

# Power-Constrained Low-Complexity Coding of Compressed Sensing Measurements

Ahmad Abou Saleh, Wai-Yip Chan, and Fady Alajaji  
Queen's University, Kingston, ON, K7L 3N6

Email: ahmad.abou.saleh@queensu.ca, chan@queensu.ca, fady@mast.queensu.ca

**Abstract**—We study a low delay and low complexity sensor-communication system based on compressed sensing (CS) and scalar coding for transmission. The proposed scheme uses a  $1 : r$  channel dimension expansion on the CS measurements for protection against channel noise. Simulation results show that optimizing the choice of  $r$  and the power allocation between the  $r$  transmissions significantly improve the system performance when compared to existing CS-communication schemes. Moreover, we consider the asymptotic behaviour of our CS system as the channel signal-to-noise ratio grows without bound and show that the proposed scheme achieves the optimal scaling exponent.

## I. INTRODUCTION

Low delay and complexity requirements are becoming more relevant with the increasing popularity of wireless sensors networks (WSNs). A wide range of natural and manmade sensed signals are sparse in appropriate transform domains. The sensor nodes, conceived as having limited lifetime and sensing capabilities, communicate their sensed field information to a fusion centre (FC) over wireless channels. To meet these challenges, we use low delay/complexity coding with compressed sensing (CS) in WSNs.

The sensor inputs are treated as discrete-time analog sources. The conventional approach for analog source transmission in point-to-point communications systems is to use separate source and channel coders. This approach, which is proven to be asymptotically optimal by Shannon, results in very high encoding/decoding complexity and significant delays, which are not desirable in WSNs. Low delay joint source-channel coding (JSCC) schemes have been shown to achieve a good performance under low delay and complexity constraints [1]–[9]. In this work, we propose to use a nonlinear mapping that acts as a joint source-channel encoder on the CS measurements. In [10], a low delay JSCC system was studied with distributed compressed sensing over noisy channels; more precisely a 1:2 expansion system, where each CS measurement is expanded into two dimensions, was used. In our previous work [11], we considered a purely-analog 1:2 spiral mapping within the CS context. In [12], [13], the authors studied JSCC with CS based on linear coding. Recently, in [14], the authors used a purely digital approach and studied vector-quantized transmission of CS measurements over noisy channels. Different from our previous work [11], we study a coding scheme, that is able to achieve any (integer) expansion rate on the CS measurements and investigate its asymptotic behaviour (as the channel signal-to-noise ratio goes to infinity) in terms of the scaling exponent. The functionality of achieving a

greater expansion ratios in our new scheme is shown to be useful under certain system settings; in such case, greater dimension expansion is more gainful than transmitting more measurements (with less dimension expansion for each). This coding scheme uses a multi-stage uniform quantizer and a linear map that scales the quantization error from the last stage; we refer to such low delay/complexity coding approach as “scalar coding.” Our contributions are twofold.

First, we adapt and improve a low delay scalar coding scheme considered in [8] for CS measurements over noisy channels; this coding scheme is chosen for its low delay and low complexity characteristics. Simulations show that when the scheme is optimized over the number of measurements and the expansion rate, the resulting performance surpasses (in most cases) several reference CS systems that account for channel noise. Moreover, we derive a closed form expression for the optimized power allocation between the channel inputs to further enhance the performance and reduce the complexity.

Second, using results from [8], we derive an upper bound on the system’s distortion that is used in the optimization process. We also show that the proposed scheme achieves the optimal scaling exponent while having low complexity and delay.

The rest of the paper is organized as follows. In Section II, we review the theory of compressed sensing. In Section III, we introduce the system model, derive an upper bound on its distortion and develop the system optimization. In Section IV, we show that the proposed scheme achieves the optimal scaling exponent. We present simulation results in Section V.

## II. OVERVIEW OF COMPRESSED SENSING

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 [15], [16]. The signal  $\mathbf{x}$  is assumed to be sparse in some orthonormal basis  $\Psi \in \mathbb{R}^{N \times N}$ ; in other words,  $\mathbf{x} = \Psi \mathbf{u}$ , where  $\mathbf{u}$  is a sparse vector. The sparsity assumption means that there are only  $K$  ( $K \ll N$ ) nonzero elements in  $\mathbf{u}$ .

In CS, the encoder records  $M < N$  linear measurements given by  $\mathbf{y} = \Phi \mathbf{x}$ , where  $\Phi \in \mathbb{R}^{M \times N}$  is a measurement matrix such that  $\Phi \Psi$  satisfies the restricted isometry property (RIP). The matrix  $\Phi \Psi$  satisfies the RIP of order  $K$  if there exists  $\delta_k \in (0, 1]$  such that, for any  $K$ -sparse  $\mathbf{u}$ ,  $(1 - \delta_k) \|\mathbf{u}\|^2 \leq \|\Phi \Psi \mathbf{u}\|^2 \leq (1 + \delta_k) \|\mathbf{u}\|^2$ , where  $\delta_k$  is the RIP constant of order  $K$  and  $\|\cdot\|$  denotes the  $\ell_2$  norm. This property is satisfied when the elements of the matrix  $\Phi$  are independent and identically distributed (i.i.d.) Gaussian variables with  $M$  of the order of  $K \log_2(N/K)$ . In this case, recovery of  $\mathbf{x}$  from  $\mathbf{y}$  works by solving

