

Compressed Sensing With Nonlinear Analog Mapping in a Noisy Environment

Ahmad Abou Saleh, Wai-Yip Chan, and Fady Alajaji

Abstract—We propose a low delay and low complexity sensor system based on the combination of Shannon–Kotel’nikov mapping and compressed sensing (CS). The proposed system uses nonlinear analog mappings on the CS measurements to increase their immunity against channel noise. Numerical results show that the proposed purely-analog system outperforms the state-of-the-art purely CS systems in terms of signal-to-distortion ratio. In addition to sparsity knowledge, we use a statistical characterization of the observed signal to further improve system performance.

Index Terms—Compressed sensing, Shannon–Kotel’nikov mapping.

I. INTRODUCTION

WIRELESS sensors networks (WSNs) monitor the physical world through distributed sensor nodes. These nodes, often conceived as having limited lifetime and sensing capabilities, communicate their sensed field information to a fusion center (FC) over power and bandwidth constrained noisy wireless channels. To meet these challenges, in this paper, we investigate using low delay and low complexity source-channel mapping with compressed sensing (CS) in WSNs.

The sensor inputs are treated as samples from an analog source. The traditional approach for analog source transmission is to use separate (tandem) source and channel coders. This approach is proven to be asymptotically optimal by Shannon. In tandem coding, the continuous source can be first compressed using a powerful vector quantizer, and then a capacity approaching channel code, such as a turbo or a low-density-parity-check code, is applied for channel noise protection. This method, however, results in very high encoding/decoding complexity and significant delays, which are not desirable in WSNs. The approach used in this work is analog joint source-channel coding which has been shown to achieve a good performance under low delay and complexity constraints [2]–[4]. More precisely, we propose to use a nonlinear analog mapping that acts as an analog joint source-channel encoder on the CS measurements. In [5], a hybrid digital-analog system is used with distributed compressed sensing over noisy channels. In this letter we consider a purely-analog spiral mapping within the CS context which is inherently analog. For reference, we compare the proposed analog system with state-of-the-art CS

systems that account for channel noise at the decoder. In addition to using sparsity knowledge and spiral mapping as in [1], we account for the statistical characteristics of the signal and use higher dimension expansion mapping to further improve system performance. The rest of the paper is organized as follows. In Section II, we briefly review the theory of compressed sensing. Section III describes Shannon–Kotel’nikov mapping using the 1:2 double Archimedes’ spiral. In Section IV, we develop the system structure and its optimization. Simulation results are included in Section V. Finally, conclusions are drawn in Section VI.

II. OVERVIEW OF COMPRESSED SENSING THEORY

The theory of compressed sensing has been developed in [6], [7]. In essence, CS exploits prior knowledge about the sparsity of a signal $\mathbf{x} \in \mathbb{R}^N$ in order to provide efficient signal sampling and reconstruction [7]. The signal \mathbf{x} is assumed to be sparse in some orthonormal basis $\Psi \in \mathbb{R}^{N \times N}$ (i.e., $\mathbf{x} = \Psi \mathbf{u}$, where \mathbf{u} is a transform coefficient vector). The sparsity assumption means that there are only K ($K \ll N$) nonzero elements in \mathbf{u} .

In CS we record $M < N$ linear measurements given by

$$\mathbf{y} = \Phi \mathbf{x} \quad (1)$$

where $\Phi \in \mathbb{R}^{M \times N}$ is a measurement matrix that is incoherent with the basis matrix Ψ (i.e., $\mu(\Psi, \Phi) \approx 1$, where $\mu(\Psi, \Phi)$ measures the largest correlation between any two elements in Ψ and Φ), and satisfies the restricted isometry property which is equivalent to requiring that all subsets of K columns of Φ are nearly orthogonal. This property can be achieved when the entries of the matrix Φ are i.i.d. Gaussian variables. In this case, with M on the order of $K \log_2(N/K)$, recovery of \mathbf{x} from \mathbf{y} works with overwhelming probability and is conducted by solving the convex optimization problem

