COMPRESSIVE SENSING FOR WIRELESS AND SENSOR SYSTEMS

by

JI WU

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2013

To my family

ACKNOWLEDGEMENTS

I would like to thank my supervising professor Dr. Qilian Liang for motivating and inspiring me, and also for his advice during the course of my doctoral studies. I wish to thank my academic advisors Dr. Alan Davis, Dr. William Dillon, Dr. Ioannis Schizas, Dr. Saibun Tjuatja for their interest in my research and for taking time to serve in my dissertation committee.

I would also like to thank my family for providing support for my doctoral studies. Without them, I cannot finish my journey.

I am grateful to all the teachers who taught me during the years. I would like to thank Dr. Zheng Zhou from Beijing University of Posts and Telecommunications for encouraging and inspiring me to pursue Ph.D. in United States.

I am also extremely grateful to my wife for her sacrifice, encouragement and patience. She always stands behind me and support me for whatever I want to do.

August 12, 2013

ABSTRACT

COMPRESSIVE SENSING FOR WIRELESS AND SENSOR SYSTEMS

Ji Wu, Ph.D.

The University of Texas at Arlington, 2013

Supervising Professor: Qilian Liang

Compressive sensing (CS) is an emerging field based on the revelation that a small collection of linear projections of a sparse signal contains enough information for stable, sub-Nyquist signal acquisition. It provides a potential way to acquire the sparse data efficiently, or equivalently, highly accurate recovery of sparse data from undersampled measurements.

Huffman coding and compressive sensing are adopted to compress real-world wind tunnel data. Both uniform and non-uniform Huffman coding are evaluated in terms of the number of quantization levels, mean square error, codeword length and compression ratio. The main drawback of Huffman coding is that it requires calculating the probability of each symbol before encoding. It means it may not be appropriate for real-time compression. We applied CS to wind tunnel data and compared its performance against the theoretical error bound.

Due to limited energy and physical size of the sensor nodes, the conventional security mechanisms with high computation complexity are not feasible for wireless sensor networks (WSNs). A compressive sensing-based encryption is proposed for distributed WSNs, which provide both signal compression and encryption guarantees,

without the additional computational cost of a separate encryption protocol. The computational and information-theoretical secrecy of the compressive sensing algorithm is also investigated. For the proposed distributed WSNs, if only a fraction of randomizer bits is stored by an eavesdropper, then the eavesdropper cannot obtain any information about the plaintext.

We studied a compressive sensing-based Ultra-WideBand (UWB) wireless communication system. Compared with the conventional UWB system, it can jointly estimate the channel and compress the data. No information about the transmitted signal is required in advance as long as the channel follows the autoregressive model. The performance of compressive sensing-based data encryption scheme shows that the original data could never be reconstructed when the measurement matrix is not available. Hence, compressive sensing can be implemented as a data encryption scheme with good secrecy.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Compressive Sensing Overview

Compressive Sensing (CS) is a recently emerging field based on the revelation that a small collection of linear projections of a sparse signal contains enough information for stable, sub-Nyquist signal acquisition. It has attracted much attention in the signal processing community in the last few years in view of its application in medical imaging [1]-[3], analog-to-information conversion [4]-[6], compressive radar [7]-[9], and communications [10]-[13]. CS can be viewed as a scheme for simultaneously sensing and compression, and the required data acquisition rate is only proportional to the sparsity of the original signal, which has been subjected to extensive research [14]-[22]. There is also discussion on some interesting observations on the recovery of sparse signals [14], i.e., signals which have only a few nonzero terms, from limited measurements. In [15], the author shows that it is possible to reconstruct a sparse signal with a very high accuracy from various types of random measurement ensembles, e.g., binary, Gaussian and Fourier ensembles. It shows that a greedy algorithm called Orthogonal Matching Pursuit (OMP) can recover the original signal with far less samples compared with matching pursuit (MP) [16] . The author discusses the number of measurements that are sufficient for recovery with high accuracy in [17].

1.1.1 Sparsity and Undersampled Signal Recovery

Compressive sensing (CS) provides a framework for integrated sensing and compression of discrete-time signals that are sparse in a known basis or frame. Many

natural signals have concise representations when expressed in the proper basis [21]. Mathematically speaking, consider a discrete signal $f \in \mathbb{R}^N$ which can be expanded in an orthonormal basis $\Psi = [\psi_1 \psi_2 \ \cdots \psi_n]$ as follows:

$$f(t) = \sum_{i=1}^{N} x_i \psi_i(t), \tag{1.1}$$

where x is the coefficient sequence of f. It can be computed from signal f:

$$x_i = \langle f, \psi_i \rangle, i = 1, 2, \cdots, N. \tag{1.2}$$

It will be convenient to express f as Ψx (where Ψ is the $n \times n$ matrix with $\psi_1 \psi_2 \ \cdots \psi_n$ as columns). The discrete signal f is K-sparse in the domain Ψ, $K << N$, only if K out of N coefficients in the sequence x are nonzero. Sparsity of signal is a fundamental principle used in the compressive sensing as well as in most modern lossy coders such as JPEG-2000 and many others, since a simple way for image compression would be to compute x from f and then only encode the values and locations of the largest K coefficients.Unfortunately, those compression schemes require computing all N coefficients of signal f and the locations of the significant coefficients, which may not be known in advance.

Compressive sensing suggests information in the signal f can be captured using a small number of random measurements of the signal. It provides a potential way to acquire the sparse data efficiently, or equivalently, highly accurate recovery of sparse data from undersampled measurements.

The new M-length observation vector y can be expressed by:

$$y = \Phi f, \tag{1.3}$$

where Φ is an $M \times N$ measurement matrix. Equation (1.3) can be written as

$$y = \Phi\Psi x = \Theta x, \tag{1.4}$$

where Θ is given by:

$$\Theta = \Phi\Psi, \tag{1.5}$$

The signal f can be perfectly recovered from M measurements, if Θ satisfies the so-called Restricted Isometry Property (RIP) [17].

It has been shown in [19] that choosing an independent and identically distributed (i.i.d) Gaussian random matrix as a sensing matrix Φ, Θ is also i.i.d Gaussian for various orthonormal bases, Ψ, such as spikes, sinusoids, wavelets, Gabor functions, curvelets, and so on.sible with high probability. It is noted that coefficients of f are not known in advance.

With the new observation matrix y, the signal f is recovered by ℓ_1–norm minimization; the proposed reconstruction f^* is given by $f^* = \Psi x^*$, where x^* is the solution to the convex optimization program($\|x\|_{\ell_1} \equiv \sum_i |x|$)

$$\min_{\tilde{x}\in\mathbb{R}^N} \|\tilde{x}\|_{\ell_1} \quad \text{subject to} \quad y = \Phi\Psi\tilde{x}, \tag{1.6}$$

That is, among all the objects $\tilde{f} = \Psi\tilde{x}$ consistent with the data, we choose the one whose coefficient sequence has minimal ℓ_1–norm. However, ℓ_1–minimization is not the only way to recover sparse solutions; other methods, such as the greedy algorithm [16], has also been proposed.

1.1.2 Partial Measurement of Compressive Sensing

We already know how to recover signals from far fewer samples. The remaining question is that how many measurements we need to have to get the original signal

fully recovered. It has been shown in [21] that for a fixed signal support T of size $|T| = S$, equation (1.6) recovers the overwhelming majority of x supported on T and observation subsets Ω of size

$$\|\Omega\| \geq C \cdot \mu^2(\Theta) \cdot S \cdot \log n, \tag{1.7}$$

where $\mu(\Theta)$ is simply the largest magnitude among the entries in Θ:

$$\mu(\Theta) = \max_{k,j} \|\Theta_{k,j}\| . \tag{1.8}$$

The results in (1.7) demonstrate the relationship between the sensing modality (Φ) and signal model (Ψ) affects the number of measurements required to reconstruct a sparse signal. The parameter μ can be rewritten as

$$\mu(\Phi\ \Psi) = \max_{k,j} |\langle \phi_k, \psi_j \rangle| , \tag{1.9}$$

and serves as a rough characterization of the degree of similarity between the sparsity and measurement systems. To emphasize this relationship, $\mu(\Theta)$, is often referred to as the mutual coherence [14]. The bound (1.7) tells us that a S-sparse signal can be reconstructed from $\sim S \log n$ samples in any domain in which the test vectors are 'flat', i.e., the coherence parameter is $O(1)$.

1.2 UWB and UWB Noise Radar Signal Overview

1.2.1 UWB Communication System

The UWB impulse radio communication systems provide a revolutionary approach to radio communications, supporting location aware applications such as smart homes/offices and object tracking and detection, wireless connectivity and sensor networks, intelligent transportation systems, or even machine controlled robots.

Rather than looking for still available but possibly unsuitable new bands, UWB radio technology is based on sharing already occupied spectrum resources by means

of the overlay principle. Unlike conventional narrowband techniques that use a combination of amplitude, frequency, and phase modulation of a continuous sinusoidal waveform to carry the information, UWB impulse radios can transmit the information using long sequences of pulses and modulate the data information onto certain parameters of the transmitted pulses, such as the impulse position, amplitude or orientation. UWB has several features that is totally different from conventional narrowband systems [56]:

- Large instantaneous bandwidth enables fine time resolution for network time distribution, precision location capability, or use as a radar.
- Short duration pulses are able to provide robust performance in dense multipath environments by exploiting more resolvable paths.
- Low power spectral density allows coexistence with existing users and has a Low Probability of Intercept (LPI).
- Data rate may be traded for power spectral density and multipath performance. It sends out very short duration pulses to convey information which means the duty cycle is very low.

1.2.2 UWB Noise Radar Signal

The time-frequency model for the UWB noise radar signal is expressed by [50]:

$$x(t) = a(t)\cos\{[\omega_0 + \delta\omega(t)]t\}, \tag{1.10}$$

where $a(t)$ is the Rayleigh distributed amplitude which describes amplitude fluctuations, and $\delta\omega(t)$ is the uniformly distributed frequency fluctuations over the $\pm\Delta\omega$ range, i.e., $[-\Delta\omega \leq \delta\omega \leq +\Delta\omega]$ [50]. Assuming that the random variables $a(t)$ and $\delta\omega(t)$ are uncorrelated, we can show that the average power of the signal x is $\langle a^2(t)\rangle / 2R_0$, where $\langle\cdot\rangle$ denotes time average and R_0 is the system impedance. The

center frequency, f_0, and the bandwidth, B, can be defined as $\omega_0/2\pi$ and $\Delta\omega/\pi$, respectively.

An alternative time–frequency representation of UWB noise radar signal is given by:

$$s(t) = s_I(t)\cos(\omega_0 t) - s_Q(t)\sin(\omega_0 t), \tag{1.11}$$

where $s_I(t)$ and $s_Q(t)$ are zero-mean Gaussian processes and $f_0 = \omega_0/2\pi$ is the center frequency. This can also be written as:

$$s(t) = a(t)\cos[\omega_0 t + \phi(t)], \tag{1.12}$$

where $a(t) = \sqrt{s_I^2(t) + s_Q^2(t)}$ is the Rayleigh distributed amplitude and $\phi(t) = \tan^{-1}(s_Q(t)/s_I(t))$ is the uniformly distributed phase.

CHAPTER 2

COMPRESSIVE SENSING FOR WIND TUNNEL DATA

In this Chapter, Huffman coding and compressive sensing are employed to compress real-world wind tunnel data. Both uniform and non-uniform Huffman coding are evaluated in terms of the number of quantization levels, mean square error, codeword length and compression ratio. The performance of Huffman coding is also compared with that of compressive sensing. The main drawback of Huffman coding is that it requires calculating the probability of each symbol before encoding. It means that it may not be appropriate for real-time compression. We applied CS to wind tunnel data compression and compared it against theoretical error bound.

2.1 Experimental Setup

Wind tunnel data used in this paper is available for download online (http://fris2.nist.gov/winddata/uwo-data/ss20-test1/ee1-ee2.html). Wind pressure is measured at hundreds of taps (sensors) on a structure, as shown in Figure 2.1. At each tap, the pressure is sampled 500 times per second for about 100 seconds. There are 37 wind angles over the range between 1800 and 3600 at 50 increments. We use the data set "ee1, ee2" from website . Based on a nominal full scale roof height (see the documentation on the website) wind speed of 84 MPH (approximated Hurricane Andrew condition), the sampled data are equivalent to about 22 samples per second for 0.64 hours in full scale for the open exposure tests and equivalent to about 29 samples per second for 0.48 hours in full scale for the suburban exposure tests. A 650-tap building yields a 40 MB file when sampled at 500 Hz for 60 seconds

or over 1.3 GB for a typical test with 36 wind directions. All of the samples were stored which means the amount of data is huge. Figure 2.2 shows the received signal at the first tap. The data looks like noise at the first glance, which indicates that the conventional compression scheme may not work for this data. In this paper, both Huffman coding and compressive sensing are employed to compress the wind tunnel dataset. Each dataset contains 49792 samples. To analyze the dataset in later experiments, it is parsed into small subsequence each with a length of 1000 samples. This is done because of the limitation of the memory and the CPU.

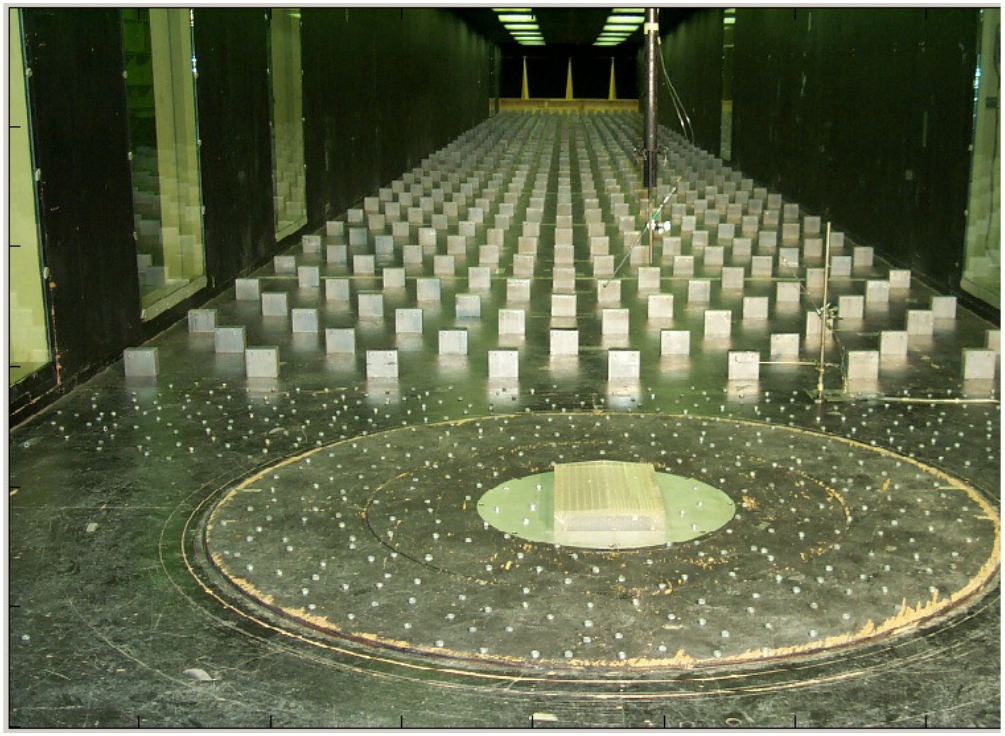


Figure 2.1. Demonstration of the wind tunnel sensor in the building.

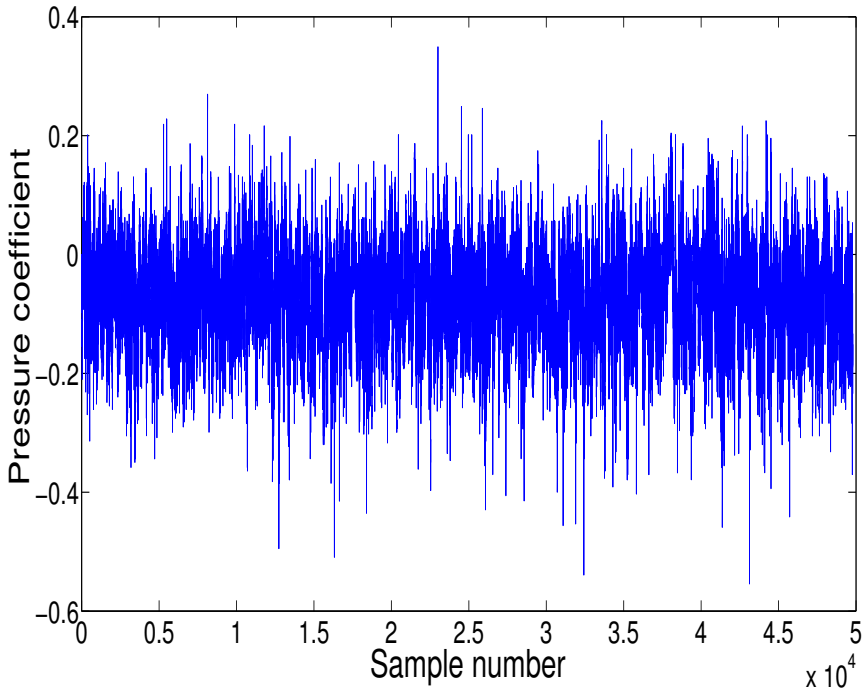


Figure 2.2. Time series of pressure coefficients.

2.2 Data Compression Analysis

Data compression is the process of encoding information using fewer bits than the original representation would use due to the limitation of bandwidth, power and storage, etc. All the images available on the website are compressed, typically in the JPEG or GIF formats, and HDTV will be compressed using MPEG-2, and several file systems automatically compress files when stored. Figure 2.3 shows a basic compression scheme. The raw data (X) is processed by an encoder and the result (B) is the compressed data, whose size is usually much smaller than original data size. To reconstruct the data, the compressed data is processed by a decoder. If the reconstructed data (X') is different from the original data, the compression is lossy; otherwise, the compression is lossless. The compression ratio is defined by the ratio between compressed size and uncompressed size, i.e.,

$$r = \frac{\text{size(B)}}{\text{size(X)}} \tag{2.1}$$

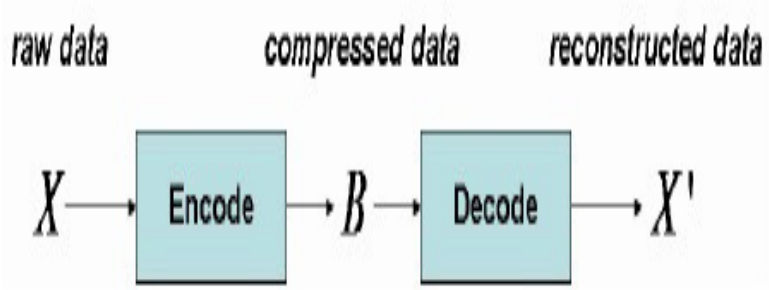


Figure 2.3. Compression scheme.

For lossy compression, there is a tradeoff between the compression ratio and the result quality. The higher the quality is, the larger the ratio will be. A common quality measurement is a peak signal to noise ratio (PSNR)

$$\text{PSNR} = 20\log_{10}[\frac{\text{MAX}}{\text{RMSE}}] \tag{2.2}$$

where MAX is the maximum possible value of the data and RMSE is the root mean squared error.

One of the key ideas of compression is to first transform the data to a new domain and then only keep the large coefficients. Usually, after transformation, most of the coefficients are close to zero. Therefore, the data in the new domain are more suitable for compression (due to lots of zeros).

All of the transformations share the same basic model:

$$x = Dw \tag{2.3}$$

where x is the signal to be approximated, D is the dictionary and w is the set of coefficients. There are lots of transformations such as FFT, DCT, Wavelet, etc. The difference between them is the dictionaries they use. Dictionaries based on the FFT, DCT and Wavelet are fixed and independent of data. These transformations do not require any prior knowledge of data.

2.2.1 Compression Using Huffman Coding

Huffman coding is a well-known lossless compression algorithm. The quantization process is introduced before compression in order to increase the compression ratio. Therefore, the compression is lossy due to the quantization error. Although the implementation of Huffman coding is quite simple, one major drawback of Huffman coding is that we need to calculate the probability of each symbol before the encoding process, which means it may not be appropriate for real-time processing. This is because a large amount of calculation and considerable time required. Another disadvantage of Huffman coding is that prior knowledge of data is required to generate the codebook: hence the compression process is not uniform.

8-level and 16-level uniform quantizations are first applied to the original data. Figure 2.4 depicts the waveform of an 8-level and a 16-level quantized data versus the original data. After the original data passes through the quantizer, each output signal can be represented by 8 or 16 symbols. Huffman coding is implemented by calculating the probability of appearance of each symbol. The average length and mean square error (MSE) of Huffman coding is shown in Table 2.1. On the other hand, if Huffman coding is not employed, the length required to represent the original data should be 13000 bytes (For each symbol, 1 byte for sign "+" or "-" and 12 bytes to represent the amplitude range from 0 to 4096.).

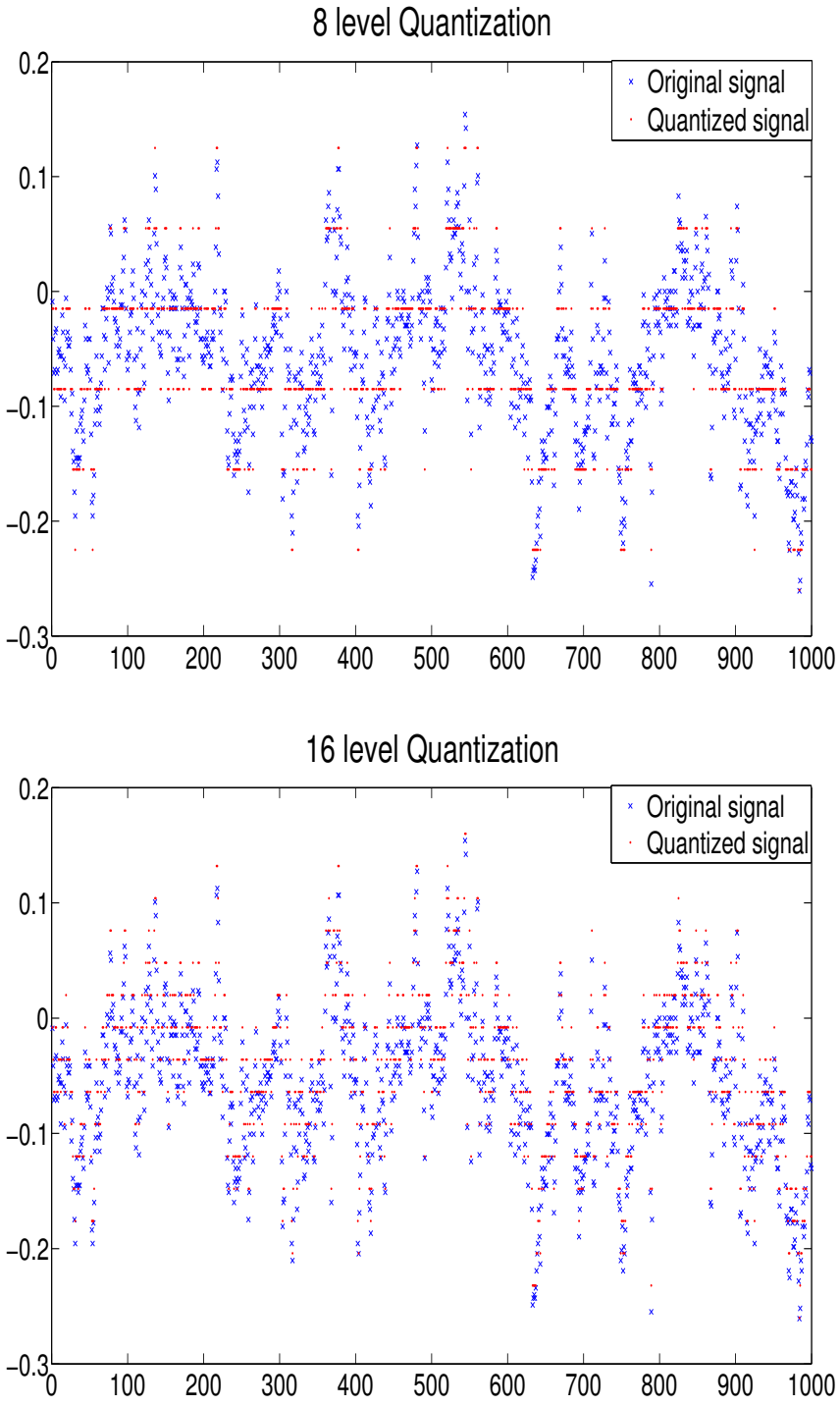


Figure 2.4. Quantized data versus the original data..

