Non-Linear Coding for Improved Performance in Compressive Sensing

Yichuan Hu, Zhongmin Wang, Javier Garcia-Frias, and Gonzalo R. Arce
Department of Electrical and Computer Engineering
University of Delaware
Newark, DE 19716
E-mail: {yhu, zhongmin, jgarcia, arce}@ee.udel.edu

***Abstract*—We propose a system based on the combination of compressive sensing and non-linear processing that shows excellent robustness against noise. The key idea is the use of non-linear mappings that act as analog joint source-channel encoders, processing the compressive sensing measurements proceeding from an analog source and producing continuous amplitude samples that are transmitted directly through the noisy channel. As we will show in our simulation results, the proposed framework is readily applicable in practical systems such as imaging, and clearly outperforms systems based on stand-alone compressive sensing.**

I. Introduction

Recently, the new field of compressive sensing (CS) [1], [2], [3], [4], [5] has emerged with the promise to revolutionize digital processing broadly. In CS, randomized nonadaptive linear projections are used to acquire an efficient, dimensionally reduced representation of a sparse signal or image directly employing just a few measurements. The idea is to replace samples by inner products with random vectors. The signal is then reconstructed by solving an inverse problem such as a linear program or a greedy pursuit in a basis where it admits a sparse representation. Surprisingly, Nyquist-rate sampling can be leaped over through CS theory, which may have a profound impact broadly, including applications in spectroscopy, imaging, communications, as well as in consumer electronics.

Most of the work in CS deals with deterministic signals and no other prior knowledge is assumed. In many applications, however, additional *a priori* information on the underlying signals is available, in addition to their sparsity. As shown in our ongoing research, such *a priori* information can be used in the general framework of CS [6], [7], [8], [9], [10], [11]: by exploiting prior knowledge, either at the encoder site [11] or by modifying the recovery process, reconstruction can be performed with much less measurements than expected in conventional CS, even in a basis in which the signal of interest is not sparse [10], making CS more flexible and applicable. Thus, we can think of an "extended" CS scenario that removes all existing redundancy in the signal of interest, thereby acting as an analog source encoder that represents the original signal with a small number of non-redundant measurements.

In this paper, we expand the aforementioned "extended" CS framework to make it an integral part of a discrete-time all-analog-processing communications system that completely skips the digital domain (bits are never utilized) and shows excellent robustness against noise. The key idea is the use of non-linear mappings that act as analog channel encoders, processing the CS measurements proceeding from an analog source and producing continuous amplitude samples that are transmitted directly through the noisy channel. Thus, we can think of the proposed system as the concatenation of an extended CS block, which acts as an analog source encoder, and a non-linear analog joint source-channel encoder. As we will see in our simulation results, this non-linear processing of the CS output makes the resulting system very robust against noise, clearly outperforming stand-alone CS. This is very significant, since even the best methods in the literature dealing with noisy CS measurements experience significant degradation. Intuitively, this degradation can be explained by realizing that CS can be interpreted as an analog linear code, and, as described in the literature [12], linear codes are extremely poor at high SNR.

In addition to the theoretical interest of the proposed all-analog framework, its relevance in practical systems is substantial. This may not be obvious at a first look, since digital communications systems based on separation between source and channel coding are optimal from a theoretical perspective [13]. However, in such standard digital communications systems the continuous source has to be source encoded up to the desired rate/distortion pair (e.g., using very powerful vector quantization), and then capacity approaching channel codes such as turbo codes or LDPC codes should be applied. Therefore, the price to pay to achieve near-theoretical performance is a very high encoding/decoding complexity and significant delays, since any capacity approaching channel code (and any quasi-optimal vector quantizer) requires long block lengths. Moreover, such digital systems have to be specifically designed for the desired rate and distortion: if the desired rate/distortion pair changes, the system has to be completely re-designed. All these problems are alleviated in our proposed purely analog framework. As we will see in our simulation results, contrary to digital communications systems, the use of long blocks is not necessary to achieve near-optimal performance, and processing complexity is much less than in quasi-optimal digital systems.

This work was supported in part by NSF Award EECS-0725422.

