AN EFFICIENT DIFFERENTIAL ENCODING BASED COMPRESSED SENSING CODEC DESIGN FOR HIGHLY CORRELATED DATA

3.1 INTRODUCTION

To increase the sparsity of the signal, differential encoding of data is suggested for highly correlated data, since CS performs better only when the signal is sparse in nature. Input signal of highly correlated nature are pre-processed by differential encoding and compressed using standard compressive sensing algorithm. Figure 3.1(a) shows the basic steps involved in differential encoding based CS algorithm.

![Diagram](image)

Figure 3.1(a) : Differential encoding based CS algorithm
In the proposed differential encoding based CS algorithm, the first measurement is kept as same and the differences between the consecutive measurements are taken to construct a new input signal. This new input signal is then compressed using CS algorithm. The compressed measurements are then applied to a quantizer which works on nearest neighbour coding principle. The quantized values are encoded using codebooks and appropriate quantization parameters are transmitted to the decoder of the receiver for reduction of the number of transmission. The design of codebook is done using Lloyd’s algorithm. Transmissions of quantization parameters are done using different modulation schemes namely QPSK and QAM. The quality of the suggested algorithm is measured by calculating the mean squared error (MSE) and percentage of root-mean-square difference (PRD) value.

3.2 RELATED WORK

In [50], regularized maximum likelihood and regularized least square methods for carrying out quantized measurement based Compressed Sensing are analyzed. The numerical simulation results of both the methods are verified in this work. Both the methods perform equally well. However the effectiveness of the methods is not discussed with respect to Signal to Noise Ratio (SNR) and Mean squared error (MSE).

A simple and efficient encoder device for measurement of signals within sensor nodes in a WBAN is proposed in [51]. An analog CS based encoder and a model of digital version of CS based encoder are presented in the same work. Comparison of a hardware prototype with a SPICE model is also performed. The result of this paper shows that compressed sensing encoder is much useful for biomedical signals however, it does not discusses the effectiveness of CS encoder for other types of signals.

For wireless multimedia sensor networks, an effective encoder design named as Compressed Sensing based Prediction Measurement (CSPM) is presented in [52]. CSPM encoders are much suitable for wireless multimedia sensor networks. A considerable reduction of storage of data and
transmission energy saving is achieved by CSPM encoding. In order to measure the compression performance of CSPM method, the parameters like compression ratio and bit rate are used. The video signal is used for analysis of the method proposed. However this paper does not analyze the signals other than video signal.

In [53], compressed sensing based compression algorithms for data compression in WSNs are used for addressing the energy and telemetry bandwidth constraints. As WSNs are resource constrained networks and hence implementations of CS systems must be energy efficient. Analog and digital circuit models of CS based systems are presented in this work. These systems enables analysis of power based performance costs in much simpler manner. However there is no analysis regarding qualitative performance cost such as SNR, MSE, PRD etc.

For upgrading the quality of CS reconstruction, In [54] a proposal is suggested for CS based reconstruction. This includes optimized CS reconstruction with combination block along with gradient spare operators. The main advantage of the suggested method is that, it avoids changes in encoder side rather than improvement on the decoder side. This suggested work is fully concentrated on decoder design however no encoder design is explained here.

### 3.3 DIFFERENTIAL ENCODING BASED CS CODEC

The model for the proposed compressed sensing CODEC is given in Figure 3.1(b). Explanation of the model is presented in following sections of this chapter. In Figure 3.1(b), the input signal is denoted by symbol X whereas the reconstructed signal is denoted by X'.


3.3.1 Differential encoding for enhancement of level of sparsity

The input signals considered for testing are temperature data obtained from the online database [55]. The dataset chosen is having very high mean correlation coefficient of 0.93. Differential sequence is generated by calculating the difference between the consecutive samples of input data to enhance the sparsity of the signal. After obtaining the difference signal the CS based compression is performed.