$$\min_{\hat{\mathbf{x}}} \|\Psi^T \hat{\mathbf{x}}\|_{\ell_1}, \quad \text{subject to } \Phi \hat{\mathbf{x}} = \mathbf{y} \quad (2)$$

where $\|(\cdot)\|_{\ell_1}$ is the ℓ_1 norm ($\|\mathbf{x}\|_{\ell_1} \triangleq \sum_{i=1}^N |x_i|$) and $(\cdot)^T$ denotes the transpose operator. Several CS reconstruction methods based on convex relaxation (such as basis pursuit (BP) [8]) and greedy search (such as orthogonal matching pursuit (OMP) [9]) were developed.

In practice, the collected measurements are usually disturbed by noise \mathbf{n} ; thus they are modeled as $\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}$. For CS to be widely applicable, signal recovery should be robust against noise; a small distortion in the measurements should result in a small distortion in the signal recovery. Using a noise-aware version of (2), the signal is recovered as follows

$$\min_{\hat{\mathbf{x}}} \|\Psi^T \hat{\mathbf{x}}\|_{\ell_1}, \quad \text{subject to } \|\Phi \hat{\mathbf{x}} - \mathbf{y}\|_{\ell_2} \leq \epsilon \quad (3)$$

where ϵ bounds the total amount of noise in the measurements. Following [7], and assuming a Gaussian noise, ϵ is chosen such that $\epsilon^2 = \sigma_n^2 (M + \zeta \sqrt{2M})$, where $\zeta = 2$ and σ_n^2 is the noise power which can be estimated at the decoder. The problem in

Manuscript received October 07, 2011; accepted November 14, 2011. Date of publication November 22, 2011; date of current version December 01, 2011. This work was supported in part by NSERC of Canada. The material in this paper was presented in part at the European Signal Processing Conference-Barcelona, Spain, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Eduard A. Jorswieck.

The authors are with the Department of Electrical and Computer Engineering, Queen’s University, Kingston, ON K7L 3N6 Canada (e-mail: ahmad.abou.saleh@queensu.ca; chan@queensu.ca; fady@mast.queensu.ca)

Digital Object Identifier 10.1109/LSP.2011.2177086

(3) is solved in [8] using basis pursuit with denoising (BPDN). (3) can be expressed using Lagrange multiplier λ as

$$\min_{\hat{\mathbf{x}}} \|\Psi^T \hat{\mathbf{x}}\|_{\ell_1} + \lambda \|\Phi \hat{\mathbf{x}} - \mathbf{y}\|_{\ell_2}^2. \quad (4)$$

This has the form of a least absolute shrinkage and selection operator (LASSO) problem. The term λ is tuned to control the tradeoff between the sparsity and the approximation error in order to achieve the best possible performance. Note that the CS properties can be generalized to the case of non-exactly sparse signals (e.g., signals with transform coefficients in \mathbf{u} that decay rapidly). Hence, the results in this work are extendable to the case of not exactly sparse signals.

III. A SHANNON-KOTEL'NIKOV MAPPING

In this section, a 1:2 double Archimedes' spiral mapping is described for a Gaussian memoryless source y with variance σ_y^2 . Bandwidth expansion is performed by mapping each source sample $y \in \mathbb{R}$ to a two-dimensional channel symbol, which is a point on the double Archimedes' spiral, given by

$$\mathbf{s}(y) = \begin{bmatrix} z_1(y) \\ z_2(y) \end{bmatrix} = \frac{1}{\pi} \begin{bmatrix} \text{sgn}(y) \Delta \varphi(y) \cos \varphi(y) \\ \text{sgn}(y) \Delta \varphi(y) \sin \varphi(y) \end{bmatrix} \quad (5)$$

