3.1 INTRODUCTION

In this chapter, different types of data compression techniques, used for biomedical signals, have been discussed in detail.

Data storage and transmission costs money. As the storage and transmission requirements increase with the increase in the amount of information, so does the cost. Thus the need arises for data compression for storage and transmission of digital data. Data compression techniques have been utilized in wide spectrum of applications e.g. speech, audio, images, video, telemetry etc. A typical computerized medical signal processing system acquires a large amount of data that is difficult to store and transmit. It is required to reduce the data storage space while preserving the significant clinical content for signal reconstruction. A data reduction algorithm seeks to minimize the number of code bits stored by reducing the redundancy present in the original signal. Compression ratio is obtained by dividing the number of bits in the original signal by the number saved in the compressed signal. Generally, higher compression ratio is desirable. The data reduction algorithm must also represent the data with acceptable fidelity [17][18]. In biomedical data reduction, usually the clinical acceptability of the reconstructed signal is determined through visual inspection. For signal coding and compression, there are two categories of data compression techniques given as follows.

**Lossless compression technique**

Lossless compression techniques mean that the restored data file is identical to the original. In lossless compression, redundancy in the data representation is targeted. This technique is necessary for some data types e.g. executable codes, word processing, tabulated numbers etc. Here no information is lost but compression is quite less usually in the range of 3:1. Examples of lossless coding techniques are Huffman encoding, PCM, DPCM.
**Lossy compression technique**

In lossy compression technique the restored file is not identical to the original. Slight degradation of information is acceptable to restore the information in the decoder. Here the compression is much more, approximately 10 to 20 times. As the compression increases, the quality of the restored signal decreases. Various biomedical signals, speech signals, images are compressed using lossy technique therefore most of the techniques discussed in this chapter are lossy techniques.

There are three types of compression techniques used for biomedical signals.

### 3.2 DIRECT DATA COMPRESSION TECHNIQUES

The direct data compressors base their detection of redundancies on the direct analysis of the actual signal. Most of the direct data compression techniques rely on utilizing prediction or interpolation algorithm. These techniques reduce the redundancy in data sequence by examining a successive number of neighbouring samples. The prediction algorithm uses priori knowledge of some previous samples, while interpolation uses priori knowledge of previous and future samples. Various direct data compression methods are given below.

#### 3.2.1 Tolerance Comparison Technique

This technique employs a polynomial predictor/interpolator to eliminate the samples from the data sequence that is implied by examining the preceding and succeeding samples. In polynomial predictors, a preset error threshold centred around the actual sample point is set. Whenever the difference between a sample and a succeeding future sample exceeds the preset error threshold, the data between the two samples is approximated by a line whereby only the line parameters (e.g. length and amplitude) are saved. Where as polynomial interpolators utilize both past and future data points to decide whether or not the actual sample point is redundant. In other words, all samples and the present sample point affect the interpolation. Compression algorithms based on this technique are AZTEC (Amplitude zone time epoch coding), Fan algorithm and SAPA technique (scan along polygonal approximation) algorithm and CORTES (co-ordinate reduction time encoding system) algorithm[34].

42
3.2.2 Differential Pulse Code Modulation (DPCM)

This is one of the first standard methods of speech coding but is being used for all types of signals. This method consists of coding directly the prediction error generated by linear filtering of input signals. Data samples are estimated from the previous samples. The error (residual) between the actual sample and the estimated sample is quantized and stored. The simplest DPCM system is a system which employs a polynomial predictor with zero order, which predicts the $n^{th}$ value same as $(n-1)^{th}$ value in the original data. Thus $\hat{y}_n = y_{n-1}$. Hence the difference signal $e_n = y_n - \hat{y}_n$ is substituted for the actual signal itself. Consequently, waveform redundancy reduction by DPCM coders is achieved by representing the actual correlated signal in terms of an uncorrelated signal namely the estimation error signal. Since the estimation error sequence is saved in place of actual data sequence, upon reconstruction the original signal is preserved without loss of information. Data compression based on such a system is referred to as delta encoding. The key feature is that the delta encoding has lower amplitude than the original signal. Thus less number of bits is required for quantization and hence compression occurs. In this technique, whenever the absolute value of the difference between the adjacent pair samples in a signal exceeds a preset threshold, data is saved. Otherwise, data is considered redundant and thus eliminated. Hence compression is performed. A more complex DPCM system employs the linear predictive coding (LPC). It is a more complicated form of delta encoding. Basically, a better guess is made by observing large numbers of previous samples to estimate the current sample. LPC uses various intensive mathematical techniques and shows better results than simple delta encoding [33] [34].

