can reflect various biological changes inside human′s body, caused by either diseases or external stimuli, they are widely used to diagnose or to study certain diseases. Researchers have been using EEG for decades in epilepsy[1], Parkinson's disease[2] and other neurological disorder studies. However, the long process of collecting biomedical data sometimes causes inconvenience to the patients or research subjects since it usually restricts their mobility. The wearable wireless signal recorder is an effective solution to this problem, because it frees users from the inpatient environment and promises better real-time healthcare from caregivers to patients with neurological or other diseases.

Nevertheless, there are some challenges when designing a wireless biomedical signal recorder, and power consumption is no doubt one of the critical issues. To give a better user experience, an EEG recorder has to be small, which will inevitably make the space for a battery even smaller. In order to satisfy some long-term recording requirements, the small battery has to last as long as possible, and that is often difficult to achieve.. On a wireless EEG monitor, the transceiver is always one of the biggest contributors to the power consumption, therefore reducing the transmitted data size might be a significant answer to the power issue.

Moreover, reducing the data size can further help to save power where an onboard storage medium, such as flash memory, is used as high read/write rates may be reduced.

On the other hand, a data reduction unit certainly consumes extra power on a device, such that there are some trade-offs we have to consider.

Log2 Sub-band is a data compression algorithm designed to be used on wireless EEG recorders and possibly on other biomedical signal recorders. This algorithm gives satisfying compression results without consuming too much hardware resource, and it also inherently offers some potential on seizure detection. A comprehensive analysis of the compression performance of Log2 Sub-band is given in[3], and its power analysis and seizure detection potential will be discussed in this paper.

II. Log2 sub-band encoding

Biomedical signals are highly non-Gaussian, non-stationary and non-linear, and with strict clinical requirements, many data compression techniques are ruled out despite the outstanding CRs they can give, and lossless techniques are always desirable. Many well-known compression algorithms have been tested on EEG signals, including entropy encoding[4][5], predictive coding[4][6], transform-based encoding[7], etc. Most of these algorithms are able to produce excellent compression results, but some of them are either lossy or incapable of compressing the signal in real-time, and those very sophisticated techniques might be suitable for compressing real-time EEG signals theoretically, except for their high complexity of design at the system level. For these reasons, Log2 Sub-band is proposed to tackle some of the mentioned issues, and provides lossless compression with minimal hardware.

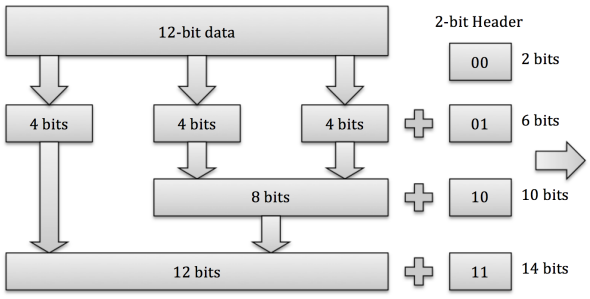

Fig. 1. Log2 Sub-band encoding

On most recorders, biomedical signals are usually digitized into 8-12 bit samples, and Log2 Sub-band is a basic technique that works on time-domain. As shown in Fig. 1, for instance, a channel of EEG signal is converted into 12 bits

per sample after Analog-to-Digital Conversion(ADC), then in this case every data sample is divided into three bands, and each band is a 4-bit nibble, and nibbles in each current sample will be compared with the same part of previous sample. Instead of transmitting or storing all nibbles of every sample, only the bands that are different from previous sample are sent to the transceiver or the memory. A header is given to every sample to indicate how many nibbles are left after this process, and it could be 3 nibbles, 2 nibbles, 1 nibble or none in the best case, so a 2-bit header is required.

The scheme of Log2 Sub-band resembles a simple prediction encoding technique known as Differential Pulse-Code Modulation (DPCM), but with some advantages. DPCM removes short-term redundancies by taking the difference of adjacent data samples, but its performance deteriorates quickly if more rapid changes occur in the signal, and causes the CR it gets from signals like seizure to be sometimes disappointing. Even with Huffman encoding, the CR can be slightly better, but the algorithm is still very sensitive to the nature of input signals, and to effects of scale and level offsets. Comparing with DPCM and Huffman coding, Log2 Sub-band is more adaptive to these biomedical signals, as will be demonstrated later.

III. Experimental EEG data

The EEG signals used in data reduction simulations are from two sources: (a) research on non-linear deterministic patterns of brain electrical signals at the University of Bonn[8] and (b) research on seizure detection at the Massachusetts Institute of Technology[9].

The data from (a) were originally categorised into different groups, and each group contains one hundred data segments of 23.6 sec duration of EEG signals. The data segments were selected and cut out from continuous multichannel EEG recordings after visual inspection and removal for artifacts such as EOG. Three groups of data from this source are chosen for the simulation, which are from five healthy people(eyes closed), five epilepsy patients during seizure-free time and seizure signal.

The EEG from (b) were collected from 24 different subjects, and signals are mostly divided into many uncategorised one-hour long data segments. Two one-hour segments from one subject that contain both non-seizure signals and a seizure event in each segment are used in the following test.

The EEG from both sources were converted into 12-bit, and retained sampling rates of 173.6 Hz and 256 Hz respectively.

IV. Compression Ratio Results

The two-stage DPCM and Huffman coding technique mentioned earlier is used here to set a benchmark for Log2 Sub-band, and the results are shown in Table I. CR is given as:

$$CR = Size_{Original}/Size_{Compressed} \quad (1)$$

The Log2 Sub-band compresses the data with a three-band '4,4,4' encoder as mentioned in our earlier example, due to the 12-bit width of target signals. The results indicate

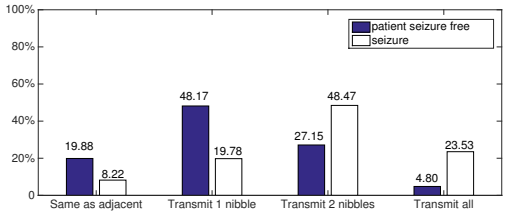

Fig. 2. Seizure-free/seizure Band Distribution[8]

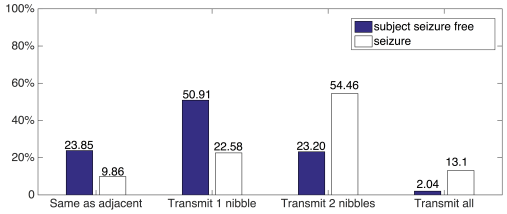

Fig. 3. Seizure-free/seizure Band Distribution of Mixed Signal 1[9]

TABLE I
Compression Ratio results

Data Type	CR of DPCM +Huffman	CR of Log2 Sub-band
Healthy people with eyes closed[8]	1.82	1.94
Patients′ seizure free[8]	2.20	2.58
Seizure[8]	1.51	1.66
Mixed Signal 1[9]	2.13	2.77
Mixed Signal 2[9]	1.91	2.32

that Log2 Sub-band clearly has a superior compression performance, and even though both the techniques give lower CRs whilst compressing seizure signals, unlike Huffman coding′s scheme, which requires a pre-generated codebook based on the frequency of occurrence of every sample, Log2 Sub-band is more adaptive and flexible.

During the compression process, different signals show various behaviours, and signals that can be largely compressed are clearly more likely to be represented with less bands under Log2 Sub-band scheme. Therefore, the distribution of the occurring frequency of each band can be used to describe some of the features of EEG signals. The band distributions of patients′s seizure-free and seizure signal from[8][9] are given in Fig. 2 and Fig. 3 respectively.

According to Fig. 2 and Fig. 3, seizure signals are mostly encoded with two or three bands whilst seizure-free signals are encoded with one or two bands. This will be shown to be useful and will be used to identify seizure in the next section.

V. Seizure detection

Most seizure detection systems involve two basic stages, and the first is to extract the features from the EEG signal, whilst the second is to set a threshold to identify the seizure based on the extracted features. As previous noted, Log2

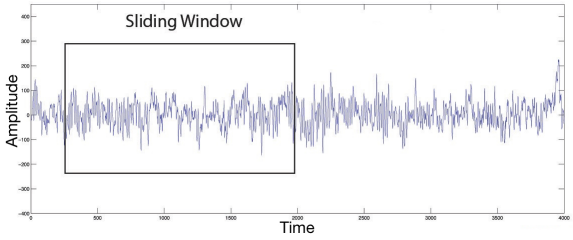

Fig. 4. Healthy Human EEG

Sub-band provides a band distribution whilst compressing each sample, and we suggest that this distribution is one of the signal features that can be applied to seizure detection. To carry out real-time seizure detection, a sliding window scheme(Fig. 4) is proposed.

A sliding window allows Log2 Sub-band encoding to analyse the band distribution of samples within the given window during the compression, and if the distribution within the window matches the seizure criteria, then the current segment can be considered as the seizure candidate. Various methods can be used to set the threshold or criteria with the help of the band distribution, and one straight-forward method will be presented here. A threshold will be set based on the total probabilities of a sample encoded with zero and one nibble, and this method should be rather effective according to Fig. 2 and Fig. 3.