$$\min_{\hat{\mathbf{x}}} \|\Psi^T \hat{\mathbf{x}}\|_{\ell_1}, \quad \text{subject to } \Phi \hat{\mathbf{x}} = \mathbf{y} \quad (1)$$

where  $\|\cdot\|_{\ell_1}$  denotes the  $\ell_1$  norm ( $\|\mathbf{x}\|_{\ell_1} \triangleq \sum_{i=1}^N |x_i|$ ) and  $(\cdot)^T$  is the transpose operator. Several CS reconstruction methods were developed in [17], [18]. In practice, the collected measurements are disturbed by noise  $\hat{\mathbf{n}}$  (i.e.,  $\hat{\mathbf{y}} = \Phi \mathbf{x} + \hat{\mathbf{n}}$ ). Using a noise-aware version of (1), the signal is decoded as

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

where  $e$  bounds the total amount of noise in the measurements. (2) can be expressed using Lagrange multiplier  $\lambda$  as

$$\min_{\hat{\mathbf{x}}} \|\Psi^T \hat{\mathbf{x}}\|_{\ell_1} + \lambda \|\Phi \hat{\mathbf{x}} - \hat{\mathbf{y}}\|^2. \quad (3)$$

This has the form of a least absolute shrinkage and selection operator (LASSO) problem where  $\lambda$  is used to control the tradeoff between the sparsity and the approximation error.

### III. CS CODING OVER NOISY CHANNELS

#### A. System Structure

Consider a group of sensors observing a discrete-time continuous amplitude source signal  $\mathbf{x} \in \mathbb{R}^N$ . This observation is assumed to be sparse in  $\Psi$ . Each sensor encodes its observation and transmits it to the FC over additive white Gaussian noise (AWGN) channels. Fig 1 shows the proposed system structure. Our aim is to reconstruct the sensor observations under a mean square error (MSE) fidelity criterion given a constraint on the total transmission power  $P_{\text{tot}}$ .

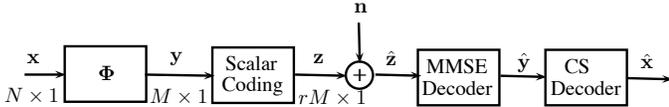


Fig. 1. System structure.

At the transmitter side, the  $M$  sensors measure the observation  $\mathbf{x}$  using a random measurement matrix  $\Phi \in \mathbb{R}^{M \times N}$ . We draw each element of  $\Phi$  from an i.i.d. Gaussian distribution with variance  $(1/M)\sigma_{\Phi}^2$ . The measurement vector is

$$\mathbf{y} = \Phi \mathbf{x}. \quad (4)$$

Next we map each element of  $\mathbf{y}$  to an  $r$ -dimensional channel input using  $1:r$  scalar coding based on a multi-stage scalar quantizer and a simple analog mapping [8] as described next.

To encode each element  $y$  into  $r$  dimensions, where  $r$  is an (integer) expansion rate, we first find the pairs  $(A_i, B_i)$

$$A_i = (\Delta_a)^i \text{int}(B_{i-1}/(\Delta_a)^i), \quad B_i = (B_{i-1} - A_i) \quad (5)$$

where  $i = 1, \dots, r-1$ ,  $B_0 = y$ ,  $0 < \Delta_a < 1$  is the quantization resolution and  $\text{int}(x)$  is the unique integer  $j$  such that  $x \in [j - \frac{1}{2}, j + \frac{1}{2})$ . Note that  $A_i$  represents the hierarchical quantization of  $y$  and  $B_i$  is the associated quantization error. To complete the description of the transmission strategy,  $A_i$ 's and  $B_{r-1}$  are scaled by a gain factor to satisfy the total power constraint  $P_{\text{tot}}/M$  of the  $r$  samples; hence, the corresponding channel inputs for the sample  $y$  can be expressed as follows

$$z_i = C_i A_i, \quad z_r = \sqrt{P_r/\sigma_E^2} B_{r-1} \quad (6)$$

where  $i = 1, \dots, r-1$ ,  $C_i = \sqrt{P_i/\sigma_{A_i}^2}$ ,  $\sigma_{A_i}^2$  and  $\sigma_E^2$  denotes the variance of  $A_i$  and  $B_{r-1}$ , respectively. Here,  $P_i$  is the power for  $i = 1, \dots, r$  with  $\sum_{i=1}^r P_i = P_{\text{tot}}/M$ . Different from [8], we optimize the power allocation between channel samples to minimize an upper bound on the MSE of the reconstructed measurements (10).