where $\text{sgn}(\cdot)$ is the signum function, Δ is the radial distance between any two neighboring spiral arms, and $\varphi(y) = \sqrt{6.25|y|/\Delta}$ is a stretching bijective function. For a given channel signal-to-noise ratio (CSNR) defined as P/σ_n^2 , where P is the average channel power, the radial distance Δ is optimized to minimize the total distortion by solving the following unconstrained optimization problem [4]

$$\Delta_{\text{opt}} = \arg \min_{\Delta} [\bar{\varepsilon}_{\text{wn}}^2(\Delta) + \bar{\varepsilon}_{\text{th}}^2(\Delta)] \quad (6)$$

where $\bar{\varepsilon}_{\text{wn}}^2$ and $\bar{\varepsilon}_{\text{th}}^2$ are, respectively, the average weak noise and threshold distortion under maximum likelihood (ML) decoding as defined in [4]. For a Gaussian source, the average weak noise distortion is given by [4]

$$\bar{\varepsilon}_{\text{wn}}^2 \approx \frac{\sigma_n^2}{\alpha^2} \quad (7)$$

where α is a gain factor related to the average channel power constraint P . The threshold distortion is approximated by [4]

$$\bar{\varepsilon}_{\text{th}}^2 \approx \left[1 - \text{erf} \left(\frac{\Delta}{2\sqrt{2}\sigma_n} \right) \right] \left[\left(\frac{\pi^4 \eta^2 \Delta^2}{\alpha^2} + 4\sigma_y^2 \right) \text{erf}(a) - \left(\frac{2\pi^2 \eta \Delta + a\sigma_y \alpha}{\sqrt{2\pi} \alpha} \right) 8\sigma_y e^{-\frac{a^2}{2\sigma_y^2}} + \frac{16\pi^2 \eta \Delta \sigma_y}{\sqrt{2\pi} \alpha} \right] \quad (8)$$

where $\text{erf}(\cdot)$ is the Gaussian error function, $a = 4\sigma_y$, and $\eta = 0.16$. Note that (7) and (8) become increasingly accurate at high CSNR levels.

At the receiver side, we use the minimum mean square error (MMSE) decoder instead of the ML decoder used in [4]. MMSE decoding has been shown to achieve a substantial performance improvement over ML decoding at low CSNRs under 2:1 bandwidth reduction [10] and 1:2 bandwidth expansion as simulation has shown. For 1:2 bandwidth expansion, the MMSE decoding rule is given by

$$\hat{y}_{\text{MMSE}} = \mathbb{E}[y|\hat{z}_1, \hat{z}_2] = \frac{\int y p(\hat{z}_1, \hat{z}_2|y) p(y) dy}{\int p(\hat{z}_1, \hat{z}_2|y) p(y) dy} \quad (9)$$

where $\mathbb{E}[\cdot]$ denotes the expectation operator, $p(\cdot)$ represents the probability density function (pdf), and $\hat{z}_i = z_i + n_i$, $i = 1, 2$, are the received channel outputs. To make the decoder implemen-

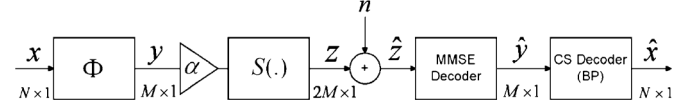


Fig. 1. Proposed system structure.

tation computationally efficient, we devise a decoder based on quantization and table-lookup, thereby avoiding having to compute a numerical integration for each received sample. This is achieved via uniform quantization of the output of the channel $\hat{\mathbf{z}} \in \mathbb{R}^2$ and looking up the decoded value \hat{y} for each quantization bin in a table. For a sufficiently large number of bins, the uniform quantization will not degrade performance; for a small number of bins, however, it might be necessary to design an optimal quantizer.

IV. SYSTEM MODEL

A. System Structure

We consider a group of sensors observing a discrete time continuous amplitude source signal $\mathbf{x} \in \mathbb{R}^N$ that is sparse in some transform basis Ψ . Each sensor encodes its observation and transmits it to the FC over additive white Gaussian noise (AWGN) channels with variance σ_n^2 . Our objective is to recover the sensor observations under a mean square error (MSE) fidelity criterion. The proposed system structure is shown in Fig. 1.