EEG signals from datasets (a) and (b) are used, and data from (a) are pre-categorised, hence 20 seizure data segments were randomly picked and then inserted into patients′ seizure free data to create a simulated signal for seizure detection test. The detection performance of Log2 Sub-band on the simulated data is measured in terms of sensitivity(SEN), false positive rate(FPR) and speed of detection(SoD), which are percentage of successfully detected seizure, percentage of false positive detection, and average time cost to identify the seizure. The window size is set to 100 samples, which is typically around 580 ms. The detection process is repeated with different threshold settings from 30 to 45, and if the percentage of samples encoded with zero and one nibble in a current window falls below the given threshold, samples will be identified as seizure cases. The results when threshold is set to 30, 35, 40 and 45 are given in Table II.

The EEG signals from[9] are uncategorised and therefore are used as the clinical signal input in the experiment. There is only one seizure event in each one-hour′s signal, and the first mixed signal was used for feature extraction(Fig. 3), then based on the acquired feature, the threshold was set for the second one-hour signal to detect the seizure. The performance of Log2 Sub-band derived detection in this case is measured merely with the SoD under difference threshold settings, and the highest SoD achieved is 0.9 sec.

The algorithm we present here is of course somewhat less sophisticated than more advanced techniques such as[10][11]. However those techniques often use computationally expensive methods, with an implication for high power cost when used continuously. We envisage the Log2 band signature technique being useful as a pre-filter, identifying possible seizure candidates with low power and then triggering a more sophisticated analysis on demand, or recording for later analysis. In many epilepsy diagnosis studies, researchers are more interested in seizure signals, and discontinuous recording has been more often applied as a result[12], and our technique is a good option for these studies since a balance between power cost and seizure detection accuracy may be facilitated by this technique with minimal hardware cost.

VI. Power Analysis

The hardware design for this case of three-band '4,4,4' encoder is conducted in 65-nm CMOS technology. We have modelled this and produced trial layouts, and estimated power whilst compressing the EEG data. The results will be given in next section. A silicon test chip is currently being fabricated.

A. Power Saving from Data Compression

The extra power consumed whilst compressing the data[8] is given in Table III. The total power consists of static and dynamic power. Obviously, static power dominates the total power consumption with the typical EEG data sample rates, which in this case is 1 kHz, but dynamic power will increase when more channels of signals are needed.

According to the results, it is quite clear that the circuit consumes more dynamic power when compressing seizure signals, as compare to non-seizure signals. A regression analysis between CR and the dynamic power is given in Fig. 5.

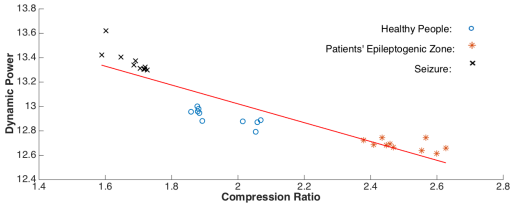

Fig. 5. regression analysis

In order to estimated the possible power saving after data compression, a simplified EEG recorder that only contains amplifier(s), ADC(s), a transceiver and a data compression unit is given[13]. The system power consumption P_{sys} can be roughly modeled as[14]:

$$P_{sys} = P_{amp} + P_{ADC} + P_{comp} + (D * P_{Tx})/CR \quad (2)$$

where P_{Amp} and P_{ADC} are the power consumption of amplifier(s) and ADC(s), and P_{Tx} and P_{Comp} are the power consumption of the transceiver and the data compressor.

Furthermore D is duty cycle rate which is defined such that:

$$D = R_{Orig}/R_{Tx}. \quad (3)$$

where R_{Orig} is the required data rate before compression, and R_{Tx} is the maximum data rate of transceiver.

TABLE II
SEIZURE DETECTION RESULTS ON[8]

Threshold	30	35	40	45
SEN	95%	100%	100%	100%
FPR	0	9.1%	31.0%	37.5%
SoD(sec)	2.9	2.5	1.4	0.6

TABLE III
POWER CONSUMPTION OF LOG2 SUB-BAND

Data Type	CR	Static Power(nW)	Dynamic Power(nW)	Total
Healthy people with eyes closed	1.94	147.23	12.92	160.15
Epileptogenic zone	2.58	147.29	12.68	159.97
Seizure	1.66	147.29	13.37	160.66

TABLE IV
POWER SAVING

Data Type	Power Saving(μW)	Extra Circuit Cost(nW)	Total Saving(μW)
Healthy people with eyes closed	127.91	160.15	127.75
Epileptogenic zone	161.67	159.97	161.51
Seizure	104.96	160.66	104.80

A typical 22mW off-the-shelf Bluetooth transceiver nRF8001 is chosen[15] in this case, and the maximum data rate is 1Mbps, and therefore we assume that to transmit 1 bit of data, it consumes 22nJ′s power.

Combining all the known quantities with (2), we derived the estimated power saving results in Table IV. We see for instance, if we transmit one channel of healthy subject EEG signals, a system with the proposed data compression unit could save 127.91μW′s power from the transceiver with only an extra cost of 160.15 nW. More details of this analysis is given in[14].

B. Power Saving from Seizure Detection

Saving power from conducting seizure detection is based on the assumption that only the seizure signals are of interest, therefore by sending or storing seizure signals only, or applying lossy compression technique on non-seizure signals, a great power saving can be achieved and also a reduction in data storage or transmission. The one-hour signal we used to test seizure detection contains 90 sec seizure signal, and with only 90 sec′s data are transmitted, a total 90% transceiver power could be saved even if that data is uncompressed. If we also use lossless Log2 Sub-band compression on these seizure data segments, then we can project power savings of over 94% (assuming 10% of data is seizure and CR = 1.66). Even though some false positive seizure signals might be encountered, Log2 Sub-band could still help to save a great amount of power.

VII. CONCLUSION

Based on all the results presented, Log2 Sub-band encoding is clearly a promising data reduction candidate for wireless EEG recording devices. However it also has potential on seizure detection. Our on-going research also suggests that our encoder could be applied to other biomedical signals such as electromyogram(EMG) and electrocardiogram(ECG) after making some slight changes on its scheme such as altering the width of each band to attune to new signals. Our research team have recently prototyped the described compression scheme in 65nm test chip which is currently being fabricated for Apr, 2015 delivery.

VIII. ACKNOWLEDGEMENTS

The 65nm test chip is being fabricated as part of a UK Technology Strategy Board founded project in collaboration with Cybula Ltd, grant number 26172-182148.

REFERENCES

[1] G. E. Chatrian, C.-M. Shaw, and H. Leffman, "The significance of periodic lateralized epileptiform discharges in eeg: an electrographic, clinical and pathological study," Electroencephalography and clinical neurophysiology, vol. 17, no. 2, pp. 177–193, 1964.
[2] R. Soikkeli, J. Partanen, H. Soininen, A. Pääkkönen, and P. Riekkinen, "Slowing of eeg in parkinson's disease," Electroencephalography and clinical neurophysiology, vol. 79, no. 3, pp. 159–165, 1991.
[3] C. Dai and C. Bailey, "A time-domain based lossless data compression technique for wireless wearable biometric devices," in SENSORCOMM 2013, The Seventh International Conference on Sensor Technologies and Applications, 2013, pp. 104–107.
[4] G. Antoniol and P. Tonella, "EEG data compression techniques." IEEE transactions on bio-medical engineering, vol. 44, no. 2, pp. 105–14, Feb. 1997.
[5] D. O'Shea and R. McSweeney, "Efficient Implementation of Arithmetic Compression for EEG," IET Proceedings of Irish, 2011.
[6] N. Sriraam, "A high-performance lossless compression scheme for EEG signals using wavelet transform and neural network predictors." International journal of telemedicine and applications, vol. 2012, p. 302581, Jan. 2012.
[7] W. Optimization, L. Brechet, M.-f. Lucas, C. Doncarli, and D. Farina, "Compression of Biomedical Signals With Mother Wavelet Packet Selection," IEEE transactions on bio-medical engineering, vol. 54, no. 12, pp. 2186–2192, 2007.
[8] R. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Physical Review E, vol. 64, no. 6, p. 061907, Nov. 2001.
[9] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 975–982.
[10] T. Kalayci, O. Ozdamar, and N. Erdol, "The use of wavelet transform as a preprocessor for the neural network detection of EEG spikes," Proceedings of SOUTHEASTCON '94, pp. 1–3, 1994.
[11] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," Journal of Neuroscience Methods, vol. 123, no. 1, pp. 69–87, 2003.
[12] P. Valenti, E. Cazamajou, M. Scarpettini, A. Aizemberg, W. Silva, and S. Kochen, "Automatic detection of interictal spikes using data mining models," Journal of Neuroscience Methods, vol. 150, no. 1, pp. 105–110, 2006.
[13] D. C. Yates and E. Rodriguez-Villegas, "A Key Power Trade-off in Wireless EEG Headset Design," 2007 3rd International IEEE/EMBS Conference on Neural Engineering, pp. 453–456, May 2007.
[14] C. Dai and C. Bailey, "Power analysis of a lossless data compression technique for wireless wearable biometric devices," submitted.
[15] "nRF8001 Bluetooth® low energy Connectivity IC," http://www.nordicsemi.com/eng/Products/Bluetooth-Smart-Bluetooth-low-energy/nRF8001, 2014, [Online; accessed 09-Dec-2014].