The received symbol, that is the  $(rM \times 1)$  expanded measurement vector  $\mathbf{z}$  disturbed by noise, is  $\hat{\mathbf{z}} = \mathbf{z} + \mathbf{n}$ , where  $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{rM})$  is AWGN and  $\mathbf{I}_{rM}$  is the identity matrix of size  $rM$ . At the receiver, we use the symbol-by-symbol minimum mean square error (MMSE) estimator to recover the measurement vector

$$\hat{y}_i = \frac{\int y_i p(\hat{z}_{(i-1)r+1}, \dots, \hat{z}_{ir} | y_i) p(y_i) dy_i}{\int p(\hat{z}_{(i-1)r+1}, \dots, \hat{z}_{ir} | y_i) p(y_i) dy_i} \quad (7)$$

for  $i = 1, \dots, M$ ,  $p(\cdot)$  and  $p(\cdot|\cdot)$  denote the probability density function (pdf) and the conditional pdf, respectively. To recover the original signal, we solve the LASSO problem in (3), where  $\lambda$  is chosen to minimize  $\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$  over a training set. For a given number of measurements  $M$  and expansion rate  $r$ ,  $\Delta_a$  and the power allocation  $\{P_i\}_{i=1}^r$  are the only parameters that need to be optimized. Since CS measurements are equally important [19], the power has to be equally allocated between them; this explains the constraint  $\sum_{i=1}^r P_i = P_{\text{tot}}/M$  (where we divide the total power by  $M$ ). Note that the same system parameters are used for each measurement.

#### B. Upper Bound on the System Distortion

To find a tractable upper bound on the system's distortion, we will assume a suboptimal decoder for estimating the measurement vector  $\hat{\mathbf{y}}$ . We follow a similar approach to [8]. Instead of using MMSE estimation as in (7), we use a maximum likelihood (ML) estimator to analytically obtain an estimate for each  $A_i$  and a linear MMSE (LMMSE) estimator to decode  $B_{r-1}$ . These estimators can be expressed as follows

$$\hat{A}_i = \Delta_a^i \arg \min_{j \in \mathbb{Z}} \left| \sqrt{\frac{P_i}{\sigma_{A_i}^2}} \Delta_a^i j - \hat{z}_i \right| \quad (8)$$

$$\hat{B}_{r-1} = \frac{\mathbb{E}[B_{r-1} \hat{z}_r]}{\mathbb{E}[\hat{z}_r^2]} \hat{z}_r = \frac{\sqrt{P_r \sigma_E^2}}{P_r + \sigma_n^2} \hat{z}_r. \quad (9)$$

Each sample in the measurement matrix is then given by  $\hat{y} = \sum_{i=1}^{r-1} \hat{A}_i + \hat{B}_{r-1}$ . The MSE in each sample of measurement vector  $\mathbb{E}[(y - \hat{y})^2]$  can be decomposed as follows

$$\mathbb{E}[(y - \hat{y})^2] = \sum_{i=1}^{r-1} \mathbb{E}[(A_i - \hat{A}_i)^2] + \mathbb{E}[(B_{r-1} - \hat{B}_{r-1})^2] \quad (10)$$

due to the fact that the error terms  $A_i - \hat{A}_i$  depend only on the channel noise and hence are independent of each other and of the quantization error. The distortion  $\mathbb{E}[(A_i - \hat{A}_i)^2]$  is

$$\Delta_a^i \sum_{j \in \mathbb{Z}} j^2 Pr(\hat{z}_i \in \mathcal{I}_j) = 2\Delta_a^i \sum_{j=1}^{\infty} j^2 Pr(\hat{z}_i \in \mathcal{I}_j) \quad (11)$$

where the above equality follows from the symmetry of the distribution of  $\hat{z}_i$ , the interval  $\mathcal{I}_j = [(j - \frac{1}{2})\sqrt{P_i/\sigma_{A_i}^2} \Delta_a^i, (j + \frac{1}{2})\sqrt{P_i/\sigma_{A_i}^2} \Delta_a^i]$ , the transition probability  $Pr(\hat{z}_i \in \mathcal{I}_j) = \mathcal{Q}\left((j - \frac{1}{2})\sqrt{P_i/\sigma_{A_i}^2} \Delta_a^i/\sigma_n\right) - \mathcal{Q}\left((j + \frac{1}{2})\sqrt{P_i/\sigma_{A_i}^2} \Delta_a^i/\sigma_n\right)$  and  $\mathcal{Q}(\cdot)$  is the Gaussian  $Q$ -function. Note that the ML decoder described in (8) gives  $j\Delta_a^i$  whenever  $\hat{z}_i \in \mathcal{I}_j$ . By using the fact that  $\mathcal{Q}(x) \leq 0.5e^{-x^2/2}$ , dropping the second term in  $Pr(\hat{z}_i \in \mathcal{I}_j)$ , and noting that  $(j - 1/2)^2 > j$  for  $j \geq 2$ , the distortion  $\mathbb{E}[(A_i - \hat{A}_i)^2]$  in (11) is upper bounded as follows

$$\Delta_a^i \exp\left(-\frac{P_i \Delta_a^{2i}}{8\sigma_n^2 \sigma_{A_i}^2}\right) + \Delta_a^i \sum_{j=2}^{\infty} j^2 \exp\left(-\frac{j P_i \Delta_a^{2i}}{2\sigma_n^2 \sigma_{A_i}^2}\right) \quad (12)$$

where the infinite sum can be written as  $\sum_{j=1}^{\infty} j^2 p^j - p = \frac{p^2+p}{(1-p)^3} - p$  with  $p = \exp(-P_i \Delta_a^{2i} / (2\sigma_n^2 \sigma_{A_i}^2))$ . The first term in (12) dominates for high values of  $P_i / \sigma_n^2$ . From LMMSE theory, the distortion in the quantization error is

$$\mathbb{E}[(B_{r-1} - \hat{B}_{r-1})^2] \stackrel{(a)}{=} \sigma_E^2 - \frac{P_r \sigma_E^2}{P_r + \sigma_n^2} \stackrel{(b)}{\leq} \sigma_E^2 \frac{\sigma_n^2}{P_r}. \quad (13)$$