On the encoder side, the M sensors measure the observation \mathbf{x} using a measurement matrix $\Phi \in \mathbb{R}^{M \times N}$ where each element is i.i.d. Gaussian variable. The measurement vector comprising the M measurements is

$$\mathbf{y} = \Phi \mathbf{x}. \quad (10)$$

Using the double Archimedes' spiral given in (5), we map each element of the measurement vector \mathbf{y} to a two-dimensional channel input. It is observed that the measurements fit well a Gaussian distribution (see Section V). Hence, the results from Section III are hereafter used; the radial spiral distance Δ is calculated using (6) given the power allocated to the channel input under a total transmission power constraint (see below).

At the receiver side, we use the MMSE decoder to obtain the measurement estimate $\hat{\mathbf{y}}$. To recover the original signal, we use BP [8] to solve the ℓ_1 minimization problem in (2), and for comparison also the minimization in (3).

B. System Optimization

We optimize the proposed system for minimal end-to-end MSE distortion $\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$ which is a function of two sources of distortion: D_{CS} from compressed sensing (without lossy transmission) and D_{exp} from channel noise. As power is a scarce resource in WSNs, we aim for efficient utilization of this resource. Given a total transmission power constraint P_{tot} , we aim to minimize the end-to-end distortion with no constraint on the number of measurements and channel use. From CS theory, it is known that as the number of measurements increases, the distortion D_{CS} decreases. However, due to the total power constraint, the average power per channel (use) will decrease. This will increase the distortion D_{exp} from bandwidth expansion transmission. Thus, for a given channel condition, our aim is to determine the optimal number of measurements which results in a minimum end-to-end distortion under the total transmission power constraint P_{tot} .

Distortion D_{exp} from channel noise is minimized by optimizing Δ using (6). In the CS literature, however, there is not yet an explicit relation between the number of measurements M and the distortion D_{CS} obtained with BP. Thus optimization is done numerically by searching for the number of measurements that minimizes the end-to-end MSE distortion $\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$. A set \mathcal{X} of T realizations of the source vector with signal dimension N is created. Each source vector is synthesized as $\mathbf{x} = \Psi\mathbf{u}$, where Ψ is the sparsity basis and \mathbf{u} is a K sparse transform coefficient vector. There are $\binom{N}{K}$ possible sparsity patterns for \mathbf{u} (i.e., the sets of indices of the nonzero components of \mathbf{u}). Each realization is drawn uniformly from these patterns. For each number of measurements M , we create a fixed measurement matrix Φ whose entries follow a Gaussian distribution, and calculate the set of CS measurement vectors $\{\mathbf{y}\}$ using (10). For a given noise level, we optimize (Δ, α) using (6) under the average power constraint $P = P_{\text{tot}}/(2M)$. Note that Δ is optimized assuming ML decoding without any noticeable performance loss. The Archimedes' spiral, given in (5), is used on each component of the measurement vector to achieve 1:2 dimension expansion. Then, the measurement estimate $\hat{\mathbf{y}}$ is calculated using the MMSE decoder according to (9) and BP is used for signal reconstruction according to (2). We evaluate the overall MSE distortion $\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$ over the data set \mathcal{X} . Starting from a small value of M , we keep increasing M until we observe an increase in the overall distortion. The design search algorithm (Algorithm 1) is shown below. In our simulations, we used $T = 3 \cdot 10^4$, $\text{Inc} = 4$, and $m = K$. Note that the algorithm is run offline.

Algorithm 1 System Optimization

Data Input: Input data set $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$, channel noise variance σ_n^2 , and transmission power constraint P_{tot} .