Sparsity measurement is an important parameter measurement in sparse signal processing. In this chapter, the sparsity of the input measurements are improved by considering the difference of input samples rather than considering the raw input samples. In general ‘\(l_1\)’ norm of a vector has the property of finding coefficients with zero values and hence ‘\(l_1\)’ norm can be used to find the level of sparsity of any measurement vector. Sparseness is an important parameter used to find the signal sparsity. If sparsity of a signal is high then better is the CS reconstruction.
3.3.2 Differential encoding based CS

The input data ‘x’ shown in system model of Figure 3.1 is pre-processed by performing differential encoding and then compressed using standard CS. Here the temperature sensor data are collected from database [55]. In order to generate sparse signal, discrete cosine transform (DCT) is used as the basis function. The sparse data is compressed using a random sensing matrix. The CS encoding is performed based on the equations 2.1 and 2.2 discussed in chapter 2. The compressed signal is then quantized in form of codebook and appropriate indices of codebook are transmitted to the decoder.

3.3.3 Quantization using nearest neighbour coding

The process of mapping samples from a continuous set of samples of a signal into a discrete alphabet using fixed codebooks is termed as quantization. In this chapter the quantization used is based on the concept of nearest neighbor coding principle. In the nearest neighbour coding quantization process, coding of each entry of a measurement vector is done by its nearest neighboring code point value. A fixed codebook is designed to map the measurement vectors of a signal. The entries in the code book are chosen in such a way that MSE is minimized. However this fixed type of codebook design does not ensure the minimization of end to end MSE value. As the sparse reconstruction techniques are non linear and also CS sensing matrix are non-orthogonal, hence end to end MSE cannot be guaranteed. In Figure 3.1, the fixed sized codebooks are used for CS encoding and decoding.

3.3.4 Codebook design based on Lloyd’s algorithm

A data compression technique based on differential encoding and codebook design is proposed for highly correlated data compression. To achieve high compression ratio, the compressed signal is quantized and the quantization parameters are transmitted from transmitter to the receiver. In quantization process the large numbers of samples of input signal are
mapped into output values in a smaller set. Partition and codebook are two important quantization parameters. It is very common to have signal distortion in quantization process and the distortion can be reduced by choosing appropriate partition and codebook parameters. Quantization parameters are optimized using the Lloyd’s function. The Lloyd’s function works based on Lloyd algorithm [56]. Lloyd’s function optimizes the partition and codebook parameters.

An example is used to explain the quantization partition and codebook. As it is well known that, partition in quantization is nothing but several neighboring, non-overlapping ranges of values within the set of real numbers. The following example is used to explain the concept of partition. If the partition separates the input signal or real number line into the four sets as:

\[
\begin{align*}
(1) & : [x(n) \leq 0] \\
(2) & : [0 < x(n) \leq 1] \\
(3) & : [1 < x(n) \leq 3]
\end{align*}
\]

Then partition as the three-element vector can be represented as: partition = [0, 1, 3].

A codebook provides the knowledge to the quantizer which common value to assign to inputs that fall into each range of the quantization partition. In the example explained here, the codebook is represented by [-1, 0.5, 2, 3]. This codebook represented here is one of the possible codebook for the partition [0, 1, 3]. There are many more possible codebooks for the example given here but Lloyd’s algorithm is used to design an optimum codebook. The steps involved in Lloyd's algorithm are discussed in Figure 3.2.

Assuming the codebook ‘\(y_i\)’ which is to be generated using Lloyd’s algorithm are fixed. The input ‘\(x\)’ is obviously nearest to one of the representative levels. The quantization result of the input ‘\(x\)’ can be at most as small as that minimum distance.
Algorithm : Lloyd’s algorithm for codebook generation

Step 1: Start with initial representative levels \( y_i \)

Step 2: Find the optimum interval boundaries \( (x_i) \) corresponding to the \( y_i \)'s according to equation (3.2):

\[
x_{i, \text{opt}} = \frac{1}{2}[y_{i, \text{opt}} + y_{i-1, \text{opt}}].
\]  

(3.2)

Step 3: Recalculate new representative levels's according to the equation (3.3):

\[
y_{i, \text{opt}} = \frac{\int_{x_{i, \text{opt}}^{x_{i+1, \text{opt}}}} x f(x) \, dx}{\int_{x_{i, \text{opt}}^{x_{i+1, \text{opt}}}} f(x) \, dx}
\]  

(3.3)

In above equation 3.3, \( f(x) \) is the probability density function (pdf) of Input signal ‘x’.