Adding (12) and (13)-(a) gives an upper bound on the distortion  $\mathbb{E}[(y - \hat{y})^2]$ . Assuming enough measurements are available to satisfy the RIP condition, the distortion from reconstructing the source signal using the CS decoder depends on the amount of noise affecting the measurements [16]. In particular, the distortion from recovering the source is upper bounded by the noise in the measurement vector up to a scaling factor

$$\|\mathbf{x} - \hat{\mathbf{x}}\|^2 \leq c^2 \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \quad (14)$$

where the constant  $c$  depends on the CS decoder and is related to the RIP constant  $\delta_k$ .  $\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$  can be expressed similarly.

### C. System Design and Optimization

From (14), we can notice that the upper bound on the distortion  $\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$  assumes that the RIP condition is satisfied, i.e., there is a sufficient number of measurements  $M$ . The behaviour of a CS system is not understood well when not enough measurements are collected. Hence, we will use a semi-analytical approach to optimize our scheme.

The end-to-end MSE distortion  $\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$  of the proposed scheme comprises two sources of distortion:  $D_{CS}$  from compressed sensing and  $D_{exp}$  from the use of dimension expansion over the AWGN channel. Given a total power constraint  $P_{tot}$  and a maximum expansion rate  $r_{max}$ , we aim to minimize the end-to-end distortion with no constraint on the number of measurements. From CS theory, it is known that the distortion  $D_{CS}$  decreases with increasing number of measurements. However, due to the total power constraint, the average power per channel (use) will decrease; this will increase the distortion  $D_{exp}$ . Hence, we need to determine the optimal number of measurements and expansion rate that should be used to minimize the overall distortion.

Distortion  $D_{exp}$  from channel noise is minimized by optimizing the quantization resolution  $\Delta_a$  and the power allocation between channel samples (see Sec. III-B). However, in the CS literature, there is not yet an explicit equation that relates  $D_{CS}$  to the number of measurements  $M$ . Thus optimization is done numerically by searching for the number of measurements that minimizes  $\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$ . This is achieved by forming a set  $\mathcal{X}$  of  $T$  realizations of the source  $\mathbf{x}$ . Each source vector is synthesized as  $\mathbf{x} = \Psi \mathbf{u}$ , where  $\Psi$  is the sparsity basis and  $\mathbf{u}$  is a  $K$  sparse vector where each nonzero element is drawn from an i.i.d. Gaussian distribution with variance  $\sigma_u^2$ . 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. The set  $\mathcal{X}$  will be used in order to evaluate (numerically) the end-to-end performance. For a given noise level  $\sigma_n^2$ , we optimize the system parameters  $\Delta_a$  and  $\{P_i\}_{i=1}^r$ . Power is allocated to minimize the upper bound on the distortion  $\mathbb{E}[(y - \hat{y})^2]$ . To find a closed form expression for the power allocation, we approximate the distortion  $\mathbb{E}[(A_i - \hat{A}_i)^2]$  by the first term

of (12) (this is accurate for low noise levels). Adding the first term of (12) with (13)-(a), we obtain a good approximation of the bound on  $\mathbb{E}[(y - \hat{y})^2]$ , denoted by  $\tilde{D}_y$ . The problem now is to find  $\{P_i\}_{i=1}^r$  that minimize  $\tilde{D}_y$  under the constraints  $\sum_{i=1}^r P_i = P_{tot}/M$  and  $P_i \geq 0$ . Using the Karush-Kuhn-Tucker (KKT) conditions [20], we can find a closed form expression for the power allocation of the first  $r - 1$  stages (i.e.,  $\{P_i\}_{i=1}^{r-1}$ ) as follows (see the Appendix for the proof)

$$P_i = \frac{8\sigma_n^2 \sigma_{A_i}^2}{\Delta_a^{2i}} \left( \frac{P_{tot}/M - P_r - 8\sigma_n^2 \log F}{8\sigma_n^2 \sum_{i=1}^{r-1} \sigma_{A_i}^2 / \Delta_a^{2i}} + \log \frac{\Delta_a^{3i}}{8\sigma_n^2 \sigma_{A_i}^2} \right)^+ \quad (15)$$

where  $i = 1, \dots, r - 1$ ,  $F = \prod_{i=1}^{r-1} \left( \frac{\Delta_a^{3i}}{8\sigma_n^2 \sigma_{A_i}^2} \right)^{\frac{\sigma_{A_i}^2}{\Delta_a^{2i}}}$  and

$(x)^+ = \max(0, x)$ . It is hard to find a closed form expression for  $P_r$ ; instead we perform a numerical search to find the ‘‘best’’  $P_r$ . Note that  $\{P_i\}_{i=1}^{r-1}$  are calculated using (15). The design (offline) optimization (Algorithm 1) is shown below.