Appendix B

Presented posters

Two posters have been presented to the following conference and workshop.

C. Dai and C. Bailey, "Trade-offs of EEG data reduction on wearable device," in *Building Bridges to Build Brains Meeting*, Nov 2012.

C. Dai and C. Bailey, "A lossless data reduction technique for wireless eeg recorders and its use in selective data filtering for seizure monitoring," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, Aug 2015.

Trade-offs of EEG data reduction on wearable device

Chengliang Dai, Dr Christopher Crispin-Bailey

Department of Computer Science, University of York

Introduction

Electroencephalogram (EEG) has been proved as a noninvasive and reliable way for collecting a person's brain wave data which can be used in numerous areas such as epilepsy diagnosis [1], brain-computer interface (BCI), etc.

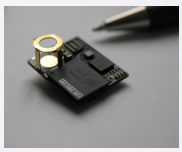

Figure 1. A wearable EEG recorder with a 4-channel input designed by Advanced Computer Architecture Group in University of York

Wearable EEG recorder has drawn a lot of attention in recent years since it saves much more time and money for researchers comparing to traditional data collecting equipment. A wearable EEG recorder can be as small as a 10p coin (Figure 1) so that the object of recording will not even notice it. Such device is preferable to be able to transmit the data via wireless transceiver to achieve the real–time monitoring and data analysis. However, wireless transmission apparently shortens device's running time because of the extra power consumed by the transceiver. The most obvious way to extend running time is to compress the EEG data before transmission, and with less data to be transmitted, the overall power consumption will be largely reduced (Figure 3).

Figure 2. Current solution of reading in the EEG data, a docking station.

Meanwhile, adding on a data compressor to the device also increases the complexity of the system, and compression unit itself is one of the contributors to recorder's power consumption, so exploring a best set of trade-offs between compression ratio (CR) and the power consumption of compressing the data will be helpful to solve all these problems, and that is the purpose of doing this research.

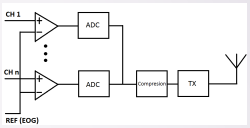

Figure 3. A simplified architecture of wearable EEG recorder

Experiments

Several compression method candidates have been simulated so far including most well–known lossless Huffman coding [2], a lossy encoding technique called Discrete Wavelet Transform (DWT) [3], as well as a lossless log2 sub-band algorithm we designed which will be introduced in detail below.

Currently, EEG signals are usually quantized to no more than 16 bits in most experiments, and take this resolution as an example, in log2 sub-band algorithm scheme, we first compare the first 4 bits from 16 bits of the current sample with the same part in previous sample, and if they are identical, next 4 bits will be further compared, and the transceiver only transmits the bits that are different from previous sample. A flag of 2 bits is introduced to indicate how many nibbles are transmitted in a sample (Figure 4), which can be used to decompress the data.

The dataset used for simulation come from an EEG experiment taken in University Hospital of Bonn [4], and the sampling rate of data is 173.61 Hz, and are quantized into 10 bits for the simulation.

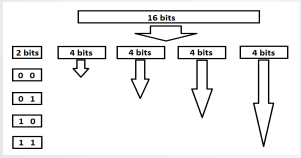

Figure 4. '00' indicates that 4 nibbles of this sample are transmitted, '01' indicates 3, '10' indicates 2, and '11' indicates 1 nibble.

Results

The CR is defined as

$$\text{CR} = \frac{size\ of\ compressed\ data}{size\ of\ original\ data}$$

The data set for testing the Huffman coding is divided into 2 subsets, and each contains 10 files with 4096 data samples in every file. One subset is used to generate the Huffman tree (dictionary) so that the other subset can be encoded based on the previous dictionary. The result of log2 sub-band algorithm will also shown in the Figure 5.

According to [5], the DC-bias of EEG signal degrades the inter-channel decorrelation performance, and it

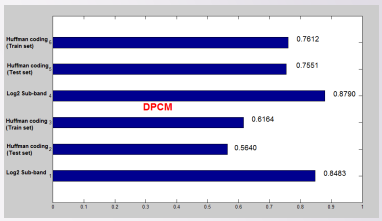

Figure 5. Compression Ratio of presented techniques

can be simply removed by using the differential pulse-code modulation (DPCM) which calculates the difference between adjacent samples. The performance of Huffman coding and log2 sub-band algorithm are both increased after applying DPCM (Figure 5).

As for DWT, it expresses a signal as a weighted sum of basis functions, and these basis are derived from dilated and translated versions of a function which is called mother wavelet [6]. So the original signal can be defined with coefficients of the set of basis functions, and this process causes some loss of the original signals, but as long as a optimal mother wavelet is chosen, the discrepancies between original and reconstructed signals are still acceptable. One file with 4096 data samples is used for simulation, and the mother wavelet applied are Haar (Figure 6), Daubechies 10 (Figure 7), and Coiflet (Figure 8) [3]. The reconstructed signal is recovered with 50 percent coefficients of original signal, so the CR is 0.5 in this case, and as they are shown in the figures, using Haar as mother wavelet gives worst reconstruction quality, and the results of using Daubechies 10 and Coiflet are close.

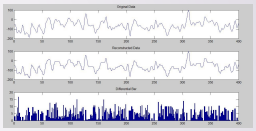

Figure 6. Reconstructed with Harr, and from top to bottom are original signal, reconstructed signal and differentials

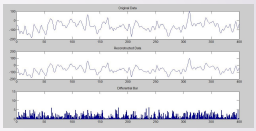

Figure 7. Reconstructed with Daubechies 10

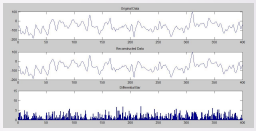

Figure 8. Reconstructed with Coiflet

Conclusions

For a simplified wearable EEG recorder, which is only consisted with amplifiers, ADCs, and the transceiver, and power consumption of DSPs and the processor are neglected, the overall power consumption of the recorder can be given as

$$P_{sys} = P_{Amp} + P_{ADC} + P_{Comp} + CR * P_{Tx}$$

Where P_{Amp} and P_{ADC} are the power consumption of amplifiers and ADCs, and P_{Tx} and P_{Comp} are the power consumption of transceiver and doing data compression. Then the power reduction is

$$(1 - CR) * P_{Tx}$$

By implementing a popular off-the-shelf GFSK transceiver nRF24L01 [7](TX power consumption is 23 mW), the power saving for using Huffman coding is at least 5.5mW, 11.5mW for using DWT, and 3.45mW for using log2 sub-band algorithm. As a result, these are power budgets for future hardware implementation of data compression techniques, and any technique requires lower power than its saving is considered as a proper method which leads to a longer working time of wearable EEG recorder. Furthermore, possible data compression techniques may also applicable to other biomedical data recorders as these signals have many similarities.

Bibliograph

[1] E. Waterhouse, "New horizons in ambulatory electroencephalography," IEEE Eng. Med. Biol. Mag., vol. 22, no. 3, pp. 74–80, 2003.

[2] G. Antoniol and P. Tonella, "EEG data compression techniques," *IEEE Trans. Biomed. Eng.*, vol. 44, pp. 105–114, 1997.

[3] S. Mallat, A Wavelet Tour of Signal Processing, 2nd ed. San Diego, CA: Academic, 1999.

[4] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke,, P. David, & C. E. Elger, "Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Phys. Rev. E, vol. 64, no. 6, pp. (061907)1–8, 2001.

[5] Y. Wongsawat, S. Oraintara, T. Tanaka, and K. R. Rao, "Lossless multi-channel EEG compression," in Proceedings of IEEE Int. Symp. on Circ. and Syst. (ISCAS '06), pp. 1611–1614, Island of Kos, Greece, 2006.

[6] A. N. Akansu and M. J. T. Smith, Subband and Wavelet Transforms: Design and Applications, Kluwer Academic Publishers, 1995.

[7] http://www.nordicsemi.com/eng/Products/2.4GHz-RF/nRF24L01

THE UNIVERSITY of York

Contact : Chengliang Dai
cd633@york.ac.uk
Advanced Computer Architecture Group,
Department of Computer Science,
University of York, Heslington, York, YO10 5GH, UK

Figure B.1: Trade-offs of EEG data reduction on wearable device

A LOSSLESS DATA REDUCTION TECHNIQUE FOR WIRELESS EEG RECORDERS AND ITS USE IN SELECTIVE DATA FILTERING FOR SEIZURE MONITORING

Chengliang Dai, Christopher Bailey

Advanced Computer Architecture Group, Department of Computer Science, University of York, UK

EMBC 2015 MILANO

Introduction

EEG signals are known to have great potential in multiple applications such as Brain-Computer Interface, Prosthetic Devices, Physical Activity Monitors, etc. Since these signals can reflect various biological changes inside human's body, caused by either diseases or external stimuli, they are widely used to diagnose or to study certain diseases. Researchers have been using EEG for decades in epilepsy, Parkinson's disease and other neurological disorder studies. However, the long process of collecting biomedical data sometimes causes inconvenience to the patients or research subjects as it usually restricts their mobility. The wearable wireless biomedical signal recorder is one straightforward solution to these problems but with a drawback, which is the battery lifetime. Signal recorders like the one shown in Figure 1 don't have enough space for the battery, therefore the operating time is sometimes disappointing.