Adaptive Differential Pulse Code Modulation (ADPCM) is similar to DPCM but the prediction coefficients and the quantization levels are dependent on the past reconstructed signal. ADPCM shows better performance than DPCM for the compression of ECG/EMG signals [34].

3.2.3 Entropy Encoding

This technique was given by Huffman in 1950’s. Basically, data compression by entropy encoding is obtained by means of assigning variable length code words to a given quantized data sequence according to their frequency of occurrence. This compression method attempts to remove signal redundancy that arises whenever the
quantized signal levels do not occur with equal probability. Values occurring with higher frequency are assigned shorter code length as compared to less frequent ones. This results in the minimization of the mean code length or optimum code.

3.3 TRANSFORM BASED DATA COMPRESSION TECHNIQUES

Transform based data compression techniques use signal in frequency domain to find some hidden information of the signal. The transformed signal (expansion coefficients) is then encoded to reduce the amount of data needed to represent the original signal. Upon signal reconstruction, an inverse transform is applied to obtain the original signal, with some amount of error. Transformational compression method mainly utilize spectral and energy distribution analysis for detecting redundancies. Various transformation methods include Fourier Transform, Wavelet Transform, Discrete cosine transform, Karhunen-Loeve Transform (KLT).

Karhunen-Loeve Transform (KLT) considers that least number of orthogonal functions is needed to represent the input signal for a given RMS error. The KLT transform leads to the de-correlated transform coefficients in the form of a diagonal covariance matrix. The computation is extremely intensive due to the Eigen values and the Eigen vectors of the covariance matrix (which can be extremely large symmetric matrix) [33]. In Discrete Cosine Transform (DCT), a sequence of N samples are considered in the form of a vector X. Orthogonal transformation of these N samples is obtained which compact energy in first M transformed vector components (such that M<N). Therefore only these M significant components are encoded, rest N-M are discarded which leads to compression. It gives a good compression and a time complexity of O[log2N] [36][37].

Karhunen-Loeve transform is supposed to be the optimum one but its only drawback lies in its prohibited computation time. The Discrete Cosine Transform (DCT) is a suboptimal solution with computational advantages. Another transform method which has received a great deal of attention over the past several years, is the Wavelet Transform. Wavelets have been applied extensively and specific applications in the field of biomedical signal analysis and compression have been made. Thus, the wavelet scheme is taken under consideration to perform EMG data compression. DWT has become a popular technique due to good compression and acceptable quality of the reconstructed signal without being computationally intensive.
Results are almost identical when applied to surface or needle EMG, although signal shapes are very different. In both cases, transform-based methods outperform linear predictive techniques [34]. Moreover, in the needle EMG case, the DWT gives the best results (Fig 3.1(a)). This can be explained by the fact that the energy of the needle EMG signals is concentrated in a few subbands. Using a wavelet-based subband coding, a compression ratio of 6 has been achieved with generally excellent reconstructed signal quality.
Various wavelet-based algorithms are given as follows. These techniques use wavelet transform for further signal compression.

3.3.1 Embedded Zero Tree Wavelet (EZW) Algorithm

The first step of EZW encoding algorithm is wavelet transform, which de-correlates the signal and combines the signal information in few large coefficients. This is followed by quantization of the coefficient vector to generate the quantized coefficients, which are further encoded.

The EZW algorithm has two very interesting properties. The first is that the EZW algorithm creates a connection between the wavelet coefficients from different sub-bands. This is the reason why several quantized coefficients from different sub-bands can be encoded using only one symbol. The second property is that the important coefficients are encoded using the successive approximation technique, which puts the most important parts of the coefficients at the beginning of a bit stream. Therefore, the encoder can easily stop encoding at any desired target rate. In multi-resolution analysis approach, Daubechies filters and ‘wrap around’ technique is used to calculate the wavelet coefficients of the finite length segments. All the detail coefficients and the approximation coefficients are assembled together in a coefficient vector $c$. The length $N$ of the original signal segment corresponds to the length of the coefficient vector $c$.