**Data Input:** Input data set  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$ , channel noise variance  $\sigma_n^2$ , total transmission power constraint  $P_{tot}$  and maximum expansion rate  $r_{max}$ .

**Initialization:** Set the incremental step  $Inc$  for the number of measurements, the overall distortion  $D = 10^{20}T$ .

**for**  $M = K : Inc : M_{max}$

- 1: Create a random Gaussian CS matrix  $\Phi$  and  $D_y = 10^{19}T$ .
- 2: Obtain  $\mathbf{y}$  for each observation in  $\mathcal{X}$  using (4).

**for**  $r = 2 : r_{max}$

- 3: Search over a discrete set for  $\Delta_a$  and  $P_r$  with  $\{P_i\}_{i=1}^{r-1}$  as given in (15) so that the upper bound on  $\mathbb{E}[(y - \hat{y})^2]$  (Sec. III-B) is minimized.

**if**  $D_y > \mathbb{E}[(y - \hat{y})^2]$

- 4: Set  $D_y = \mathbb{E}[(y - \hat{y})^2]$ ,  $r_b = r$ ,  $(\Delta_a)_b \leftarrow \Delta_a$  and  $(\{P_i\}_{i=1}^r)_b \leftarrow \{P_i\}_{i=1}^r$ .

**end if**

**end for**

- 5: Apply  $(1 : r_{opt})$  scalar coding on  $\mathbf{y}$  using (5) and (6).

- 6: Estimate  $\hat{\mathbf{y}}$  using (7), and  $\hat{\mathbf{x}}$  using (3).

- 7: Evaluate  $\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$  over the data set  $\mathcal{X}$ .

**if**  $D > \mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$

- 8: Set  $D = \mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$ ,  $M_{opt} \leftarrow M$ ,  $r_{opt} = r_b$ ,  $(\Delta_a)_{opt} \leftarrow (\Delta_a)_b$  and  $(\{P_i\}_{i=1}^r)_{opt} \leftarrow (\{P_i\}_{i=1}^r)_b$ .

**end if**

**end for**

**Return**  $(M_{opt}, D, (\Delta_a)_{opt}, r_{opt}, (\{P_i\}_{i=1}^r)_{opt})$ .

In simulations, we used  $T = 3 \times 10^4$ ,  $M_{max} = \frac{N}{2}$  and  $Inc = 4$ .

## IV. SCALING EXPONENT FOR THE CS SYSTEM

In this section, we assume equal power allocation between channel inputs denoted by  $P = P_{tot}/(rM)$ . We refer to a CS system to be in the large system regime if  $N$ ,  $M$  and  $K$  tend to infinity under the constraint that  $K/N$  and  $K/M$  converge to a constant. In such a large system regime, the CS measurement  $\mathbf{y}$  converges to an i.i.d. Gaussian distribution with variance  $\sigma_y^2 = (K/M)\sigma_u^2\sigma_\Phi^2$ . This can be proved using the central limit theorem following similar approach as in [21]. Since  $\Psi$  is an orthonormal matrix,  $\mathbf{G} = \Phi\Psi$  is a random matrix with i.i.d. entries drawn from  $\mathcal{N}(0, (1/M)\sigma_\Phi^2)$ . Using the fact that  $\mathbf{G}$  and  $\mathbf{u}$  are independent and that there are only  $K$  nonzero

elements in  $\mathbf{u}$ , each sample in the vector  $\mathbf{y}$  converges to a Gaussian distribution with variance  $(K/M)\sigma_u^2\sigma_\Phi^2$ .

#### A. Optimal Scaling Exponent

Considering the CS system (in the large system regime) over an AWGN channel, and using the lossy joint source channel coding theorem on the measurement vector  $\mathbf{y}$ , we have

$$\sigma_y^2/\mathbb{E}[(y - \hat{y})^2] = (1 + P/\sigma_n^2)^r \quad (16)$$

where we equate the rate distortion function  $R(D) = \frac{M}{2} \log(\sigma_y^2/\mathbb{E}[(y - \hat{y})^2])$  of an i.i.d. Gaussian vector  $\mathbf{y}$  to the capacity of an AWGN channel  $C(P) = \frac{rM}{2} \log(1 + P/\sigma_n^2)$ . Now assuming knowledge of the sparsity pattern  $S$  of  $\mathbf{u}$  at the receiver, the MMSE estimator is linear and the reconstruction (for the nonzero elements in  $\mathbf{u}$ ), denoted by  $\hat{\mathbf{u}}_S$ , is given by

$$\hat{\mathbf{u}}_S = \Sigma_U \mathbf{G}_S^T (\mathbf{G}_S \Sigma_U \mathbf{G}_S^T + \sigma_n^2 \mathbf{I}_M)^{-1} \hat{\mathbf{y}} \quad (17)$$

where  $\Sigma_U = \sigma_u^2 \mathbf{I}_K$  and  $\mathbf{G}_S$  is a sub-matrix of  $\mathbf{G}$  formed by the columns of  $\mathbf{G}$  with indices  $s \in S$ . The estimated signal is then found as  $\hat{\mathbf{x}} = \Psi \hat{\mathbf{u}}$ . Denoting the smallest singular value of the matrix  $\Sigma_U \mathbf{G}_S^T (\mathbf{G}_S \Sigma_U \mathbf{G}_S^T + \sigma_n^2 \mathbf{I}_M)^{-1}$  by  $c_{lb}$ , hence  $\|\mathbf{x} - \hat{\mathbf{x}}\|^2 \geq c_{lb}^2 \|\mathbf{y} - \hat{\mathbf{y}}\|^2$ . The average distortion is then

