Sparse Event Detection in Wireless Sensor Networks using Compressive Sensing

Jia Meng[1], Husheng Li[2], and Zhu Han[1]
[1] Department of Electrical and Computer Engineering Department, University of Houston, Houston, TX
[2] Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN

***Abstract*— Compressive sensing is a revolutionary idea proposed recently to achieve much lower sampling rate for sparse signals. For large wireless sensor networks, the events are relatively sparse compared with the number of sources. Because of deployment cost, the number of sensors is limited, and due to energy constraint, not all the sensors are turned on all the time. In this paper, the first contribution is to formulate the problem for sparse event detection in wireless sensor networks as a compressive sensing problem. The number of (wake-up) sensors can be greatly reduced to the similar level of the number of sparse events, which is much smaller than the total number of sources. Second, we suppose the event has the binary nature, and employ the Bayesian detection using this prior information. Finally, we analyze the performance of the compressive sensing algorithms under the Gaussian noise. From the simulation results, we show that the sampling rate can reduce to 25% without sacrificing performance. With further decreasing the sampling rate, the performance is gradually reduced until 10% of sampling rate. Our proposed detection algorithm has much better performance than the l_1-magic algorithm proposed in the literature.**

I. Introduction

The dogma of signal processing maintains that a signal must be sampled at a Nyguist rate at least twice its bandwidth in order to be represented without error. However, in practice, we often compress the data soon after sensing, trading off signal representation complexity (bits) for some error (consider JPEG image compression in digital cameras, for example). Clearly, this is wasteful of valuable sensing/sampling resources. Over the past few years, a new theory of "compressive sensing" [1–3] has begun to emerge, in which the signal is sampled (and simultaneously compressed) at a greatly reduced rate. Very recently, there are emerging applications [4–7] for wireless communication and networking.

In this paper, we investigate how to employ compressive sensing in wireless sensor networks [8], [9], which mostly involve a large number of sensor nodes. Specifically, we target on two problems of wireless sensor networks. First, there are a very limited number of active sensors compare with the total number of sensors in the network. Moreover, the number of events is much less compared to the number of all sources. Second, different events may happen simultaneously and cause interference to detect them individually. As a result, the received signals are superimposed all together, and an efficient algorithm is needed to separate the signals.

To overcome the above two problems, in this paper we propose a sparse event detection scheme in wireless sensor networks by employing compressive sensing. Our contributions are listed as follows:

1) Most compressive sensing work formulate the problem in image processing, especially bio image processing. Little work has studied the wireless networking problem. We formulate the compressive sensing problem using sparse nature of wireless sensor networks.
2) To improve the performance, we employ the Bayesian detection and a heuristic, using the prior information that the events are binary. So the estimation probability can be substantially increased, compared with the l_1-magic algorithm in the literature [15].
3) Most compressive sensing schemes suffer susceptibility under Gaussian noise environment, since Gaussian noise can be unbounded. We conduct the simulations to investigate the effects of the noise. We show that the performance decays with the signal to noise ratio (SNR) approaching 20dB.

From the simulation results, we show that the sampling rate can reduce to 25% without sacrificing performance. With further decreasing the sampling rate, the performance is gradually reduced until 10% of sampling rate. As a result, the cost and energy efficiency of wireless sensor networks can be greatly improved.

The rest of this paper is organized as follows: Section II presents the system model. Section III formulates the problem, conducts analysis and exposes the proposed algorithm using compressive sensing. Simulation results are presented and analyzed in Section IV. Finally, conclusions are drawn in Section V.