Step 4: If new \( y_i \)'s are very different from old \( y_i \)'s, then go to step 2.

The above mentioned steps are repeated until the optimum quantization parameters are achieved.

Figure 3.2: Lloyd’s algorithm for codebook generation

The step by step procedure for finding the quantization parameters using well known Lloyd’s algorithm is shown in Figure 3.2.

3.3.5 Modulation

A very important and essential step involved in any communication system is modulation. It is the technique of converting the information bits into a form which can be transmitted via radio in WSN. In a wireless sensor environment, the data collected from the sensor node has to be modulated and transmitted to a sensor head or any base station for further monitoring.
and processing of the sensor data. The reconstruction quality of signal is also
dependent on the modulation scheme to be chosen in WSN and hence it is
important to choose the modulation techniques which are best suited for
transmission in WSN.

The following two modulation schemes are used here as in literature
these two techniques are mentioned as suitable modulation scheme in WSN
[57] :

(i) Quadrature amplitude modulation (QAM)

(ii) Quadrature Amplitude phase shift keying (QPSK)

(i) Quadrature Amplitude Modulation (QAM)

QAM scheme may be of analog or digital. In QAM, there are two carriers
and both the carriers are having same frequency but differed by $90^0$ phase
shift which is similar to phase modulation (PM) or phase shift keying (PSK).
The resultant QAM waveform is a mixture of both phase-shift keying (PSK)
and amplitude-shift keying (ASK) in case of digital modulation. In the analog
case, it is the combination of phase modulation (PM) and amplitude
modulation (AM).

In case of QAM, a finite number of at least two phases and at least two
amplitudes are used. Selecting a suitable constellation size in QAM,
high spectral efficiencies can be achieved. The limitation parameters of QAM
modulation are noise level and linearity of the communications channel. As it
is required to use two RF carriers in QAM modulation, these two RF carriers
are referred to as the I - In-phase and Q - Quadrature components.

The representation of I and Q components of QAM in terms of
trigonometric functions can be given by the equations below:

$$I = A \cos(\psi)$$  \hspace{1cm} (3.4)

and  \hspace{1cm} $$Q = A \sin(\psi)$$  \hspace{1cm} (3.5)
In equation (3.4), ‘I’ component is represented by cosine term. In equation (3.5), ‘Q’ component is represented by sine term. These two RF carriers are 90° out of phase with one another.

The QAM modulated carrier in analog form can be represented using the expression:

\[ A \cos(2\pi f t + \Psi) = I \cos(2\pi f t) - Q \sin(2\pi f t) \]  

(3.6)

In above expression (3.6), ‘f’ is the carrier frequency. QAM expression shown in equation 3.6 is a periodic waveform in which the phase can be adjusted with the help of changing the amplitude in ‘I’ and ‘Q’ components. This phase change may also result in amplitude change as well.

(ii) Quadrature Phase Shift Keying (QPSK)

Quadrature Phase Shift Keying (QPSK) is a phase modulation technique where the numbers of symbols or states used are four. If the numbers of states are two, then it is referred as BPSK. BPSK refers to PSK with only two states where as PSK with 4 states can be termed as QPSK.