![EMG signal and the quantized wavelet coefficients of iteration level 2.](image)
The next step is the quantization of the coefficient vector \( c \), and the encoding of the quantized coefficients. Due to the quantization, perfect reconstruction is not possible, and reconstruction errors occur. The quantization is performed iteratively by a three-level quantizer. The encoding of the quantized coefficients is performed at each iteration level of the quantization procedure. The main idea behind the encoding algorithm is to find a connection between the quantized coefficients from different sub-bands. A so-called zero-tree exists, if those quantized coefficients, which describe the same signal segment, are all zero. A zero-tree and its members are encoded with only one symbol, namely with its root. Modifications to the quantization and to the encoding parts of the EZW algorithm was proposed in order to improve its performance in the analysis of EMG signals. In contrast to the conventional EZW algorithm, more coefficients are allowed to be a root of a zero-tree. Second, segmentation algorithms for EMG signal have been included. It suppresses the fluctuations and detects the active segments with threshold criteria and reduces the reconstruction error in active segments.

A DB filter of order 11 and segments of length 4096 samples gave the best results. Simulations have shown that the best wavelet packet bases for EMG signal segments always have high resolution in low frequency bands. This can be explained by the energy density distribution of the EMG signal. The EZW algorithm with the modified encoding part gives better results than the original EZW algorithm. Given a target
bitrate, the algorithm is capable of compressing EMG data rapidly and with little
distortion. EZW algorithm has been used for EMG signals and has shown maximum
compression ratio of 12:1 at PRD of 11%[8].

3.3.2 Set Partitioning in Hierarchical Trees (SPIHT) Algorithm
This model was given by W.A. Pearlman, Z. Lu, D.Y Kim. The SPIHT algorithm
has achieved notable success in still image coding but has been modified to be used
for one dimensional signals such as biomedical and speech signals. It exploits the
spatial self-similarity property of wavelet coefficients across sub bands. The property
implies that if a node coefficient is insignificant with respect to a given threshold,
probably all nodes descending from that node are also insignificant. SPIHT assumes
that the decomposition structure is the octave band structure and then uses the fact
that subbands at different levels but of the same orientation display similar
characteristics.

After the wavelet transform, partial ordering of the transformed coefficients is done
by magnitude with a set partitioning sorting algorithm. A tree structure is formed in
wavelet domain. Every point in layer I correspond to 2 points in the next layer i+1
with parent-offspring relation. SPIHT makes the hypothesis that if a parent coefficient
is below a certain threshold then it is likely that all its descendents are below the
threshold too. If this prediction is successful then SPIHT can represent the parent and
all its descendents with a single symbol called a zero tree. To predict energy of
coefficients in lower level sub-bands (children) using coefficients in higher level sub-
bands (parents) makes sense since there should be more energy per coefficient in
these small bands, than in the bigger ones. SPIHT consists of two passes, the ordering
pass and the refinement pass. In the ordering pass SPIHT attempts to order the
coefficients according to their magnitude. In the refinement pass the quantization of
coefficients is refined. The ordering and refining is made relative to a threshold. The
threshold is appropriately initialised and then continuously made smaller with round
of the algorithm. After each sorting pass, the nth most significant bit of every
coefficient, found significant at a higher threshold, is sent to the decoder. By
transmitting the bit stream in this ordered bit plane fashion, only most valuable
remaining bit is transmitted. SPIHT gives excellent coding performance in
biomedical signal compression, better than the direct compression methods [11].