II. System Model

We consider the system model as shown in Figure 1. There are a total of N sources randomly located in a field. Those source randomly generate the events to be measured. We denote K as the number of events that the sources generate. K is a random number, and is much smaller than N. We denote $\mathbf{X}_{N\times 1}$ as the event vector, in which each component has a binary value, i.e., $X_n \in \{0, 1\}$. Obviously $\mathbf{X}$ is a sparse vector since $K << N$. In the system, there are M active monitoring sensors trying to capture these events. There are two challenges for those monitoring sensors. First, all those events happen simultaneously. As a result, the received signals are interfering with each other. Second, the received signal

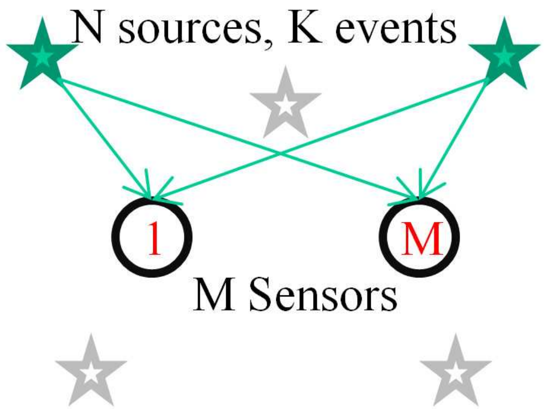

Fig. 1. System Model for Sensor Network

is deteriorated by propagation loss and thermal noise. The received signal vector can be written as

$$\mathbf{Y}_{M\times 1} = \mathbf{G}_{M\times N}\mathbf{X}_{N\times 1} + \epsilon_{M\times 1}, \tag{1}$$

where $\epsilon_{M\times 1}$ is the thermal noise vector whose component is independent and has zero mean and variance of σ^2. $\mathbf{G}_{M\times N}$ is the channel response matrix whose component can be written as

$$G_{m,n} = (d_{m,n})^{-\alpha/2}|h_{m,n}|, \tag{2}$$

where $d_{m,n}$ is the distance from the n^{th} source to the m^{th} sensing device, α is the propagation loss factor, and $h_{m,n}$ is the Raleigh fading modeled as complex Gaussian Noise with zero mean and unit variance.

Notice that the number of events, the number of sensors, and total number of sources have the following relation $K < M << N$. Consequently, the received signal vector $\mathbf{Y}$ is an condensed representation of the event. In other words, vector $\mathbf{Y}$ has aliasing of vector $\mathbf{X}$, due to the low sampling rate M. From the algorithm proposed in the next section, we can estimate $\mathbf{X}$ from $\mathbf{Y}$.

III. Compressive Sensing Algorithm

In this section, we first formulate the compressive sensing problem. Next, we investigate how to use the prior information to improve the recovery performance.

A. Problem Formulation and Analysis

The problem is to obtain the K information of $\mathbf{X}$ using the limited number of sensors (M). The first question is that whether or not the information of K-sparse signal is damaged by the dimensionally reduction from $\mathbf{X} \in \mathbb{R}^N$ down to $\mathbf{Y} \in \mathbb{R}^M$. In general, if $\mathbf{X}$ is not sparse enough, as long as $M < N$, the signal is damaged since there are fewer equations than unknowns. On the other hand, for the K-sparse signal, $\mathbf{Y}$ is just a linear combination of K columns of $\mathbf{G}$. A necessary and sufficient condition to ensure that this $M \times K$ system can be compressed and reconstructed is listed as the following property:

Definition 1: Restricted Isometry Property (RIP) [1–3]: For any vector $\mathbf{V}$ sharing the same K nonzero entries as $\mathbf{X}$, if

$$1-\epsilon \le \frac{\|\mathbf{GV}\|^2}{\|\mathbf{V}\|^2} \le 1+\epsilon, \tag{3}$$

for some $\epsilon > 0$, then the matrix $\mathbf{G}$ preserves the information of the K-sparse signal. A sufficient condition for stable inverse in practice is that $\mathbf{G}$ satisfies (1) for an arbitrary $3K$-sparse vector $\mathbf{V}$.

In [1–3], it has been proved that if $\mathbf{G}$ is an iid Gaussian matrix or random ± 1 entry matrix, then the K-sparse signal is compressible with high probability if $M \le cK\log(N/K) << N$, where c is a constant. For our specific problem, we will show in the simulation that the signal can be reconstructed, which is due to randomness introduced by random locations and Rayleigh fading.