Initialization: Set $M = m$, the incremental step Inc for the number of measurements, the overall MSE distortion $D^{(0)} = 10^{20}\text{T}$, $D^{(1)} = 10^{19}\text{T}$, $\Delta = \infty$, $\alpha = \infty$, and $i = 1$.

while $D^{(i)} < D^{(i-1)}$

1: $i \leftarrow i + 1$.

2: Set $\Delta_{\text{opt}} \leftarrow \Delta$, $\alpha_{\text{opt}} \leftarrow \alpha$, and $D_{\text{opt}} \leftarrow D^{(i-1)}$.

3: Create a random Gaussian CS matrix $\Phi^{(i)}$.

4: Obtain \mathbf{y} for each observation in \mathcal{X} using (10).

5: Scale the average channel power constraint as $P = P_{\text{tot}}/2M$.

6: Optimize (Δ, α) for the given channel noise variance σ_n^2 according to (6) under the power constraint $\mathbb{E}[z^2] = P$.

7: Apply 1:2 dimension expansion on \mathbf{y} using (5).

8: Decode $\hat{\mathbf{y}}$ using MMSE as in (9), and $\hat{\mathbf{x}}$ using BP as in (2).

9: Evaluate $D^{(i)} = \mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$ over the data set \mathcal{X} .

10: $M \leftarrow M + \text{Inc}$.

end while

Return $(M - \text{Inc}, D_{\text{opt}}, \Delta_{\text{opt}}, \alpha_{\text{opt}})$.

V. NUMERICAL RESULTS

In this section, we assume \mathbf{x} to be a sparse source in the discrete cosine transform basis Ψ with signal length $N = 100$. The signal \mathbf{x} is synthetically generated as $\Psi\mathbf{u}$, where \mathbf{u} has $K = 6$ nonzero elements. The results presented here are for the case where the nonzero elements in \mathbf{u} are i.i.d. zero-mean Gaussian with unit variance ($\mathcal{N}(0, 1)$) and the sparsity pattern

is uniformly distributed. We use the spiral mapping, discussed in Section III, to apply 1:2 dimension expansion, and BP to recover the source signal $\hat{\mathbf{x}}$ from the received measurements.

For benchmarking, we consider two CS-based systems BPDN (“CS-BPDN”) and LASSO (“CS-LASSO”), that were previously described in conjunction with (3) and (4), respectively. These reference systems use uncoded transmission over AWGN channels. Also CS decoding based on OMP is considered [9]. The number of measurements M is optimized for all systems under the total transmission power P_{tot} . This is realized using Algorithm 1 for the proposed system, whereas for the reference systems, we search over a range of M to obtain the one that produces the minimum end-to-end distortion. Since M varies with the channel noise variance, source signal-to-distortion ratio ($\text{SDR} \triangleq \mathbb{E}[\|\mathbf{x}\|^2]/\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$) is plotted against total signal-to-noise ratio ($\text{TSNR} \triangleq P_{\text{tot}}/\sigma_n^2$). From Fig. 2, it can be seen that the proposed system “CS-Mapping” outperforms the CS-BPDN for all TSNR levels, and CS-LASSO and CS-OMP from moderate to high TSNRs ($\text{TSNR} > 29$ dB). Notice that the gain from CS-Mapping as well as its gap to the reference systems gets more prominent as TSNR increases. This benefit from the spiral mapping contrasts with the two reference systems’ SDR which increases at a rate of 1 dB per dB increase in TSNR. We also simulate the proposed system using BPDN instead of BP on the noisy decoded measurement $\hat{\mathbf{y}}$. This gives around 1 dB gain in SDR over CS-Mapping with BP. In addition, we studied the CS system using sawtooth instead of spiral mapping. However, simulations have shown that our proposed system gives better performance. This can be attributed to the fact that \mathbf{y} fits well a Gaussian distribution; hence the structure of the optimized sawtooth approaches a linear mapping. Next, we show the result of using 1:3 analog mapping with CS system. In [11], a 3:1 dimension reduction was introduced; here we use the same parametric analog mapping to investigate 1:3 dimension expansion for the CS system. As shown from Fig. 2, we can notice that there is around 4 dB gain over CS-Mapping with 1:2 spiral mapping. The number of measurements and the mapping parameters were optimized numerically to give the best possible performance. Note that finding better parametric analog mappings for 1:3 dimension expansion is still an open problem; hybrid digital-analog mapping may be incorporated in the proposed system to increase its error resilience.