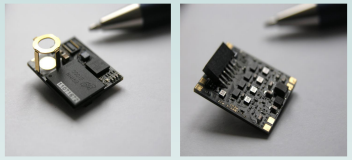

Figure 1. NAT-1 multi-functional biomedical signal recorder, Cybula Ltd

A simplified hardware model of an EEG signal recorder is introduced in [1] and it is shown in Figure 2, and as the transceiver always consumes a big proportion of the system power, introducing a data compressor that reduces the data sent or stored will certainly help to save the power and extend the battery lifetime. A data compression unit also brings a set of design trade-offs that needs to be considered, such as

- Compression ratio vs Algorithm complexity
- Power consumed vs Power saved, etc

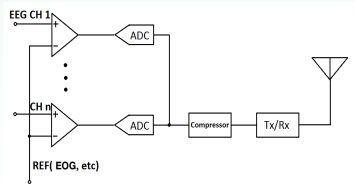

Figure 2. Simplified wireless EEG signal recorder

Methodology

The research flow is given in Figure 3, and to achieve power saving, a number of data compression/reduction techniques are considered for wearable wireless EEG recorders, and the differential pulse code modulation (DPCM) and the Huffman coding are chosen as a 2-stage data compression process for further evaluation since this process is simple, lossless and reversible, and it also gives promising performance according to literatures.

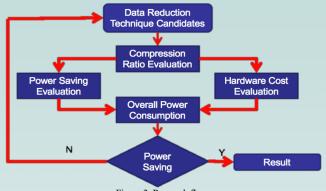

Figure 3. Research flow

An improved technique called Log2 sub-band that does not require prior knowledge of the original data like the Huffman coding does is then designed based on the mentioned 2-stage technique. The scheme of Log2 sub-band is given in Figure 4 assuming the original data is digitized into 12-bit. Every data sample is divided into three bands, and each band is a 4-bit nibble, and nibbles in each current sample will be compared with the same part of previous sample. Instead of transmitting or storing all nibbles of every sample, only the bands that are different from previous sample are sent to the transceiver or the memory. A header is given to every sample to indicate how many nibbles are left after this process, and it could be 3 nibbles, 2 nibbles, 1 nibble or none in the best case, so a 2-bit header is required.

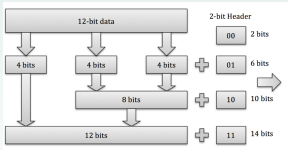

Figure 4. Log2 sub-band compression

Results

The datasets used for evaluating the compression performance of two techniques are from two sources: (a) research on brain electrical signals at the University of Bonn[2] and (b) research on seizure detection at the Massachusetts Institute of Technology[3]. The data from (a) are pre-categorized, and the data from (b) are the mixture of signals during seizure and seizure-free periods.

The compression results of two techniques are given in Table 1, and the compression ratio is defined as

$$CR = Size_{Original}/Size_{Compressed}$$

Data Type	CR of DPCM +Huffman	CR of Log2 Sub-band
Healthy people with eyes closed[8]	1.82	1.94
Patients' seizure free[8]	2.20	2.58
Seizure[8]	1.51	1.66
Mixed Signal 1[9]	2.13	2.77
Mixed Signal 2[9]	1.91	2.32

Table 1. Compression results

It is clear that Log2 sub-band yields slightly better results comparing with the 2-stage technique when compressing all sorts of signals. Both techniques perform better when compressing stable signals like EEG from health people.

Data Type	Power Saving(μW)	Extra Circuit Cost(nW)	Total Saving(μW)
Healthy people with eyes closed	127.91	160.15	127.75
Epileptogenic zone	161.67	159.97	161.51
Seizure	104.96	160.66	104.80

Table 2. Power results

The power analysis of the Log2 sub-band compressor is based on the hardware model given in Figure 2. An off-the-shelf Bluetooth transceiver nRF8001 is chosen for this model, and the power results are listed in Table 2. The results shows that the Log2 sub-band compressor can bring a significant power saving to a wearable wireless device with a very low extra power cost, and it tends to consumes more power when compressing signals with rapid changes like seizure signals. A regression analysis between the compression ratio the Log2 sub-band achieved and the power it consumed is shown in Figure 5.

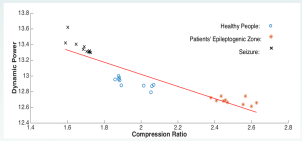

Figure 5. Regression analysis

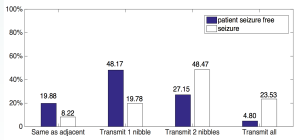

Figure 6. Seizure-free/seizure band distribution

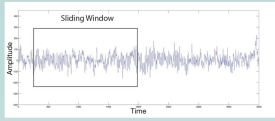

Figure 7. Sliding window

During the process of compressing the simulated signals from [2] with Log2 sub-band, a distribution of the band can be acquired as it is given in Figure 6, and the distribution shows that seizure signals need more nibbles to send compare to seizure free EEG signals. This difference allows a sliding window scheme as it is shown in Figure 7 to be used to identify the seizure event, and therefore the scheme can further reduces the data size if seizure-free data are not wanted in some cases.

Threshold	30	35	40	45
Sensitivity	95%	100%	100%	100%
FPR	0	9.1%	31.0%	37.5%
SoD(sec)	2.9	2.5	1.4	0.6

Table 3. Seizure detection results

Conclusions

Based on all the results acquired, Log2 Sub-band encoding is clearly a promising data reduction candidate for wireless EEG recording devices. However it also has potential on seizure detection. Our on-going research also suggests that our encoder could be applied to other biomedical signals such as electromyogram(EMG) and electrocardiogram(ECG) after making some slight changes on its scheme such as altering the width of each band to attune to new signals. Our research team have recently prototyped the described compression scheme in 65nm test chip which has been fabricated in Apr, 2015.

References

[1] D. C. Yates and E. Rodriguez-Villegas, "A key power trade-off in wireless EEG head- set design," Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering, pp. 453–456, May 2007.

[2] R. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Physical Review E, vol. 64, no. 6, p. 061907, Nov. 2001.

[3] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 975–982.

Acknowledgements

The 65nm test chip is being fabricated as part of a UK Technology Strategy Board founded project in collaboration with Cybula Ltd, grant number 26172-182148.

Contacts:
Chengliang Dai, cd633@york.ac.uk
Christopher Bailey, christopher.crispin-bailey@york.ac.uk

THE UNIVERSITY of York

Figure B.2: A lossless data reduction technique for wireless eeg recorders and its use in selective data filtering for seizure monitoring

Appendix C

Log2 sub-band encoding testing code

C1 shows the algorithm of the Log2 sub-band written in MATLAB that has been used for software simulation in this research. C.2 gives the hardware design of the Log2 sub-band modelled in VHDL and the test-bench prepared ub Verilog that were used to test the design.

C.1 MATLAB code

```
function [compdatasize,transmit0, transmit1, transmit2,transmit3]=Nibble(data) %input binary data
compdata = bitget(data(1),12:-1:1);
transmit0=0;
transmit1=0;
transmit2=0;
transmit3=0;
parfor i=2:length(data);
if not(xor(bitget(data(i),12:-1:9),bitget(data(i-1),12:-1:9)))
    if not(xor(bitget(data(i),8:-1:5),bitget(data(i-1),8:-1:5)))
        if not(xor(bitget(data(i),4:-1:1),bitget(data(i-1),4:-1:1)))
            transmit0=transmit0+1; % transmit 0 band
        else
            compdata=[compdata,bitget(data(i),4:-1:1)];
            transmit1=transmit1+1; % transmit 1 band
        end
    else
        compdata=[compdata,bitget(data(i),8:-1:1)];
        transmit2=transmit2+1; % transmit 2 bands
    end
```

```
else
    compdata=[compdata, bitget(data(i),12:-1:1)];
    transmit3=transmit3+1; % transmit 3 bands
end
end
compdatasize=length(compdata) + 2 * length(data);
end
```