From above, we know that under a certain condition the signal is still preserved in M dimensions. The next question is how to develop a reconstruction algorithm to recover $\mathbf{X}$ from the measurement $\mathbf{Y}$. Since $M < N$ there are infinite number of $\hat{\mathbf{X}}$ satisfy $\mathbf{Y} = \mathbf{G}\hat{\mathbf{X}}$. All solutions lie on the $N-M$ dimension hyperplane $\mathcal{H} := \mathcal{N}(\mathbf{G}) + \mathbf{X}$ which corresponding to the null space $\mathcal{N}(\mathbf{G})$ translated to $\mathbf{X}$. So the problem is to find the sparse reconstructed signal $\hat{\mathbf{X}}$ in the translated null space as:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{Y}=\mathbf{G}\hat{\mathbf{X}}} |\hat{\mathbf{X}}|_{\mathbf{1}}, \tag{4}$$

where $|\cdot|_1$ is norm one. Notice that it has been shown in the literature [15] that for norm two, there might be many solutions, and for norm zero, the complexity is NP hard [1–3].

The problem in (4) is a convex optimization problem, which can reduce to a linear program. Define $\hat{\mathbf{X}} = \mathbf{U} + \mathbf{W}$. We can change (4) to

$$\min(\mathbf{U}+\mathbf{W}) \tag{5}$$
$$\text{s.t.} \begin{cases} (\mathbf{GU} - \mathbf{GW}) = \mathbf{Y}, \\ \mathbf{U}, \mathbf{W} > 0. \end{cases}$$

The above optimization is called the l_1-magic in the literature. The complexity is $O(N^3)$. But much simple algorithms such as simplex algorithm [10], [11] can be easily employed.

B. Bayesian Detection

For the problem formulation in (4), we do not consider the fact that the components of $\mathbf{X}$ are either 0 or 1. To utilize this prior information, in this subsection instead of using the traditional signal recovery algorithm like the l_1-magic or linear programming, we adopt the Bayesian compressive sensing [12–14], which is fully probabilistic and introducing a set of hyper-parameters which is viewed as a prior over the signal, and the most probable values are iteratively estimated from the received data. The main reason why this algorithm fits our needs in the sensor network compressive sensing is that, the posterior distributions of many of the signals are sharply peaked around zero, which matches exactly our sparse binary signal system model. By exploiting such a probabilistic Bayesian framework, we can achieve accurate reconstruction with dramatically fewer samples than using other recovering algorithms, which will be shown in simulations in Section IV.

In the following, we first propose the model, then the iterative marginal likelihood maximization, and finally we propose a heuristic algorithm based on some observations.

1) Model Specification: In (1), the noise in the system is composed of propagation loss with zero mean and variance σ^2. The probability density function can be approximated as Gaussian distribution as:

$$p(\epsilon) = \prod_{i=1}^{M} \mathcal{N}(\epsilon_i|0, \sigma^2). \quad (6)$$

Due to the assumption of independence of Y_n, the likelihood of the complete data set can be written as:

$$p(\mathbf{Y}|\mathbf{G}, \sigma^2) = (2\pi\sigma^2)^{-M/2} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{Y} - \mathbf{G}\mathbf{X}\|^2\right). \quad (7)$$

By adopting a Bayesian perspective, we constrain the parameters by defining an explicit prior probability distribution over them.

The real distribution of **X** is Bernoulli distribution. However, the close form solution in our problem is hard to be obtained. Instead, we assume a zero-mean Gaussian prior distribution over the signal **X**:

$$\begin{aligned} p(\mathbf{X}|\alpha) &= \prod_{n=1}^{N} \mathcal{N}(X_i|0, \alpha_i^{-1}) \quad (8) \\ &= (2\pi)^{-N/2} \prod_{n=1}^{N} \alpha_n^{1/2} \exp\left(-\frac{\alpha_n x_n^2}{2}\right), \end{aligned}$$

where α is a vector of N independent hyper-parameters. We will show in the simulations that this assumption can improve the performance. For the Bernoulli distribution case, we will investigate at the end of this section to further utilize the prior information.