To further improve the performance of the proposed system, we exploit knowledge of the statistical along with the sparsity characteristics of the signal. This is referred to as “Statistical CS-Mapping.” The optimal decoding in the MSE sense is

$$\hat{\mathbf{u}}_{\text{MMSE}} = \mathbb{E}[\mathbf{u}|\hat{\mathbf{z}}] = \mathbb{E}_{p(I_{\mathbf{u}})} [\mathbb{E}[\mathbf{u}|\hat{\mathbf{z}}, I_{\mathbf{u}}]] \quad (11)$$

where $p(I_{\mathbf{u}})$ is the probability of the sparsity pattern $I_{\mathbf{u}}$ which follows a uniform distribution. The conditional expectation $\mathbb{E}[\mathbf{u}|\hat{\mathbf{z}}, I_{\mathbf{u}}]$ is calculated as follows:

$$\hat{\mathbf{u}}' \triangleq \mathbb{E}[\mathbf{u}|\hat{\mathbf{z}}, I_{\mathbf{u}}] = \frac{\int \mathbf{u} p(\hat{\mathbf{z}}|\mathbf{u}, I_{\mathbf{u}}) p(\mathbf{u}|I_{\mathbf{u}}) d\mathbf{u}}{\int p(\hat{\mathbf{z}}|\mathbf{u}, I_{\mathbf{u}}) p(\mathbf{u}|I_{\mathbf{u}}) d\mathbf{u}} \quad (12)$$

where $p(\mathbf{u}|I_{\mathbf{u}})$ is calculated from the fact that $p(u_i|I_{\mathbf{u}})$ is

$$\begin{cases} \delta_0, & \text{if } i \notin I_{\mathbf{u}} \\ \mathcal{N}(0, 1), & \text{if } i \in I_{\mathbf{u}} \end{cases} \quad (13)$$

with δ_0 is the dirac distribution, and $u_i, i = 1, \dots, N$, represents the i -th element in \mathbf{u} . Note that there are $\binom{N}{K}$ sparsity patterns

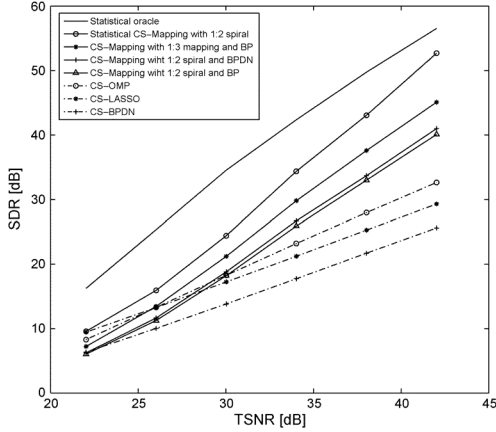


Fig. 2. Performance of CS-Mapping and CS reference systems with sparsity level $K = 6$ and signal length $N = 100$. The nonzero components of \mathbf{u} have i.i.d. $\mathcal{N}(0, 1)$ distribution. The number of measurements used by CS-Mapping at the asterisk marks are: [38 38 42 42 42 42], which correspond to the following CSNR = TSNR/(2M) [dB] levels: [3.2 7.2 10.7 14.7 18.7 22.7]. The performance of statistical CS-Mapping and oracle sparsity-pattern selection are also plotted. Note that the number of measurements M used for the reference systems are around twice of that used for our proposed system.

and thus in practice, it is hard to use the MMSE decoder as in (11) for a reasonable choice of N and K .