$$\mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}})^2] \geq c_{lb}^2 M \sigma_y^2 / (1 + \text{CSNR})^r \quad (18)$$

where the channel signal-to-noise ratio  $\text{CSNR} = \frac{P}{\sigma_n^2}$ . Following the definition of [8], the scaling exponent is

$$E_{sc} \triangleq \lim_{\text{CSNR} \rightarrow \infty} \frac{\log \text{SDR}}{\log \text{CSNR}} \leq r \quad (19)$$

where the signal-to-distortion ratio  $\text{SDR} = \frac{\mathbb{E}[\|\mathbf{x}\|^2]}{\mathbb{E}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]}$ . The optimal  $E_{sc}$  is then  $r$ . We refer to the large regime system, which has high delay/complexity and uses the sparsity patterns as side information, by ‘‘CS-Oracle.’’

**Remark 1** Numerical results show that the large system regime (i.e.,  $\mathbf{y}$  converges to a Gaussian distribution) is valid for dimension values  $(K, M, N)$  of practical interest.

#### B. Scaling Exponent for the Proposed Scheme

Assuming equal power allocation  $P$  per channel input, the result in Sec. III-B can be still used and an upper bound on the MSE distortion can be found by replacing  $P_i$  and  $P_r$  in  $\mathbb{E}[(A_i - \hat{A}_i)^2]$  and  $\mathbb{E}[(B_{r-1} - \hat{B}_{r-1})^2]$  with  $P$ . Note that as  $\Delta_a$  goes to zero, the quantization error  $B_i$  is uniformly distributed over the interval  $[-\Delta_a^i/2, \Delta_a^i/2]$ . Hence, the variance of  $B_{r-1}$  from the last quantization stage is  $\sigma_E^2 = \Delta_a^{2(r-1)}/12$ .

From (12), we can see that the distortion  $\mathbb{E}[(A_i - \hat{A}_i)^2]$  decreases exponentially when  $\text{CSNR} \Delta_a^2$  tends to  $\infty$ . This happens for increasing CSNR when  $\Delta_a^2 = 1/\text{CSNR}^{(1-\epsilon)}$ , for some  $\epsilon > 0$ . Hence, the distortion  $\mathbb{E}[(y - \hat{y})^2]$  is dominated by the error from decoding  $B_{r-1}$ . Using (13)-(b) and the fact that  $\sigma_E^2 = \Delta_a^{2(r-1)}/12$ , we have the following

$$\mathbb{E}[(y - \hat{y})^2] \leq a \text{CSNR}^{-(r-\epsilon')} \quad (20)$$

where  $\epsilon' = (r-1)\epsilon$  and  $a > 0$ . Now, using (14) with (20), the scaling exponent of the proposed scheme is

$$\lim_{\text{CSNR} \rightarrow \infty} \frac{\log \text{SDR}}{\log \text{CSNR}} \geq r - \epsilon' \quad (21)$$

and  $\epsilon$  can be chosen so that the scaling exponent is optimal (i.e.,  $r$ ) as in [8]. While the limiting exponent is optimal, the

SDR of the proposed scheme is always inferior to that of the CS-Oracle (for any CSNR). This is due to the fact that the latter scheme knows the true sparsity pattern (at the receiver).

## V. NUMERICAL RESULTS AND DISCUSSION

In this section, we consider  $\mathbf{x}$  to be sparse in the discrete cosine transform basis  $\Psi$  with signal length  $N = 350$ . The signal is generated as  $\Psi \mathbf{u}$ , where  $\mathbf{u}$  has  $K = 5$  nonzero elements. The nonzero elements in  $\mathbf{u}$  are chosen to be i.i.d. zero-mean Gaussian with unit variance and the sparsity pattern is uniformly distributed. Fig. 2 shows the SDR performance against the  $\text{TSNR} = P_{\text{tot}}/\sigma_n^2$ . Note that all schemes in Fig. 2 are optimized in terms of the number of measurements as in Algorithm 1 for  $r_{\text{max}} = 3$ . We can notice that the proposed CS-Scalar scheme outperforms CS-Spiral of [11] for moderate to high TSNR levels and CS-LASSO (described at the end of Section II) for all levels. Note that for low TSNRs, the proposed scheme underperforms the CS-Spiral scheme due to the fact that we restrict the quantizer to be uniform in our scheme. Moreover, the use of optimized power allocation gives a few dBs gain in SDR over the one with equal power allocation. When optimizing the power allocation, the best expansion rate for all TSNRs is  $r = 3$ ; for equal power allocation, the optimized expansion rate is  $r = 2$  for low to moderate TSNR levels ( $\leq 45$  dB), and  $r = 3$  for high TSNRs. The new degree of freedom  $r$  improves the performance of the CS system. We have also verified that the derived power allocation provides the same performance as exhaustive (discrete) search for the optimized power allocation. Note that as TSNR increases, using higher expansion rate is more beneficial than using more measurements (which behaves as a linear mapping after certain threshold  $M_{th}$ ); this explains the downward trend of  $M$  with TSNR levels in Fig. 2.