Given α, the posterior parameter distribution conditioned over the signal is given by combining the likelihood and prior with Bayes' rule:

$$p(\mathbf{X}|\mathbf{Y}, \alpha, \sigma^2) = \frac{p(\mathbf{Y}|\mathbf{X}, \sigma^2)p(\mathbf{X}|\alpha)}{p(\mathbf{Y}|\alpha, \sigma^2)}, \quad (9)$$

which is a Gaussian distribution $\mathcal{N}(\mu, \mathbf{\Sigma})$ with covariance and mean of

$$\mathbf{\Sigma} = (\mathbf{A} + \sigma^{-2}\mathbf{G}^T\mathbf{G})^{-1} \quad (10)$$

and

$$\mu = \sigma^{-2}\mathbf{\Sigma}\mathbf{G}^T\mathbf{Y}, \quad (11)$$

respectively, where $\mathbf{A} = diag(\alpha_1, \ldots, \alpha_N)$.

2) Marginal Likelihood Maximization: A most-probable point estimate α_{MP} may be found via a type-II maximum likelihood procedure [18]. The sparse Bayesian model is formulated as the local maximization with respect to α of the marginal likelihood, or equivalently its logarithm:

$$\begin{aligned} L(\alpha) &= \log p(\mathbf{Y}|\alpha, \sigma^2) \quad (12) \\ &= \log \int_{-\infty}^{\infty} p(\mathbf{Y}|\mathbf{X}, \sigma^2)p(\mathbf{X}|\alpha)dX \\ &= -\frac{1}{2}\left(M \log 2\pi + \log|\mathbf{C}| + \mathbf{Y}^T\mathbf{C}^{-1}\mathbf{Y}\right) \end{aligned}$$

with

$$\mathbf{C} = \sigma^2 + \mathbf{I} + \mathbf{G}\mathbf{A}^{-1}\mathbf{G}^T. \quad (13)$$

A point estimate μ_{MP} for the parameters is then obtained by evaluating (11) with $\alpha = \alpha_{MP}$, giving a posterior mean approximator $\mathbf{G}\mathbf{X} = \mathbf{G}\mu_{MP}$.

However, marginal likelihoods are generally difficult to compute, i.e., values of α and σ^2 which maximize $L(\alpha)$ cannot be obtained in closed form. Thus, we need to re-estimate them iteratively.

For the updating of α, following the approach in [18], we differentiate (12), and then equate it to 0. After rearranging, we have

$$\alpha_i^{new} = \frac{\gamma^i}{\mu_i^2}, \quad (14)$$

where μ_i is the i^{th} posterior mean signal from (11), and γ_i is defined as

$$\gamma_i = 1 - \alpha_i N_{ii} \quad (15)$$

with N_{ii} being the i^{th} diagonal element of the posterior signal covariance from (10) computed with current α and σ^2 values. Each γ_i can be treated as a measure of how well-determined its corresponding parameter X_i is by the data. For the variance σ^2, differentiation leads to re-estimate:

$$\sigma_{new}^2 = \frac{\|\mathbf{Y} - \mathbf{G}\mu\|^2}{M - \sum_i \gamma_i}. \quad (16)$$

We repeat calculation of α and σ^2 with iteratively updating μ and $\mathbf{\Sigma}$ until certain convergence criteria have been reached. This procedure leads to the maximization the marginal likelihoods.

Then at the convergence of α estimation procedure, we make predictions based on the posterior distribution over the signal, conditioned on the maximizing values α_{MP} and σ_{MP}^2. In other words, by doing this, we could pick up those entries in the projection matrix **G** which after projection preserves the information of the signal in **Y**. With the utilizing of the correspondent elements in the measurements **Y** and projection matrix **G**, we could reconstruct our signal with an overwhelming probability. Experimental data in the following Section show that, compared with l_1-minimization, sampling rate could be reduced dramatically.