C.2 VHDL+Verilog code

```
----------------------------------------------------------------------------------
-- Company:
-- Engineer:
--
-- Create Date:     19:11:07 11/21/2014
-- Design Name:
-- Module Name:     Log2 - Behavioral
-- Project Name:
-- Target Devices:
-- Tool versions:
-- Description:
--
-- Dependencies:
--
-- Revision:
-- Revision 0.01 - File Created
-- Additional Comments:
--
----------------------------------------------------------------------------------
```

```
library IEEE;
use IEEE.STD_LOGIC_1164.ALL;

-- Uncomment the following library declaration if using
-- arithmetic functions with Signed or Unsigned values
use IEEE.NUMERIC_STD.ALL;

-- Uncomment the following library declaration if instantiating
-- any Xilinx primitives in this code.
--library UNISIM;
--use UNISIM.VComponents.all;

entity Log2 is
    Port ( CLK : in  STD_LOGIC;
           RESET : in  STD_LOGIC;
           DataIN : in  STD_LOGIC_VECTOR (11 downto 0);
           DataOUT : out  STD_LOGIC;
           DataRDY : in  STD_LOGIC;
           SYNC : out  STD_LOGIC);
end Log2;
```

```
architecture Behavioral of Log2 is

SIGNAL COUNT : STD_LOGIC_VECTOR(3 DOWNTO 0);
SIGNAL NewVal : STD_LOGIC_VECTOR(11 DOWNTO 0);
SIGNAL OldVal : STD_LOGIC_VECTOR(11 DOWNTO 0);
SIGNAL Shifter : STD_LOGIC_VECTOR(15 DOWNTO 0);

begin

        PROCESS(CLK,RESET)
        BEGIN

                IF(RESET='0') THEN

                        COUNT <= X"0";

                ELSIF(rising_edge(CLK)) THEN

                        COUNT <= std_logic_vector(signed(COUNT) + 1 );

                END IF;
```

```
END PROCESS;

PROCESS(CLK,RESET)
BEGIN

        IF(RESET='0') THEN

                        NewVal <= X"000";
                        OldVal <= X"FFF";
                        DataOUT <= '1';

                        SYNC   <= '0';
                        Shifter <= "1" & "0000" & "0000" & "0000" & "11" & "0";

        ELSIF(rising_edge(CLK)) THEN

                        IF(COUNT = X"0") THEN

                                -- flag state ----------------
```

```
            SYNC <='1';

-- update storage registers --

            OldVal <= NewVal;
            NewVal <= DataIN;

-- do comparison ---------------

            IF(OldVal=NewVal) THEN

                  Shifter <= "1111" & "1111" & "1111" & "1" & "00
                     " & "0";

            ELSIF(OldVal(11 DOWNTO 4) = NewVal(11 DOWNTO 4)) THEN

                  Shifter <= "1" & "1111" & "1111" & NewVal(3
                     DOWNTO 0) & "01" & "0";

            ELSIF(OldVal(11 DOWNTO 8) = NewVal(11 DOWNTO 8)) THEN
```

```
                Shifter <= "1" & "1111" & NewVal(7 DOWNTO 0) &
                    "10" & "0";

            ELSE

                Shifter <= "1" & NewVal & "11" & "0";

            END IF;

            DataOUT <= '0';
        ELSE

            SYNC <='0';
            Shifter <= "1" & Shifter(15 DOWNTO 1);
            DataOUT <= Shifter(1);

        END IF;

    END IF;

END PROCESS;
```

```
end Behavioral;
```

```
`timescale 1ns / 1ps

////////////////////////////////////////////////////////////////////////////////
// Company:
// Engineer:
//
// Create Date:   19:19:34 11/21/2014
// Design Name:   Log2
// Module Name:   C:/Documents and Settings/chrisb/Desktop/Log2/Log2Tester.v
// Project Name:  Log2
// Target Device:
// Tool versions:
// Description:
//
// Verilog Test Fixture created by ISE for module: Log2
//
// Dependencies:
//
// Revision:
// Revision 0.01 - File Created
```

```
// Additional Comments:
//
//////////////////////////////////////////////////////////////////////////////////

module Log2Tester;

        // Inputs
        reg CLK;
        reg RESET;
        reg [11:0] DataIN;
        reg DataRDY;

        reg [15:0]RECIEVED;
        reg [11:0]RVAL;

        // Outputs
        wire DataOUT;
        wire SYNC;

        // Instantiate the Unit Under Test (UUT)
        Log2 uut (
```

```
        .CLK(CLK),
        .RESET(RESET),
        .DataIN(DataIN),
        .DataOUT(DataOUT),
        .DataRDY(DataRDY),
        .SYNC(SYNC)
);

integer     data_file     ; // file handle
integer DataValue;
integer DataCount;

localparam TCLK   = 20;
localparam TSETUP = 16;


initial begin

        // init files //

                $dumpfile("F10.vcd");
```

```
        data_file = $fopen("../data/F10.dat", "r");

        if (data_file == 0) begin
                $display("-+- ERROR data_file Not Found");
                $finish;
        end


// Initialize Inputs
CLK = 1;
RESET = 0;
DataIN = 12'h000;
DataRDY = 0;

// Wait 100 ns for global reset to finish
#TSETUP;
#(TCLK*4);

RESET = 1;
```

```
        #(TCLK*50000);

        // Add stimulus here

end

        always
        begin
                #(TCLK/2);
                        CLK = ! CLK;
                        $dumpvars;
        end

        always @(posedge(CLK))
        begin
                if(RESET == 0)
                begin
                        RECIEVED=16'hFFFF;
                        RVAL = 12'h000;
                        DataCount = 0;
                end
```

```
if (SYNC==1)
begin

        $display ("--- %dns ----------------------------------" ,$stime);
        $display ("DATA RECIEVED = %d" ,RECIEVED);

        $display ("HEADER = %2b"  ,  RECIEVED[2:1]);
        $display ("NUMERIC = %3xh --- BINARY  = [%b][%12b][%2b][%b] ",  RECIEVED
            [14:3],RECIEVED[15],RECIEVED[14:3],RECIEVED[2:1],RECIEVED[0]);
        $display ("----------------------------------------");

        if (RECIEVED[2:1]==2'b00)
        begin
        end

        if (RECIEVED[2:1]==2'b01)
        begin

                RVAL[3:0] = RECIEVED[6:3];
```

```
        end

        if(RECIEVED[2:1]==2'b10)
        begin

                RVAL[7:0] = RECIEVED[10:3];

        end

        if(RECIEVED[2:1]==2'b11)
        begin

                RVAL[11:0] = RECIEVED[14:3];

        end

        $display("RVAL = %xh -- %d.dec",RVAL,RVAL);
        $display("-----------------------------------");
end

RECIEVED = { DataOUT, RECIEVED[15:1] };
```

```
end

always @(negedge(SYNC))
begin



        if($feof(data_file))
        begin
                $display("-- DATA RUN COMPLETED, %d values tested --",DataCount);
                $stop();
        end

        $fscanf(data_file, "%b\n", DataValue);

        $display("%dns -- data value %d =   [%d] ",$stime, DataCount, DataValue);

        DataIN = DataValue;

        DataCount = DataCount +1;
```

```
    end

endmodule
```

Appendix D

Bit shifter codes for Huffman coding

This part gives the hardware design of bit shifter prepared for DPCM+Huffman coding.

```
‘timescale 1ns / 1ps

////////////////////////////////////////////////////////////////////////////////
// Company:
// Engineer:
//
// Create Date:   19:19:34 11/21/2014
// Design Name:   Log2
// Module Name:   C:/Documents and Settings/chrisb/Desktop/Log2/Log2Tester.v
// Project Name:  Log2
// Target Device:
// Tool versions:
// Description:
//
// Verilog Test Fixture created by ISE for module: Log2
//
// Dependencies:
//
// Revision:
// Revision 0.01 - File Created
// Additional Comments:
```

```
//
////////////////////////////////////////////////////////////////////////////////

module Huf2Tester;

        // Inputs
        reg CLK;
        reg RESET;
        reg [12:0] DataIN;
        reg DataRDY;

        reg [19:0]RECIEVED;
        reg [11:0]RVAL;

        // Outputs
        wire DataOUT;
        wire SYNC;
        wire [4:0]BC;
```

```
// Instantiate the Unit Under Test (UUT)
Huffman uut (
        .CLK(CLK),
        .RESET(RESET),
        .DataIN(DataIN),
        .DataOUT(DataOUT),
        .DataRDY(DataRDY),
        .BC(BC),
        .SYNC(SYNC)
);

integer     data_file     ; // file handle
integer DataValue;
integer DataCount;

localparam TCLK   = 20;
localparam TSETUP = 16;


initial begin
```

```
// init files //

        $dumpfile("healthy.vcd");

        data_file = $fopen("../data/healthy_x.dat", "r");

        if (data_file == 0) begin
                $display("-+- ERROR data_file Not Found");
                $finish;
        end


// Initialize Inputs
CLK = 1;
RESET = 0;
DataIN = 12'h000;
DataRDY = 0;

// Wait 100 ns for global reset to finish
#TSETUP;
#(TCLK*4);
```

```
        RESET = 1;

        $display("---- RESET COMPLETED ----------");

        #(TCLK*50000);

        // Add stimulus here

end

        always
        begin
                #(TCLK/2);
                        CLK = ! CLK;
                        $dumpvars;
        end

        always @(posedge(CLK))
        begin
                if(RESET == 0)
```

```
                begin
                        RECIEVED=19'h03FFFF;

                        DataCount = 0;
                end


$display("-(C%02d)(SH%20b)-<%8x>--[%20b]--S[%d]--BC[%d]--",uut.COUNT,uut.Shifter[19:0],DataIN,RECIEVED,
    SYNC,BC);

                if(SYNC==1)
                begin

                        $display("--- %dns ----------------------------",$stime);
                        $display("DATA RECIEVED = <%xh><%bb>, STARTBIT[%d],CODELEN = %d",
                            RECIEVED[19:1],RECIEVED[19:1],RECIEVED[0],BC);



                                $display("----------------------------------------");
```

```
                RECIEVED = -1;
            end

            RECIEVED = { DataOUT, RECIEVED[19:1] };

        end

        always @(posedge(SYNC))
        begin



            if($feof(data_file))
            begin
                $display("-- DATA RUN COMPLETED, %d values tested --",DataCount);
                $stop();
            end

            $fscanf(data_file, "%b\n", DataValue);
```

```
            $display("%dns -- Input from file value %b<%d>, values read = %d",$stime,
                DataValue,DataValue,DataCount);

            DataIN = DataValue;

            DataCount = DataCount +1;


        end

endmodule
```