To make the decoder implementation computationally efficient, we use a two-stage implementation in a similar approach to [12]. In the first stage, we use the decoder shown in Fig. 1 to estimate the sparsity pattern $\hat{I}_{\mathbf{u}}$ by finding the K elements in $\hat{\mathbf{u}}_{\text{subopt}}$ with the K largest magnitudes, where $\hat{\mathbf{u}}_{\text{subopt}} = \Psi^T \hat{\mathbf{x}}_{\text{subopt}}$, with $\hat{\mathbf{x}}_{\text{subopt}}$ referring to the estimate calculated using the decoder shown in Fig. 1. The second stage is to apply the MMSE estimator using (12) in order to find a better estimate of $\hat{\mathbf{u}}$. Calculation of (12) can be simplified to only multiplication and addition operations by discretizing the vector \mathbf{u} using a uniform quantization step. However, the computation complexity will increase exponentially with the dimension of \mathbf{u} , hence making it not applicable in our case.

Monte carlo techniques can be used to lower the complexity of decoding. Given $\hat{I}_{\mathbf{u}}$, (12) can be expressed as follows:

$$\hat{\mathbf{u}}' = \frac{\mathbb{E}_{p(\mathbf{u}|\hat{I}_{\mathbf{u}})} [\mathbf{u}p(\hat{\mathbf{z}}|\mathbf{u}, \hat{I}_{\mathbf{u}})]}{\mathbb{E}_{p(\mathbf{u}|\hat{I}_{\mathbf{u}})} [p(\hat{\mathbf{z}}|\mathbf{u}, \hat{I}_{\mathbf{u}})]} \quad (14)$$

$$= \frac{\sum_{j=1}^L \mathbf{u}_j p(\hat{\mathbf{z}}|\mathbf{u}_j, \hat{I}_{\mathbf{u}})}{\sum_{j=1}^L p(\hat{\mathbf{z}}|\mathbf{u}_j, \hat{I}_{\mathbf{u}})} \quad (15)$$

where the samples \mathbf{u}_j follow the distribution in (13), and the last equation is valid due to the strong law of large numbers. Due to the high dimension of \mathbf{u} , the convergence of (15) is still slow. An alternative to sampling from the distribution $p(\mathbf{u}|\hat{I}_{\mathbf{u}})$ is to use importance sampling. This is conducted by using samples from another distribution $q(\mathbf{u}|\hat{I}_{\mathbf{u}})$. After some manipulations, (14) can be written as follows:

$$\hat{\mathbf{u}}' = \frac{\sum_{j=1}^L \frac{\mathbf{u}_j p(\hat{\mathbf{z}}|\mathbf{u}_j, \hat{I}_{\mathbf{u}}) p(\mathbf{u}_j|\hat{I}_{\mathbf{u}})}{q(\mathbf{u}_j|\hat{I}_{\mathbf{u}})}}{\sum_{j=1}^L \frac{p(\hat{\mathbf{z}}|\mathbf{u}_j, \hat{I}_{\mathbf{u}}) p(\mathbf{u}_j|\hat{I}_{\mathbf{u}})}{q(\mathbf{u}_j|\hat{I}_{\mathbf{u}})}}$$

where L is set to 10^4 . The distribution $q(\mathbf{u}|\hat{I}_{\mathbf{u}})$ has to be chosen to improve the convergence speed. The approach used here is

inspired by [10]. Since the result from the suboptimal decoding shown in Fig. 1 gives some information about the estimate $\hat{\mathbf{u}}'$, we choose $q(\mathbf{u}|\hat{I}_{\mathbf{u}})$ such that

$$\begin{cases} u_i \sim \delta_0, & \text{if } i \notin \hat{I}_{\mathbf{u}} \\ u_i \sim \mathcal{N}((\hat{u}_i)_{\text{subopt}}, \sigma^2), & \text{if } i \in \hat{I}_{\mathbf{u}} \end{cases}$$

where $(\hat{u}_i)_{\text{subopt}}$ denotes each element in the estimated signal $\hat{\mathbf{u}}_{\text{subopt}}$ found using the decoder shown in Fig. 1, and σ^2 is properly chosen to increase the speed of convergence, and gets smaller in relation to the variance of the channel noise.