3) Heuristic using Prior Information: After the reconstruction of $\hat{\mathbf{X}}$, if the algorithm converges to wrong results (which will be shown in the following section), there are two possible situations. First, the algorithm can converge to either around 0 and 1, but with the wrong position for the sparse events. This type of errors could not be easily distinguished. The other type of error would let $\hat{\mathbf{X}}$ have values deviating from 0 or 1. Under this condition, it is very easy to find the error using threshold methods. Then the system can wait until the next time slot (which has different fading parameters) to make a decision, hoping the channel matrix **G** could be changed so that the reconstructed $\bar{\mathbf{X}}$ can be improved. Here we propose a heuristic algorithm to achieve the above ideas as shown in Table I. Here δ is a small positive constant less than one. If

TABLE I

HEURISTIC ALGORITHM USING PRIOR INFORMATION

If $1-\delta \leq \max(\bar{\mathbf{X}}) \leq \mathbf{1}+\delta$
report the decision of $\bar{\mathbf{X}}$;
Exit.
Else
Wait until next time slot to obtain $\mathbf{Y}$.
Go to First Step. Exit after a certain number of unsuccessful trials.

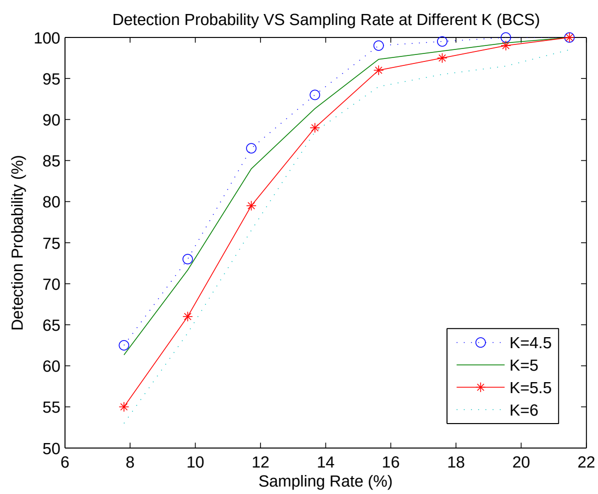

Fig. 2. Proposed Scheme Performance

the maximal value of $\bar{\mathbf{X}}$ is within $[1-\delta, 1+\delta]$, we assume the reconstructed signal is correct. Otherwise, we obtain the new $\mathbf{Y}$. Notice that the cost for the heuristic algorithm is the possible delay for the responses, since the new $\mathbf{Y}$ needs to be obtained after the channel states have been changed.

IV. SIMULATION RESULTS AND ANALYSIS

In this section, we set up some preliminary simulation results. There are a total of $N = 256$ events randomly located within 500m-by-500m area. The M wireless sensors are also randomly located within this area. The minimal distance between a event and a sensor is 5m. The propagation loss factor is 3. The transmitted power is normalized to 1 and the thermal noise is 10^{-12}. The number of random events is K which is a small number.

Figures 2 shows the detection probability of the proposed algorithm. We define the sampling rate as M/N. We can see that if the sampling rate is higher than 25%, the detection probability is almost 100%. The performance gradually reduces as the sampling rate reduces and as K (the number of event) increases. For comparison, Figure 3 shows the traditional l_1 algorithm. Compared with the conventional linear programming recovery algorithm for compressive sensing, the proposed Bayesian framework with fast marginal likelihood maximization algorithm brings much higher detection probability especially at extremely low sampling rate. For example, with sampling rate as 10%, the proposed scheme is more than 2 times better than the l_1 algorithm. Notice that this performance gain is based on the assumption that the source is