In order to minimize the quantization error, we further explore the relationship between quantization levels, mean square error and compression ratio. Figure 2.5 shows the waveform of 64-level, 128-level and 256-level quantized data compared with the original data. Mean square error (MSE), codeword length, and compression ratio for each case are shown in Tables 2.1 and 2.2. Non-uniform quantization using the Lloyd algorithm is also considered in our investigation. Simulation results are given in Table 2.3 and 2.4. From these tables, it is obvious that the performance of non-uniform quantization is much better than that of uniform quantization. In order to achieve an aggressive compression ratio, 256-level non-uniform quantization is adopted in our investigation.

Table 2.1. Performance of 8 and 16 level uniform quantization

Quantization level	8	16
Mean square error	5.21%	3.54%
Codeword length	2112	3365
Compression ratio	32.41%	51.83%

Table 2.2. Performance of uniform quantization using various quantization levels

Quantization level	64	128	256
Mean square error	0.17%	0.04187%	0.011217%
Codeword length	5602	7106	73100
Compression ratio	61.36%	51.18%	61.76%

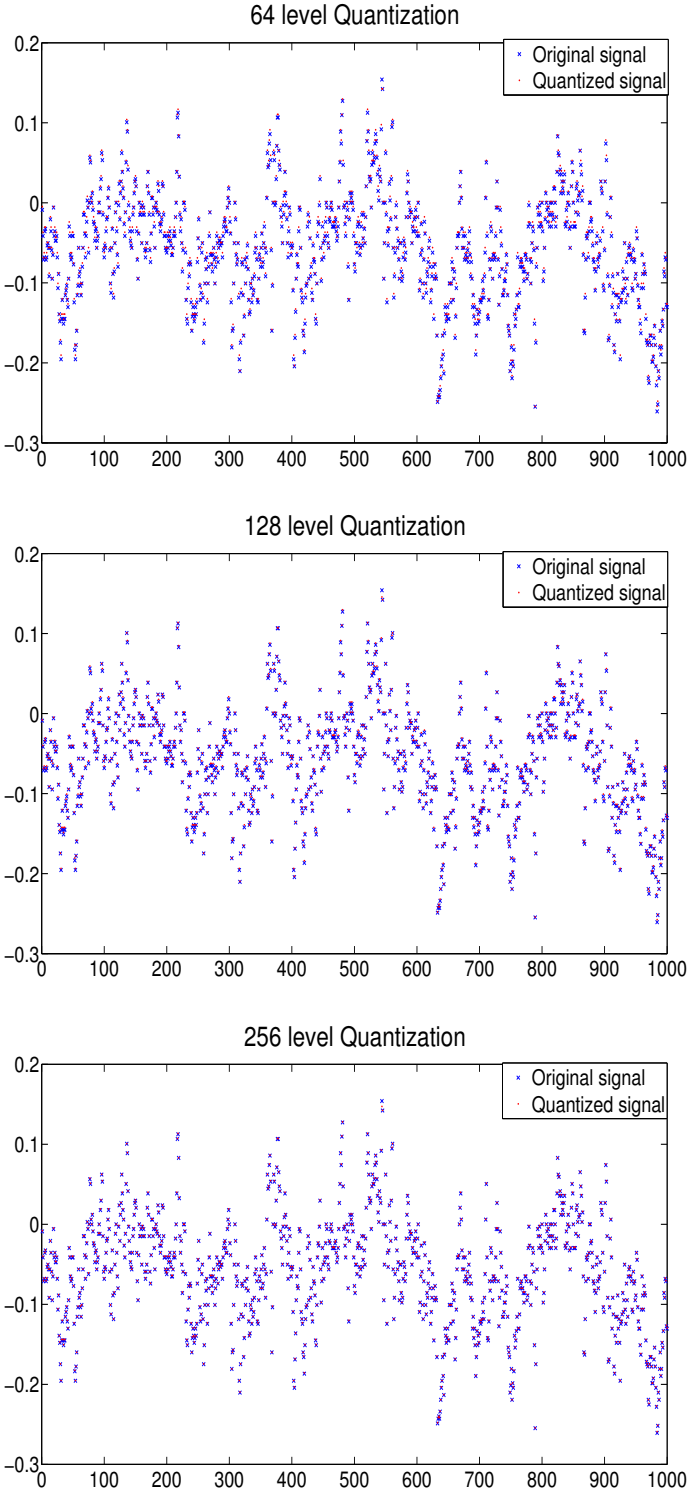


Figure 2.5. Quantized data versus the original data using various quantization levels.

Table 2.3. Performance of 8 and 16 level non-uniform quantization

Quantization level	8	16
Mean square error	1.85%	0.61%
Codeword length	3958	6334
Compression ratio	70.29%	55%

Table 2.4. Performance of non-uniform quantization using various quantization levels

Quantization level	64	128	256
Mean square error	3.6×10^{-4}	4.1×10^{-5}	3.0×10^{-34}
Codeword length	9226	15572	117259
Compression ratio	57.72%	39.69%	5.4%

2.2.2 Compression Using DCT-based Compressive Sensing

In this section, we employ compressive sensing to address the data compression problem using the same data as in Huffman coding. The first step is to choose the proper basis for the original data. We choose, in our case, the Discrete Cosine Transform (DCT) as the basis function. Then, we decompose the data in terms of atoms from the DCT dictionary. The DCT is an example of a frequency dictionary. It converts data into sets of frequencies, and compress the data by deleting the frequencies that are less meaningful. The dictionary elements are:

$\frac{1}{\sqrt{n}} \cos \frac{2\pi ml}{n} \quad m, l = 0, 1, 2 \ldots n-1$ (the odd columns in the sensing matrix)

and

$\frac{1}{\sqrt{n}} \sin \frac{2\pi ml}{n} \quad m, l = 0, 1, 2 \ldots n-1$ (the even columns in the sensing matrix),

where $n = 1000$ in our case.

After decomposing the data in the basis functions, the coefficients are sorted in a descending order and small ones are discarded. Figure 2.6 shows the remaining

coefficients. It is noted that less than 50 out of 1000 coefficients are kept and it is a truly sparse signal .

Since there is noise when we receive the data, we would like to use the noisy version of the data, b', to estimate coefficient, x, of the original signal. Basis Pursuit (BP) is adopted in our simulation. We also modify the BP technique to solve the above problem:

$$\min \frac{1}{2}(\|b' - Ax\|_2 + \gamma\|x\|_1) \text{ subject to } Ax = b' \tag{2.4}$$

Figure 2.7 demonstrates all coefficients are perfectly recovered, which means all information that the original data contains is successfully retrieved. Figure 2.8 explains the role of probability that plays in the compressive sensing. The probability of perfect recovery decreases while increasing the compression ratio. If stable and accurate reconstruction is required, at least 200 samples are needed to represent the original data. Hence, the highest compression ratio we can achieve is 5.55:1. The Mean square error (MSE) for compressive sensing is 3.75% which is at the same level as the 16-level quantization. Since the data we used here is real world data which contains noise, it is acceptable to have part of the data distortion when we perform data reconstruction. It could also be treated as "signal denoising".

2.3 Statistical Distribution Analysis of Wind Tunnel Data

Before we consider the lower bound of the probability of error in the reconstruction process of compressive sensing, more statistical information of wind tunnel data is investigated. For example, in general, what kind of distribution could best characterize the original data as well as how to estimate the value of parameters for

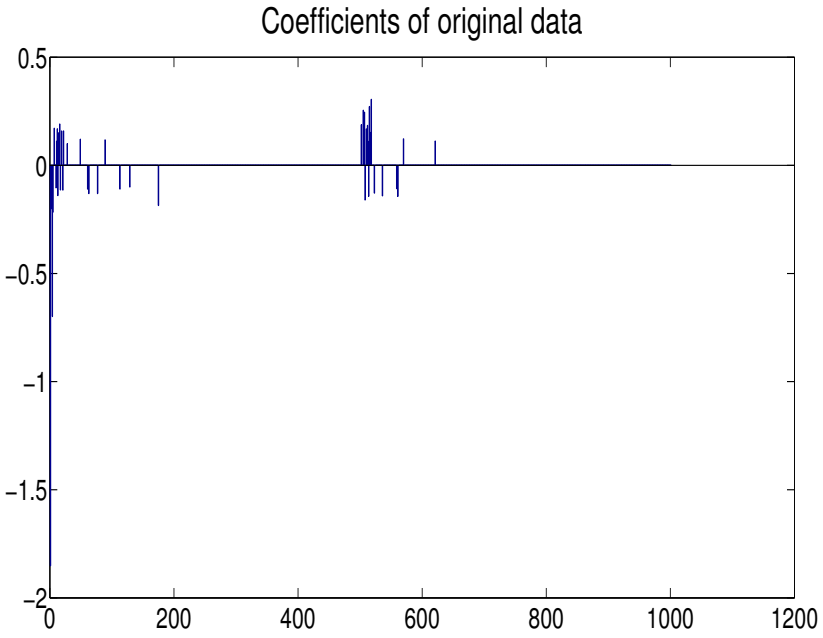


Figure 2.6. Coefficients of the original data.

these distributions. Four well-known distribution models and Maximum-Likelihood Estimation (MLE) are considered in our case [23].

2.3.1 Statistical Models

The Probability Density Function (PDF) for the well-known normal distribution [45] with parameters μ and σ is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x > 0, \quad \sigma > 0 \tag{2.5}$$

Similarly, the PDF for log-normal distribution [46] with parameters μ and σ is

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, \quad x > 0, \quad \sigma > 0 \tag{2.6}$$

The Rayleigh distribution, whose real and imaginary components are Gaussian variables, has the PDF as follows:

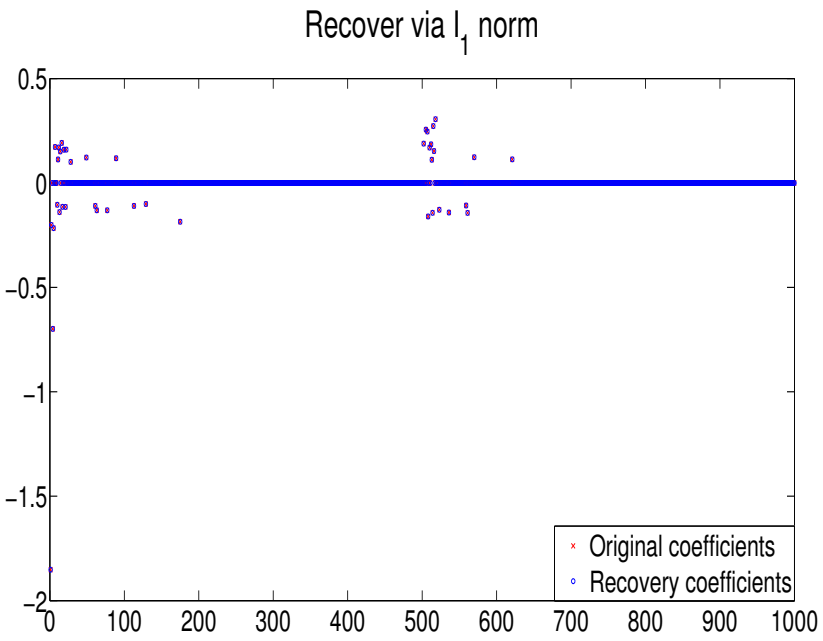


Figure 2.7. Recovery via ℓ_1 norm.

$$f(x) = \frac{x}{b^2} e^{-\frac{x^2}{2b^2}}, \quad b > 0 \tag{2.7}$$

The Weibull distribution [47] can be made to fit measurements that lie between the Rayleigh and log-normal distributions. The PDF is represented as

$$f(x) = ba^{-b}x^{b-1}e^{-(\frac{x}{a})^b}, \quad x > 0, \quad a > 0 \quad b > 0 \tag{2.8}$$

where b is the shape parameter, and a is the scale parameter.

If a and b are the parameters for the Weibull distribution, then the Rayleigh distribution with parameter b is equivalent to the Weibull distribution with parameters $a = \sqrt{2}b$ and $b = 2$.

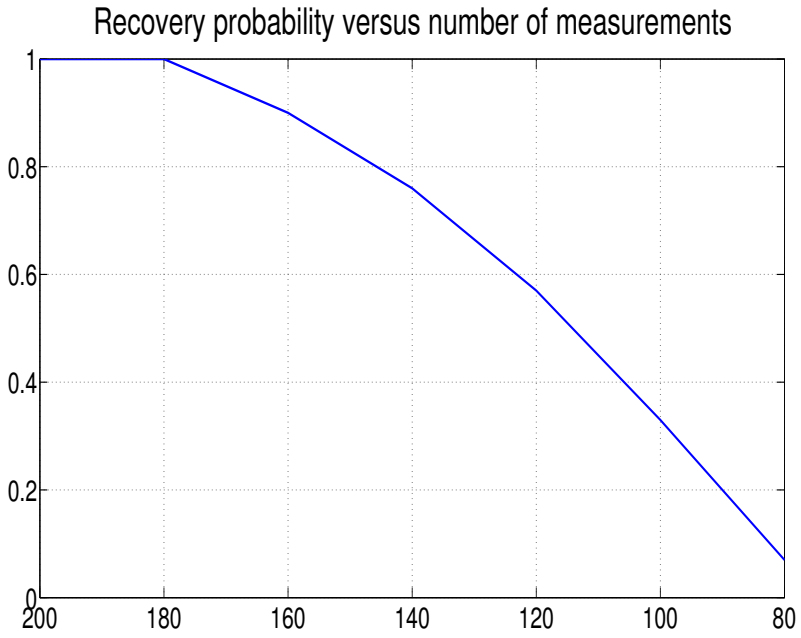


Figure 2.8. Recovery probability versus the number of measurements.

2.3.2 Maximum-Likelihood Estimation

On the basis of accessibility to large amount of wind tunnel data, we employ the maximum-likelihood estimation (MLE) approach to estimate the preceding parameters for Normal, Log-normal, Weibull, and Rayleigh models, respectively. MLE is often used when the sample data is known and the parameters of the underlying probability distribution are to be estimated [45][48]. It is generalized as follows. Let $y_1, y_2, \ldots, y_N$ be N independent samples drawn from a random variable Y with m parameters $\theta_1, \theta_2, \ldots, \theta_m$, where $\theta_i \in \theta$, then the joint PDF of $y_1, y_2, \ldots, y_N$ is [23]

$$L_N(\mathbf{Y}|\theta) = f_{Y|\theta}(y_1|\theta_1, \ldots, \theta_m) \ldots f_{Y|\theta}(y_N|\theta_1, \ldots, \theta_m) \tag{2.9}$$

When expressed as the conditional function of Y that depends on the parameter θ, the likelihood function is [23]

$$L_N(\mathbf{Y}|\theta) = \prod_{k=1}^{N} f_{Y|\theta}(y_k|\theta_1, \ldots, \theta_m) \tag{2.10}$$

The MLE of $\theta_1, \theta_2, \ldots, \theta_m$ is the set of values $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ that maximize the likelihood function $L_N(Y|\theta)$.

Since the logarithmic function is monotonically increasing, maximizing $L_N(Y|\theta)$ is equivalent to maximizing $\ln(L_N(Y|\theta))$. Hence, it can be shown that a necessary, but not sufficient condition to obtain the MLE is to find $\hat{\theta}$ which solves the likelihood equation (2.11) [23]

$$\frac{\partial}{\partial\theta}\ln(L_N(\mathbf{Y}|\theta)) = 0 \tag{2.11}$$

We obtain the different parameters for normal, log-normal, Weibull and Rayleigh distribution, respectively, which are shown in Table 2.5. The STandard Deviation (STD) for each parameter is also derived in the form ε_x, where x denotes the parameter for different models. From Table 2.5, it is obvious that the normal distribution model provides the smallest STD among four different statistical models that we have considered.

Table 2.5. Estimated parameters for different statistical models

PDF	Normal	Log-normal
	$\hat{\mu}$= -0.0024	$\hat{\mu}$= -4.1350
	$\hat{\sigma}$= 0.1223	$\hat{\sigma}$= 1.3029
	$\varepsilon_\mu = 1.09 \times 10^{-5}$	$\varepsilon_\mu = 0.0044$
	ε_σ= 0.0002	ε_σ=0.0020
PDF	Weibull	Rayleigh
	$\hat{a}$= 0.0294	
	$\hat{b}$= 0.8052	$\hat{b}$= 0.0865
	$\varepsilon_a = 0.0003$	ε_b= 0.0012
	ε_b= 0.0009	

2.3.3 Root Mean Square Error

We may also observe to what extent does the PDF curve of the statistical model matches that of the data by means of Root Mean Square Error (RMSE). Let i $(i = 1, 2, \ldots, n)$ be the sorted sample index of data, r_i is the corresponding probability density value, whereas $\hat{r}_i$ is the probability density value of the statistical model with estimated parameters shown in Table 2.5. The RMSE is obtained as follows [23]

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \hat{r}_i)^2} \tag{2.12}$$

where n is the total sample number. The RMSEs for different statistical models are listed in Table 2.6. It also demonstrates that normal distribution fits data the best.

Table 2.6. RMSE for different statistical models

PDF	Normal	Log-normal	Rayleigh	Weibull
RMSE	0.2983	1.8439	2.4502	1.9753

2.4 Information Theoretic Lower Bound of The Probability of Error

In this section, we evaluate the lower bound of the probability of error in the reconstruction process [24]. Consider the flow of the entire CS scheme. The information x is first compressed, corresponding to (1.4), and the observation y is obtained. Then, from the observation, the algorithm of the similar form to (1.6) is executed to obtain the estimator $\hat{x}$. It is observed that $x \to y \to \hat{x} = g(y)$ forms a Markov chain. Define the probability of error P_e as $\Pr(\hat{x} \neq x)$ and an error random variable E as

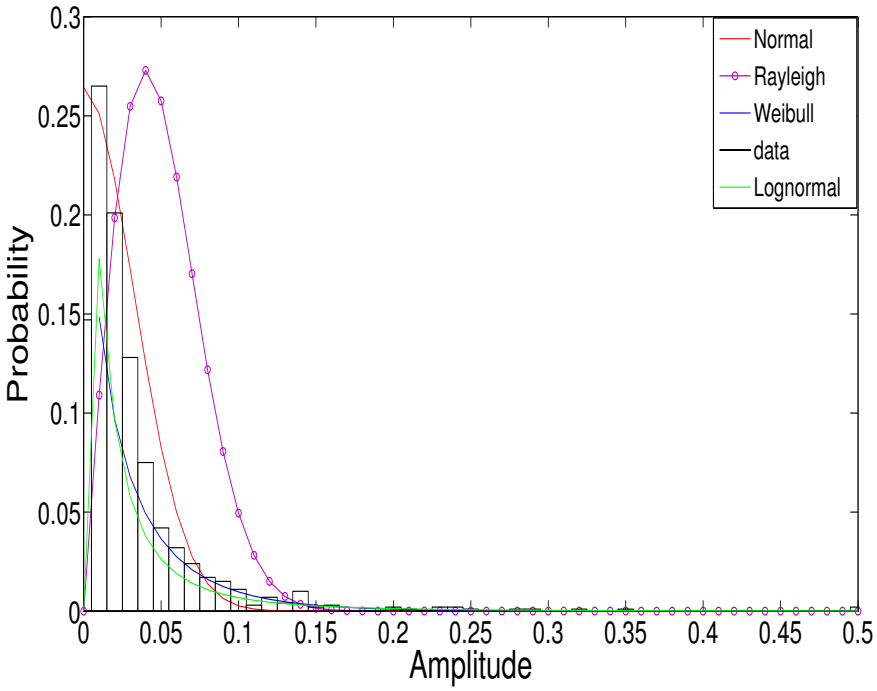


Figure 2.9. RMSE for different statistical models.

$$E = \begin{cases} 1 & \text{if } \hat{\mathbf{x}} \neq \mathbf{x}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.13}$$

Finally, we obtain the Fano's inequality, which provides the lower bound of the probability of error

$$P_e \geq \frac{h(\mathbf{x}|\mathbf{y}) - 1}{h(\mathbf{x})} \tag{2.14}$$

From Section 2.3, we already know wind tunnel data follows normal distribution. Therefore, we can calculate the entropy of the original data and derive the lower bound of the probability of error [24].

$$h(\mathbf{x}) = \frac{1}{2} \log \left[(2\pi e \sigma_x^2)^K \right] \tag{2.15}$$

$$h(\mathbf{y}) = \frac{1}{2}\log\left[(2\pi e)^M |\Sigma_y|\right] \tag{2.16}$$

where

$$\Sigma_y = \begin{pmatrix} \mathrm{COV}(y_1, y_1) & \dots & \mathrm{COV}(y_1, y_M) \\ \vdots & \dots & \vdots \\ \mathrm{COV}(y_M, y_1) & \dots & \mathrm{COV}(y_M, y_M) \end{pmatrix} \tag{2.17}$$

Since each entry y_m of y are previously known to have zero mean and variance $K\sigma_x^2$, the covariance $\mathrm{COV}(y_m, y_{m'})$ between the mth and the m'th entries of y can then be expressed as

$$y = \begin{cases} K\sigma_x^2 & \text{if } m = m', \\ (\sigma_x^2 + \mu_x^2)\sum_{u=1}^{K} \phi_{m,u}\phi_{m',u} & \\ +\mu_x^2 \sum_{v=1}^{K} \phi_{m,v} \sum_{\substack{w=1 \\ w\neq v}}^{K} \phi_{m',w} & \text{otherwise.} \end{cases} \tag{2.18}$$

Because x and y are joint $(K + M)$-dimensional Gaussian random variables, the entropy of (x, y) can be expressed as [24]

$$h(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\log[(2\pi e)^{K+M}|\Sigma_{xy}|] \tag{2.19}$$

where

$$\Sigma_{xy} = \begin{pmatrix} \Sigma_y & A \\ A^T & \sigma_x^2 I_K \end{pmatrix} \tag{2.20}$$

The submatrix A in σ_{xy} denotes the matrix of the covariance $\mathrm{COV}(y_m, x_n)$ between the mth entry of y and the nth entry of x, which can be calculated from

$$\mathrm{COV}(y_m, x_n) = \phi_{m,n}(\sigma_x^2 + \mu_x^2) + \mu_x^2 \sum_{\substack{k=1 \\ k\neq n}}^{K} \phi_{m,k} \tag{2.21}$$

Obviously, to calculate h(y) and h(x,y), we need to consider the determinant of Σ_y and the determinant of Σ_{xy}. Since these covariance matrices are symmetric positive definite matrices, we can, therefore, apply the SVD technique to decompose and find the eigenvalues of the matrices. In other words, the determinants of the covariance matrices can be expressed as [24]

$$|\Sigma_y| = \prod_{i=1}^{M} \lambda_y^i \tag{2.22}$$

and

$$|\Sigma_{xy}| = \prod_{j=1}^{K+M} \lambda_{xy}^j \tag{2.23}$$

If the information is normalized to unit variance, i.e., $\sigma_x^2 = 1$, we can, based on Fano's inequality, rewrite the lower bound as

$$P_e \geq 1 - \frac{1}{4}\log\left(\frac{\lambda_{y\_max}}{\lambda_{xy\_min}}\right)^{\frac{M}{K}+1} \tag{2.24}$$

Hence, the logarithm term in (2.24) is always non-negative and the lower-bound value is always less than or equal to 1. Finally, for a given information vector x, one should consider the following scenarios when selecting the measurement matrix Φ:

- If Φ is generated such that $\lambda_{y\_max} = \lambda_{xy\_min}$, then the error is certain, i.e., $P_e = 1$. This means the information can never be perfectly recovered from the reconstruction process.
- If Φ is chosen such that $\frac{\lambda_{y\_max}}{\lambda_{xy\_min}} \geq 2^{\frac{4}{\frac{M}{K}+1}}$, then the lower bound is non-positive; which implies that the perfect reconstruction of information is possible.