As in QPSK there are four phases, hence it can be encoded as two bits per symbol, which in turn increases the data rate. The original data stream In QPSK can be represented as:

\[ a_k(t) = [a_0, a_1, a_2, ... ] \]  

(3.7)

Equation (3.7) is divided into a quadrature stream, ‘aq(t)’ and an in-phase stream, ‘ai(t)’ and can be given as below.

\[ a_i(t) = [a_0, a_2, a_4, ... ] \]  

(3.8)

\[ a_q(t) = [a_1, a_3, a_5,...] \]  

(3.9)
Each of ‘ai(t)’ and ‘aq(t)’ have half the bit rate of ‘ak(t)’. Orthogonal realization of a QPSK waveform, S(t), is mainly achieved by amplitude modulating the in-phase and quadrature data streams onto the cosine and sine functions of a carrier wave as follows.

\[
S(t) = [0.707 \times a_i(t) \cos(2\pi ft + \pi/4)] + [0.707 \times a_q(t)\sin\left(2\pi ft + \frac{\pi}{4}\right)] \tag{3.10}
\]

The above equation can also be represented using trigonometric identities as follows:

\[
S(t) = \cos\left[ 2\pi ft + \varphi(t) \right] \tag{3.11}
\]

The value of \( \varphi(t) \) will correspond to one of the four possible combinations of \( a_i(t) \) and \( a_q(t) \) in the equation (3.11).

\[
\varphi(t) = 0^\circ, \pm 90^\circ, \text{ or } 180^\circ.
\]

In this chapter, the sensor readings are quantized using codebooks which are generated using Lloyd’s algorithm and the corresponding indices from the codebooks are transmitted to the decoder. The decoder uses the indices and decodes the values using the fixed codebooks available at the receiver side. The indices are modulated using any one of the modulation scheme mentioned above and transmitted to decoder. The decoder demodulates the received indices values and decodes the transmitted data using codebooks. Figure 3.1 describes the system along with modulation.

The performance of the modulation schemes are analyzed using the ideal channel with additive white Gaussian noise (AWGN). The quality of the reconstructed signal is measured in terms of mean squared error.

### 3.3.6 CS demodulation and decoding

As the design of the encoding algorithm is generic hence any suitable reconstruction algorithm can be used. In this chapter, Orthogonal Matching Pursuit (OMP) is used as CS reconstruction algorithm.
3.3.7 Orthogonal Matching Pursuit (OMP) reconstruction

OMP [47] is a class of algorithm that iteratively builds up a sparse solution. In this algorithm the signal is decomposed into linear expansion of functions that creates a dictionary. In the first step of OMP algorithm, identification of non zero coefficient is done by finding the maximum inner product with the residue of dictionary elements. In second step, orthogonal projection of the residue over the set of selected element is done to find the best match of the signal.

3.4 RESULTS AND DISCUSSION

The performance evaluation of the suggested differential encoding based CS algorithm is done in terms of mean squared error for different sizes of codebooks. The codebooks are designed and optimized using Lloyd’s algorithm. Total three datasets with size of 1000 samples (dataset1), 1200 samples (dataset2) and 200 samples (dataset3) are considered as input data to the proposed algorithm. The datasets are obtained from the database available in the literature [55]. For measurement of level of sparsity there is no direct formula available but the level of sparsity can be measured by finding the ‘l_1’- norm of the signal. If the value of ‘l_1’- norm is less, then the level of sparsity is more [58] which is shown below in Table 3.1. One more parameter is used to define the sparsity [59] of the signal or vector given by equation (3.12).

\[
\text{Sparseness of } (x) = \frac{\sqrt{k} - (\| x \|_1 / \| x \|_2)}{\sqrt{k-1}}
\]  

(3.12)

Where ‘x’ is the vector whose sparseness to be measured and ‘k’ is the length of the vector. The ‘l_1’- norm is denoted by \( \| x \|_1 \), and \( \| x \|_2 \) is the ‘l_2’-norm. ‘l_1’- norm of a vector x can be expressed by:

\[
\| x \|_1 = \sum_{i=1}^{n} | x |
\]  

(3.13)
The $l_2$ norm or two norm of a vector $x$ is given by the equation 3.14.

$$\| x \|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

(3.14)

Whenever the vector is dense, the sparseness approaches to zero and if Sparseness of $(x)$ approaches to 1, the vector is sparse [59].