Abbreviations

EEG	ElectroEncephaloGram
EMG	ElectroMyoGram
ECG	ElectroCardioGram
BCI	Brain Computer Interface
BAN	Body Area Network
PDF	Probability Density Function
EOG	ElectroOculoGram
i.i.d	independent identically distributed
CR	Compression Ratio
PRD	Root-Mean-Square Difference
RLE	Run-Length Encoding
LZ77	Lempel-Ziv 77
LPC	Linear Predictive Coding
AR	AutoRegressive
FFT	Fast Fourier Transform
DFT	Discrete Fourier Transform
DCT	Discrete Cosine Transform

KLT Karhunen-Loeve Transform

DWT Discrete Wavelet Transform

ADC Analog-to-Digital Converter

DPCM Differential Pulse-Code Modulation

LMS Least Mean-Square

ICA Independent Component Analysis

REACT Real-Time EEG Analysis for Event Detection

EHG ElectroHysteroGram

ERP Event-Related Potentials

DOT diffuse optical tomography

CMOS Complementary Metal-Oxide Semiconductor

NAT Neural Acquisition Tracker

vcd value change dump

RAM Random-Access Memory

SRAM Static Random-Access Memory

HP High Performance

LS Low Static Power

LO Low Operating Power

SISO Serial-In Serial-Out

PIPO Parallel-In Parallel-Out

VLSI Very-Large-Scale Integration

GDSII Graphic Database System II

IC Integrated Circuit

References

[1] G. E. Chatrian, C.-M. Shaw, and H. Leffman, "The significance of periodic lateralized epileptiform discharges in EEG: An electrographic, clinical and pathological study," *Electroencephalography and Clinical Neurophysiology*, vol. 17, pp. 177–193, Aug. 1964.

[2] C. Stam, B. Jelles, H. Achtereekte, S. Rombouts, J. Slaets, and R. Keunen, "Investigation of EEG non-linearity in dementia and Parkinson's disease," *Electroencephalography and Clinical Neurophysiology*, vol. 95, pp. 309–317, Nov. 1995.

[3] R. N. Khushaba, S. Kodagoda, M. Takruri, and G. Dissanayake, "Expert Systems with Applications Toward improved control of prosthetic fingers using surface electromyogram (EMG) signals," *Expert Systems With Applications*, vol. 39, no. 12, pp. 10731–10738, 2012.

[4] U. Hoffmann, J. M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects," *Journal of Neuroscience Methods*, vol. 167, pp. 115–125, Jan. 2008.

[5] "EEG test cost." `http://health.costhelper.com/eeg.html`, 2015. [Online; accessed 22-June-2015].

[6] E. Waterhouse, "New Horizons in Ambulatory Electroencephalography," *IEEE Engineering in Medicine and Biology Magazine*, vol. 22, no. 3, pp. 74–80, 2003.

[7] R. Rieger and J. T. Taylor, "An adaptive sampling system for sensor nodes in body area networks.," *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 17, pp. 183–189, Apr. 2009.

[8] "Emotiv® headset." `https://emotiv.com/`, 2015. [Online; accessed 29-Apr-2015].

[9] "Myo® Gesture Control Armband." https://www.thalmic.com/en/myo/, 2015. [Online; accessed 29-Apr-2015].

[10] "Apple Watch®." http://www.apple.com/watch/health-and-fitness, 2015. [Online; accessed 22-Apr-2015].

[11] J. Austin, C. Bailey, and A. Moulds, "A Miniaturized 4-Channel, 2KSa/sec Biosignal Data Recorder With 3-Axis Accelerometer and Infra-red Timestamp Function," in *SENSORCOMM 2013, The Seventh International Conference on Sensor Technologies and Applications*, pp. 213–219, 2013.

[12] "Myo® Battery Life." https://medium.com/@allingeek/magic-in-a-box-my-first-day-using-myo-aded707c2042, 2015. [Online; accessed 29-Apr-2015].

[13] "Emotiv® Insight headset." https://emotiv.com/insight.php, 2015. [Online; accessed 29-Apr-2015].

[14] "Google® Glass." http://www.google.com/glass/start/, 2015. [Online; accessed 29-June-2015].

[15] J. S. Barlow, *The electroencephalogram: its patterns and origins*. MIT press, 1993.

[16] H. Berger, "Über das elektrenkephalogramm des menschen," *European Archives of Psychiatry and Clinical Neuroscience*, vol. 87, no. 1, pp. 527–570, 1929.

[17] C. D. Binnie, A. J. Rowan, and T. Gutter, *A manual of electroencephalographic technology*. No. 6, CUP Archive, 1982.

[18] C. Andrew and G. Pfurtscheller, "Event-related coherence as a tool for studying dynamic interaction of brain regions," *Electroencephalography and clinical neurophysiology*, vol. 98, no. 2, pp. 144–148, 1996.

[19] D. L. Schomer, "The normal eeg in an adult," in *The clinical neurophysiology primer*, pp. 57–71, Springer, 2007.

[20] G. H. Klem, H. O. Lüders, H. Jasper, and C. Elger, "The ten-twenty electrode system of the international federation," *Electroencephalogr Clin Neurophysiol*, vol. 52, no. suppl., p. 3, 1999.

[21] A. Gevins, H. Leong, M. E. Smith, J. Le, and R. Du, "Mapping cognitive brain function with modern high-resolution electroencephalography," *Trends in neurosciences*, vol. 18, no. 10, pp. 429–436, 1995.

[22] L. Sörnmo and P. Laguna, *Bioelectrical signal processing in cardiac and neurological applications*. Academic Press, 2005.

[23] J. D. Bronzino, *Biomedical engineering handbook*, vol. 2. CRC press, 1999.

[24] G. Antoniol and P. Tonella, "EEG data compression techniques," *IEEE Transactions on Biomedical Engineering*, vol. 44, pp. 105–114, Feb. 1997.

[25] Y. Wongsawat, S. Oraintara, T. Tanaka, and K. Rao, "Lossless multi-channel EEG compression," *2006 IEEE International Symposium on Circuits and Systems*, vol. 6, pp. 1611–1614, 2006.

[26] D. O'Shea and R. McSweeney, "Efficient Implementation of Arithmetic Compression for EEG," *IET Proceedings of Irish ...*, 2011.

[27] N. Memon, X. Kong, and J. Cinkler, "Context-based lossless and near-lossless compression of EEG signals," *IEEE Transactions on Information Technology in Biomedicine*, vol. 3, no. 3, pp. 231–238, 1999.

[28] M. Klimesh, "Quantization considerations for distortion-controlled data compression," *The Telecommunications and Mission Operations Progress Report: TMO PR*, pp. 42–139, 1999.

[29] N. Sriraam and C. Eswaran, "Performance evaluation of neural network and linear predictors for near-lossless compression of EEG signals," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, pp. 87–93, Jan. 2008.

[30] N. Ahmed and K. R. Rao, *Orthogonal transforms for digital signal processing*. Springer Science & Business Media, 2012.

[31] S. Mallat, *A wavelet tour of signal processing*. Academic press, 1999.

[32] A. N. Akansu and M. J. Smith, *Subband and wavelet transforms: design and applications*, vol. 340. Springer Science & Business Media, 2012.

[33] G. Higgins, S. Faul, R. P. McEvoy, B. McGinley, M. Glavin, W. P. Marnane, and E. Jones, "EEG compression using JPEG2000 how much loss is too much?," *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'10*, vol. 2010, pp. 614–617, Jan. 2010.

[34] H. GholamHosseini, H. Nazeran, and B. Moran, "Ecg compression: evaluation of fft, dct, and wt performance.," *Australasian physical & engineering sciences in medicine/supported by the Australasian College of Physical Scientists in Medicine and the Australasian Association of Physical Sciences in Medicine*, vol. 21, no. 4, pp. 186–192, 1998.

[35] R. Mahajan and D. Bansal, "Hybrid multichannel eeg compression scheme for tele-health monitoring," in *Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2014 3rd International Conference on*, pp. 1–6, IEEE, 2014.

[36] E. J. Candè and M. B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, 2008.