From Figure 2.9, it is obvious that the normal distribution provides the best goodness-of-fit performance, while the Rayleigh provides the worst performance due to high kurtosis and tails.

All the above investigations show that normal distribution is more appropriate to characterize the statistic of the original data than other widely used Rayleigh, Weibull and Log-normal distribution.

In order to evaluate the performance of the derived lower bound of the probability of error, we use the same wind tunnel data as in Section 2.2.2. The simulation results are shown in Table 2.7. From the above analysis, the number of measurements required for stable and accurate reconstruction is 160. From Figure 2.8, we can see that, when the number of measurements is 160, the probability of stable reconstruction is around 90%. Although it does not provide the optimum solution to the number of measurements, it still presents a meaningful guideline when we choose the measurement matrix. The inaccuracy may caused by the error introduced in the parameter estimation process. When the number of processed data increases, the estimation error will decrease. Therefore, the derived lower bound we obtained will be more accurate. In order to compare the performance with that of the simulation results in Section 2.2.2, the length of samples we processed is fixed to 1000.

Table 2.7. Lower bound of the probability of error of the wind tunnel data

No. of measurement	100	120	140	160	180
P_e	0.3489	0.2064	0.0873	-0.0469	-0.1341

2.5 Conclusions

We employ DCT-based compressive sensing to compress real-world wind tunnel data. Simulation results show the maximum compression ratio we can achieve in stable reconstruction case is 5:1. Mean square error (MSE) for compressive sensing is 3.75% which is at the same level as16-level quantization Huffman coding. Com-

pared with compressive sensing, the main drawback of Huffman coding is that it requires calculating the probability of each symbol before encoding. It means it may not be appropriate for real time compression. It also requires prior knowledge of the data to generate the codebook, so that the compression process is not uniform. We also investigate the statistical information of the wind tunnel data. Based on the Maximum-Likelihood Estimation (MLE) and Root Mean Square Error (RMSE) criteria, the normal distribution characterizes the original data best. Then, we provide the lower bound of the probability of error for compressive sensing, using Fano's inequality. Since we already know the original data is normal distributed, we can easily compute the lower bound and provide the condition for choosing the measurement matrix, Φ.

It has been shown that if the measurement matrix, Φ, is chosen such that $\lambda_{y\_max} = \lambda_{xy\_min}$, then the reconstruction error is inevitable; therefore, it is unwise to perform the data compression on the choice of the measurement matrix. However, if the selected measurement yields $\frac{\lambda_{y\_max}}{\lambda_{xy\_min}} \geq 2^{\frac{4}{\frac{M}{K}+1}}$, the probability of error is lower bounded by a non-positive value; which implies that there is a potential for the information can be perfectly recovered. In order to evaluate the performance of the derived lower bound of the probability of error, we use the same wind tunnel data as in Section 2.2.2. Although the simulation result does not provide the optimum solution to the number of measurements, it still presents a meaningful guideline when we choose the measurement matrix. The inaccuracy may caused by the error introduced in the parameter estimation process. This lower bound analysis could be easily extended to data follows other distributions.

CHAPTER 3

COMPRESSIVE SENSING FOR UWB WIRELESS COMMUNICATION SYSTEM

In this Chapter, we propose a compressive sensing-based compression and recovery UWB communication system. Compared to the conventional UWB system, it can jointly estimate the channel and compress the data which also reduces the hardware complexity. No information about the transmitted signal is required in advance as long as the channel follows autoregressive model. As an application example, real-world UWB signal is collected and processed to evaluate the performance of the proposed system.

3.1 Problem Formulation

Since the duty cycle of impulse UWB is very low, it is sparse in the time domain. Then the UWB signal $s(t)$ is filtered by discrete time stable linear filter H and

$$x(t) = (Hs)(t) = \sum_i s(\tau_i)h(t - \tau_i) \tag{3.1}$$

is contaminated by Gaussian noise, i.e., $y(t) = x(t)+n(t)$ for $t = 0, 1, \ldots, N$. The goal is to estimate the channel $h(t)$ and recover the transmitted UWB signal $s(t)$. The block diagram of the conventional UWB communication and compression system is illustrated in Figure 3.1. We propose a novel compressive sensing-based UWB communication and compression system, which is shown in Figure 3.2.

Now, our problem now becomes how can we recover the original UWB signal from the compressed data $y = GHs$. At first glance the problem seems impossible to solve. However, the main idea is that we apply some transformations to $y = Gx =$

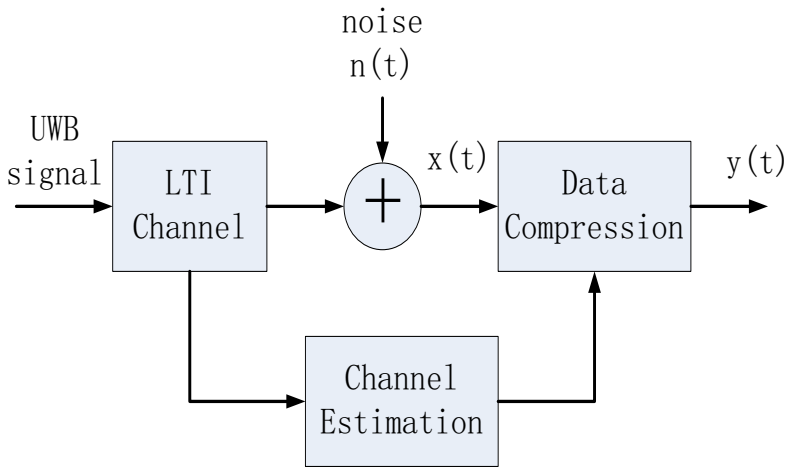


Figure 3.1. Block diagram of the conventional UWB communication and compression system.

GHs, and set up a new standard CS equation. In our investigation, the channel is assumed to follow autoregressive model.

Then, our objective is to reconstruct x_t which can be represented by using an autoregressive model from a number of random measurements. A typical p-order autoregressive model can be represented as follows

$$x_t + \sum_{i=1}^{p} \alpha_i x_{t-i} = s_t \tag{3.2}$$

where α_i is the ith corresponding coefficient and s_t is a sparse signal. We assume the vector $s = [s_0, s_1, \cdots, s_{n-1}]^T$ is k-sparse, that is, there are only k entries in s that are nonzero. In order to recover x_t, we need to find out the AutoRegressive (AR) coefficients $\alpha = [\alpha_0, \alpha_1, \cdots, \alpha_{n-1}]^T$ and sparse vector $s = [s_0, s_1, \cdots, s_{n-1}]^T$ from the original signal x_t and the new measurement y.

Note that in the standard compressive sensing (CS) scenario, the signal x_t is assumed to be sparse in some known orthogonal basis. Sometimes it is difficult to

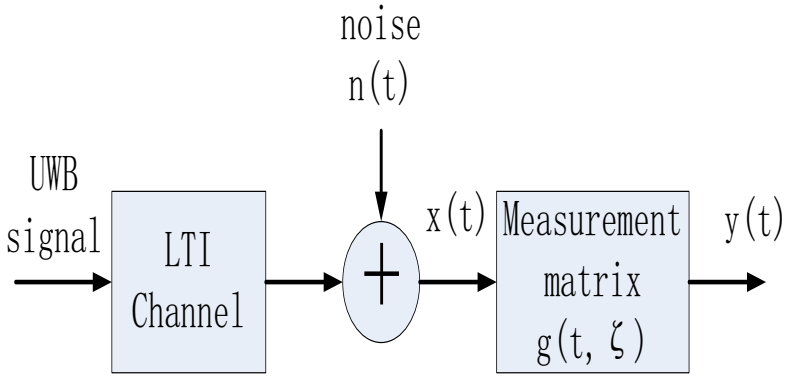


Figure 3.2. Block diagram of proposed compressive sensing-based UWB communication and compression system.

find the appropriate basis matrix. However, in our problem, both the autoregressive model and the basis are not known and we try to solve them simultaneously.

Refer to the standard compressive sensing equation, we have:

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{m-1} \end{bmatrix} = \begin{bmatrix} \phi_{1,1}\phi_{1,2}\cdots & \phi_{1,n} \\ \phi_{2,1}\phi_{2,2}\cdots & \phi_{2,n} \\ \vdots\ddots & \ddots\vdots \\ \phi_{m,1}\phi_{m,2}\cdots & \phi_{m,n} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} \tag{3.3}$$

where each entry, $\phi_{i,j}$, is independent Gaussian random variable or an independent Bernoulli entry.

In the above standard compressive sensing equation, vector x is the sparse signal. While in our case, s is the sparse signal of interest. Based on (3.2), s can be represented by a linear combinations of vector x multiplied by some coefficients α_i

$$
\begin{bmatrix} s_0 \\ s_1 \\ \vdots \\ s_p \\ \vdots \\ s_{n-1} \end{bmatrix} = \begin{bmatrix} x_0 \\ x_1 + \alpha_1 x_0 \\ \vdots \\ x_p + \alpha_1 x_{p-1} + \cdots + \alpha_p x_0 \\ \vdots \\ x_{n-1} + \alpha_1 x_{n-2} + \cdots + \alpha_p x_{n-p} \end{bmatrix}
$$

$$
= \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_p \\ \vdots \\ x_{n-1} \end{bmatrix} + \alpha_1 \begin{bmatrix} 0 \\ x_0 \\ \vdots \\ x_{p-1} \\ \vdots \\ x_{n-2} \end{bmatrix} + \cdots + \alpha_p \begin{bmatrix} 0 \\ 0 \\ \vdots \\ x_0 \\ \vdots \\ x_{n-p-1} \end{bmatrix}
$$

Since all the vectors on the right hand side are just the shift version of vector $x = [x_0, x_1, \cdots, x_{n-1}]^T$ except some entries are replaced by zero, we can design a new sensing matrix Φ' which is composed of entries of Φ. When we apply Φ' to the sparse signal s, the new observation vector y' could also be represented by a linear combinations of y.

Based on the above information, we can a new sensing matrices Φ as follows

$$
\Phi = \begin{bmatrix} g_{n-m} & g_{n-m-1} & \cdots & g_0 & \cdots & 0 \\ g_{n-m+1} & g_{n-m} & \cdots & g_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & 0 \\ g_{n-1} & g_{n-2} & \cdots & \cdots & g_1 & g_0 \end{bmatrix}
$$

and Φ' is a submatrix of Φ which is composed of the last $(m-p)$ rows of Φ. By applying the sensing matrix Φ' to the sparse signal s_t, we can obtain a new observation vector y' as follows

$$y' = \begin{bmatrix} y_p & y_{p-1} & \cdots y_0 \\ y_{p+1} & y_p & \cdots y_1 \\ \vdots & \vdots & \ddots \vdots \\ y_{m-1} & y_{m-2} & \cdots y_{m-p-1} \end{bmatrix} \begin{bmatrix} 1 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} = Y\alpha' \tag{3.4}$$

where $\alpha' = [1, \alpha_1, \cdots, \alpha_p]^T$. α_p is the coefficients of autoregressive model which will be discussed in the next section.

Finally, we can formulate a new compressive sensing problem as follows, where $s = [s_0, s_1, \cdots, s_{n-1}]^T$ is the sparse signal of interest.

$$\Phi' s = y' \tag{3.5}$$

It is a standard CS equation, which can be solved efficiently by a linear programming algorithm.

Based on (1.7), we know that the number of measurements we need to fully recover the original signal is proportional to $\mu(\Theta)$, which is the largest magnitude among the entries in μ. Compared the sensing matrix Φ in (3.3) with our proposed sensing matrix Φ', it is easy to figure out that with the same sparse signal x in some known basis Ψ, $\mu'(\Theta)$ is less than or equal to $\mu(\Theta)$ because some of the entries in the row of Φ' are zeros instead of Gaussian or Bernoulli entries. Therefore, less space is required to store observation vectors which also makes sensing process more efficiently.

3.2 Autoregressive Coefficients Estimation

One key point in our UWB communication system is to estimate the coefficients of the autoregressive process. We first start with the simplest case $p = 1$. The direct inversion method is used to estimate the coefficients. Then we discuss the general p coefficients case.

3.2.1 $p = 1$ Case [53]

$$x_t = \alpha_1' x_{t-1} + s_t \tag{3.6}$$

which can be re-written as

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{bmatrix} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-2} \end{bmatrix} \alpha_1' \tag{3.7}$$

or

$$b = A\Lambda \tag{3.8}$$

which can be solved by

$$\begin{aligned} \Lambda = \alpha_1' &= (A^T A)^{-1} A^T b \\ &= \frac{\sum_{t=0}^{N-2} x_t x_{t+1}}{\sum_{t=0}^{N-2} x_t^2} = \frac{c_1}{c_0} = r_1 \end{aligned} \tag{3.9}$$

where c_i and r_i are the ith autocovariance and autocorrelation coefficients, respectively.

3.2.2 General p Coefficients Case [53]

Let us consider the general p-order autoregressive model represented by

$$x_t = \alpha'_1 x_{t-1} + \alpha'_2 x_{t-2} + \cdots + \alpha'_p x_{t-p} + s_t \tag{3.10}$$

When lag equals to 1, we can multiply both sides of (3.10) by x_{t-1},

$$x_{t-1}x_t = \sum_{j=1}^{p} (\alpha'_j x_{t-1}x_{t-j}) + x_{t-1}s_t \tag{3.11}$$

where t and j are the time and term indices, respectively. Then we take the expectation on both sides,

$$< x_{t-1}x_t >= \sum_{j=1}^{p} (\alpha'_j < x_{t-1}x_{t-j} >) + < x_{t-1}s_t > \tag{3.12}$$

where the α'_j are kept outside the expectance operator because they are deterministic, rather than statistical, quantities.

Note that $< x_{t-1}s_t >= 0$ because the shock (or random perturbation) s of the current time is unrelated to- and thus uncorrelated with- previous values of the process,

$$< x_{t-1}x_t >= \sum_{j=1}^{p} (\alpha'_j < x_{t-1}x_{t-j} >) \tag{3.13}$$

we divide (3.13) through by $(N-1)$, and use the evenness of the autocovariance, e.g., $c_{-l} = c_l$, and then divide it through by c_0. Finally we can get

$$r_1 = \sum_{j=1}^{p} \alpha'_j r_{j-1} \tag{3.14}$$

$r_2, \cdots, r_k, \cdots, r_p$ can also be derived by multiplying $x_{t-2}, \cdots, x_{t-k}, \cdots, x_{t-p}$ on both sides of (3.10), respectively.

Rewriting all the equations together yields

$$
\begin{aligned}
r_1 &= \alpha_1' r_0 + \alpha_2' r_1 + \alpha_3' r_2 + \cdots + + \alpha_{p-1}' r_{p-2} + + \alpha_p' r_{p-1} \\
r_2 &= \alpha_1' r_1 + \alpha_2' r_0 + \alpha_3' r_1 + \cdots + + \alpha_{p-1}' r_{p-3} + + \alpha_p' r_{p-2} \\
&\vdots \\
r_{p-1} &= \alpha_1' r_{p-2} + \alpha_2' r_{p-3} + \alpha_3' r_{p-4} + \cdots + + \alpha_{p-1}' r_0 + + \alpha_p' r_1 \\
r_p &= \alpha_1' r_{p-1} + \alpha_2' r_{p-2} + \alpha_3' r_{p-3} + \cdots + + \alpha_{p-1}' r_1 + + \alpha_p' r_0
\end{aligned}
$$

Recalling that $r_0 = 1$, the above equation can be written as

$$
\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_{p-1} \\ r_p \end{bmatrix} = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{p-2} & r_{p-1} \\ r_1 & 1 & r_1 & \cdots & r_{p-3} & r_{p-2} \\ & \vdots & & & \vdots & \\ r_{p-2} & r_{p-3} & r_{p-4} & \cdots & 1 & r_1 \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1' \\ \alpha_2' \\ \vdots \\ \alpha_{p-1}' \\ \alpha_p' \end{bmatrix} \tag{3.15}
$$

or succinctly

$$
r = R\Lambda \tag{3.16}
$$

It is noted that R is full-rank and symmetric, so that invertibility is guaranteed,

$$
\hat{\Lambda} = R^{-1} r \tag{3.17}
$$

Then we propose the novel ℓ_1 minimization algorithm

$$
\min_{s \in \mathbb{R}^N, \alpha' \in \mathbb{R}^P} \|s\|_{\ell_1} \quad \text{subject to} \quad y' = Y\alpha' = \Phi' s \tag{3.18}
$$

When the observation measurement y is contaminated by noise, e.g., $y = \Phi x + n$ where n is i.i.d Gaussian noise, we have

$$\min_{s\in\mathbb{R}^N,\alpha'\in\mathbb{R}^P} \|s\|_{\ell_1} \quad \text{subject to} \quad \|\Phi' s - y'\|_{\ell_2} < \epsilon \tag{3.19}$$

where ϵ bounds the amount of noise in y'. It is again a convex problem (a second-order cone program) and can be solved efficiently. Once s is found from (3.18) and (3.19), α' is derived and the interest signal x_t can be recovered through (3.2).

An alternative method for dealing with absolute values in a linear program problem in (3.18) is to introduce new variables s^+, s^-, constrained to be nonnegative, and let $s_i = s_i^+ + s_i^-$ [54] (Our intention is to have $s_i = s_i^+$ or $s_i = -s_i^-$, depending on whether s_i is positive or negative.). We then replace the occurrence of $|s|$ with $s_i^+ + s_i^-$ and obtain the alternative formulation

$$\begin{aligned} &\min \ \sum_{i=1}^{N}(s_i^+ + s_i^-) \\ &\text{subject to} \ \ \Phi' s_i^+ - \Phi' s_i^- = y' \\ &s^+, s^- \geq 0, \end{aligned} \tag{3.20}$$

where $s^+ = (s_1^+, s_2^+, \cdots, s_n^+)$ and $s^- = (s_1^-, s_2^-, \cdots, s_n^-)$.

The relations $s_i = s_i^+ + s_i^-$, $s^+ \geq 0$, $s^- \geq 0$, are not enough to guarantee that $|s| = s_i^+ + s_i^-$, and the validity of this reformulation may not be entirely obvious. At an optimal solution to the reformulated problem, and for each i, we must have either $s_i^+ = 0$ or $s_i^- = 0$, because otherwise we could reduce both s_i^+ and s_i^- by the same amount and preserve feasibility, while reducing the cost, in contradiction of optimality. Having guaranteed that either $s_i^+ = 0$ or $s_i^- = 0$, the desired relation $|s| = s_i^+ + s_i^-$ now follows.

Let A be the m by $2n$ matrix $[\Phi' \ -\Phi']$. (3.18) can be written as:

$$Ag = y', \ \ s \geq 0. \tag{3.21}$$

It has a solution g^* which is a vector in R^{2n} and can be partitioned as $g^* = [u^* \, v^*]$; then $s^* = u^* - v^*$ solves (3.18).

3.3 Numerical Results

The UWB noise radar signal used in this simulation is real-world data. The frequency of the transmitted signal is 400-700 MHz and the sampling rate is 1.5 GHz. The property and waveform of the UWB noise radar signal is illustrated in Figure 3.3.

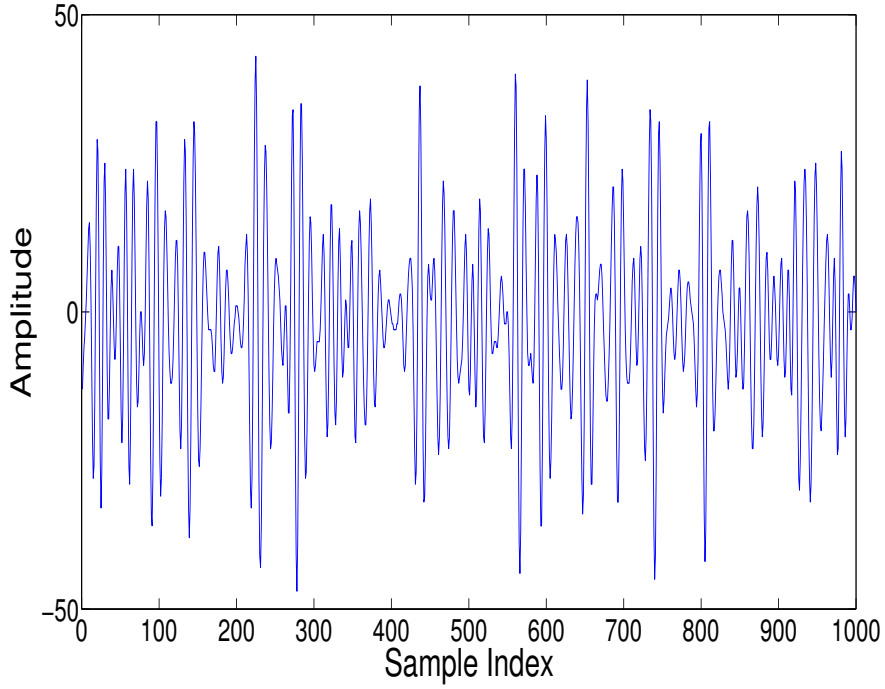


Figure 3.3. UWB noise waveforms of received signal.

We already know the sensing matrix Φ' in our algorithm has a better RIP property than the conventional Gaussian sensing matrix, which means $O(k\log(\mathrm{n}))$ measurements are sufficient to reliably recover the original signal. The first step is to estimate autoregressive coefficients. In this example, we consider a $p = 4$ case: $\alpha'_1 = 3.2$, $\alpha'_2 = -5.4$, $\alpha'_3 = 3.8$, $\alpha'_4 = -0.5$. Then we can get a sparse signal s_t based on (3.6) which is depicted in Figure 3.4. It is obvious that the signal is really sparse since less than 50 samples out of 1000 are nonzero. Figure 3.5 shows the recovered sparse signal s_t from noisy measurement. As we can see, a large number of the reconstructed samples have small value, hence we can setup a threshold to filter s_t and it will become sparse again The only distortion occurs when the amplitude of the sample is large. Since few samples have large amplitude, its effect on our signal recovery procedure is truly negligible. We choose the Bernoulli sensing matrix Φ and let the number of measurements vary from 100 to 500. For each selection, Monte Carlo simulation is performed 2000 times to get the probability of successful recovery. We also choose Φ to be Gaussian and repeat the experiment. The results are shown in Figure 3.6 and Figure 3.7. We can see that the performance of the Bernoulli sensing matrix is better than that of the Gaussian sensing matrix. From the above investigation, we can conclude that based on our proposed UWB communication system, we can jointly estimate the channel and compress the data. The data could be perfectly recovered if the compression ratio does not exceed 2.5:1 while the random Bernoulli matrix is chosen as the sensing matrix.