To increase the sparsity of the signal, differential sequence or delta sequence is generated by finding the difference between the consecutive samples of input data. Hence the numbers of zeros have been increased as the WSN dataset consist similar data value.

Table 3.1: Comparison of $l_1$ norm value and the level of sparsity

<table>
<thead>
<tr>
<th>Data Set Name</th>
<th>Input Signal (x)</th>
<th>$l_1$ norm value</th>
<th>Level of Sparsity or Sparseness of $(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set-1</td>
<td>Input Data without differential encoding</td>
<td>$4.59 \times 10^4$</td>
<td>$2.1486 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>Input Data with differential encoding</td>
<td>$0.0023550 \times 10^4$</td>
<td>$0.3209$</td>
</tr>
<tr>
<td>Data Set-2</td>
<td>Input Data without differential encoding</td>
<td>$2249$</td>
<td>$2.1425 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>Input Data with differential encoding</td>
<td>$28.27$</td>
<td>$0.3199$</td>
</tr>
</tbody>
</table>

The size of the input is 1000 samples, the MSE is 2.0750 for normal input data with QAM modulation, whereas with differential encoding the MSE is 0.0161 which is shown in Table 3.2 and Table 3.3. Thus the results shows that when 1000 measurements are compressed to 500 measurements, there
is an reduction in MSE and quality of the reconstructed signal is improved with differential encoding based compressed sensing technique. In case of another temperature dataset (dataset2) mentioned earlier also shows the better result in case of differential encoding scheme with codebook size 8.

Table 3.2: Performance comparison in terms of mean squared error for normal data and difference input data with size of the codebook=8 with QAM modulation

<table>
<thead>
<tr>
<th>Data Set Name</th>
<th>Modulation Type</th>
<th>Compression Ratio</th>
<th>MSE without differential encoding</th>
<th>MSE with differential encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set-1</td>
<td>QAM</td>
<td>50%</td>
<td>2.075</td>
<td>0.0161</td>
</tr>
<tr>
<td>Data Set-2</td>
<td>QAM</td>
<td>50%</td>
<td>4.9079</td>
<td>0.0140</td>
</tr>
</tbody>
</table>

Table 3.3 Performance comparison in terms of MSE with size of the codebook=8 with QPSK modulation

<table>
<thead>
<tr>
<th>Data Set Name</th>
<th>Modulation Type</th>
<th>Compression Ratio</th>
<th>MSE without differential encoding</th>
<th>MSE with differential encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set-1</td>
<td>QPSK</td>
<td>50%</td>
<td>4.5443</td>
<td>0.0191</td>
</tr>
<tr>
<td>Data Set-2</td>
<td>QPSK</td>
<td>50%</td>
<td>2.1303</td>
<td>0.0187</td>
</tr>
</tbody>
</table>
Table 3.4 Performance comparison in terms of mean squared error with size of the codebook=16 with QAM modulation

<table>
<thead>
<tr>
<th>Data Set Name</th>
<th>Modulation Type</th>
<th>Compression Ratio</th>
<th>MSE without differential encoding samples</th>
<th>MSE with differential encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set-1</td>
<td>QAM</td>
<td>50%</td>
<td>2.0383</td>
<td>0.0146</td>
</tr>
<tr>
<td>Data Set-2</td>
<td>QAM</td>
<td>50%</td>
<td>4.8775</td>
<td>0.0139</td>
</tr>
</tbody>
</table>

Table 3.5 Performance comparison in terms of mean squared error with size of the codebook=16 with QPSK modulation