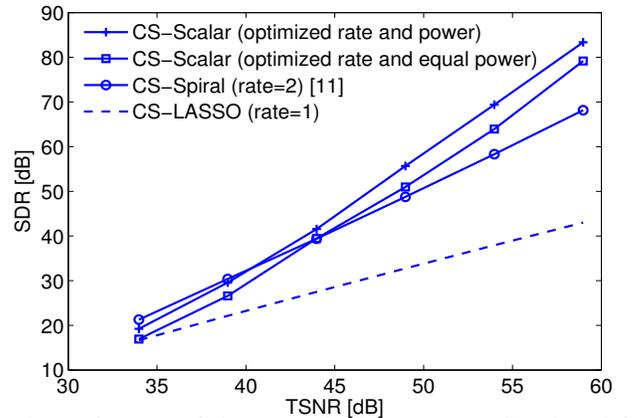


Fig. 2. Performance of the proposed scheme over AWGN channel for  $N = 350$ ,  $K = 5$  and  $P_{\text{tot}} = 250$ . The number of measurements (for all schemes) and the rate ( $r \leq 3$ ) are optimized using Algorithm 1. The number of measurements (for increasing TSNR) used at the plus marks are [97, 69, 73, 81, 65, 61], at the square marks are [97, 81, 65, 73, 65, 65]. The TSNR levels used in this graph correspond to CSNRs between 9 and 35 dB.

Having chosen the optimal  $\Delta_a$  in Fig. 2, we next shed some light on the importance of  $\Delta_a$  to both the SDR and the scaling exponent. Fig. 3 shows the SDR against  $\text{CSNR} = \frac{P}{\sigma_n^2}$  for different  $\Delta_a$ . For small  $\Delta_a$ , the error from the quantization parts vanishes slowly but the scaling exponent in the limit is larger. For larger  $\Delta_a$ , the error due to the discrete parts vanishes quickly but the resulting exponent is smaller. If  $\Delta_a$  is

chosen optimally in terms of CSNR, the resulting performance is the upper hull of the collection of all curves in Fig. 3.

**Remark 2** *The proposed scheme exhibits a similar trend as in Fig. 2 when decreasing  $P_{\text{tot}}$  to 100. In this case, using a rate threshold  $r_{\text{max}} = 4$  in Algorithm 1 is shown to be beneficial for  $\text{TSNR} \geq 45$  dB. With no constraint on the number of channel inputs, it is always beneficial to use higher rates for moderate to high  $\text{CSNR} = \text{TSNR}/(r_{\text{max}}M)$  levels.*

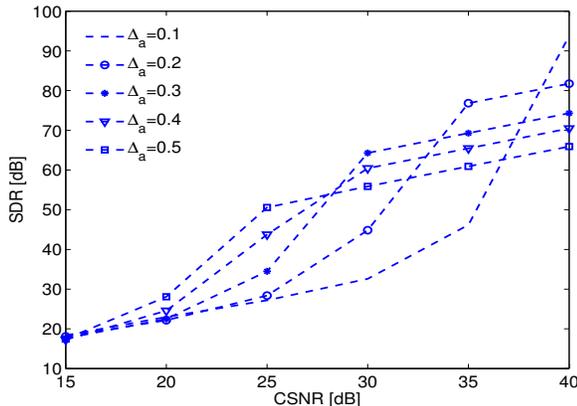


Fig. 3. Performance of CS-Scalar system over AWGN channel for different  $\Delta_a$ :  $P = 1$ ,  $N = 350$ ,  $K = 5$ ,  $M = 90$ .

#### APPENDIX

In this section, we aim to derive a closed form expression for the power allocation  $\{P_i\}_{i=1}^r$ . Recall, the problem is to minimize  $\tilde{D}_y = \Delta_a^i \exp\left(-\frac{P_i \Delta_a^{2i}}{8\sigma_n^2 \sigma_{A_i}^2}\right) + \frac{\sigma_E^2 \sigma_n^2}{P_r + \sigma_n^2}$  such that  $\sum_{k=1}^r P_k = P_{\text{tot}}/M$  and  $P_i \geq 0$ . This constrained optimization can be recast into an unconstrained minimization problem via the Lagrange cost function

$$J(P_i, \mu, \lambda_i) = \Delta_a^i \exp\left(-\frac{P_i \Delta_a^{2i}}{8\sigma_n^2 \sigma_{A_i}^2}\right) + \frac{\sigma_E^2 \sigma_n^2}{P_r + \sigma_n^2} + \mu \left(\sum_{i=1}^r P_i - \frac{P_{\text{tot}}}{M}\right) - \sum_{i=1}^r \lambda_i P_i \quad (22)$$

where  $\mu$  and  $\lambda_i$  denote the Lagrangian multipliers. Using the KKT conditions [20], we can write the following

$$\frac{\partial J}{\partial P_i} = 0, \quad P_i \lambda_i = 0, \quad \text{for } i = 1, \dots, r, \quad (23)$$