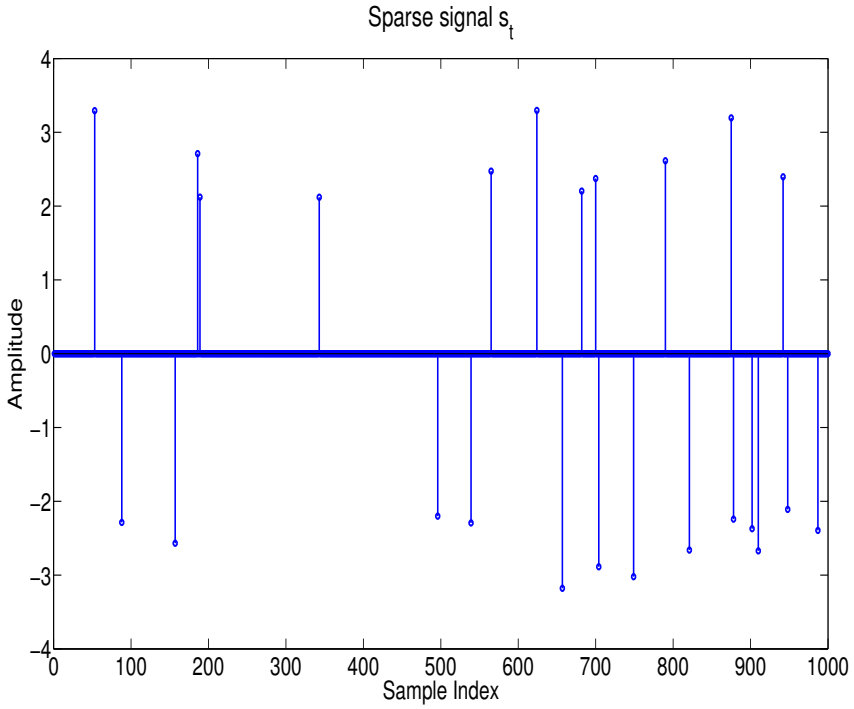


Figure 3.4. Sparse signal s_t obtained from an UWB signal .

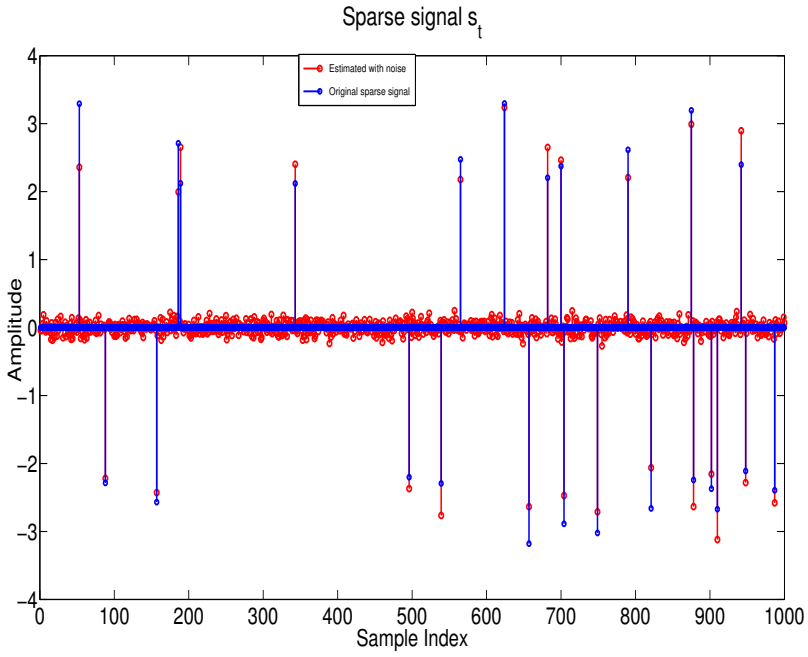


Figure 3.5. Comparison of estimated sparse singal s_t from noisy measurements with the original signal.

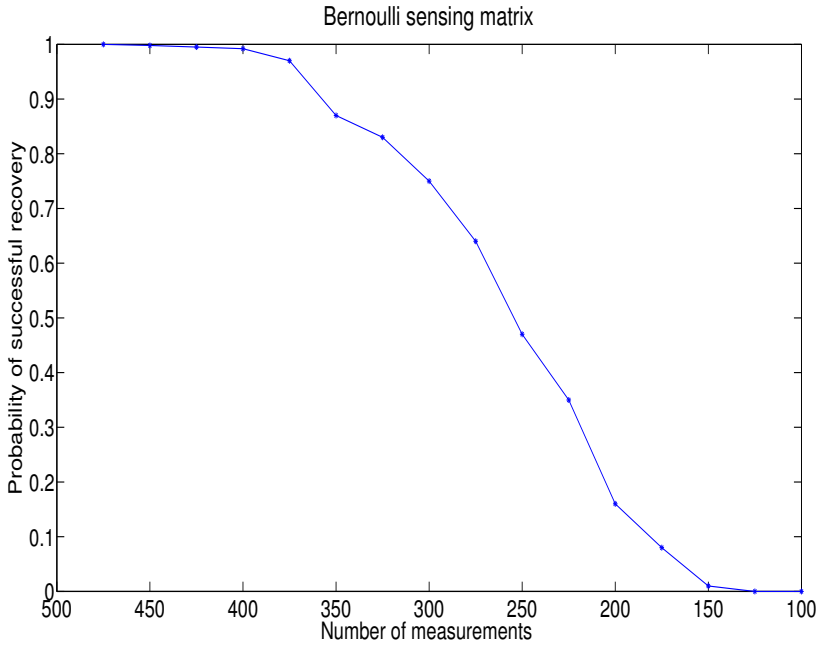


Figure 3.6. Probability of successful recovery as a function of the number of measurements (Bernoulli sensing matrix).

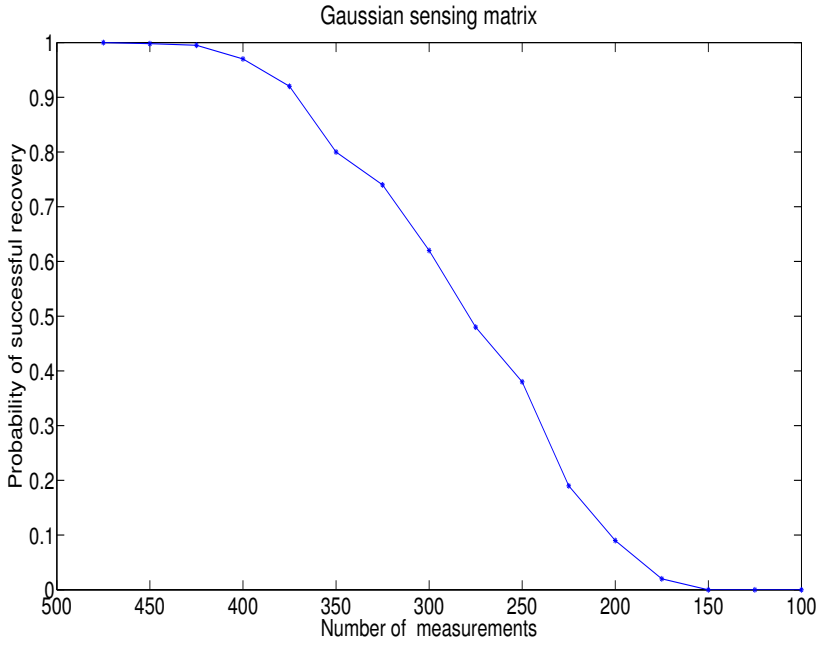


Figure 3.7. Probability of successful recovery as a function of the number of measurements (Gaussian sensing matrix).

CHAPTER 4

SECURITY ANALYSIS OF COMPRESSIVE SENSING-BASED ENCRYPTION

In this Chapter, both information-theoretic and computational secrecy of compressive sensing are investigated. Compressive sensing-based encryption is quite simple since it provides both signal compression and encryption guarantees, without the additional computational cost of a separate encryption protocol.

4.1 Information-Theoretic Secrecy of Compressive Sensing

An encryption method provides perfect secrecy only if the ciphertexts appear sufficiently random to the eavesdropper whose computation capability is unbounded. Shannon addressed the security problem by introducing the idea of perfect secrecy [37]. The encryption E is perfectly secret if ciphertext, c, and message, m, are independent, i.e., $\text{prob}(m, c) = \text{prob}(m) \cdot \text{prob}(c)$.

Shannon also proved the uncertainty of the keys cannot be smaller than the uncertainty of the messages if the encryption is perfectly secret [37]. Compressive sensing-based encryption can be interpreted as a block cipher, which takes mn-bit keys and n-bit messages as inputs and outputs a m-bit ciphertexts. The value of m is determined by the sparsity of messages k, typically $\sim 4k$. Therefore, compressive sensing-based encryption could achieve perfect secrecy due to the length of key is much longer than that of the message.

Theorem 1 states compressive sensing-based encryption is perfectly secret if the length of message X goes to infinity.

Theorem 1. *If message $X = [X_1, X_2, \ldots, X_n]$ are independent and the length of X goes to infinity, then perfect secrecy can be achieved by compressive sensing.*

Proof. By the Central Limit Theorem, we know that message X follows Gaussian distribution. Suppose the measurements vector $\mathbf{Y} = \Phi\mathbf{X}$, then we have

$$\begin{aligned}
I(\mathbf{X};\mathbf{Y}) &= H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}) \\
&= H(\mathbf{Y}) - H(\mathbf{Y}|\mathbf{X}=\mathbf{0})P_{\mathbf{x}}(\mathbf{X}=\mathbf{0}) - \sum_{\mathbf{x}\in\chi,\mathbf{x}\neq\mathbf{0}} H(\mathbf{Y}|\mathbf{X}=\mathbf{x})P_{\mathbf{x}}(\mathbf{X}=\mathbf{x}) \\
&\overset{(a)}{=} H(\mathbf{Y}) - \sum_{\mathbf{x}\in\chi,\mathbf{x}\neq\mathbf{0}} H(\mathbf{Y}|\mathbf{X}=\mathbf{x})P_{\mathbf{x}}(\mathbf{X}=\mathbf{x}) \\
&\overset{(b)}{\leq} \log|T| - \sum_{\mathbf{x}\in\chi,\mathbf{x}\neq\mathbf{0}} H(\mathbf{Y}|\mathbf{X}=\mathbf{x})P_{\mathbf{x}}(\mathbf{X}=\mathbf{x}) \\
&\overset{(c)}{\leq} \log|T| - \log|T-1| \\
&\overset{(d)}{=} 0
\end{aligned}$$

where

(a) follows from the fact that if message $\mathbf{X} = \mathbf{0}$, then $\mathbf{Y} = \mathbf{0}$ with probability one. Thus, $H(\mathbf{Y}|\mathbf{X} = \mathbf{0}) = 0$.

(b) follows from the maximum value of $H(Y)$ equals to $\log|T|$ only if Y is uniformly distributed, i.e., $H(\mathbf{Y}) \leq \log|T|$.

(c) follows from the fact that let Φ be a random $m \times n$ matrix and $y = \Phi x$. If $m > 2k$ (in our case $m \approx 4k$), then any k- sparse signal has a unique projection with probability one (see Appendix in [25]).

(d) follows that the value of $I(\mathbf{X};\mathbf{Y})$ is always non-negative.

Therefore, we can conclude that as the length of message X goes to infinity ($T \to +\infty$), $I(\mathbf{X};\mathbf{Y}) = 0$, and perfect secrecy is achieved by compressive sensing. □

However, Theorem 2 shows that compressive sensing-based encryption cannot achieve perfect secrecy.

Theorem 2. *For any message $x \in \mathbb{R}^n$, $p_{\mathbf{X}}(\mathbf{x}) > 0$, and Φ is a $m \times n$ measurement matrix, $\mathbf{Y} = \Phi\mathbf{X}$. Then, perfect secrecy can never be achieved.*

Proof. Since $\mathbf{x} = \mathbf{0}$ and $\mathbf{Y} = \Phi\mathbf{X}$, we have $\mathbf{y} = \mathbf{0}$ and $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y} = \mathbf{0}|\mathbf{X} = \mathbf{0}) = 1$. It is noted that $\mathbf{y} = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$. Since $p_{\mathbf{X}}(\mathbf{x}) > \mathbf{0}$ for any message $\mathbf{x} \in \mathbb{R}^n$, then we can conclude that $p_{\mathbf{Y}}(\mathbf{Y} = \mathbf{0}) < 1$. Therefore, $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y} = \mathbf{0}|\mathbf{X} = \mathbf{0}) \neq p_{\mathbf{Y}}(\mathbf{Y} = \mathbf{0})$ and the secrecy of compressive sensing-based encryption is not perfect. □

Although compressive sensing-based encryption cannot achieve perfect secrecy, the security of our proposed compressive sensing-based encryption could approach perfect information-theoretic security, which will be proven in Section 5.2.2.

4.2 Computational Secrecy of Compressive Sensing

The secrecy of many currently used encryption is based on the hardness of an underlying computational problem, such as factoring integers or computing discrete logarithm. In our approach, the computational secrecy of compressive sensing-based encryption relies on the computation hardness of finding the correct measurement matrix among a large number of candidates.

Assume an eavesdropper, Eve, has the information of ciphertext y and the sparsity of x. One possible way for Eve is to try all possible measurement matrix Φ and attempt to recover the original data x. If the recovered data is k-sparse, Eve could claim he/she successfully decrypted the plaintext. Lemma 1 shows rank($\Phi\Psi$) = rank(Φ) if Ψ is a Bernoulli basis matrix.

Lemma 1. *If Φ is a $m \times n$ measurement matrix of rank m and Ψ is a $n \times n$ Bernoulli basis matrix of rank n, then rank($\Phi\Psi$) = rank(Φ).*

Proof. Since rank(Ψ)=n, there exist Ψ^{-1} such that $\Psi\Psi^{-1} = I$. rank(Φ)=rank($\Phi\Psi\Psi^{-1}$) =rank (($\Phi\Psi$)Ψ^{-1})$\leq$ rank ($\Phi\Psi$). The reverse inequality rank ($\Phi\Psi$) $\leq$ rank(Φ) follows directly from the theorem rank(AB) $\leq$ rank(A). Hence, rank($\Phi\Psi$)=rank(Φ). $\square$

Theorem 3 states if an eavesdropper has the wrong measurement matrix, he/she can never recover the original data x. Here we only consider the case using the Bernoulli sensing matrix, the Gaussian sensing matrix case is discussed in [26].

Theorem 3. *Let Φ and Φ' be the original and generated $m \times n$ matrices with Bernoulli ensembles. For the original k-sparse data x, we have $\mathbf{x} = \Psi\theta$ and $\mathbf{y} = \Phi\mathbf{x}$. Then, for all recovered data $\mathbf{x}'$ and generated matrix Φ' such that $\mathbf{x}' = \Psi\theta'$ and $\mathbf{y} = \Phi'\mathbf{x}'$ satisfy $||\theta'||_0 = m$ if $m > k$, which indicates the original data x can never be recovered.*

Proof. Since Ψ is a orthonormal Bernoulli sensing matrix, rank(Ψ)=n, rank($\Phi\Psi$)=rank($\Phi'\Psi$)=m if $m \leq n$, which directly follows from lemma 1. Without loss of generality, we assume Ψ is the identity matrix and $\mathbf{y} = \Phi\theta$.

Let Φ'_m be a matrix obtained by taking m columns of Φ', which are linearly independent since rank (Φ'_m)=m. We can use matrix inversion to uniquely determine m entries of θ' that satisfy $\mathbf{y} = \Phi'\theta'$.

Then, we prove a p-sparse ($p < m$) solution ϑ that satisfies $\mathbf{y} = \Phi'\vartheta$ does not exist by way of contradiction. If $\mathbf{y} = \Phi'\vartheta$, then we have $\mathbf{y} = \Phi'\theta' = \Phi'\vartheta$. Since, by assumption, θ' and ϑ are m-sparse and p-sparse solutions ($p < m$), it is easy to show that m columns of the sensing matrix Φ' can be represented by only p columns of Φ', where $p < m$. It contradicts the conclusion in Lemma 1 that rank(Φ')=m and this happens with probability zero. Hence, a p-sparse ($p < m$) solution x' that satisfies $\mathbf{y} = \Phi'\mathbf{x}'$ does not exist.

Assume $\mathbf{y} = \Phi'\mathbf{x}' = \Phi\mathbf{x}$ with $||\theta'||_0 = m$ and $||\theta||_0 = k$ ($m > k$). It implies that $(m-k)$ columns of Φ' are linearly dependent, which contradicts the assumption

we made above. Therefore, $\mathbf{x}' \neq \mathbf{x}$ and the eavesdropper cannot recover the original data x. □

The amount of computation required to find the correct measurement matrix is proportional to the number of candidates. Once an eavesdropper recovers a k-sparse vector using a measurement matrix, he/she knows it is correct. Thus, compressive sensing-based encryption only provides computational secrecy.

CHAPTER 5

SECURITY ANALYSIS OF COMPRESSIVE SENSING-BASED DISTRIBUTED WIRELESS SENSOR NETWORKS

5.1 Introduction

Sensor nodes in WSNs are inherently resource-constrained. These battery-operated nodes have limited processing capability and very low storage capacity. These limitations are due to limited energy and physical size of the sensor nodes. In most practical situations, the sensor nodes are unattended and even deployed in the hostile environments, which demand careful security consideration in the design of WSN. Because of those constraints, the conventional security mechanisms with high computation complexity are not feasible for WSNs. In order to design encryption s suitable for WSNs, it is necessary to be aware about the constraints of the sensor nodes [27] such as energy constraint [28], memory limitations [29] and high latency in communication and synchronization [30].

A rich literature has been published to deal with this challenging problem in the WSNs scenario. The author proposed a key management protocol for WSNs based on symmetric key algorithms in [31]. It uses different keying mechanisms for different packets depending on their security requirements. A common initial key is loaded into each node before deployment. Then, a master key which depends on the common key and its unique identifier is derived. A BROadcast Session Key (BROSK) negotiation protocol is proposed in [32] . BROSK assumes a master key shared by all the nodes in the network. To establish a session key with its neighbor node B, a sensor node A broadcasts a key negotiation message and both arrive at a shared

session key. A deterministic key management protocol is proposed to facilitate key establishment between every pair of neighboring nodes in a WSN in [33]. A random key pre-distribution scheme is introduced for WSNs that relies on probabilistic key sharing among nodes of a random graph in [34].

The difference in computational capacity and energy between sensor nodes and the fusion center is also investigated in proposed WSN. We assume that an individual sensor node possesses far less computational power and energy than the fusion center. Then we place the major computations and public key broadcast on the fusion center. On the sensor side, simple compressive sensing-based encryption is deployed, which requires the sensor nodes only been active in every s intervals, where s is the sparsity of the measurement matrix. Hence, the energy consumption in sensor nodes is tremendously reduced. In addition, node addition and revocation is also supported due to each sensor node works in a distributed manner.

In our proposed WSN, every sensor node and the fusion center share a short randomly-selected secret key, which is independently chosen by each sensor node. Then, the randomizer, which is a burst of random data, is broadcasted by the fusion center, and the key used by each sensor node is determined by the secret key and a few randomizer bits. Each sensor node stores the product of key bit with received data if the corresponding key bit is nonzero. This process is repeated until all data is received. Then, every sensor node simply sends the aggregated data to the fusion center.

5.2 System Design and Security Analysis

5.2.1 System Model

Due to the energy constraint and limited computation capability of sensor nodes, our proposed compressive sensing-based encryption for WSNs is quite simple. The system diagram is shown in Figure 5.1.

Consider an $n \times n$ sparse measurement matrix Φ' [38]

$$\Phi'_{ij} = \sqrt{s} \begin{cases} +1 & \text{with probability } \frac{1}{2s}, \\ 0 & \text{with probability } 1 - \frac{1}{s}, \\ -1 & \text{with probability } \frac{1}{2s}. \end{cases} \tag{5.1}$$

We assume entries Φ'_{ij} within each row are four-wise independent, i.e., every subset of size 4 of Φ'_{ij} within each row is independent. Φ'_{ij} across different rows are totally independent. The parameter s indicates sparsity of the random measurement matrix . If $\frac{1}{s} = 1$, the new sparse random measurement matrix Φ' is identical to the conventional measurement matrix Φ in compressive sensing.

Our distributed algorithm is described as follows.

1. Each sensor j generates a set of independent random variables $\{\Phi'_{1j}, \Phi'_{2j}, \ldots, \Phi'_{nj}\}$. If $\Phi'_{ij} \neq 0$, then sensor j stores the product of Φ'_{ij} with received signal x_i. Repeat the process for all $1 \leq j \leq n$.
2. After receiving all x_i, each sensor node simply sends the aggregated data $\sum_{i=1}^{n} \Phi'_{ij} x_i$ to the fusion center, where n is the length of received signal x.

The signal received at the l-th sensor is a delayed and scaled version of the original signal $z(t)$ at the source [41]

$$x_l(t) = \alpha_l z(t - \tau_l) \tag{5.2}$$

where α_l and τ_l are the attenuation and delay at the l-th sensor, respectively.

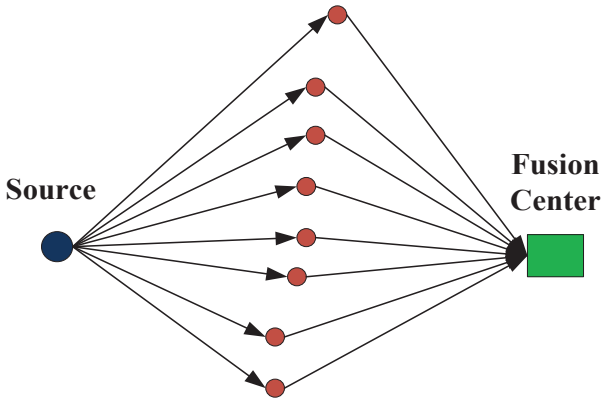


Figure 5.1. System model of proposed WSN.

The delay at the l-th sensor is given by

$$\tau_l = \frac{d_l}{c} \tag{5.3}$$

where d_l is the distance between the source and l-th sensor, and c is the speed of light. Assuming a point source model, the attenuation at the l-th sensor is given by [41]

$$\epsilon_l = \frac{1}{4\pi d_l^2} \tag{5.4}$$

Since $P(\Phi'_{ij} \neq 0) = 1/s < 1$, less energy is consumed by sensor node than that of the conventional sensor network because the sensor node is active only at n/s time slots. The proposed compressive sensing-based encryption is quite simple and a small amount of memory and storage are required to perform the algorithm. Each sensor node only sends the aggregated data to the fusion center at every n symbol interval, which tremendously reduces the overall network throughput. Hence, latency and congestion is less likely to occur. It is noted that sensor node works in a distributed manner and synchronization is not required.

5.2.2 Unconditional Security of Proposed WSN

Shannon's security model provides unconditional security, which does not rely on the hardness of a computational problem, or limits of the computational power of an eavesdropper. Although perfect secrecy cannot be achieved in most situations, there are various practical cryptosystems whose security provably come close to perfect information-theoretic security. In our approach, the unconditional security is guaranteed by the limited storage capacity of an eavesdropper, which is first proposed by [39].

Due to the energy constraint and other limitations in sensor nodes, encryption employed in our proposed WSN is quite simple. Every sensor node and the fusion center share two short randomly-select secret keys, which are independently chosen by each sensor node. Then, the fusion center broadcasts the randomizer R, which is a burst of random bits over the insecure communication channel prior to the transmission of the actual ciphertext. Hence, the randomizer can be accessed by an eavesdropper as well. The ciphertext is a function of the plaintext, two secret keys, and the randomizer. It can be uniquely determined by the ciphertext, two secret keys, and the randomizer. The key idea is the ciphertext depends on a few randomizer bits whose positions determined by two secret keys. Without two secret keys it is impossible to obtain any information about the plaintext without examining a very large number of randomizer bits.

Assume for each sensor node, the plaintext $X = [X_1, \ldots, X_N]$, the ciphertext $Y = [Y_1, \ldots, Y_N]$ and the keystream $W = [W_1, \ldots, W_N]$ are sequences of length N. The randomizer matrix R consists of K rows and T columns and thus has total length $L = KT$ bits and $R = \{-1, +1\}^{KT}$. Each row is denoted by $R[k, 0], \ldots, R[k, T-1]$ $(1 \leq k \leq K)$. The secret key $Z \in \{0, \ldots, T-1\}$ specifies a position within each

row of R, and is chosen to be uniformly distributed over the key space. The other secret key $Z' = [Z_1, \ldots, Z_{\lceil \frac{N}{s} \rceil}]$, where $Z_k \in \{0, \ldots, N\}$ for $1 \leq k \leq \lceil \frac{N}{s} \rceil$, specifies $\lceil \frac{N}{s} \rceil$ positions within each row of R, and is chosen to be uniformly distributed over the space $\{1, \ldots, N\}$, where $\lceil X \rceil$ is the largest integer not greater than X.