<table>
<thead>
<tr>
<th>Data Set Name</th>
<th>Modulation Type</th>
<th>Compression Ratio</th>
<th>MSE without differential encoding samples</th>
<th>MSE with differential encoding samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set-1</td>
<td>QPSK</td>
<td>50%</td>
<td>4.5195</td>
<td>0.0136</td>
</tr>
<tr>
<td>Data Set-2</td>
<td>QPSK</td>
<td>50%</td>
<td>4.587</td>
<td>0.0177</td>
</tr>
</tbody>
</table>

The size of the codebooks is based on the application and availability of resources. There is a tradeoff between the size of code book and error.
Based on the requirement and error criterion, the size of the code book is decided. The performance results using 16 code points are tabulated in Table 3.4 and in Table 3.5. As the size of the codebook increases there is further reduction in the error. Figure 3.3 and Figure 3.4 show the reconstructed data without and with differential encoding with QAM modulation.

![Graph showing input and reconstructed signal](image)

**Figure 3.3:** Input and reconstructed signal without differential encoding and QAM modulation

In Figure 3.4, the reconstructed data with differential encoding shows better reconstruction quality compared to without differential encoding shown in Figure 3.3.
Figure 3.4: Input and reconstructed signal with differential encoding and QAM modulation

Dataset -3 of 200 samples is also considered and reconstructed data with and without differential encoding is plotted below in Figure 3.5 and Figure 3.6. In this case the result shows improved performance in proposed method.

Figure 3.5: Input and reconstructed signal without differential encoding and QAM modulation
In above Figure 3.5, the reconstructed signal does not match the input original signal. The input and reconstructed signal with differential encoding based CS are plotted in Figure 3.6.

![Figure 3.6: Input and reconstructed signal with differential encoding and QAM modulation](image)

The Percentage of Root-mean-square Difference (PRD) between the reconstructed and the original signal is used as the evaluation index for reconstruction quality [60]. PRD is defined as:

\[
PRD = \frac{\| \| (x - \hat{x}) \|_2 \| .100}{\| x \|_2}
\]  

(3.15)
Where \( \hat{x} \) denotes the reconstructed signal and \( x \) denotes the original signal. \( \| \cdot \|_2 \), defines the \( \| \cdot \|_2 \) -norm of the vector. The smaller is PRD value, the better is the reconstruction quality [61].

Table 3.6 shows the comparison of PRD values of data with and without differential encoding. It is very clear that when the PRD value is nearer to zero it ensures the perfect reconstruction in case of differentially encoded data.

Table 3.6 Comparison of the PRD value for normal data and differentially encoded data with size of the codebook=16 with QAM Modulation

<table>
<thead>
<tr>
<th>Data Set Name</th>
<th>Modulation type</th>
<th>Compression Ratio</th>
<th>PRD in case of normal input samples</th>
<th>PRD in case of difference input samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set-2</td>
<td>QAM</td>
<td>50%</td>
<td>12.4002</td>
<td>0.0339</td>
</tr>
<tr>
<td>Data Set-3</td>
<td>QAM</td>
<td>50%</td>
<td>14.01</td>
<td>0.2823</td>
</tr>
</tbody>
</table>

3.5. CONCLUSION

The proposed differential encoding based CS is analyzed and the proposed algorithm improves the quality of the reconstructed data by reducing the mean squared error value. Reconstructed signal quality is also measured by the percentage of root-mean-square difference (PRD) value. The PRD value between the reconstructed and the original signal is used as
the evaluation index for reconstruction quality which is explained in this chapter. The lesser the PRD value the better is reconstruction quality. The simulation result shown here confirms that with differential input signal the PRD is minimum whereas without differential input the PRD value is high. The number of measurements which are transmitted are made sparse using compressive sensing technique. The compressed measurements are then quantized using Lloyd’s algorithm and quantization parameters are transmitted using suitable modulation schemes. Not only the PRD value, the MSE is also found minimum in case of proposed technique. The comparison is done with normal data and differential data by keeping the compression ratios as constant. It is found from the result that, the reconstruction quality is improved using differential encoding based compressed sensing technique.