As shown in Fig. 2, Statistical CS-Mapping gives large (3~10) dB gains in SDR over CS-Mapping. Notice that the gap between the two curves widens as TSNR increases. This can be attributed to the fact that at low TSNRs, CS-Mapping has poor performance which leads to a poor estimate of the sparsity pattern $\hat{I}_{\mathbf{u}}$. This is clearly shown from the ‘‘statistical oracle’’ curve which assumes perfect knowledge of the sparsity pattern at the decoder side. Note that the gap between the statistical oracle curve and the statistical CS-Mapping is from not knowing the indices of the nonzero components in the sparse signal. Hence, the performance of the proposed CS system cannot reach the oracle curve. Moreover, We have noticed that CS-Mapping without statistical knowledge outperforms CS-BPDN with statistical knowledge for moderate to high TSNRs.

VI. SUMMARY AND CONCLUSION

In this letter, we have presented a system which combines nonlinear analog mapping and compressed sensing over AWGN channels. The proposed purely-analog system is optimized for minimal overall MSE distortion under a transmission power constraint. Simulation results have shown that the system outperforms the state-of-the-art CS-based systems. In addition to sparsity knowledge, statistical characteristics of the signal is used to further improve the system.

REFERENCES

- [1] A. Abou Saleh, W.-Y. Chan, and F. Alajaji, ‘‘Compressed sensing with Shannon-Kotel’nikov mapping in the presence of noise,’’ in *Proc. Eur. Signal Processing Conf.*, Barcelona, Spain, 2011.
- [2] C. E. Shannon, ‘‘Communication in the presence of noise,’’ *Proc. IRE*, pp. 10–21, 1949.
- [3] V. A. Kotel’nikov, *The Theory of Optimum Noise Immunity*. New York: McGraw-Hill, 1959.
- [4] F. Hekland, P. A. Floor, and T. A. Ramstad, ‘‘Shannon-Kotel’nikov mappings in joint source-channel coding,’’ *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 94–105, Jan. 2009.
- [5] A. Kim and F. Hekland, ‘‘Dimension reduction and expansion: Distributed source coding in a noisy environment,’’ in *Proc. Data Compression Conference DCC*, Snowbird, UT, Mar. 2008, pp. 332–341.
- [6] D. Donoho, ‘‘Compressed sensing,’’ *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [7] E. Candes, J. Romberg, and T. Tao, ‘‘Stable signal recovery from incomplete and inaccurate measurements,’’ *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [8] S. Chen, D. Donoho, and M. Saunders, ‘‘Atomic decomposition by basis pursuit,’’ *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 31–61, 1998.
- [9] J. A. Tropp and A. C. Gilbert, ‘‘Signal recovery from random measurements via orthogonal matching pursuit,’’ *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [10] Y. Hu, J. Garcia-Frias, and M. Lamarca, ‘‘MMSE decoding for analog joint source channel coding using monte carlo importance sampling,’’ in *Proc. IEEE Workshop Signal Process. Adv. Wireless Commun.*, Italy, Jun. 2009.
- [11] P. A. Floor, ‘‘On the Theory of Shannon-Kotel’nikov Mappings in Joint Source-Channel Coding,’’ Ph.D. dissertation, Norwegian Univ. Sci. Technology, Trondheim, Norway, 2008.
- [12] S. Chartterjee, D. Sundman, and M. Skoglund, ‘‘Statistical postprocessing improves basis pursuit denoising performance,’’ in *Proc. IEEE Int. Symp. on Sig. Proc. and Infor. Tech.*, Luxor, Egypt, Dec. 2010.