The keystream W for the kth sensor node is a function of the secret key Z and the randomizer R, i.e., W is N consecutive bits in the randomizer chosen starting at the position specified by the secret key Z as follows [39]

$$W_n^k = R[k, (n - 1 + Z^k) \mod \mathrm{T}] \qquad 1 \leq n \leq N$$

The other keystream W' for kth sensor node can be expressed as

$$W_n'^k = \begin{cases} 1 & \text{if } n \in \mathbf{Z}', 1 < n < N, \\ 0 & \text{otherwise.} \end{cases}$$

Then, our proposed encryption can be decomposed into two steps as follows.

1. For the kth sensor node, generate the keystream W_n^k by the secret key Z and the randomizer R. The ciphertext $Y_n^k = W_n^k \cdot X_n^k$ for $1 < n < N$.
2. For the kth sensor node, $Y_n'^k = Y_n^k$ if $n \in Z', 1 < n < N$; otherwise, $Y_n'^k = 0$.

The following theorem proves our encryption scheme can achieve perfect secrecy for each sensor node.

Theorem 4. *Let* $\mathbf{W} = \{-1, +1\}^N$, $\mathbf{W}' = \{0, +1\}^N$, $\mathbf{X} = \mathbf{Y} = \{-1, +1\}^{N \log_2 |\mathbf{X}|}$, $\mathbf{Y}' = \{-1, 0, +1\}^{N \log_2 |\mathbf{X}|}$, *where* $|X|$ *denotes the cardinality of* X, *and let* E *be the encryption scheme as mentioned above, which can be decomposed into* E_1 *and* E_2. *Then,* E *is perfectly secret.*

Proof. For encryption E_1: Since W is constructed by bits from R whose positions determined by the secret key Z, it is totally independent from X . Then, we have

$$P_{\mathbf{XY}}(X, Y) = P_{\mathbf{XW}}(X, \frac{Y}{X}) = P_{\mathbf{X}}(X) \cdot P_{\mathbf{W}}(\frac{Y}{X}).$$

Since W is uniformly distributed, then

$$\begin{aligned} P_{\mathbf{Y}}(Y) &= \sum_{X\in\mathbf{X}} P_{\mathbf{XW}}(X,W) = \sum_{X\in\mathbf{X}} P_{\mathbf{X}}(X)\cdot P_{\mathbf{W}}(\frac{Y}{X}) \\ &= \sum_{X\in\mathbf{X}} P_{\mathbf{X}}(X)\cdot \frac{1}{2^{N\log_2|\mathbf{X}|}} = \frac{1}{2^{N\log_2|\mathbf{X}|}} \end{aligned}$$

Hence, Y is also distributed uniformly, and we obtain:

$$\begin{aligned} P_{\mathbf{X},\mathbf{Y}}(X,Y) &= P_{\mathbf{XW}}(X,W) = P_{\mathbf{XW}}(X,\frac{Y}{X}) \\ &= P_{\mathbf{X}}(X)\cdot P_{\mathbf{W}}(\frac{Y}{X}) \\ &= P_{\mathbf{X}}(X)\cdot \frac{1}{2^{N\log_2|\mathbf{X}|}} \\ &= P_{\mathbf{X}}(X)\cdot P_{\mathbf{Y}}(Y) \end{aligned}$$

Thus, X and Y are independent.

For encryption E_2: Since W' is constructed by the secret key Z', which independent from X and Y. Then

$$P_{\mathbf{YY'}}(Y,Y') = P_{\mathbf{YW'}}(Y,\frac{Y'}{Y}) = P_{\mathbf{Y}}(Y)P_{\mathbf{W'}}(\frac{Y'}{Y}).$$

From (5.1), the distribution of the secret key W' can be denoted by

$$W_n{}' = \begin{cases} 1 & \text{with prob. } \frac{1}{s}, \\ 0 & \text{with prob. } 1-\frac{1}{s}. \end{cases}$$

Since Y is uniformly distributed, then

$$Y_n{}' = \begin{cases} -1 & \text{with prob. } \frac{1}{2s}, \\ 0 & \text{with prob. } 1-\frac{1}{s}, \\ +1 & \text{with prob. } \frac{1}{2s}. \end{cases}$$

Combining W and W' to give

$$\overline{W}_n = W_n {W_n}' = \begin{cases} -1 & \text{with prob. } \frac{1}{2s}, \\ 0 & \text{with prob. } 1 - \frac{1}{s}, \\ +1 & \text{with prob. } \frac{1}{2s}. \end{cases}$$

Therefore, $P_{\mathbf{Y}'}(Y') = P_{\overline{W}}(\frac{Y'}{X})$ for all $X \in \{-1, +1\}$. Multiplying by $P_X(X)$ on both sides

$$P_{\mathbf{X}}(X)P_{\mathbf{Y}'}(Y') = P_{\mathbf{X}}(X)P_{\overline{\mathbf{W}}}(\frac{Y'}{X}) = P_{\mathbf{X}\overline{\mathbf{W}}}(X, \frac{Y'}{X}) = P_{\mathbf{X}\mathbf{Y}'}(XY')$$

Thus, X and Y' are independent and E is perfectly secret. □

Although the encryption is perfectly secret for each sensor node, there is still some chance for an eavesdropper to extract the original data. One possible way is to collect the aggregated data $\sum_{i=1}^{n} \overline{W}_{ij} X_i = \sum_{i=1}^{n} Y_i'$ for n sensor nodes, where n is the length of the signal X and the eavesdropper can obtain

$$\begin{aligned} &\alpha_1((\overline{W}_{11}X_1) + (\overline{W}_{21}X_2) + \ldots + (\overline{W}_{n1}X_n)) = \sum_{i=1}^{n} Y_1' \\ &\alpha_2((\overline{W}_{12}X_1) + (\overline{W}_{22}X_2) + \ldots + (\overline{W}_{n2}X_n)) = \sum_{i=1}^{n} Y_2' \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ &\alpha_n((\overline{W}_{1n}X_1) + (\overline{W}_{2n}X_2) + \ldots + (\overline{W}_{nn}X_n)) = \sum_{i=1}^{n} Y_n' \end{aligned}$$

which can also be expressed in the matrix form

$$\alpha\overline{\mathbf{W}}\mathbf{X} = \mathbf{Y}'$$

Then

$$\mathbf{X} = (\alpha\overline{\mathbf{W}})^{-1}\mathbf{Y}'$$

The following theorem states that if only a fraction of the randomizer bits is stored by an eavesdropper, then the probability that he/she could obtain any information about the plaintext approaches zero [39].

Theorem 5. *There exists an event ξ such that, for all possibly probabilistic strategies for examining M bits of randomizer R, then*

$$I(\mathbf{X}; \mathbf{Y}'|\alpha, \mathbf{W}', \xi) = 0 \quad \textit{and } P(\xi) > 1 - \sigma^{N^2},$$

where $\sigma = M/NT$ is the fraction of randomizer bits stored by an eavesdropper and n is the length of original signal X.

Proof. Define ξ as the event that at least one bit of W is not contained in e^M. For each sequence $e^M = [e_1, e_2, \ldots, e_M]$ of length M, we assume the m_k randomizer bits are specified by e^M that within the kth column of R, where $1 \leq k \leq T$. Let P_k be the possibility of the secret key W^k for sensor node k is contained by e^M, then $P_k = \prod_{k=1}^{N} (\frac{m_k}{N})^N$. Under the condition $\sum_{k=1}^{T} m_k = M$, the maximum value of P_k is $(\frac{M}{NT})^N$ if $m_1 = m_2 = \ldots = \frac{M}{T}$.

Since the keystream W' and α may not known by the eavesdropper, then

$$P(\xi) > 1 - (M/NT)^{N^2}$$

□

Example 1. *For proposed WSN, assume $K = 2000$, the length of plaintext is 1000 symbols, and an eavesdropper store a fraction $\sigma = 99.99\%$ of all bits of randomizer. The maximum chance the eavesdropper could have to obtain any new information about the plaintext is less than or equal to $(0.9999)^{1000^2} \leq 3.7015 \times 10^{-44} \to 0$.*

5.3 Data Reconstruction at the Fusion Center

Before stating our main theorems, we first explore some useful properties of the measurement vector and introduce a lemma that would facilitate later proof of proposed theorem.

Lemma 2. *[42] Consider a measurement matrix $\Phi \in \mathbb{R}^{m\times n}$ with entries as shown in (5.1). For the data vector $u \in \mathbb{R}^n$, let $\mathbf{y}$ be a measurement vector of u, which can be expressed by*

$$\mathbf{y} = \frac{1}{\sqrt{m}}\Phi\mathbf{u} \in \mathbb{R}^m$$

Then,

$$E[\mathbf{y}^T\mathbf{y}] = \mathbf{u}^T\mathbf{u} = ||\mathbf{u}||_2^2$$

$$Var[\mathbf{y}^T\mathbf{y}] = \frac{1}{m}\left(2(||\mathbf{u}||_2^4 + (s-3)\sum_{j=1}^{n} u_j^4\right) \tag{5.5}$$

Chernoff Bounds is introduced in Lemma 3, which will be used in the proof of Theorem 6 and Theorem 7.

Lemma 3. *Let $X_1, X_2, \ldots, X_n$ be independent Poisson trials with $P[X_i = 1] = p_i$. Then if X is the sum of the X_i and μ is $E[X]$, for any $\delta > 0$:*

$$P[X > (1+\delta)\mu] < \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu$$

This bound is difficult to compute. A simple and weaker bound can be expressed by

$$P[X > (1+\delta)\mu] < \exp\left(-\mu\sigma^2/4\right) \quad for \quad 0 < \delta < 1. \tag{5.6}$$

Proof. By the Taylor expansion of nature log, we can have

$$\ln(1+\delta) = \delta - \delta^2/2 + \delta^3/3 - \delta^4/4 \cdots$$

multiplying by $(1+\delta)$ gives:

$$\begin{aligned}(1+\delta)\ln(1+\delta) &= \delta+\delta^2/4+(\delta^2/4-\delta^3/6)+\text{all positive terms}\\ &> \delta+\delta^2/4\end{aligned}$$

Then

$$(1+\delta)^{(1+\delta)} > e^{\delta+\delta^2/4}$$

Hence,

$$P[X>(1+\delta)\mu] < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} < \left(\frac{e^{\delta}}{e^{\delta+\delta^2/4}}\right)^{\mu} < e^{-\mu\sigma^2/4}$$

□

Theorem 6 states that if data satisfies certain conditions, then one could estimate its inner product within a small error.

Theorem 6. *Suppose data* $\mathbf{u}\in\mathbb{R}^n$ *and* $\frac{||\mathbf{u}||_\infty}{||\mathbf{u}||_2}\le M$, *where* $||\mathbf{u}||_\infty=\max(|x_1|,|x_2|,\ldots,|x_n|)$. *Consider a measurement matrix* $\Phi\in\mathbb{R}^{m\times n}$ *with entries as shown in (5.1). Then* $y=\frac{1}{\sqrt{m}}\Phi\mathbf{u}$ *can produce an estimate with m sensor nodes which satisfies* $(1-\epsilon)||\mathbf{u}||_2^2\le||\frac{1}{\sqrt{m}}\Phi\mathbf{u}||^2\le(1+\epsilon)||\mathbf{u}||_2^2$ *with probability* $\ge 1-\delta$, *where* $\delta=e^{-\frac{\epsilon^2\tau m}{8+4(s-3)M^2}}(0<\tau<\frac{1}{4})$.

Proof. For any vector $\mathbf{u}\in\mathbb{R}^n$ which satisfies $\frac{||\mathbf{u}||_\infty}{||\mathbf{u}||_2}\le M$. Define $\mathbf{x}=\frac{1}{\sqrt{m}}\Phi\mathbf{u}$ and random variable $z=\mathbf{u}^T\mathbf{u}$. By use of Chebyshev's inequality

$$\begin{aligned}&P(|z-\mathbf{u}^T\mathbf{u}|\ge\epsilon||\mathbf{u}||_2^2)\\ &\le\frac{Var(z)}{\epsilon^2||\mathbf{u}||_2^4}\\ &=\frac{1}{\epsilon^2 m}\left(2\frac{||\mathbf{u}||_2^4}{||\mathbf{u}||_2^4}+(s-3)\frac{\sum_{j=1}^n u_j^4}{||\mathbf{u}||_2^4}\right)\\ &\le\frac{1}{\epsilon^2 m}\left(2\frac{||\mathbf{u}||_2^4}{||\mathbf{u}||_2^4}+(s-3)\frac{\sum_{j=1}^n u_j^2||\mathbf{u}||_\infty^2}{||\mathbf{u}||_2^4}\right)\\ &\le\frac{1}{\epsilon^2 m}(2+(s-3)M^2)\end{aligned}\tag{5.7}$$

Therefore, $P(|z - \mathbf{u}^T\mathbf{u}| \leq \epsilon||\mathbf{u}||_2^2) \geq 1 - \frac{1}{\epsilon^2 m}(2 + (s-3)M^2)$. Let $1 - \frac{1}{\epsilon^2 m}(2 + (s-3)M^2) = 1 - \delta$ and we obtain $\delta = \frac{2+(s-3)M^2}{\epsilon^2 m}$. In many situations, this bound is impractical since ϵ is chosen to be small (even close to zero).

A tighter bound can be derived by application of Theorem 3. Partition measurement matrix Φ into m_2 sub-matrices $\{\Phi_1, \Phi_2, \ldots, \Phi_{m_2}\}$ with size of $m_1 \times n$. We define $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{m_2}] = [\frac{1}{\sqrt{m_1}}\Phi_1\mathbf{u}, \frac{1}{\sqrt{m_1}}\Phi_2\mathbf{u}, \ldots, \frac{1}{\sqrt{m_1}}\Phi_{m_2}\mathbf{u}]$ and $z_l = \mathbf{x}_l^T\mathbf{x}_l$, where $1 \leq l \leq m_2$. From (5.7) we have:

$$P(|z_l - \mathbf{u}^T\mathbf{u}| \geq \epsilon||\mathbf{u}||_2^2) \leq \frac{1}{\epsilon^2 m_1}(2 + (s-3)M^2) \stackrel{\triangle}{=} \rho \tag{5.8}$$

Note that ρ is a constant if ϵ is fixed and $m_1 = (2 + (s-3)M^2)/(\epsilon^2\rho)$.

Let χ_l be the indicator random variable of the event $\{|z_l - \mathbf{u}^T\mathbf{u}| \geq \epsilon||\mathbf{u}||_2^2\}$ and $\chi = \sum_{l=1}^{m_2}\chi_l$. Then, $E[\chi] = m_2\rho$ and by application of Lemma 3

$$P\big(\chi > (1+c)m_2\rho\big) < e^{-m_2\rho c^2/4} \tag{5.9}$$

where c is a constant and $0 < c < 1$.

Set $1 - e^{-m_2\rho c^2/4} = 1 - \delta$ and we can get $m_2 = \frac{4}{\rho c^2}\log_e(1/\delta)$. $m = m_1 m_2 = \frac{2+(s-3)M^2}{\epsilon^2\rho} \cdot \frac{4}{\rho c^2}\log_e(1/\delta) = \frac{8+4(s-3)M^2}{\epsilon^2\rho^2 c^2}\log_e(1/\delta)$ and $\delta = e^{-\frac{\epsilon^2\rho^2 c^2 m}{8+4(s-3)M^2}}$. Therefore, the event $(1-\epsilon)||\mathbf{u}||_2^2 \leq ||\frac{1}{\sqrt{m}}\Phi\mathbf{u}||^2 \leq (1+\epsilon)||\mathbf{u}||_2^2$ occurs with probability $\geq 1 - \delta$, where $\delta = e^{-\frac{\epsilon^2\rho^2 c^2 m}{8+4(s-3)M^2}} = e^{-\frac{\epsilon^2\tau m}{8+4(s-3)M^2}} (0 < \tau < \frac{1}{4})$.

□

Theorem 7 states that if the aggregated data received by sensor nodes is sparse in the original data domain, one can produce an estimate of the original data within a small error in the existence of bogus data.

Theorem 7. *Suppose data* $\mathbf{u} \in \mathbb{R}^n$, $\frac{||\mathbf{u}||_\infty}{||\mathbf{u}||_2} \leq M$, *where* $||\mathbf{u}||_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|)$ *and a measurement matrix* Φ *with vectors* $(\mathbf{v}_j)_{j\in J} \in \mathbb{R}^n$ *as columns. The entries of*

Φ *are generated according to (5.1). Let* $\mathbf{h}$ *be the bogus data that disseminated by the attacker and the corresponding transformation coefficients* $\mathbf{e} = \Psi\mathbf{h} = [e_1, e_2, \ldots, e_n]$. *The aggregated data* $\mathbf{g}$ *is a vector supported on a set* $T_0 \subset J$ *obeying* $|T| = K < J$. *Assume* $\frac{||\theta||_2}{|e_i|} = \frac{1}{\gamma}$, $\mathbf{y} = \frac{1}{\sqrt{m}}\Phi\mathbf{u} = \frac{1}{\sqrt{m}}\Phi\Psi\theta$ *and* $\delta_s < 0$. *One can produce an estimate with* m *uncompromised sensor nodes such that*

$$||\mathbf{u}' - \mathbf{u}||_2^2 = ||\theta' - \theta||_2^2 \leq K\epsilon^2||\mathbf{u}||_2^2 \tag{5.10}$$

with probability $\geq 1 - \delta$, *where* $\delta = ne^{-\frac{(\epsilon-\gamma)^2\tau m}{8+4(s-3)M^2}} (0 < \tau < \frac{1}{4})$.

Proof. $\mathbf{g} = \frac{1}{\sqrt{m}}\Phi(\mathbf{u} + \mathbf{h}) = \frac{1}{\sqrt{m}}\Phi\Psi(\theta + \mathbf{e})$. We first show that $u + h$ is the unique solution to

$$\min ||\mathbf{f}||_{\ell_1} \quad s.t. \quad \frac{1}{\sqrt{m}}\Phi\mathbf{f} = \mathbf{g} \tag{5.11}$$

We know that $u + h$ is one possible solution to equation (5.11). Based on Theorem 1.3 in [44] and assuming there exists another solution f such that

$$||\mathbf{f}||_{\ell_1} \leq ||\mathbf{u} + \mathbf{h}||_{\ell_1} = \sum_{j\in T_0} |u_j + h_j|$$

$$\begin{aligned}
||\mathbf{f}||_{\ell_1} &= \sum_{j\in T_0} |u_j + h_j + (f_j - u_j - h_j)| + \sum_{j\notin T_0} |f_j| \\
&\geq \sum_{j\in T_0} \operatorname{sgn}(u_j + h_j)(u_j + h_j + (f_j - u_j - h_j)) + \sum_{j\notin T_0} |f_j| \\
&= \sum_{j\in T_0} |u_j + h_j| + \sum_{j\in T_0} (f_j - u_j - h_j)\langle \mathbf{w}, \mathbf{v}_j\rangle + \sum_{j\notin T_0} f_j\langle \mathbf{w}, \mathbf{v}_j\rangle \\
&= \sum_{j\in T_0} |u_j + h_j| + \langle \mathbf{w}, \sum_{j\in J} f_j\mathbf{v}_j - \sum_{j\in T_0}(u_j + h_j)\rangle \\
&= \sum_{j\in T_0} |u_j + h_j| + \langle \mathbf{w}, \mathbf{g} - \mathbf{g}\rangle \\
&= \sum_{j\in T_0} |u_j + h_j|
\end{aligned}$$

Hence $||\mathbf{f}||_{\ell_1} = \sum_{j\in T_0} |u_j + h_j|$. Since $|\langle \mathbf{w}, \mathbf{v}_j\rangle|$ is strictly less than 1 for all $j \notin T_0$, this forces $f_j = 0$ for all $j \notin T_0$. Thus,

$$\sum_{j \in T_0} (f_j - u_j - h_j)\mathbf{v}_j = \mathbf{g} - \mathbf{g} = 0$$

Since $\delta_s < 1$, it is conclude that $f_j = u_j + h_j$ for all $j \in T_0$ and thus $\mathbf{f} = \mathbf{u} + \mathbf{h}$.

We also define $\mathbf{x} = \frac{1}{\sqrt{m}}\Phi\Psi$ and $\mathbf{z} = \mathbf{x}^T\mathbf{y}$. From (5.8) and (5.9) we have

$$P(|z_l - \mathbf{u}^T\Psi| \geq \epsilon||\mathbf{u}||_2||\Psi||_2) \leq \frac{1}{\epsilon^2 m_1}(2 + (s-3)M^2) \stackrel{\triangle}{=} \rho \tag{5.12}$$

and

$$P\big(\chi > (1+c)m_2\rho\big) < e^{-m_2\rho c^2/4} \tag{5.13}$$

where c is a constant and $0 < c < 1$.

For any pair of vectors u and $\Psi_i \in \{\Psi_1, \ldots, \Psi_n\}$, the probability that z_l lies out the approximation interval is at most $e^{-m_2\rho c^2/4}$. Taking the union bound over all $\Psi_i \in \{\Psi_1, \ldots, \Psi_n\}$, the probability that at least one z_l lies outside the approximation interval is upper bounded by $p_e < ne^{-m_2\rho c^2/4}$. Set $1 - ne^{-m_2\rho c^2/4} = 1 - \delta$ and we can get $m_2 = \frac{4}{\rho c^2}\log_e(n/\delta)$. Then, $m = m_1 m_2 = \frac{2+(s-3)M^2}{\epsilon^2\rho} \cdot \frac{4}{\rho c^2}\log_e(n/\delta) = \frac{8+4(s-3)M^2}{\epsilon^2\rho^2 c^2}\log_e(n/\delta)$ and $\delta = ne^{-\frac{\epsilon^2\rho^2 c^2 m}{8+4(s-3)M^2}}$. Therefore,

$$|z_l - \mathbf{u}^T\Psi_i| \leq \epsilon||\mathbf{u}||_2||\Psi_i||_2$$

which can be rewritten as

$$|\theta_i' - \theta_i| \leq \epsilon||\theta||_2 \tag{5.14}$$

which occurs with probability $\geq 1 - \delta$, where $\delta = ne^{-\frac{\epsilon^2\tau m}{8+4(s-3)M^2}}(0 < \tau < \frac{1}{4})$. Here we use the property $||\Psi_i||_2 = 1$ and $||\mathbf{u}||_2 = ||\theta||_2$. Then we have

$$|\theta_i' - \theta_i - e_i| \leq \epsilon||\theta||_2 \tag{5.15}$$

By triangle inequality

$$||\theta_i' - \theta_i| - |e_i|| \leq \epsilon||\theta||_2$$

and

$$|\theta_i' - \theta_i| \leq \epsilon||\theta||_2 + |e_i| = (\epsilon + \gamma)||\theta||_2$$

Summing all K items together,

$$||\theta' - \theta||_2^2 \leq K(\epsilon + \gamma)^2||\theta||_2^2$$

with probability $\geq 1 - \delta$, where $\delta = ne^{-\frac{\epsilon^2 \tau m}{8+4(s-3)M^2}} (0 < \tau < \frac{1}{4})$. Let $\epsilon' + \gamma = \epsilon$, so that

$$||\mathbf{u}' - \mathbf{u}'||_2^2 = ||\theta' - \theta||_2^2 \leq K\epsilon^2||\mathbf{u}||_2^2$$

with probability $\geq 1 - \delta$, where $\delta = ne^{-\frac{(\epsilon-\gamma)^2 \tau m}{8+4(s-3)M^2}} (0 < \tau < \frac{1}{4})$.

□

Theorem 8 states that if nearly sparse data satisfies certain conditions, one can produce an estimate of data within a small error in the existence of bogus data, and the performance is comparable to that of the best k-term approximation.

