

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.

Keywords—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.

In the rest of this paper, the basic idea of Log2 Sub-band algorithm will be given in section II, and followed by an introduction to test data in section III and the result analysis in section IV. Preliminary hardware design will be given in section V, and finally the conclusion in section VI.

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. There will be a 2-bit header added before transmission to guide

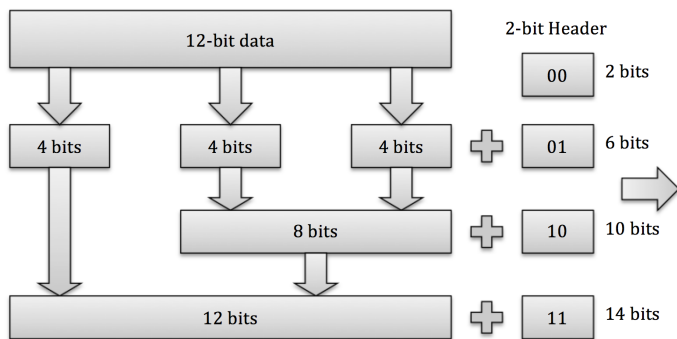


Figure. 1. Log2 Sub-band encoding

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 generating area) 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 1.51 was observed with unstable seizure signals. And the results are shown in 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 Log2 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 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

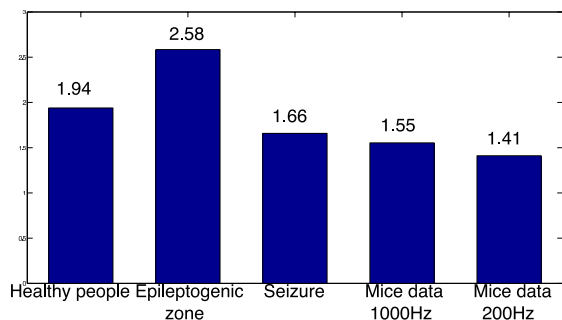


Figure. 2. Compression ratios

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

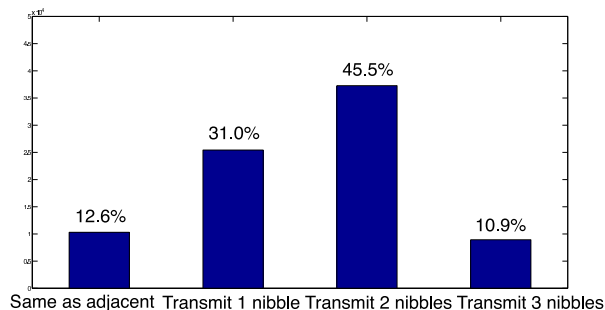


Figure. 3. The distribution of the length of transmitted healthy peoples data samples

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.

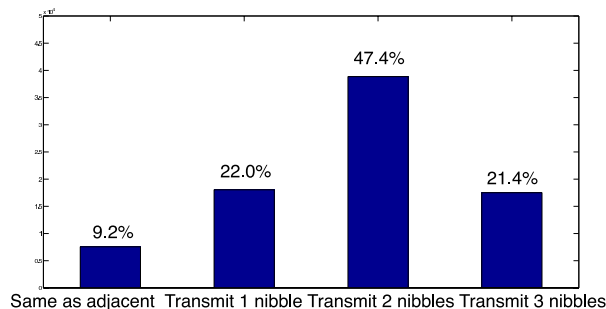


Figure. 4. The distribution of the length of transmitted seizures data samples

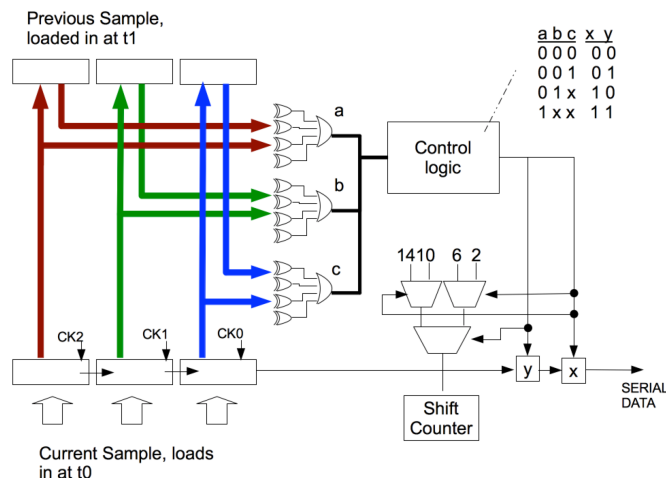


Figure. 5. Simplified Hardware Design

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. 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 with 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. Meanwhile, 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, University 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] 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.
- [4] G. Antonioli 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. Orantara, 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 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.
- [10] "nRF8001 Bluetooth® low energy Connectivity IC," <http://www.nordicsemi.com/eng/Products/Bluetooth-R-low-energy/nRF8001>, 2012, [Online; accessed 09-April-2013].