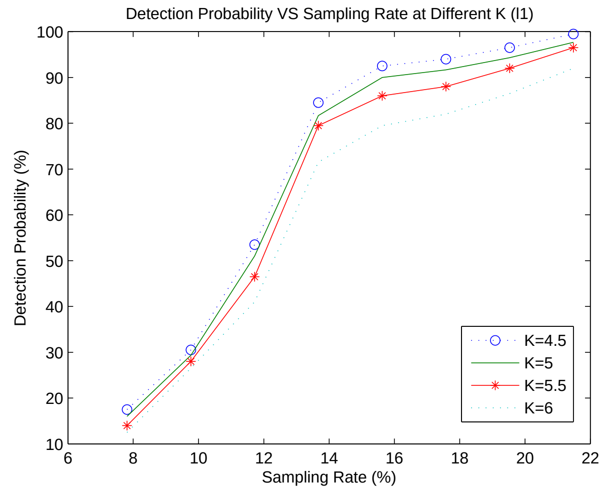

Fig. 3. l_1 Magic Scheme Performance

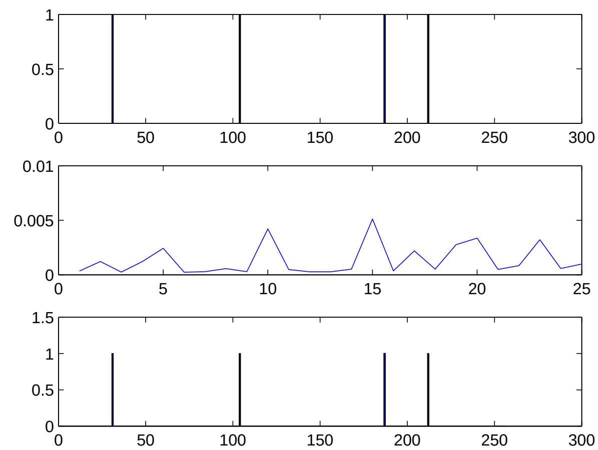

Fig. 4. Illustration of Correct Detection

Gaussian distribution. If we model the distribution to be more realistic Bernoulli distribution, the performance can be further improved.

Figure 4 shows an illustrative example with $M = 20$ and $N = 256$. We can see that the original signals are very sparse ($K = 4$). The received measurement from different sensors could not tell the original signals, which is shown in the middle of the figure. With the compressed sensing algorithm proposed in the previous sections, the reconstructed signal is shown to match the original signal perfectly. Figure 5 shows a possible mistake, in which the reconstructed signals do not match the original signals. However, we can see that the converged results conflict with the prior information of binary sources. As a result, we can easily tell the results are not valid. In Figure 6, we show the performance improvement of our proposed heuristic scheme. Here $\delta = 0.5$, and maximal number of unsuccessful trail is 20. We can see that we can reduce the sampling rate further. For example for detection probability 80%, we can reduce sampling rate about 20% more.

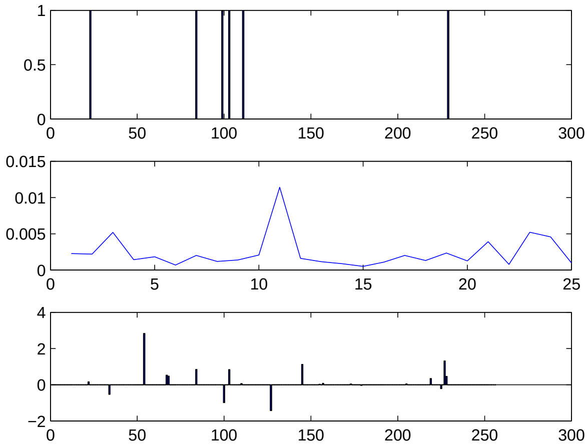

Fig. 5. Illustration of Incorrect Detection

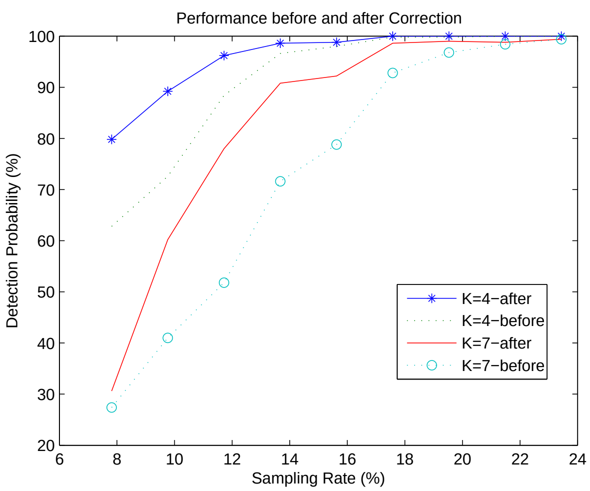

Fig. 6. Heuristic Improvement

Figure 7 shows the effect of noise on the proposed scheme. We can see that when the SNR reduces, the detection probability is significantly reduced. The detection probabilities for different sampling rates are similarly bad when the SNR is reduced to around 20dB. This is due to the reason that Gaussian distribution is unbounded and can cause significant detection error. In the future work, we will propose some denoise method. In addition, we will develop simulated annealing method to avoid undesired solutions so as to improve the detection performance.