Theorem 8. *Suppose data* $\mathbf{u} \in \mathbb{R}^n$, $\frac{||\mathbf{u}||_\infty}{||\mathbf{u}||_2} \leq M$, *where* $||\mathbf{u}||_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|)$ *and measurement matrix* $\Phi \in \mathbb{R}^{m \times n}$ *with entries as shown in (5.1). h be the bogus data that disseminated by the attacker and the corresponding transformation coefficients* $e = [e_1, e_2, \ldots, e_n]$. *Let* $\frac{||\theta||_2}{|e_i|} = \frac{1}{\gamma}$, $\mathbf{x} = \frac{1}{\sqrt{m}}\Phi\mathbf{u}$ *and* $\theta = \Psi\mathbf{u}$. *If the k largest transform coefficients in magnitude gives an approximation error* $||\theta - \theta_{app}||_2^2 \leq \eta||\theta||_2^2$, *then one can produce an estimate with m uncompromised sensor nodes such that*

$$||\mathbf{u} - \hat{\mathbf{u}}||_2^2 \leq (1 + \epsilon)\eta||\mathbf{u}||_2^2$$

with probability $\geq 1 - \delta$, *where* $\delta = ne^{-\frac{(\sqrt{3+\frac{\epsilon\eta}{k}} - \gamma - \sqrt{3})^2 \tau m}{8+4(s-3)M^2}} (0 < \tau < \frac{1}{4})$.

Proof. Suppose an orthonormal basis matrix Ψ consisting of n basis vectors $\{\psi_1, \psi_2, \ldots, \psi_n\} \subset \mathbb{R}^n$ and $\theta = [\mathbf{u}^T\psi_1, \mathbf{u}^T\psi_2, \ldots, \mathbf{u}^T\psi_n]^T$. Then, we sort the magnitude of coefficients θ in decreasing order, i.e., $|\theta|_1 > |\theta|_2 > \ldots > |\theta|_n$ and the approximation error by taking

the k coefficients with the largest magnitude and discarding the remaining coefficients is $\sum_{i=k+1}^{n} |\theta|_i^2$ and $||\theta - \theta_{app}||_2^2 \leq \eta ||\theta||_2^2$.

Then, from equation (5.15) in Theorem 7

$$|\theta_i' - \theta_i - e_i| \leq \alpha ||\theta||_2 \tag{5.16}$$

which occurs with probability $\geq 1 - \delta$, where $\delta = ne^{-\frac{\alpha^2 \tau m}{8+4(s-3)M^2}} (0 < \tau < \frac{1}{4})$. By the triangle inequality, (5.16) becomes

$$|\hat{\theta}_i - \theta_i| \leq \alpha ||\theta||_2 + |c|_i \leq (\alpha + \gamma) ||\theta||_2 \tag{5.17}$$

The above condition implies

$$|\theta_i| - (\alpha + \gamma)||\theta||_2 \leq |\hat{\theta}_i| \leq |\theta_i| + (\alpha + \gamma)||\theta||_2$$

We define our approximation $\hat{\theta}$ as keeping the k largest coefficients of $\hat{\theta}$ in magnitude, and setting the remaining coefficients to zero. Let $\hat{\Omega}$ be the index set of the k largest estimates of $\hat{\theta}_i$ which we keep and Ω be the index set of the k largest coefficients θ.

$$\begin{aligned} ||\theta - \hat{\theta}||_2^2 &= \sum_{i \in \hat{\Omega}} |\theta_i - \hat{\theta}_i|^2 + \sum_{i \in \hat{\Omega}^C} |\theta_i|^2 \\ &\leq k(\alpha + \gamma)^2 ||\theta||_2^2 + \sum_{i \in \hat{\Omega}^C} |\theta_i|^2 \end{aligned}$$

In the ideal case, $\Omega = \hat{\Omega}$. If $\Omega \neq \hat{\Omega}$, there exists some i and j that $|\hat{\theta}_i| > |\hat{\theta}_j|$, but $|\theta_i| < |\theta_j|$. From equation (5.17), we have $|\theta_i| - |\theta_j| \leq 2(\alpha + \gamma)||\theta||_2$. Moreover, $|\theta_i|^2 + |\theta_j|^2 \leq ||\theta||_2^2$ implies that $|\theta_i| + |\theta_j| \leq \sqrt{3}||\theta||_2$. Therefore, $|\theta_i|^2 - |\theta_j|^2 \leq \sqrt{3}||\theta||_2 \cdot 2||\theta||_2 = 2\sqrt{3}(\alpha + \gamma)||\theta||_2^2$. Then,

$$\sum_{i \in \hat{\Omega}^C} |\theta_i|^2 \leq \sum_{i \in \Omega^C} |\theta_i|^2 + 2\sqrt{3}k(\alpha + \gamma)||\theta||_2^2$$

and

$$\begin{aligned} ||\theta - \hat{\theta}||_2^2 &= \sum_{i \in \hat{\Omega}} |\theta_i - \hat{\theta}_i|^2 + \sum_{i \in \hat{\Omega}^C} |\theta_i|^2 \\ &\leq k(\alpha+\gamma)^2 ||\theta||_2^2 + 2\sqrt{3}k(\alpha+\gamma)||\theta||_2^2 + ||\theta - \theta_{app}||_2^2 \\ &\leq (1+\epsilon)\eta||\theta||_2^2 \end{aligned}$$

Let $k(\alpha+\gamma)^2 + 2\sqrt{3}k(\alpha+\gamma) = \epsilon\eta$, we can find the positive root, which is $\alpha = \sqrt{3+\frac{\epsilon\eta}{k}} - \gamma - \sqrt{3}$. Therefore, we can have

$$||\mathbf{u} - \hat{\mathbf{u}}||_2^2 = ||\theta - \hat{\theta}||_2^2 \leq (1+\epsilon)\eta||\theta||_2^2 = (1+\epsilon)\eta||\mathbf{u}||_2^2$$

with probability $\geq 1 - \delta$, where $\delta = ne^{-\frac{(\sqrt{3+\frac{\epsilon\eta}{k}} - \gamma - \sqrt{3})^2 \tau m}{8+4(s-3)M^2}} (0 < \tau < \frac{1}{4})$. □

5.4 Numerical Results

We use simulation to investigate the property of received data and the effect of the various parameters on the data reconstruction process. Figure 5.2 shows the received signal at the sensor node in the time domain. It is collected by PulsOn P220 UWB radar. This real-world data is not truly sparse in any transformation domain. However, it can be well approximated by the optimal k-term approximation, i.e., the data can be decomposed into a discrete cosine transform (DCT) domain and the coefficients are sorted in decreasing order of the magnitude, only the largest k coefficients are kept and the remaining is set to zero. Figure 5.3 illustrates the relation between the number of coefficients we keep and the approximation error. When $k = 200$, the approximation error is about 2%. Since only 1/10 of the original coefficients are kept, we can claim the received signal is nearly sparse in the DCT domain.

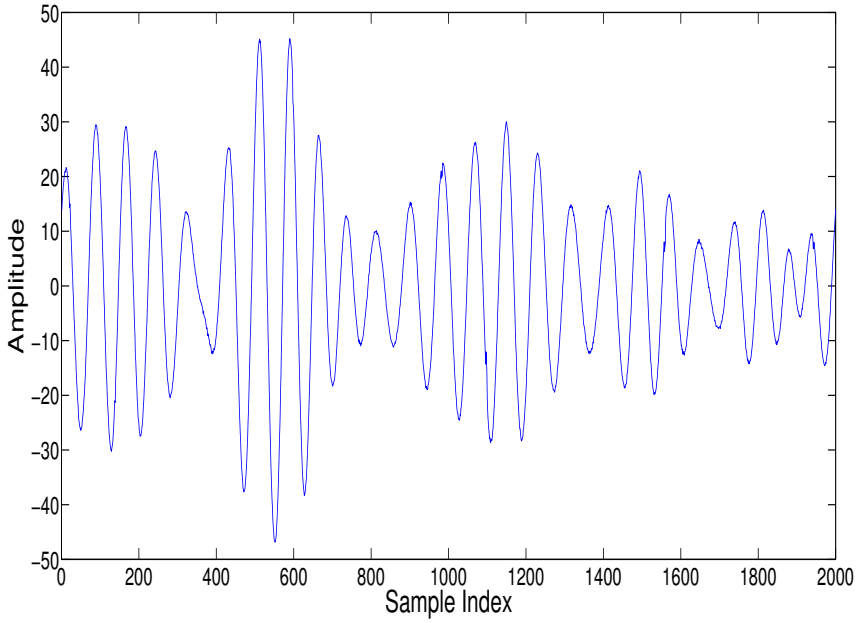


Figure 5.2. Waveform of received signal at the sensor node.

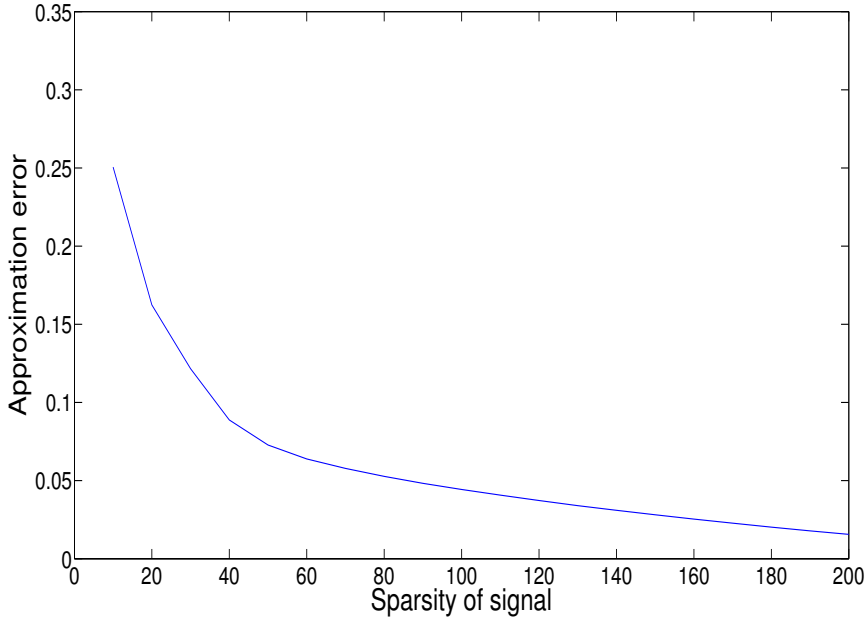


Figure 5.3. The approximation error of the data $\frac{||u-\hat{u}||_2^2}{||u||_2^2}$ versus sparsity of signal k.

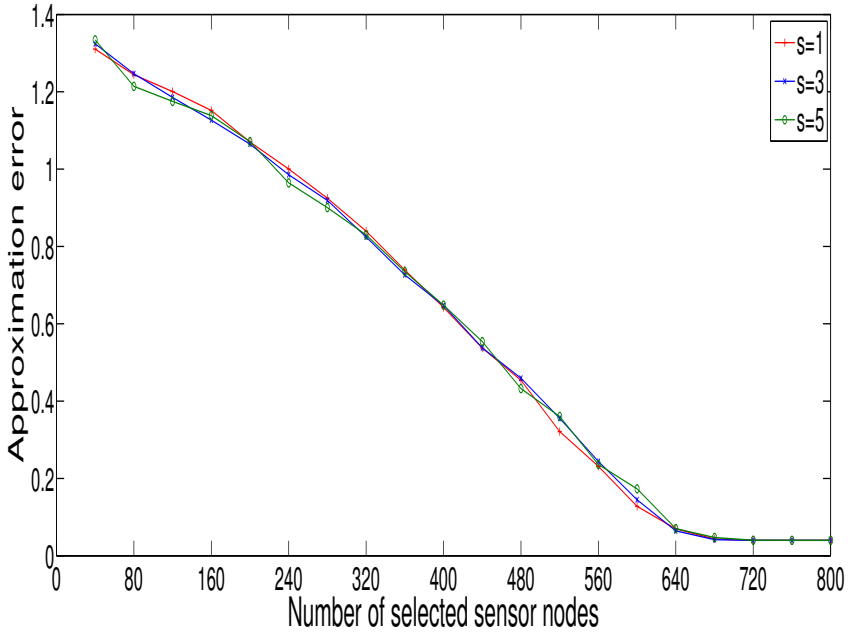


Figure 5.4. The approximation error of the data $\frac{||u-\hat{u}||_2^2}{||u||_2^2}$ versus the number of selected sensor nodes with various sparsity of random measurement matrix.

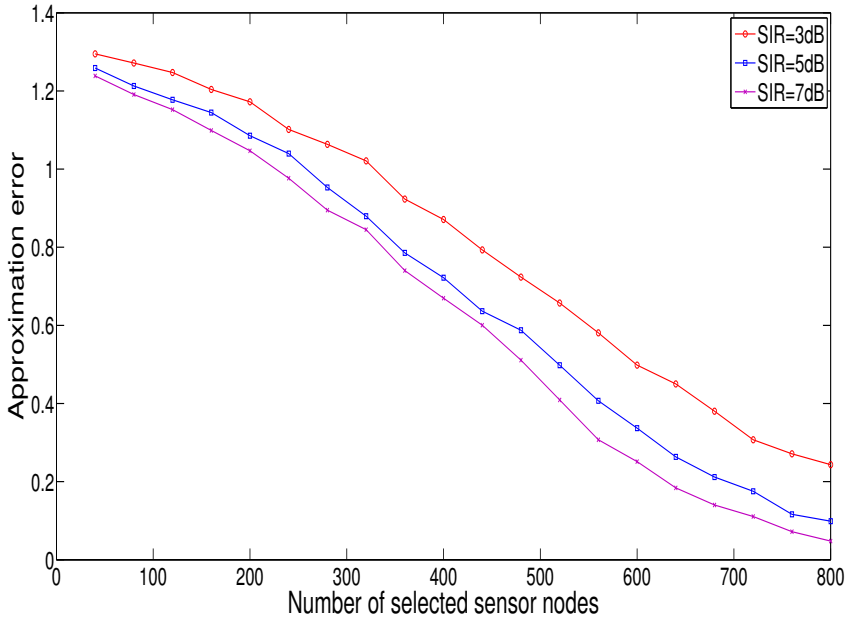


Figure 5.5. The approximation error of the data $\frac{||u-\hat{u}||_2^2}{||u||_2^2}$ versus the number of selected sensor nodes. $s = 3$ and 10% sensor nodes are compromised.

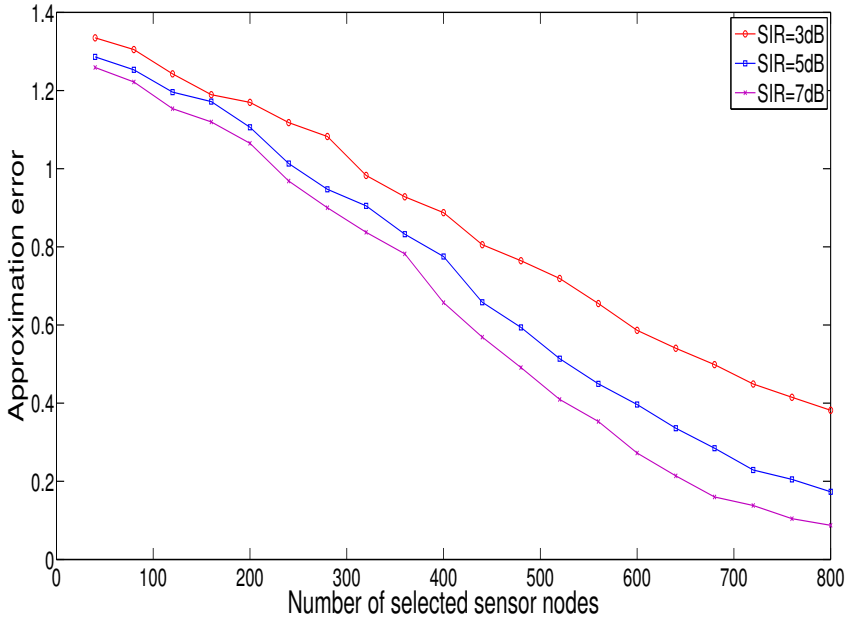


Figure 5.6. The approximation error of the data $\frac{||u-\hat{u}||_2^2}{||u||_2^2}$ versus the number of selected sensor nodes. $s = 3$ and 30% sensor nodes are compromised.

Figure 5.4 compares the approximation error of the data versus the number of selected sensor nodes with various sparsity s of a random measurement matrix . Note that $1/s = 1$ implies the conventional compressive sensing with Bernoulli ensemble is employed. The approximation error is comparable to the one shown in Figure 5.3 if the number of selected sensor nodes is sufficient large. It is noted that the performance is almost the same for sparsity $s = 1, 3$ and 5. This is because in our case $M = 0.458$, which is very small and the length of data is 2000, which is comparablely large. From equation (5.7), we can see the extra approximation error introduced by the sparsity of the measurement matrix could be ignored. Figure 5.4 also implies the trade-off between the sparsity of random measurement matrix and the number of sensor nodes used to reconstruct the original signal, which can be shown by Theorem 6.

Figure 5.5 and Figure 5.6 compare the approximation error of the data versus the number of selected sensor nodes when $s = 3$ and a different number of compromised sensor nodes. As we can expect from the theoretical analysis, the fusion center could achieve better performance with larger SIR, smaller value of sparsity, s, and less sensor nodes that is compromised by the attacker.

Finally, the average network throughput is reciprocal to the length of data n and the compressive sensing decoding has $O(n^3)$ computational complexity. Therefore, there is also a trade-off between the overall throughput of proposed WSN and the computation complexity at the fusion center.

5.5 Conclusions

Due to limited energy and physical size of the sensor nodes, the conventional security mechanisms with high computation complexity are not feasible for WSNs. Compressive sensing provides both signal compression and encryption guarantees, without the additional computational cost of a separate encryption protocol. We also studied the computational and information-theoretic secrecy of compressive sensing scheme. This motivates us to design a compressive sensing-based encryption scheme for WSN with simple structure.

The proposed encryption scheme is quite simple and only a small amount of memory and storage are required to perform the encryption. Each sensor node sends the aggregated data to the fusion center at every n symbol interval, which significantly reduces the overall throughput. It is noted that each sensor node works in a distributed manner and synchronization is not required. We prove the encryption is perfectly secret and if only a fraction of randomizer bits is stored by an eavesdropper, then the probability that he/she could obtain any information about the plaintext approaches zero.

Our approach is also scalable and flexible: a trade-off can be made between the sparsity of a random measurement matrix and the number of uncompromised sensor nodes used to reconstruct the original signal. The simulation results indicate that our scheme is superior to the traditional distributed WSN. There is also a trade-off between the overall throughput of proposed WSN and the computation complexity at the fusion center.

CHAPTER 6

DATA ENCRYPTION-BASED COMPRESSIVE SENSING

Security of data is an issue that is of significant interest. In this Chapter, we propose a compressive sensing-based data encryption scheme, which can represent the original signal with far fewer samples than the conventional Nyquist sampling-based system. Compressive sensing could also be used as an encryption algorithm with good secrecy.

6.1 Introduction

Security of data is an issue that is of significant interest. There are two fundamental issues:

- Security of data during transmission;
- Security of stored data.

The security of data during wireless transmission has been well addressed by different coding and modulation schemes such as spectrum spreading. The second issue, security of stored data is less studied than other security issues of data and of greater relevance. If sensitive data is stored in a database, then the user needs to find a way to preserve the security of the data. There needs to be a security measure in place so that even if the data is stolen, the eavesdropper cannot obtain further information about the plaintext.

Encryption is the perfect technique to solve this problem. Prior work does not address the critical issue of performance. But in this work, for the first time, we have

employed the novel compressive sensing which method is usually treated as a data compression method for the success of encryption in data as well.

One aspect of encryption is the granularity of data to be encrypted or decrypted. A small sample group may appear to be the best choice, because it would minimize the number of bytes encrypted. However, practical methods of embedding encryption entail a significant start up cost for an encryption operation. A large sample group encryption amortizes this cost over larger data. Since the encryption process of compressive sensing is quite simple, it can easily solve this problem by increasing the number of encrypted information bits without introducing any extra start up cost.

In this approach the sender of the encrypted data supplies the key, and only authorized users who are given the key can decrypt the data using the decryption algorithm. Since the key is owned by the sender, an unauthorized person who may get hold of encrypted data cannot access to the secret key. From the above investigation it is clear that there is a great need for a new approach to encryption that is both highly secured and efficient to avoid the high penalty over the encrypted data.

6.2 Compressive Sensing-Based Security

In data transmission system, security is a major concern. The sender of data wants to ensure that only authorized users may gain access to the data, but not unauthorized users. Figure 6.1 illustrates the conventional data communication system with encryption, data transmitted by a sender to an authorized user is coupled to a receiver while ensuring that the data remains unavailable to other users who are coupled to the same receiver. The data is encrypted by the sender using a key. The encrypted data is then scrambled and transmitted by the sender. All users who are coupled to the receiver will have access to the descrambled data, but the data is

unreadable because it is encrypted. Only the authorized user with a proper key may decrypt the encrypted data to obtain clear data.

In most cryptographic methods, the security of the message is based on the fact that we have transformed the message into something that is not immediately decipherable. When attempts are made to break the code, if a message is extracted that makes sense, it is assumed that the message has been successfully decoded. When using compressive sensing (CS) as the encryption scheme, this is not necessarily a valid assumption.

We compress the message x into a short message y by $y = \Phi x$, where Φ is the sensing matrix. In order to decrypt the message, an attacker would need to find both Φ and x that gives y. So, the question becomes does there exist Φ' and x' such that $y = \Phi' x'$. Assuming that any sensing matrix is possible, we have

$$\Phi' = \|x'\|_2^{-2} \Phi x (x')^{\mathrm{T}} \tag{6.1}$$

is a possible solution for message x'. Since the system is under-determined, i.e., we know fewer variables than we need to find, it is impossible that we find a solution x' that is not identical to the original data x. The important part of this is Φ' being a valid sensing matrix. This is certainly the case if we assume that we convey the entire sensing matrix.

Since it is impossible to decrypt the original message without the entire sensing matrix, the next problem we should handle is how could we assure that only the authentic user can access to it. In our case, Direct-Sequence Spread Spectrum (DSSS) is employed to modulate the sensing matrix and information sequence, and the modulated message is transmitted over the air. DSSS modulates the information bit pseudorandomly with a continuous string of pseudonoise (PN) code symbols called

"chips", each of which has a much shorter duration than information bit. That is, each information bit is modulated by a sequence of much faster chips. This chip looks like a noise signal. It is a pseudorandom sequence of 1 and -1 values, at a frequency much higher than that of the original signal. Pseudonoise (PN) code is generated by a pseudo random generator with an initial random seed which is shared between transmitter and receiver. The receiver can then use the same PN sequence to counteract the effect of the PN sequence on the received signal in order to reconstruct the information sequence. The modulated signal resembles white noise. However, this noise-like signal can be used to exactly reconstruct the original data at the receiving end, by multiplying it by the same pseudorandom sequence (because $1 \times 1 = 1$, and $-1 \times -1 = 1$). This process, known as "de-spreading", mathematically constitutes a correlation of the transmitted PN sequence with the PN sequence that the receiver believes the transmitter is using. The resulting effect of enhancing signal to noise ratio on the channel is called process gain. This effect can be made larger by employing a longer PN sequence and more chips per bit, but physical devices used to generate the PN sequence impose practical limits on attainable processing gain.

We further investigate the secrecy of compressive sensing from the information point of view. Shannon introduces the idea of perfect secrecy in [37]. An encryption scheme achieves perfect secrecy if the probability of a message conditioned on the cryptogram is equal to the a priori probability of the message, $P(X = x|Y = y) = P(X = x)$. Alternatively, this condition can be interpreted as $I(X;Y) = 0$. If an eavesdropper does not have the correct PN sequence, he will only have partial information of the original message x, which can be described as $y_p = h \otimes y$, where h is the impulse response of the bandpass filter. Since basis matrix Φ is linearly dependent, we can conclude that the original message x and y_p are dependent, which

indicates $I(X;Y) \neq 0$. Therefore, compressive sensing cannot be an encryption algorithm with perfect secrecy.