[37] A. M. Abdulghani, A. J. Casson, and E. Rodriguez-Villegas, "Compressive sensing scalp EEG signals: Implementations and practical performance," *Medical and Biological Engineering and Computing*, vol. 50, pp. 1137–1145, Nov. 2012.

[38] E. C. Ifeachor, B. W. Jervis, E. L. Morris, E. M. Allen, and N. R. Hudson, "New online method for removing ocular artefacts from EEG signals," *Medical and Biological Engineering and Computing*, vol. 24, no. 4, pp. 356–364, 1986.

[39] P. He, G. Wilson, and C. Russell, "Removal of ocular artifacts from electroencephalogram by adaptive filtering," *Medical and biological engineering and computing*, vol. 42, no. 3, pp. 407–412, 2004.

[40] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[41] J. T. Gwin, K. Gramann, S. Makeig, and D. P. Ferris, "Removal of movement artifact from high-density EEG recorded during walking and running.," *Journal of neurophysiology*, vol. 103, pp. 3526–3534, June 2010.

[42] R. LIN, R.-G. LEE, C.-L. TSENG, H.-K. ZHOU, C.-F. CHAO, and J.-A. JIANG, "A new approach for identifying sleep apnea syndrome using wavelet transform and neural networks," *Biomedical Engineering: Applications, Basis and Communications*, vol. 18, no. 03, pp. 138–143, 2006.

[43] D. Bai, T. Qiu, and X. Li, "The sample entropy and its application in eeg based epilepsy detection.," *Journal of biomedical engineering*, vol. 24, no. 1, pp. 200–205, 2007.

[44] J. Gotman, J. Ives, P. Gloor, L. Quesney, and P. Bergsma, "Monitoring at the montreal neurological institute.," *Electroencephalography and clinical neurophysiology. Supplement*, vol. 37, pp. 327–340, 1984.

[45] A. Subasi, "Eeg signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.

[46] M. K. Kiymik, M. Akin, and A. Subasi, "Automatic recognition of alertness level by using wavelet transform and artificial neural network," *Journal of neuroscience methods*, vol. 139, no. 2, pp. 231–240, 2004.

[47] A. Temko, E. Thomas, G. Boylan, W. Marnane, and G. Lightbody, "An svm-based system and its performance for detection of seizures in neonates," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 2643–2646, IEEE, 2009.

[48] E. Waterhouse, "New horizons in ambulatory electroencephalography," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 22, no. 3, pp. 74–80, 2003.

[49] J.-J. Vidal, "Toward direct brain-computer communication," *Annual review of Biophysics and Bioengineering*, vol. 2, no. 1, pp. 157–180, 1973.

[50] Z. Keirn, J. I. Aunon, *et al.*, "A new mode of communication between man and his surroundings," *Biomedical Engineering, IEEE Transactions on*, vol. 37, no. 12, pp. 1209–1214, 1990.

[51] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, "Brain-computer interfacea new communication device for handicapped persons," *Journal of Microcomputer Applications*, vol. 16, no. 3, pp. 293–299, 1993.

[52] "iFocusBand headset." http://www.ifocusband.com/, 2015. [Online; accessed 22-Apr-2015].

[53] D. C. Yates and E. Rodriguez-Villegas, "A key power trade-off in wireless EEG headset design," *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering*, pp. 453–456, May 2007.

[54] R. R. Harrison and C. Charles, "A low-power low-noise cmos amplifier for neural recording applications," *Solid-State Circuits, IEEE Journal of*, vol. 38, no. 6, pp. 958–965, 2003.

[55] J. Sauerbrey, D. Schmitt-Landsiedel, and R. Thewes, "A 0.5-v 1-μw successive approximation adc," *Solid-State Circuits, IEEE Journal of*, vol. 38, no. 7, pp. 1261–1265, 2003.

[56] "nRF8001 Bluetooth® low energy Connectivity IC." http://www.nordicsemi.com/eng/Products/Bluetooth-Smart-Bluetooth-low-energy/nRF8001, 2014. [Online; accessed 09-Dec-2014].

[57] "Duracell 2025 Battery." http://www.duracell.com/en-us/products/button-batteries/duracell-2025-battery, 2014. [Online; accessed 29-Dec-2014].

[58] P. Desnoyers, "Empirical evaluation of NAND flash memory performance," *ACM SIGOPS Operating Systems Review*, vol. 44, no. 1, p. 50, 2010.

[59] J. Huang, A. Badam, R. Chandra, E. B. Nightingale, J. Huang, A. Badam, R. Chandra, and E. B. Nightingale, "WearDrive : Fast and Energy-Efficient Storage for Wearables," 2015.

[60] G. Mathur, P. Desnoyers, P. Chukiu, D. Ganesan, and P. Shenoy, "Ultra-low power data storage for sensor networks," *ACM Transactions on Sensor Networks*, vol. 5, no. 4, pp. 1–34, 2009.

[61] S. Mittal, "A Survey of Architectural Techniques For DRAM Power Management," *International Journal of High Performance Systems Architecture*, vol. 4, no. 2, pp. 110–119, 2012.

[62] S. Mittal, J. S. Vetter, and S. Member, "A Survey Of Architectural Approaches for Data Compression in Cache and Main Memory Systems," *IEEE Transactions of Parallel and Distributed Systems (TPDS)*, no. c, pp. 1–14, 2015.

[63] M. Qazi, M. Clinton, S. Bartling, and A. P. Chandrakasan, "A low-voltage 1mb feram in 0.13μm cmos featuring time-to-digital sensing for expanded operating margin in scaled cmos," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, pp. 208–210, IEEE, 2011.

[64] E. Chua and W. C. Fang, "Mixed bio-signal lossless data compressor for portable brain-heart monitoring systems," *IEEE Transactions on Consumer Electronics*, vol. 57, pp. 267–273, Jan. 2011.

[65] G. A. Luo, S. L. Chen, and T. L. Lin, "VLSI implementation of a lossless ECG encoder design with fuzzy decision and two-stage Huffman coding for wireless body sensor network," *ICICS 2013 - Conference Guide of the 9th International Conference on Information, Communications and Signal Processing*, pp. 1–4, 2013.

[66] C. J. Deepu, X. Zhang, W. S. Liew, D. L. T. Wong, and Y. Lian, "An ECG-SoC with 535nW/channel lossless data compression for wearable sensors," *Proceedings of the 2013 IEEE Asian Solid-State Circuits Conference, A-SSCC 2013*, pp. 145–148, 2013.

[67] S.-L. Chen and J.-G. Wang, "VLSI implementation of low-power cost-efficient lossless ECG encoder design for wireless healthcare monitoring application," *Electronics Letters*, vol. 49, no. 2, pp. 91–93, 2013.

[68] R. R. Harrison, P. T. Watkins, R. J. Kier, R. O. Lovejoy, D. J. Black, B. Greger, and F. Solzbacher, "A low-power integrated circuit for a wireless 100-electrode neural recording system," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 1, pp. 123–133, 2007.

[69] A. J. Casson and E. Rodriguez-Villegas, "Toward online data reduction for portable electroencephalography systems in epilepsy," *IEEE Transactions on Biomedical Engineering*, vol. 56, pp. 2816–2825, Dec. 2009.

[70] G. Higgins, B. McGinley, M. Glavin, and E. Jones, "Low power compression of EEG signals using JPEG2000," *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2010 4th International Conference on-NO PERMISSIONS*, 2010.

[71] J. Chiang and R. Ward, "Data reduction for wireless seizure detection systems," in *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*, pp. 48–52, IEEE, 2013.

[72] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state.," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 64, p. 061907, Nov. 2001.

[73] N. Sriraam, "A high-performance lossless compression scheme for EEG signals using wavelet transform and neural network predictors," *International Journal of Telemedicine and Applications*, vol. 2012, p. 302581, Jan. 2012.

[74] M. Fira and L. Goras, "Biomedical signal compression based on basis pursuit," pp. 541–545, 2009.

[75] N. Sriraam, "Context-based near-lossless compression of EEG signals using neural network predictors," *AEU - International Journal of Electronics and Communications*, vol. 63, pp. 311–320, Apr. 2009.

[76] A. Mirzaei, A. Ayatollahi, P. Gifani, and L. Salehi, "EEG analysis based on wavelet-spectral entropy for epileptic seizures detection," *Proceedings - 2010 3rd International Conference on Biomedical Engineering and Informatics, BMEI 2010*, vol. 2, no. Bmei, pp. 878–882, 2010.

[77] K. Srinivasan, J. Dauwels, and M. R. Reddy, "A two-dimensional approach for lossless EEG compression," *Biomedical Signal Processing and Control*, vol. 6, pp. 387–394, Oct. 2011.

[78] K. Srinivasan and M. Ramasubba Reddy, "Efficient preprocessing technique for real-time lossless EEG compression," *Electronics Letters*, vol. 46, no. 1, p. 26, 2010.