V. Conclusions

In this paper, we propose a compressive sensing method for sparse event detection in wireless sensor networks. We formulate the problem and propose solutions. For the signal reconstruction part, we introduced a fully probabilistic Bayesian framework which helps dramatically reduce the sampling rate while still guaranty an overwhelming detection probability. Moreover, we adopt a marginal likelihood maximization algorithm and a heuristic algorithm for the Bayesian framework, which leads to higher detection probability than the traditional linear programming solution for this problem. The noisy condition is also analyzed.

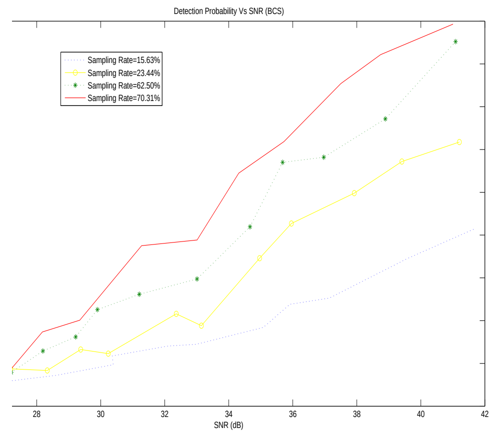

Fig. 7. Noise Effect

References

[1] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information", *IEEE Trans. on Information Theory*, 52(2) pp. 489 - 509, February 2006.

[2] E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?", *IEEE Trans. on Information Theory*, 52(12), pp. 5406 - 5425, December 2006.

[3] D. Donoho, "Compressed sensing", *IEEE Trans. on Information Theory*, 52(4), pp. 1289 - 1306, April 2006.

[4] W. U. Bajwa, J. Haupt, G. Raz, and R. Nowak, "Compressed channel sensing", *Conf. on Info. Sciences and Systems (CISS)*, Princeton, New Jersey, March 2008.

[5] G. Taubock and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots", *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Las Vegas, Nevada, April 2008.

[6] Y. Mostofi and P. Sen, "Compressed mapping of communication signal strength", *Military Communications Conference*, San Diego, CA, November 2008.

[7] Z. Tian, "Compressed wideband sensing in cooperative cognitive radio networks," *Proc. IEEE Globecom Conf.*, pp. 1-5, New Orleans, Dec. 2008.

[8] B. Krishnamachari, *Networking wireless sensors*, Cambridge University Press, UK, 2005.

[9] X. Li, *Wireless ad hoc and sensor networks*, Cambridge University Press, UK, 2008.

[10] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2006. (http://www.stanford.edu/ ~ boyd/cvxbook.html)

[11] Z. Han and K. J. R. Liu. *Resource allocation for wireless networks: basics, techniques, and applications*. Cambridge University Press, Cambridge, UK, 2008.

[12] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine", *Journal of Machine Learning Research*, vol. 1, p.p. 211-244, Sept. 2001.

[13] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models", in Proceedings of *the Ninth International Workshop on Artificial Intelligence and Statistics*, Key West, FL, Jan 3-6.

[14] S. Ji, Y. Xue and L. Carin, "Bayesian compressive sensing", *IEEE Trans. Signal Processing,* vol. 56, no. 6, June 2008.

[15] Rice University, L1-Related Optimization Project, available from http://www.caam.rice.edu/õptimization/L1/

[16] R. G. Baraniuk, "Compressive sensing", lecture notes, Rice university.

[17] Compressive sensing resource, http://www.dsp.ece.rice.edu/cs/

[18] D. J. C. MacKay, "A practical Bayesian framework for backprop networks", *Neural Computation*, 4:448-472, 1992.