It is proved in [19] that the sensing matrix Φ is said to satisfy a restricted isometry property (RIP) of order k if there exists a δ_k such that,

$$(1-\delta_k) \parallel x \parallel_2^2 \leq \parallel \Phi x \parallel_2^2 \leq (1-\delta_k) \parallel x \parallel_2^2 \tag{6.2}$$

holds for all x with sparsity k. It is verified in [55] that if $\delta_{2k} < \sqrt{2}-1$, one can recover the original message x with high probability. If the length of the PN sequence is long enough, the energy of the original message that falls in a certain bandwidth would decrease so that $||y_p||_2^2 < (1-\delta_k)||x||_2^2$. The probability to recover the original message would be quite small. In other words, compressive sensing can be an encryption algorithm with good secrecy.

In the worst case, unauthentic users are assumed to decrypt the entire sensing matrix sequence successfully. They still cannot decrypt the original message since they do not have information about the dimension of the sensing matrix. For example, 1200 information bits could construct a 12×100, 20×60, 30×40 matrix. There are so many possible combinations for the dimension of the sensing matrix. It is also shown in [55] that ck measurements (typically $c \approx 3$ or 4) are required to reconstruct the message. We assume the original sensing matrix with the dimension of $M \times N$ and the size of estimated sensing matrix is $M' \times N'$. If $M' < ck$, the probability of recovering the original message x is zero. If $M' \geq ck$, Theorem 1 in [26] demonstrates the sparsity of the estimated message is M' with probability one over the set of matrix Φ' . Hence, the original message could never be reconstructed since the sparsity of the original message x is k not M'. It is a great advantage of compressive sensing to be implemented as an encryption algorithm.

Figure 6.2 demonstrates our proposed data communication system. The transmission part is quite simple. Compressive sensing is employed to do joint data compression and encryption which can simplify the hardware complexity and reduce processing time. Pseudonoise (PN) code can be easily generated by a pseudo random generator with an initial random seed which is shared between transmitter and receiver. At the receiving part, the authorized user will have access to the clean data using the authentic measurement matrix. Since the matrix is almost random and the dimension is large, it is impossible for other users to decrypt it even if they get the encrypted data. The performance of compressive sensing as an encryption algorithm will be given in the next section.

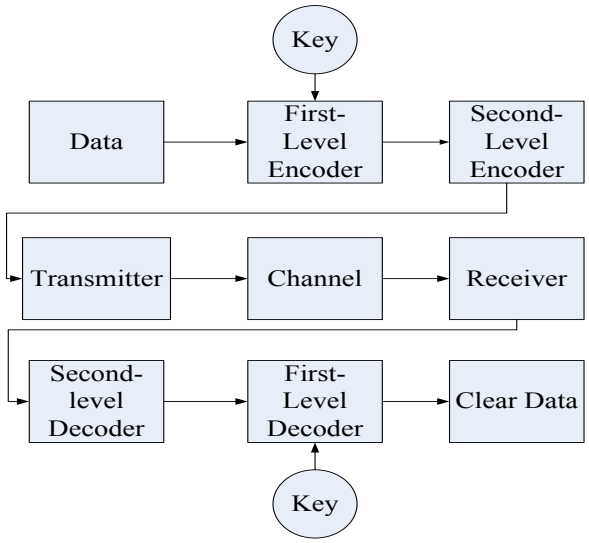


Figure 6.1. Block diagram of the conventional data communication system with encryption.

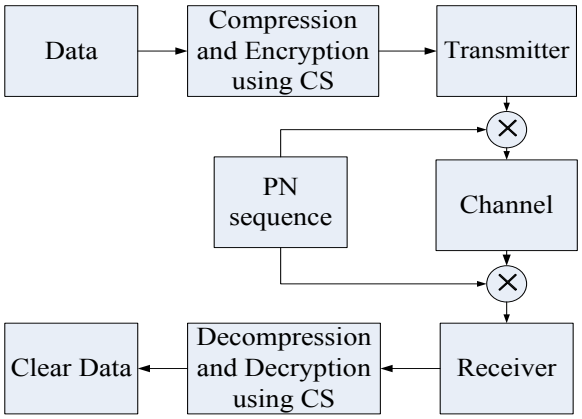


Figure 6.2. Block diagram of proposed compressive sensing-based encrypted data communication system.

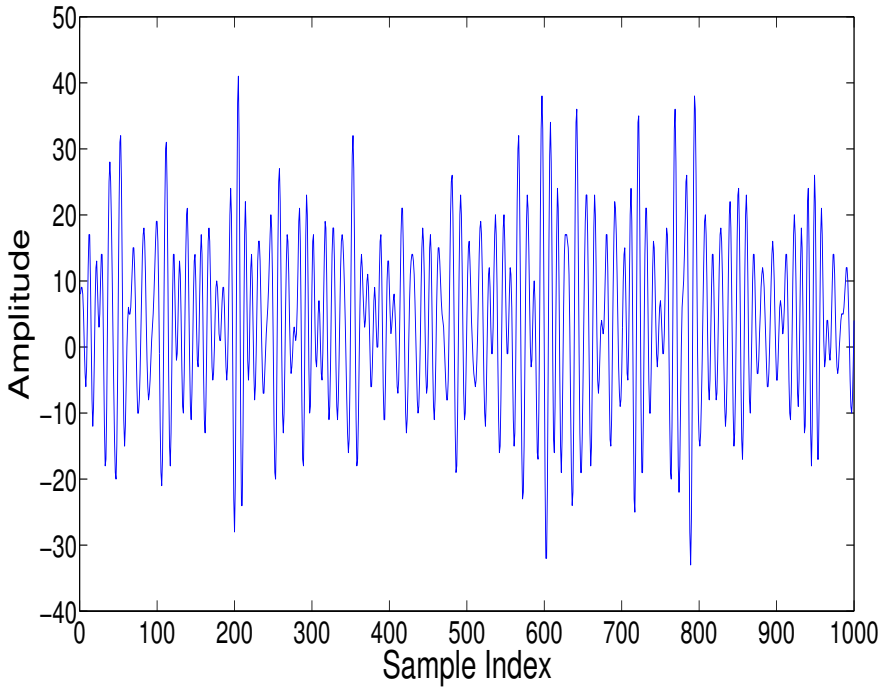
6.3 Numerical Results

UWB noise radar signal used in this simulation was collected at the Radar Imaging Lab at Villanova University . This data has also been used for through-wall imaging [56]. The frequency of the transmitted signal is 400-700 MHz and the sampling rate is 1.5 GHz. The property and waveform of UWB noise radar signal is illustrated in Figure 6.3. Our goal is to use M random measurements to exactly recover the original signal.

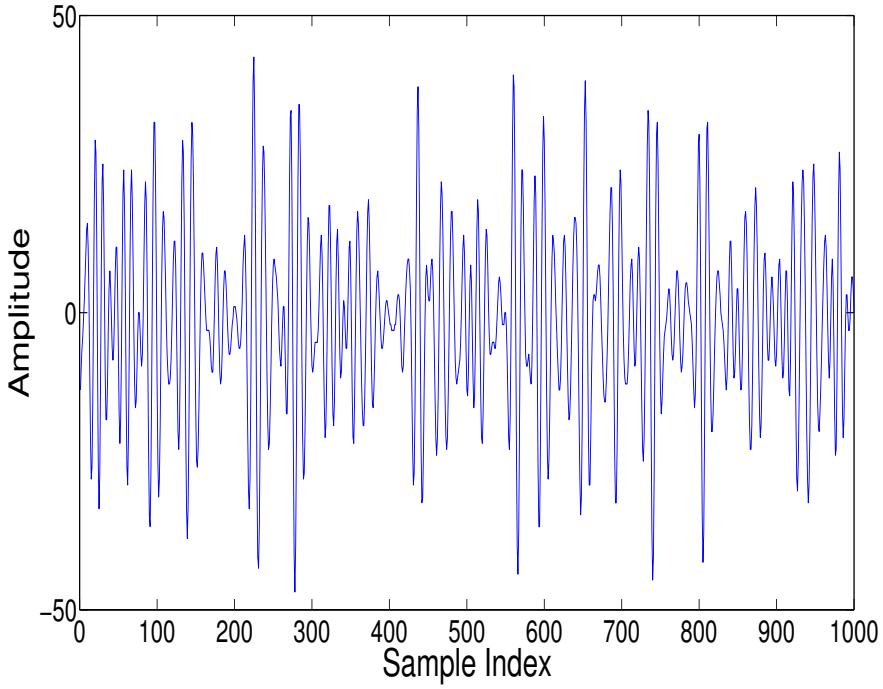
Figure 6.4(a) illustrates the sparse form of the UWB noise radar signal in cosine basis functions. We can see that only a small number of coefficients are nonzero. In other words, we can say the UWB signal is sparse when expressed in a cosine basis ($K << N$). Hence, we can apply the compressive sensing to the UWB noise radar signal. Figure 6.4(b) shows its exactly reconstruction from a new observation vector y with length of 377 samples. It suggests that we can use only 1/3 of the original data samples to represent the UWB noise radar signal without any information loss.

The performance of compressive sensing as an encryption algorithm are illustrated in Figure 6.5, 6.6 and 6.7. From the above investigation, we can see that the Bernoulli sensing matrix has a better performance, and we choose the Bernoulli matrix to run the following simulations. In Figure 6.5, we assume the only information that an unauthorized user knows is the dimension of the sensing matrix. In Figure 6.6, we assume, the unauthorized user has part of the authentic sensing matrix. In Figure 6.7, only the last row of the sensing matrix is not known to the unauthorized user. From these three Figures, we can conclude that compressive sensing is suitable to encrypt data and it is impossible to retrieve the clear data without the entire sensing matrix.

We have applied the novel concept of compressive sensing on a practical problem of sampling UWB noise radar signal. We draw the following conclusions: 1) The UWB noise radar signal is sparse when expressed in cosine basis Ψ. 2) Random Gaussian matrices are largely incoherent with any fixed basis, which can be efficiently acquiring the information from the original signal. We also propose a new compressive sensing-based data encryption and communication system. From the above investigation, we can conclude that compressive sensing is efficient to encrypt data and it is impossible to retrieve the clear data without the original sensing matrix Φ.

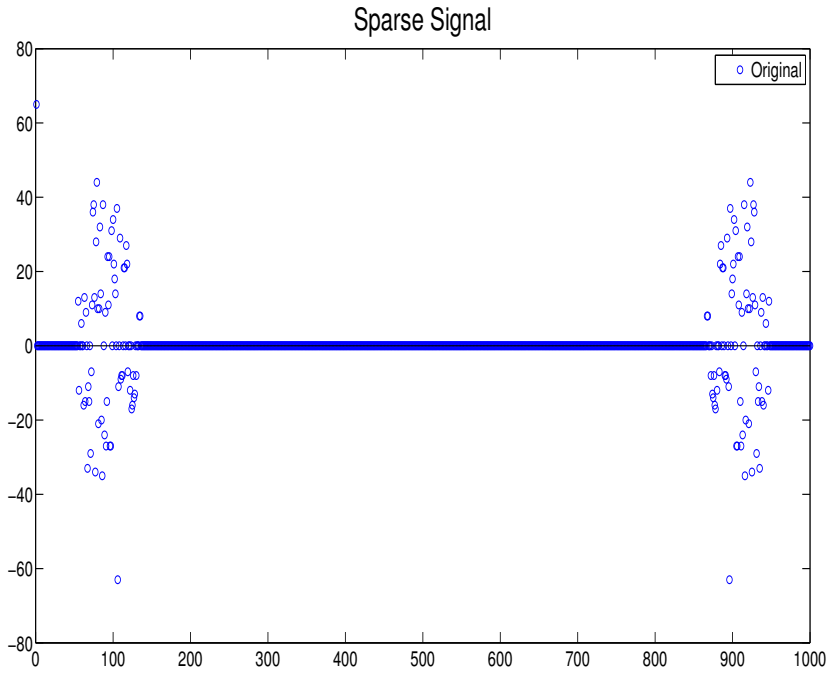


(a)

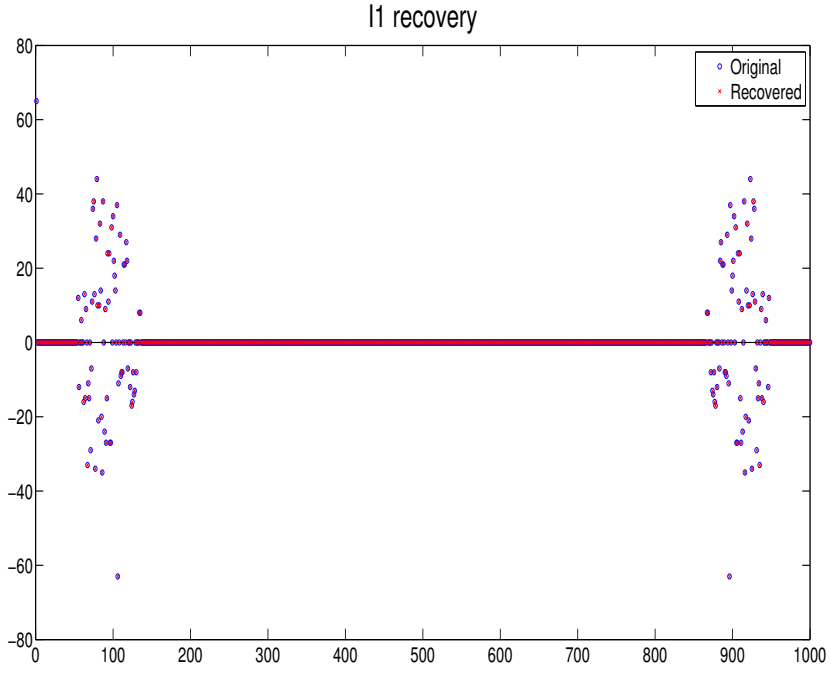


(b)

Figure 6.3. UWB noise waveforms of the transmitted signal (a) and the received signal (b) .



(a)



(b)

Figure 6.4. (a) Sparse UWB noise radar signal in cosine basis (b) reconstruction via ℓ_1 minimization. The reconstruction is perfect..

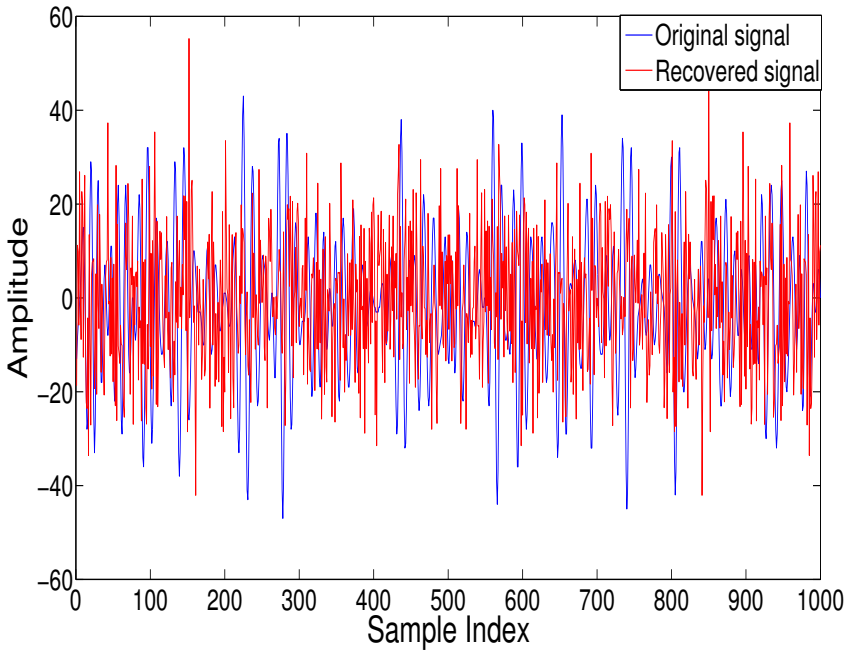


Figure 6.5. Recovery when user only knows the dimension of the sensing matrix .

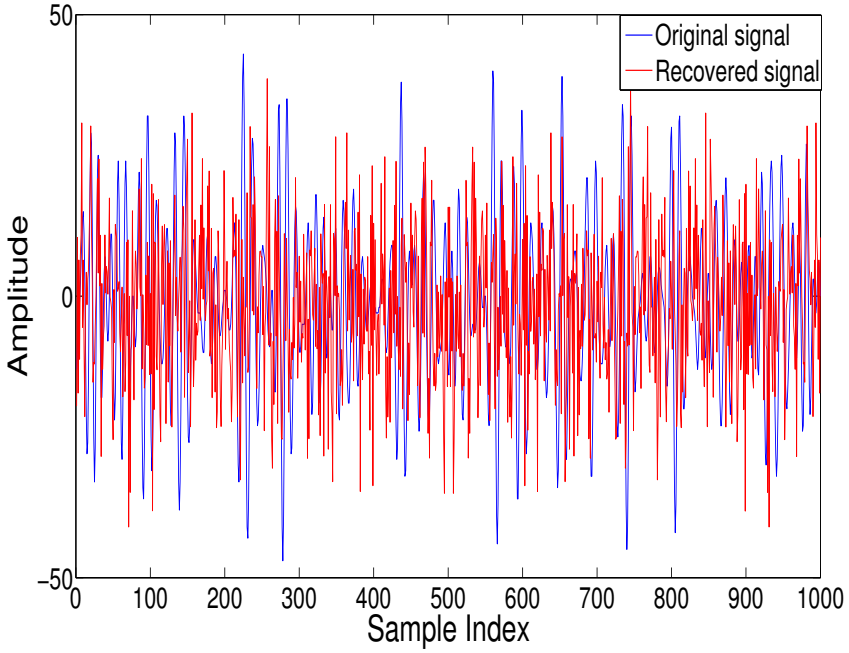


Figure 6.6. Recovery when user knows half of the sensing matrix.

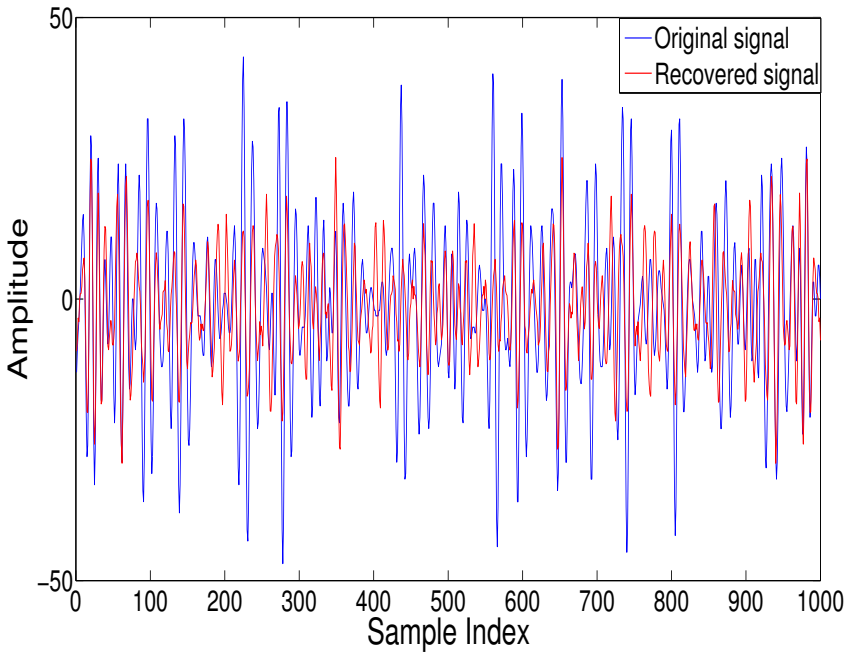


Figure 6.7. Recovery when only the last row of the sensing matrix is unknown .

CHAPTER 7

COMPRESSIVE SENSING-BASED MIMO UWB COMMUNICATION SYSTEM

The sampling rate is the bottleneck for an Ultra-WideBand (UWB) wireless communication system. Compressive sensing (CS) is a natural framework to handle this problem, which claims a small collection of linear projections of a sparse signal. It contains enough information for stable, sub-Nyquist signal acquisition. In this Chapter, we discuss a novel impulse radio UWB communication system which is sparse in the time domain. A multi-antenna PPM-modulated impulse radio UWB transceiver was first applied to broadband high-throughput 4G WLANs in [59]. Motivated by [59], we proposed a Space Time Orthogonal Block Code (STOBC), which could achieve a high diversity gain by exploiting the properties of proposed STOBC. We assume perfect channel state information is available and Maximum-Likelihood Estimation (MLE) is adopted to recover the transmitted codewords.

7.1 Space-Time Code and Maximum-Likelihood Estimation [59]

We consider a system equipped with $t = 2$ transmit and $r = 2$ receive antennas. The transmitted space-time codewords $\{S_d(-1), S_d(1)\}$ are given by the following (4×2) matrices

$$\mathbf{S}_d(-1) = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{S}_d(1) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} \tag{7.1}$$

where the rows indicate space-dimension, while columns represent time-domain.

$$\mathbf{S}_d(-1)^T\mathbf{S}_d(1) = \mathbf{0}_{2\times 2}, \tag{7.2}$$

$$\mathbf{S}_d(-1)^T\mathbf{S}_d(-1) = \mathbf{S}_d(1)^T\mathbf{S}_d(1) = 2\mathbf{I}_{2\times 2} \tag{7.3}$$

From (7.2) and (7.3), we can see the codeword in (7.1) constitutes a simple example of Space Time Orthogonal Block Code (STOBC).

It is assumed that perfect channel state information is available. Therefore, it can be shown that the Maximum Likelihood (ML) detection of the transmitted codeword can be expressed by

$$\hat{\mathbf{S}}_d \doteq \arg\max_{l\in\{-1,1\}} \mathbf{P}(\mathbf{Y}|\mathbf{S}_d(l), \mathbf{H}) \tag{7.4}$$

Then, transmitted space-time codeword $\mathbf{S}_d$ can be evaluated as

$$\hat{\mathbf{S}}_d \doteq \underset{\mathbf{S}_d}{\operatorname{argmin}} \left\{ \sum_{m=1}^{2} [\mathbf{H}_m^H \cdot \mathbf{S}_d^H \cdot \mathbf{S}_d \cdot \mathbf{H}_m - 2\Re(\mathbf{H}_m^H \cdot \mathbf{S}_d^H \cdot \mathbf{Y}_m)] \right\} \tag{7.5}$$

where H_m and Y_m are the mth column of H and Y, respectively.

7.2 Compressive Sensing-Based MIMO UWB System

Compressive sensing is usually applied to discrete signal. It can also be extended to continuous time signals. Assume there is an analog signal $x(t)$, $t \in [0, T_x]$ which is k-sparse over some basis Ψ

$$x(t) = \sum_{n=0}^{N-1} \psi_n(t)\theta_n = \Psi(t)\theta, \tag{7.6}$$

where

$$\Psi(t) = [\psi_0(t), \psi_1(t), \cdots, \psi_{N-1}(t)], \tag{7.7}$$

$$\theta = [\theta_0, \theta_1, \cdots, \theta_{N-1}]^T, \tag{7.8}$$

Note that there are only K non-zeroes in θ, ψ_i is the waveform transmitted for information. After Space-time code modulation, $x(t)$ is fed into a flat-fading channel, the received baseband analog signal can be modeled as $y_i(t)$. Note that the output discrete signal y_i has a period of $4T_c$ and N samples. Instead of sampling it directly, we multiply a $M \times N$ i.i.d Gaussian matrix with y_i, where $\frac{N}{M} = q$ and q is a positive integer. Therefore, the period of a down-sampled signal Y is $4qT_c$ and the total number of samples is M.

$$Y = [y_0, y_1, \cdots, y_{M-1}]^T, \tag{7.9}$$

Combining all the equation above, we have

$$Y = \Phi\Psi\theta = \Theta\theta, \tag{7.10}$$

Now the problem becomes recovering the $N \times 1$ vector θ from the $M \times 1$ measurement vector Y, which is exactly the same as the problem stated in (1.4). The number of measurements for successful recovery depends on the sparsity K, duration of the analog signal, T_x, and the incoherence between Φ and Ψ. Numerical results in the next section shows that when $x(t)$ is sparse, θ can be reconstructed successfully with a reduced sampling rate, requiring only $M \ll N$ measurements.