$$\text{and } \sum_{i=1}^r P_i = P_{\text{tot}}/M. \quad (24)$$

For all those  $P_i \neq 0$ , we have  $\lambda_i = 0$ ; solving the partial derivative in (23) we obtain

$$P_i = \frac{8\sigma_n^2 \sigma_{A_i}^2}{\Delta_a^{2i}} \left( \log \frac{1}{\mu} + \log \frac{\Delta_a^{3i}}{8\sigma_n^2 \sigma_{A_i}^2} \right)^+ \quad (25)$$

$$P_r = \left( \sqrt{\frac{\sigma_E^2 \sigma_n^2}{\mu}} - \sigma_n^2 \right)^+ \quad (26)$$

where  $i = 1, \dots, r-1$  and  $(x)^+ = \max(0, x)$ . Next, we need to substitute (25) and (26) into (24) and solve for  $\mu$ . Obtaining a closed form expression for  $\mu$  is intractable in this case. To overcome this problem, we assume a given  $0 < P_r < P_{\text{tot}}/M$

and solve for  $\{P_i\}_{i=1}^{r-1}$ . The new constraint is now given by  $\sum_{i=1}^{r-1} P_i = P_{\text{tot}}/M - P_r$ . Substituting (25) into the new constraint and solving for  $\mu$ , we have the following

$$\bar{\mu} \triangleq \log \frac{1}{\mu} = \frac{P_{\text{tot}}/M - P_r - 8\sigma_n^2 \log F}{8\sigma_n^2 \sum_{i=1}^{r-1} \sigma_{A_i}^2 / \Delta_a^{2i}} \quad (27)$$

where  $F = \prod_{i=1}^{r-1} \left( \frac{\Delta_a^{3i}}{8\sigma_n^2 \sigma_{A_i}^2} \right)^{\frac{\sigma_{A_i}^2}{\Delta_a^{2i}}}$ . The optimized  $\{P_i\}_{i=1}^{r-1}$  now can be easily found as given in (15) by inserting  $\bar{\mu}$  into (25). Note that  $P_r$  has to be found numerically.

#### REFERENCES

- [1] C. E. Shannon, "Communication in the presence of noise," in *Proc. IRE*, 1949, pp. 10–21.
- [2] V. A. Kotel'nikov, *The Theory of Optimum Noise Immunity*. New York: McGraw-Hill, 1959.
- [3] H. Coward and T. Ramstad, "Hybrid digital-analog transmission of analog source signals," in *Norwegian Signal Processing Symposium*, Sem Gjestegrd Asker, Norway, Sep 1999.
- [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] P. A. Floor, A. N. Kim, N. Wernersson, T. Ramstad, M. Skoglund, and I. Balasingham, "Zero-delay joint source-channel coding for a bivariate Gaussian on a Gaussian MAC," *IEEE Trans. Communications*, vol. 60, no. 10, pp. 3091–3102, Oct 2012.
- [6] Y. Hu, J. Garcia-Frias, and M. Lamarca, "Analog joint source channel coding using non-linear mappings and MMSE decoding," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3016–3026, Nov 2011.
- [7] E. Akyol, K. Rose, and K. Viswanatha, "On zero-delay source-channel coding," *IEEE Transactions on Information Theory*, in review.
- [8] M. Kleiner and B. Rimoldi, "Asymptotically optimal joint source-channel coding with minimal delay," in *Proc. Global Telecommunications Conference GLOBECOM*, Honolulu, HI, Nov 2009, pp. 332–341.
- [9] A. Abou Saleh, F. Alajaji, and W.-Y. Chan, "Low-latency source-channel coding for fading channels with correlated interference," *IEEE Wireless Communications Letter*, vol. 3, no. 2, pp. 137–140, April 2014.
- [10] A. Kim and F. Hekland, "Dimension reduction and expansion: Distributed source coding in a noisy environment," in *Proc. Data Compression Conference DCC*, Snowbird, Utah, Mar 2008, pp. 332–341.
- [11] A. Abou Saleh, W.-Y. Chan, and F. Alajaji, "Compressed sensing with nonlinear analog mapping in a noisy environment," *IEEE Signal Processing Letters*, vol. 19, no. 1, pp. 39–42, Jan 2012.
- [12] A. Hesammohseni, M. Babaie-Zadeh, and C. Jutten, "Inflating compressed samples: A joint source-channel coding approach for noise-resistant compressed sensing," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, April 2009.
- [13] S. Feizi and M. Medard, "A power efficient sensing/communication scheme: Joint source-channel-network coding by using compressive sensing," in *Proc. of the Forty-Ninth Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sep 2011.
- [14] A. Shirazinia, S. Chatterjee, and M. Skoglund, "Channel-optimized vector quantizer design for compressed sensing measurements," in *Proc. IEEE ICASSP*, Vancouver, Canada, May 2013.
- [15] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr 2006.
- [16] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [17] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 31–61, 1998.
- [18] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.
- [19] J. Laska, P. Boufounos, M. Davenport, and R. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 31, no. 3, pp. 429–443, 2011.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [21] W. Dai and O. Milenkovic, "Information theoretical and algorithmic approaches to quantized compressive sensing," *IEEE Trans. on Communications*, vol. 59, no. 7, pp. 1857–1866, Jul 2011.