48
3.3.3 Dynamic Vector Quantization using Single Codebook

This model was given by Shou Gang Mioau, H.L Yen, C.L Lin for ECG signal compression. It uses wavelet transform based vector quantization technique for the compression of biomedical signals. From the M-sample long data, n-level DWT decomposition is performed resulting in n+1 wavelet coefficient sub-bands. From these coefficients, $M/2^n$ number of Tree Vectors are extracted, composed of the coefficients from different sub-bands in a hierarchical tree order. The coefficient pairs from each subband are arranged in a tree structure. This is followed by vector quantization of tree structured data sequences. A collection of codevectors called codebook plays very important role in vector quantizer. Codebook is generated using method given by Linde Buzo Gray [42]. In this method, training vectors are obtained from DWT coefficients of some ECG signal. Then for an input tree vector, its closest tree code vector is found from the codebook. The index of the codevector (in the codebook) which can best represent the input vector forms the encoded symbol. The same codebook is used on the decoder side for decompression. For updation of the codebook, an online Distortion constrained codebook replenishment(DCCR) mechanism is used. The principle of DCCR is that distortion for any input is constrained by the predetermined threshold value i.e. new codevectors are inserted only if difference between input code vector and codevector available in the codebook is greater than the threshold value. It gives better PRD as well as good compression ratio, but this algorithm is computationally intensive due to introduction of codebook [15][22][40].

3.3.4 Optimal Wavelets for Biomedical Signal Compression

In this method, a wavelet parameter optimization is done, which significantly improves performance as wavelet is selected on a signal by signal basis i.e. signal is compressed on the basis of signal dependent wavelets. The mother wavelet is selected from a library of function or is chosen by comparing the results of a few wavelets applied to a set of experimental signals. However best wavelet depends on the specific features of the signal to be compressed and thus, reconstruction error is minimised for optimal compression. To adapt the mother wavelet to the signal, it is necessary to define (1) Family of wavelets that depend on a set of parameters and (2) quality criteria for wavelet selection(i.e. wavelet parameter optimization). A natural
performance criterion is the minimization of the signal distortion rate for the given desired compression rate. For coding the wavelet coefficients, EZW technique is used with the wavelet optimization but any coding technique can be used with the proposed wavelet optimization. The algorithm showed significant better performance with respect to EZW algorithm applied with fixed mother wavelet [41].

3.4 PARAMETER EXTRACTION BASED COMPRESSION TECHNIQUES

The parameter extraction method is an irreversible process with which a particular characteristic or parameter of the signal is extracted. The extracted parameters (e.g. measurement of the probability distribution) are subsequently utilized for classification based on a priori knowledge of the signal features. Waveform coding methods simply try to model the waveform as closely as possible. Vocoding (voice coder) techniques greatly reduce the required storage space of speech signals by encoding information based on parameter extraction compression technique. Instead of trying to encode the waveform itself, parametric modelling techniques try to determine parameters about how the speech signal was created and use these parameters to encode the signal. To reconstruct the signal, these parameters are fed into a model of the vocal system, which outputs a speech signal. Several types of parametric models exist, the oldest one being around since even 1939. They all use this simple representation of the speech production system. One of the most common vocoder is linear predictive coding (LPC). This method has found extensive use in military applications, where a high quality speech is not as important as a low bit rate to allow for heavy encryption of secret data. Following are some of the parameter extraction based techniques used mostly for speech signal compression but have been used for EMG signal compression recently.

3.4.1 Linear Predictive Coding (LPC)

LPC is one of the most powerful speech coding techniques and is also being used for biomedical signal coding. LPC models the speech as an autoregressive process and sends the parameters of the processed signal as opposed to sending the speech signal itself. LPC involves an analysis or encoding part and a synthesis or decoding part. In the encoding part, LPC takes the speech signal in blocks or frames of speech and determines the input signal and the coefficients of the filter that will be
capable of reproducing the current block of speech. This information is quantized and transmitted. A standard, which utilizes simple LPC encoding, is LPC-10. It calculates 10 coefficients per frame of the speech signal. For LPC-10, the bit rate is about 2.4 kbps. Even though this method results in an artificial sounding speech, it is intelligible [16].

Waveform coders in general do not perform well at data rates below 16 kbps. Vocoders on the other hand, can produce very low data rates while still allowing intelligible speech. However, the person producing the speech signal often cannot be recognized and the algorithms usually have problems with background noise. Hybrid coders try to exploit the advantages of both techniques. They encode speech in such a way that a low data rate is achieved while keeping the speech intelligible and the speaker recognizable. Typical bandwidth requirements lie between 4.8 kbps and 16 kbps [16]. Techniques based on Linear Predictive Coding with some modifications are given as follows.