CS-based MIMO UWB communication system is able to reduce the sampling rate to 12.5% of the Nyquist rate. The system architecture is illustrated in Figure 7.1. Two UWB signals are transmitted by feeding a sparse bit sequence through an UWB pulse generator and space-time code modulation block. Afterwards, we multiply a $M \times N$ Gaussian random matrix with the received signal and downsample

it using a low-rate A/D converter and then processed by a recovery algorithm at the receiver side. Our simulation result shows 3 to 8 GHz UWB signals can be successfully recovered by a 500 Msps A/D converter. So far, the discussion is in the baseband. If the transmitted UWB signal is a passband signal, then up-conversion is applied after the space-time code modulation block. The receiving structure remains the same. No down-conversion is required at the receiver.

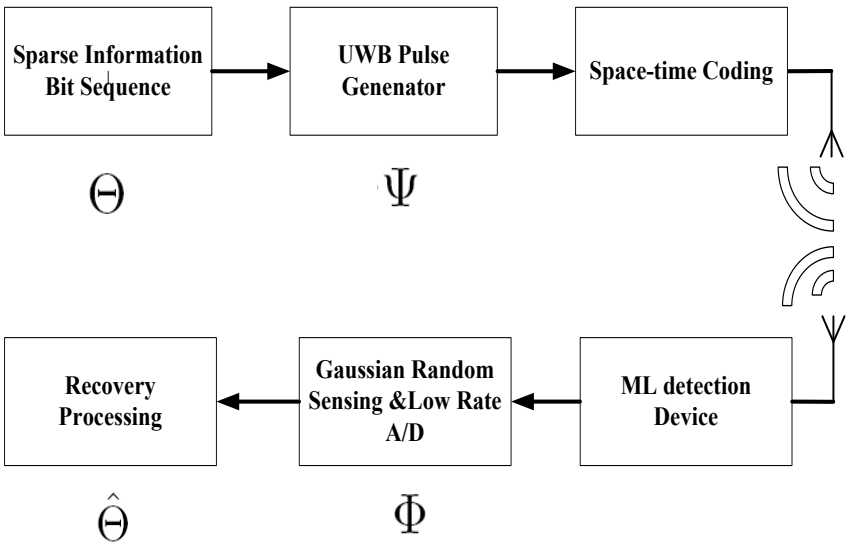


Figure 7.1. System architecture of the CS-based MIMO UWB communication system.

7.3 Numerical Results

In our simulation, the UWB system operates in 3 to 8 GHz frequency band. We also assume that the channel is time-invariant during the communication process.

For illustration purposes, we assume there is only a "1" in a 256 consecutive information sequence. More information bits can be transmitted by adding more "1"s in the information sequence. The position with maximum amplitude in $\hat{\theta}$ is

compared with the position in θ. If two positions are exactly the same, the symbol is reconstructed successfully.

Figure 7.2 shows the reconstruction result under 500 Msps sampling rate with additive white Gaussian noise ($\gamma_d = 10$). $\hat{\theta}$ reconstructed from perfect channel profile is compared with the original θ. From Figure 7.2, we can see that $\hat{\theta}$ is a little bit noisy due to the existence of Gaussian noise. However, the position of maximum amplitude is exactly the same as the one in θ. Therefore, the transmitted symbol is successfully recovered . With 10000 Monte-Carlo simulations, no error can be found due to the high SNR.

Figure 7.3 shows the BER vs SNR for 250/500/625 Msps sampling rates. Perfect channel information is available at the receiver side. We see that a high sampling rate provides a better BER performance. This is because more measurements are collected under higher sampling rates.

The performance of the receiver with different antenna configurations is shown in Figure 7.4. Figure 7.3 compared with Figure 7.4 shows the performance of the receiver with MIMO sampling at 250 Msps is much better than that of receiver with only one receiving antenna. Even at a SNR as low as 0 dB, the performance gap exceeds 1.5 orders of magnitude. More importantly, due to the higher diversity gain of the MIMO receiver, the gap increases for higher SNR values. In other words, a MIMO system can achieve same BER at a much lower sampling rate compared with that of a Single Input Single Output (SISO) system.

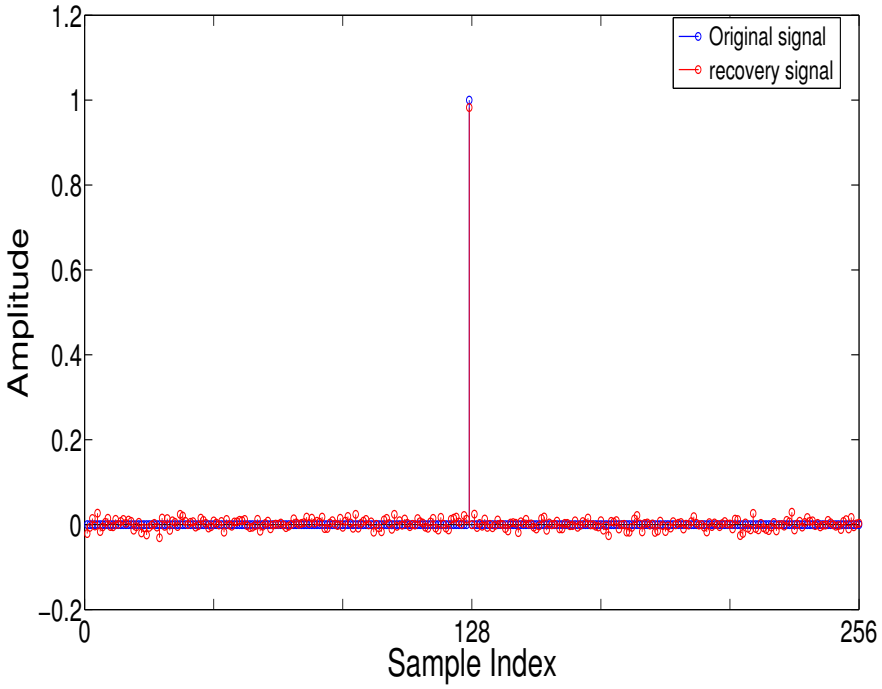


Figure 7.2. Recovery of $\hat{\theta}$ using 500 Msps A/D converter.

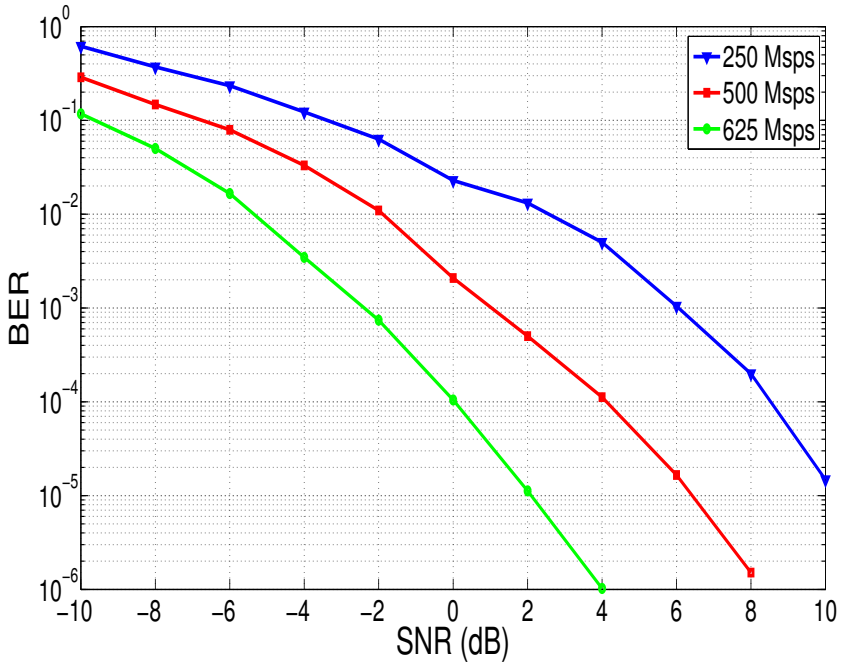


Figure 7.3. BER vs SNR at the receiver for three different sampling rates.

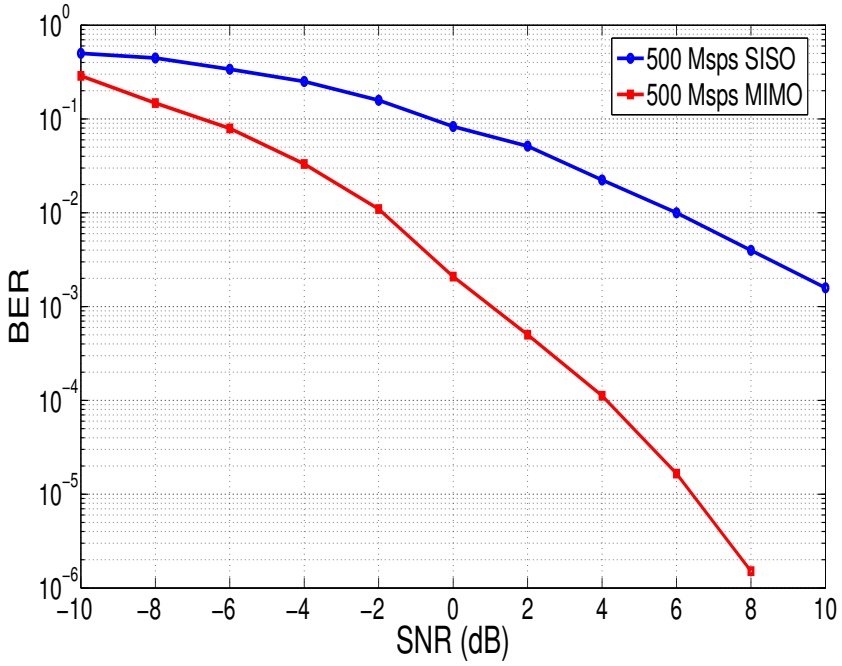


Figure 7.4. BER vs SNR at the receiver for MIMO and SISO system at a sampling rate of 500 Msps.

CHAPTER 8

CONCLUSIONS AND FUTURE RESEARCH

This dissertation set out to investigate compressive sensing for various wireless and sensor systems, and evaluate its performance. In this Chapter, we review the research contributions of this dissertation, as well as discuss directions for future research.

8.1 Contributions

The following are the main research contributions of this dissertation.

- Huffman coding and compressive sensing are employed to compress real-world data. Both uniform and non-uniform Huffman coding are evaluated from various perspectives (Chapter 2). The performance of Huffman coding is also compared with that of compressive sensing. The main drawback of Huffman coding is that it requires calculating the probability of each symbol before encoding. It means that it may not be appropriate for real-time compression. We applied CS to wind tunnel data compression and compared it against theoretical error bound. Although the theoretical error bound does not provide the optimum solution to the number of measurements, it still presents a meaningful guideline when we choose the measurement matrix. This work also provides a guide towards future research (as we will discuss in Section 8.2).
- A compressive sensing-based compression and recovery UWB communication system is proposed (Chapter 3). Compared to the conventional UWB system, we can jointly estimate the channel and compress the data which also reduces

the hardware complexity. No information about the transmitted signal is required in advance as long as the channel follows autoregressive model.

- Both information-theoretical and computational secrecy of compressive sensing are investigated (Chapter 4). It has been proven that compressive sensing-based encryption is perfectly secret if the length of message goes to infinity. If an eavesdropper has the wrong measurement matrix, he/she can never recover the original data. The amount of computation required to find the correct measurement matrix is proportional to the number of candidates.
- Due to limited energy and physical size of the sensor nodes, the conventional security mechanisms are not feasible for wireless sensor networks (WSNs) (Chapter 5). A compressive sensing-based encryption is proposed for WSNs, which provide both signal compression and encryption guarantees, without the additional computational cost of a separate encryption protocol. For our proposed distributed WSNs, if only a fraction of randomizer bits is stored by an eavesdropper, then the eavesdropper cannot obtain any information about the plaintext.
- A CS-based MIMO UWB communication system, in which the transmitted symbols are sparse in the time domain, is proposed (Chapter 7). An MIMO system can achieve same BER at a much lower sampling rate compared with that of a Single Input Single Output (SISO) system due to the higher diversity gain of the MIMO receiver.

8.2 Future Research

The research presented in this dissertation raises many new questions. There are several lines of research arising from this work which should be pursued.

- The theoretic error bound introduced in Section 2 provides a natural guide to future research. One avenue for future research is to investigate how the error introduced in the parameter estimation process affects the accuracy of the derived bound. In Chapter 2, we only provide the error bound when data is normal distributed, the error bound for data follows other distribution should also be derived. Furthermore, a binary ensemble is assumed for the measurement matrix to simplify the expressions, we should show that the result still holds if Gaussian ensemble is used to generate the measurement matrix.
- In proposed compressive sensing-based compression and recovery UWB communication system, the order of autoregressive model is assumed to be known in advance. However, this may not always be the case. In future research, we should consider the signal recovery when the order of autoregressive model is not known. Besides, we should investigate the relation between the sparsity of the original signal and the probability of perfect recovery using certain measurement matrix.
- Much research also remains to be done for CS-based MIMO UWB communication system. The channel state information is assumed to be known at the receiver in our investigation. We need to study the relation between the probability of perfect recovery of information sequence and the error introduced in channel estimation process. In Chapter 7, we only consider the binary information bit sequence. One possible direction for future research should address is the performance of information sequence with high order modulation, and the channel estimation is not perfect at the receiver.

REFERENCES

[1] J. Trzasko and A. Manduca, "Highly Undersampled Magnetic Resonance Image Reconstruction via Homotopic ℓ_0-Minimization," *IEEE Trans. on Med. Imaging*, vol. 28, no. 1, pp. 106 -121, January 2009.

[2] J. Provost and F. Lesage, "The Application of Compressed Sensing for Photo-Acoustic Tomography," *IEEE Trans. on Med. Imaging*, vol. 28, no. 4, pp. 585-594, April 2009.

[3] S. G. Lingala, H. Yue, E. DiBella and M. Jacob, "Accelerated Dynamic MRI Exploiting Sparsity and Low-Rank Structure: k-t SLR," *IEEE Trans. on Med. Imaging*, vol. 30, no. 5, pp. 1042-1054, May 2011.

[4] K. Gedalyahu and Y. C. Eldar, "Time-Delay Estimation From Low-Rate Samples: A Union of Subspaces Approach," *IEEE Trans. on Signal Processing*, vol. 58, no. 6, pp. 3017-3031, June 2010.

[5] E. Matusiak and Y. C. Eldar, "Sub-Nyquist Sampling of Short Pulses," *IEEE Trans. on Signal Processing*, vol. 60, no. 3, pp. 1134-1148, March 2012.

[6] M. Mishali and Y. C. Eldar, "From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals," *IEEE Joural of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375-391, April 2010.

[7] M. A. Herman and T. Strohmer, "High-Resolution Radar via Compressed Sensing," *IEEE Trans. on Signal Processing*, vol. 57, no. 6, pp. 2275-2284, June 2009.

[8] A. Budillon, A. Evangelista and G. Schirinzi, "Three-Dimensional SAR Focusing From Multipass Signals Using Compressive Sampling," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 488-499, January 2011.

[9] X. Zhu and R. Bamler, "Tomographic SAR Inversion by ℓ_1-Norm Regularization- the Compressive Sensing Approach," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3839-3846, October 2010.

[10] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. on Communications*, vol. 50, no. 3, pp. 374-377, March 2002.

[11] G. Taubock, F. Hlawatsch, D. Eiwen and H. Rauhut, "Compressive Estimation of Doubly Selective Channels in Multicarrier Systems: Leakage Effects and Sparsity-Enhancing Processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 255-271, April 2010.

[12] C. R. Berger, Z. Wang, J. Huang and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 164-174, November 2010.

[13] J. Meng, W. Yin, Y. Li, N. T. Nguyen and Z. Han, "Compressive Sensing Based High-Resolution Channel Estimation for OFDM System," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 1, pp. 15-25, Febuary 2012.

[14] E. Candés, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 485-509, February 2006.

[15] E. Candés and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?," *IEEE Trans. on Information Theory*, vol. 52, no. 12, pp. 5406-5424, December 2006.

[16] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. on Information Theory*, vol. 53, no. 12, pp. 4655-4666, 2007.

[17] D. Donoho, "Compressed Sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289-1306, April 2006.

[18] R. Baranuik, "Compressive Sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118-121, July 2007.

[19] E. Candés and M. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, pp. 21-30, March 2008.

[20] R. Baranuik, M. Davenport, R. Devore and M. Wakin, "A Simple Proof of The Restricted Isometry Property For Random Matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253-263, December 2008.

[21] E. Candés and J. Romberg, "Sparse And Incoherence in Compressive Sampling," *Inverse problem*, vol. 23, no. 3, pp. 969-985, 2007.

[22] D. Donoho and X. Huo, "Uncertainty Principles And Ideal Atomic Decomposition," *IEEE Trans. on Information Theory*, vol. 47, no. 7, pp. 2845-2862, November 2001.

[23] J. Liang, Q. Liang, "Outdoor Propagation Channel Modeling in Foliage Environment," *IEEE Trans on Vehicular Technology*, vol. 59, no. 3, June 2010.

[24] D. Kirachaiwanich and Q. Liang, "Compressive sensing: To compress or not to compress," *Signals, Systems and Computers (ASILOMAR)*, pp. 809-813, November 2011.

[25] Y. Weiss, "Learning compressed sensing," 2007.

[26] Y. Rachlin and D. Baron, "The secrecy of compressed sensing measurements," *46th Annual Allerton Conference on Communication, Control, and Computing* , pp. 813-817, September 2008.

[27] D. W. Carman, P. S. Krus and B. J. Matt, "Constraints and approaches for distributed sensor network security," *Technical Report 00-010, NAI Labs, Network Associates Inc.*, Glenwood, MD, 2000.

[28] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. E. Culler and K. Pister, "System architecture directions for networked sensors," *In Proceedings of the 9th International Conference on Architectural Support for Programming Languages and Operating Systems*, pp. 93-104, New York, ACM Press, 2000.

[29] A. Perrig, R. Szewczyk, V. Wen, D. E. Culler and J. D. Tygar, "SPINS: Security protocols for sensor networks, Wireless Networks," *In Proceedings of the 9th International Conference on Architectural Support for Programming Languages and Operating Systems*, vol. 8 , no. 5, pp. 521-534, September 2002.

[30] J. A. Stankovic et al, "Real-time communication and coordination in embedded sensor networks," *In Proceedings of the IEEE*, vol. 91, no. 7, pp. 1002-1022, July 2003.

[31] S. Zhu, S. Setia, and S. Jajodia, "LEAP: Efficient security mechanism for large scale distributed sensor networks," *In Proceedings of the 10th ACM Conference on Computer and Communications Security*, vol. 91, pp. 62-72, New York, NY, USA, 2003, ACM Press.

[32] B. Lai, S. Kim, and I. Verbauwhede, "Scalable session key construction protocols for wireless sensor networks," *In IEEE Workshop on Large Scale Real Time and Embedded Systems*, 2002.

[33] H. Chan and A. Perrig, "PIKE: Peer intermediaries for key establishment in sensor networks," *In IEEE INFOCOM* , 2005.

[34] L. Eschenauer and V. D. Gligor, "A Key-Management Scheme for Distributed Sensor Networks," *In Proceedings of the 9th ACM Conference on Computer and Networking*, pp. 41- 47, November 2002.

[35] Data Encryption Standard. Federal Information Processing Standards Publication 46, U. S. Department of Commerce/National Bureau of Standards, Springfield, Virginia, 1977.

[36] RSA Laboratories, http://www.rsasecurity.com/rsalabs.

[37] C. E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Techinical Journal*, vol. 28(4), pp. 656-715, October 1949.

[38] D. Achilioptas, "Datebase-friendly Random Projections: Johnson-Lindenstrauss with Binary Coins," *Journal of Computer and System Sciences*, vol. 66(4), pp. 671-687, 2003.

[39] U. M. Maurer, "Conditionally-Perfect Secrecy and a Provably-Secure Randomized Cipher," *Journal of Cryptology*, vol. 5(1), pp. 53-66, 1992.

[40] J. Spencer, "The Strange Logic of Random Graphs," *Algorithms and Combinatorics 22*, Springer-Verlag 2000, ISBN 3-540-41654-4.

[41] A. Griffin and T. Panagiotis, "Compressed Sensing of Audio Signals Using Multiple Sensors," *Reconstruction*, 3.4 (2007): 5.

[42] P. Li, T. J. Hastie and K. W. Church, "Very sparse Random Projections," *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, 2006.

[43] G. H. Golub and C. F. Van Loan, "Matrix computations," *Johns Hopkins Univ. Press, MD*, 1983.

[44] E. J. Candés, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Stable Signal Recovery from Incomplete and Inaccurate Measurements*, vol. 59, no .8, pp.1207-1223, 2006.

[45] Devore, *Probability and Statistics for Engineering and the Sciences*, 2nd, London: Artech house, 2005.

[46] E. Limpert, W. Stahel, and M. Abb, "Log-normal Distributions across the Sciences: Keys and Clues," *BioScience*, vol. 51, no. 5, pp. 341-352, May 2001.

[47] W. Weibull, "A statistical distribution function of wide applicability," *Trans. ASME, J. Appl. Mech."*, vol. 18, no. 3, pp. 293-297, 1951.

[48] M. Barkat, *Signal Detection and Estimation*, 2nd ed. London, U.K.:Artech House, 2005.

[49] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd Edition,Wiley Interscience, 2006.

[50] R. M. Narayanan, Y. Xu, P. D. Hoffmeyer and J. O. Curtis "Design, Performance, and Applications of A Coherent Ultra WideBand Random Noise Radar," *Opt.Eng* , 37(6), (1998) 1855-1869.

[51] J. M. Mendel, "Optimal Seismic Deconvolution: An Estimation Based Approach," *Academic Press, New York*, 1983.

[52] V. Saligrama, "Deterministic Designs with Deterministic Guarantees: Toeplitz Compressed Sensing Matrices, Sequence, Designs and System Identification," *preprint*, 2008.

[53] G. Eshel "The Yule Walker Equations for the AR Coefficients," *University of South Carolina*, $http://www.stat.sc.edu/\ vesselin/STAT520\_YW.pdf$.

[54] D. Bertsimas, J. N. Tsitsiklis "Introduction to Linear Optimization," *Athena Scientific* , February, 1997.

[55] E. J. Candès, "The Restricted Isometry Property and Its Implications for Compressed Sensing," *Compte Rendus de l'Academie des Sciences*, Paris, Serie I, pp. 346 589-592.

[56] R. M. Narayanan, "Through-Wall Radar Imaging Using UWB Noise Waveforms," *Journal of the Franklin Institute*, 345.6 (2008): 659-678.

[57] E. Baccarelli, M. Biagi, "Error Resistant Space-Time Coding for Emerging 4G-WLANs," *WCNC*, 2003.

[58] E. Baccarelli, M. Biagi, "Performance and Optimized Design of Space-Time Codes for MIMO Wireless Systems with Imperfect Channel-Estimates," *IEEE Trans. Signal Processing*, vol. 52, no. 10, pp. 2911-2923, 2004.

[59] E. Baccarelli, M. Biagi, C. Pelizzoni and P. Bellotti, "A Novel Multi-Antenna Impulse Radio UWB Transceiver for Broadband High-Throughput 4G WLANs," *IEEE Communication Letters*, vol. 8, no. 7, July 2004.

BIOGRAPHICAL STATEMENT

Ji Wu received his B.S. degree from Beijing University of Posts and Telecommunications, in 2005 in Electrical Engineering. He joined UTA in spring, 2009. Currently, he is a Ph.D student in the Wireless Communication Network Lab of Electrical Engineering Department at the University of Texas at Arlington. His current research interests are compressive sensing, and its application in wireless communication systems, MIMO and information theory.