[79] B. Platt, B. Drever, D. Koss, S. Stoppelkamp, A. Jyoti, A. Plano, A. Utan, G. Merrick, D. Ryan, V. Melis, H. Wan, M. Mingarelli, E. Porcu, L. Scrocchi, A. Welch,

and G. Riedel, "Abnormal cognition, sleep, eeg and brain metabolism in a novel knock-in alzheimer mouse, plb1," *PLoS ONE*, vol. 6, p. e27068, Jan. 2011.

[80] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, pp. e215–e220, jun 2000.

[81] A. Alexandersson, T. Steingrimsdottir, J. Terrien, C. Marque, and B. Karlsson, "The icelandic 16-electrode electrohysterogram database," *Scientific data*, vol. 2, 2015.

[82] G. B. Moody, R. G. Mark, and A. L. Goldberger, "Evaluation of thetrim'ecg data compressor," in *Computers in Cardiology, 1988. Proceedings.*, pp. 167–170, IEEE, 1988.

[83] L. Citi, R. Poli, and C. Cinel, "Documenting, modelling and exploiting p300 amplitude changes due to variable target delays in donchin's speller," *Journal of neural engineering*, vol. 7, no. 5, p. 056006, 2010.

[84] C. Dai and C. Bailey, "A time-domain based lossless data compression technique for wireless wearable biometric devices," in *SENSORCOMM 2013, The Seventh International Conference on Sensor Technologies and Applications*, pp. 104–107, 2013.

[85] C. Dai and C. Bailey, "Power analysis of a lossless data compression technique for wireless wearable biometric devices," in *Ph.D. Research in Microelectronics and Electronics (PRIME), 2015 11th Conference on*, pp. 97–100, June 2015.

[86] C. Dai and C. Bailey, "A lossless data reduction technique for wireless eeg recorders and its use in selective data filtering for seizure monitoring," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pp. 6186–6189, Aug 2015.

[87] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. T. Treves, and J. Guttag, "Patient-specific seizure onset detection," *Epilepsy & Behavior*, vol. 5, no. 4, pp. 483–498, 2004.

[88] A. H. Shoeb, *Application of machine learning to epileptic seizure onset detection and treatment.* PhD thesis, Massachusetts Institute of Technology, 2009.

[89] A. Ebner, D. S. Dinner, S. Noachtar, and H. Luders, "Automatisms with preserved responsiveness a lateralizing sign in psychomotor seizures," *Neurology*, vol. 45, no. 1, pp. 61–64, 1995.

[90] U. R. Acharya, C. K. Chua, T.-C. Lim, Dorithy, and J. S. Suri, "Automatic identification of epileptic eeg signals using nonlinear parameters," *Journal of Mechanics in Medicine and Biology*, vol. 9, no. 04, pp. 539–553, 2009.

[91] C. Petitmengin, M. Baulac, and V. Navarro, "Seizure anticipation: are neurophenomenological approaches able to detect preictal symptoms?," *Epilepsy & Behavior*, vol. 9, no. 2, pp. 298–306, 2006.

[92] L. D. Iasemidis, P. Pardalos, J. C. Sackellares, and D.-S. Shiau, "Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures," *Journal of Combinatorial Optimization*, vol. 5, no. 1, pp. 9–26, 2001.

[93] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of eeg records in an epileptic patient using wavelet transform," *Journal of neuroscience methods*, vol. 123, no. 1, pp. 69–87, 2003.

[94] S. Ghosh-Dastidar and H. Adeli, "Improved spiking neural networks for eeg classification and epilepsy and seizure detection," *Integrated Computer-Aided Engineering*, vol. 14, no. 3, pp. 187–212, 2007.

[95] S. Wilton and N. Jouppi, "CACTI: an enhanced cache access and cycle time model," *Solid-State Circuits, IEEE Journal of*, vol. 31, no. 5, pp. 677–688, 1996.

[96] D. Tarjan, S. Thoziyoor, and N. Jouppi, "Cacti 4.0," *HP Laboratories Palo-Alto, Tech. Rep. HPL-2006-86*, pp. 0–15, 2006.

[97] "CACTI Web Interface." `http://quid.hpl.hp.com:9081/cacti/sram.y`, 2015. [Online; accessed 22-Aug-2015].

[98] Z. H. Derafshi and J. Frounchi, "Low-noise low-power front-end logarithmic amplifier for neural recording system," *International Journal of Circuit Theory and Applications*, vol. 42, no. December, pp. 437–451, 2014.

[99] V. Majidzadeh, A. Schmid, and Y. Leblebici, "Energy efficient low-noise neural recording amplifier with enhanced noise efficiency factor," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 5, no. 3, pp. 262–271, 2011.

[100] J. Ruiz-Amaya, A. Rodriguez-Perez, and M. Delgado-Restituto, "A Low Noise Amplifier for Neural Spike Recording Interfaces," *Sensors*, vol. 15, no. 10, pp. 25313–25335, 2015.

[101] Z. H. Derafshi and J. Frounchi, "Low-noise low-power front-end logarithmic amplifier for neural recording system," *International Journal of Circuit Theory and Applications*, vol. 42, pp. 437–451, may 2014.

[102] W. Zhao, H. Li, and Y. Zhang, "A low-noise integrated bioamplifier with active DC offset suppression," *2009 IEEE Biomedical Circuits and Systems Conference, BioCAS 2009*, no. 60971084, pp. 5–8, 2009.

[103] P. Harpe, E. Cantatore, and A. Van Roermund, "A 10b/12b 40 kS/s SAR ADC with data-driven noise reduction achieving up to 10.1b ENOB at 2.2 fJ/conversion-step," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 12, pp. 3011–3018, 2013.

[104] S. Sar, Y.-h. Chung, M.-h. Wu, and H.-s. Li, "A 12-bit 8.47-fJ/Conversion-Step Capacitor- Swapping SAR ADC in 110-nm CMOS," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 1, pp. 1–9, 2014.

[105] N. Verma and A. P. Chandrakasan, "An ultra low energy 12-bit rate-resolution scalable SAR ADC for wireless sensor nodes," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 6, pp. 1196–1205, 2007.

[106] X. Zou, X. Xu, L. Yao, and Y. Lian, "A 1-V 450-nW fully integrated programmable biomedical sensor interface chip," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 4, pp. 1067–1077, 2009.

[107] B. Verbruggen, J. Craninckx, M. Kuijk, P. Wambacq, and G. V. D. Plas, "A 9.4-ENOB 1V 3.8μW 100kS/s SAR ADC with Time-Domain Comparator," *Isscc*, pp. 566–567, 2011.

[108] H. Zhang, Y. Qin, S. Yang, and Z. Hong, "A 455 nW 220 fJ/conversion-step 12 bits 2 kS/s SAR ADC for portable biopotential acquisition systems," *Journal of Semiconductors*, vol. 32, no. 1, p. 015001, 2011.

[109] "Semtech® SX1211." http://www.semtech.com/wireless-rf/rf-transceivers/sx1211/, 2015. [Online; accessed 22-Sept-2015].

[110] "Texas Instruments® CC2564." http://www.ti.com/product/CC2564/description, 2015. [Online; accessed 22-Sept-2015].

[111] "Texas Instruments® CC2420." http://www.ti.com/product/CC2420/technicaldocuments, 2015. [Online; accessed 22-Sept-2015].

[112] "Atmel® AT86RF231." http://www.atmel.com/devices/AT86RF231.aspx, 2015. [Online; accessed 22-Sept-2015].

[113] "Texas Instruments® CC1000." http://www.ti.com/product/CC1000/description, 2015. [Online; accessed 22-Sept-2015].

[114] "Mitsumi® WML-C46." http://www.mitsumi.co.jp/latest/Catalog/hifreq/bluetooth/wmlc46_e.html, 2015. [Online; accessed 22-Sept-2015].

[115] S. Park, Y. Kim, B. Urgaonkar, J. Lee, and E. Seo, "A comprehensive study of energy efficiency and performance of flash-based ssd," *Journal of Systems Architecture*, vol. 57, no. 4, pp. 354–365, 2011.

[116] V. Mohan, T. Bunker, L. Grupp, S. Gurumurthi, M. R. Stan, and S. Swanson, "Modeling power consumption of nand flash memories using flashpower," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 32, no. 7, pp. 1031–1044, 2013.

[117] "Duracell Ultra Power MN2400." http://www.battery-force.co.uk/detail_DULR03004B-Duracell-Ultra-Power-MN2400-AAA-Battery-Pack-of-4.html, 2015. [Online; accessed 22-Apr-2015].

[118] "Duracell Plus Power MN1500." http://www.battery-force.co.uk/detail_DULR6Y008B-Duracell-Plus-Power-MN1500-AA-Battery-Pack-of-5-3-FREE.html, 2015. [Online; accessed 22-Apr-2015].

[119] "Energizer CR2025." http://www.battery-force.co.uk/detail_ENCR20002B-Energizer-CR2025-3V-Lithium-Coin-Cell-Card-of-2.html, 2015. [Online; accessed 22-Apr-2015].

[120] "Energizer 363/364." http://www.battery-force.co.uk/detail_ENSR60001A-Energizer-363-364-Watch-Battery-MiniPack-of-1.html, 2015. [Online; accessed 22-Apr-2015].