**Residual Excited Linear Prediction (RELP)**

The RELP coder works in almost the same way as the LPC coder. LPC exploits the redundancies of a speech signal by modelling the speech signal as a linear filter, excited by signal called excited signal or residual signal. Speech coders process a particular group of samples called a frame or segment. The filter coefficients and the excited signal for each frame are derived in such a way that the energy at the output of the filter for that frame is minimised. This filter is called an LP analysis filter. To analyze the signal, the parameters for the vocal tract filter are determined and the inverse of the resulting filter is applied to the signal. The input signal is first filtered through LP analysis filter. The resulting signal is called residual signal for that particular frame. This residual signal acts as excitation signal for the LP synthesis filter in the decoder. The LPC coder then checks if the signal is voiced or unvoiced and uses this to model an excitation signal. In the RELP coder however, the residual is not analyzed any further, but used directly as the excitation for speech synthesis. The residual signal is compressed using waveform-coding techniques to lower the bandwidth requirements. RELP coders can allow good speech quality at bit rates in the range of 9.6 kbps [16].
**Multipulse LPC (MPC)**

Multipulse LPC is a well suited model for EMG signals. It is essentially for speech coding but EMG signals can be seen as the outputs of linear filters driven by a sequence of pulses. In this method of LPC, a stream of signals is modelled as the output of all pole filter driven by an excitation signal function consists of a pulse sequence containing a small number of pulses defined by their location and amplitude. Excitation pulses are computed for each frame of the signal. By increasing the number of excitation pulses, quality of synthesized signal can be improved. It has been shown that only a small number of pulses (4 to 10) for each sub frame are enough to produce an acceptable quality of synthesized signal. The positions and amplitudes of the pulses are determined by AbS procedure. The MPC method can produce high quality speech at rates around 9.6 kbps [19].

**Codebook Excited Linear Prediction (CELP)**

The CELP coder tries to overcome the synthetic sound of vocoders by allowing a wide variety of excitation signals, which are all captured in the CELP codebook. This is similar to Multipulse coder. The main difference is found in filter excitation. In the CELP coder, Gaussian values are used for excitation signal. Other difference is, it involves vector quantization (introduction of a codebook) on the excitation signal. To determine which excitation signal to use, the coder performs an exhaustive search. For each entry in the codebook, the resulting speech signal is synthesized and the entry, which created the smallest error, is then chosen. The excitation signal is then encoded by the index of the corresponding entry. So basically, the coder uses Vector Quantization to encode the excitation signal [20].

A CELP coder is used at 12.2 Kb/s rate for EMG signal. The EMG signal is divided into 160 samples/frame without pre-processing. Since this ensured a good trade off between coding delay and performance, each 160 samples frame was further divided into 40 samples sub-frames corresponding approximately 39 ms. Autoregressive parameters were then computed on these sub-frames. The all pole filter corresponding to AR model captures the shape of the power spectrum of the signal. In time domain, the short term correlation among samples called short term predictor (STP) filter is used as shown in Fig 3.4. Long term correlation is modelled by long term predictor. With the help of these two predictors, signal is faithfully reconstructed by using proper excitation signal in the decoder. For this purpose residual error signal from the...
two filters is vector quantized with analysis by synthesis approach. The quantization index is sent together with the filter parameters to the decoder. This method was used for surface EMG signals. It preserves the signal shape (PRD smaller than 10%) at a price of lower compression factor (87.3%).

Fig 3.4 Block diagram of the synthesis model
This technique is also called analysis-by-synthesis (AbS) technique because it analyses a signal by synthesizing several possibilities and choosing the one, which caused the least amount of error. This exhaustive search is computationally intensive. However, fast algorithms have been developed to be able to perform the search in real-time. CELP techniques allow bit rates up to 4.8 kbps [19] [20].

2.5 CONCLUSION
In this chapter, various types of data compression algorithms, for example, SPIHT, Vector quantization, Embedded Zero tree Wavelet coding algorithm, Optimal wavelet algorithm, Linear Predictive Coding, RELP, MPC, and CELP techniques used for one dimensional biomedical signals are discussed. Among all the data compression techniques, transform based compression techniques are most suitable for EMG signal compression because transform methods provide better compression ratio along with very good reconstructed signal quality. Direct data compression techniques are good as far as quality of the signal is concerned but do not give enough compression ratio. It has been discussed that parameter extraction algorithms are more suitable for speech signal but have been used for EMG signal compression also with good results.