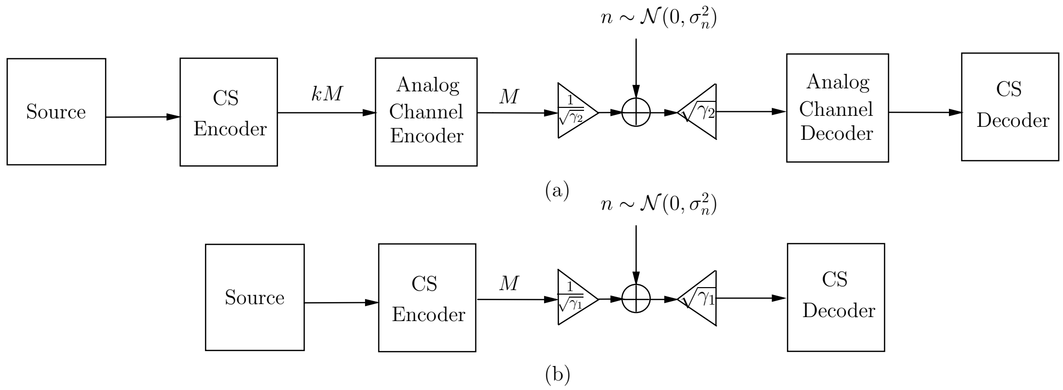


Fig. 1. (a) Proposed system model: kM measurements are acquired through CS, and the number of transmitted samples is further reduced to M using the non-linear space-filling curves described in Section II.B; (b) System based on the use of stand-alone CS: M measurements are acquired through CS and directly transmitted through the channel. Notice that in both (a) and (b) the number of transmitted samples is the same.

II. Proposed System

Fig. 1(a) depicts the proposed system, which is of great interest in practical applications such as imaging. First, data acquisition is performed through CS as described in Section II.A, which reduces the amount of samples to be processed to kM. Then, these samples are encoded using a non-linear mapping based on space-filling curves as described in Section II.B, which further reduces the amount of samples to be transmitted through the AWGN channel to M. As we will see in our simulation results, where we will focus on images, the resulting performance is much better than that of a system based on stand-alone CS, where M projections are directly obtained and transmitted through the channel (see Fig. 1(b)). In order for the proposed system to achieve the best possible performance, three aspects are critical:

- The CS block has to be able to act as an ideal analog source encoder, successfully exploiting all prior knowledge in the original source and thus generating kM i.i.d. measurements. Since this is an issue we have studied in our research [14], this will not be the main focus of this paper. However, for completeness we will present a very brief overview of our approach in Section II.A.
- The kM i.i.d. non-redundant measurements at the output of the CS block have to be optimally encoded into the M transmitted samples. Interestingly, it is well known in the literature that this can be achieved with less complexity and delay than digital systems by utilizing space-filling curves [15], [16], [17]. We will briefly review these ideas in Section II.B.
- The optimal value of k, which defines the trade-off in bandwidth reduction between CS and the non-linear mapping, has to be obtained for the SNR of interest. We will see in our simulation results that in high SNR environments parameter k has to be chosen as large as possible. From [12], the intuition for this behavior is that linear analog codes (from which CS is a particular case) are very suboptimum in such conditions. On the other hand, linear analog codes lead to good performance at low SNR values, and thus we can expect that stand-alone CS (i.e., $k = 1$) will perform well in these conditions.

A. Compressive Sensing as an Analog Source Encoder

As indicated in Section I, CS provides a new way to acquire and represent sparse signals that requires less sampling resources than traditional approaches [2], [4], [5]. Given a T sparse signal $\underline{x} \in \mathcal{R}^n$ on some basis $\Psi = [\underline{\psi}_1, \underline{\psi}_2, \ldots, \underline{\psi}_n]$, such that $\underline{x}$ can be approximated by a linear combination of T vectors from Ψ with $T \ll n$, the theory of compressive sensing shows that $\underline{x}$ can be recovered from m random projections with high probability when $m = CT \log n \ll n$, where $C \geq 1$ is the oversampling factor. The projections are given by $\underline{y} = P\underline{x}$, where P is an $m \times n$ random measurement matrix with its rows incoherent with the columns of Ψ. Commonly used random measurement matrices for CS are random Gaussian matrices ($P_{ij} \in \{\mathcal{N}(0, 1/n)\}$), Rademacher matrices ($P_{ij} \in \{\pm 1/\sqrt{n}\}$) and partial Fourier matrices. In [5], it is shown that a matrix satisfying the incoherence condition is so ubiquitous that “nearly all matrices are CS matrices”.

In order to use CS as an analog source encoder, we need to take into account the statistics of the original signal, in addition to the sparsity. In our simulation results, we focus on images, and utilize the new technique we proposed in [14], which exploits the *a priori* information that natural images exhibit in the statistical distribution of their wavelet domain representations. The basic idea is to implement CS by first performing the Hadamard transform, and then picking up the desired number of samples using a novel family of variable density sampling patterns, which are optimally designed using a simplified statistical model describing the distribution of natural images in the wavelet domain. The resulting samples can be expressed as $\underline{y} = P\underline{x}$, where matrix P depends on the Hadamard kernel and the optimally designed sampling pattern, which means that the proposed technique is indeed a CS encoder. We will skip the details here, since they can be found in [14]. However, it is important to remark that, as described in [14], the resulting performance is significantly better than that obtained with other sampling strategies.

In the case in which the measurements are not distorted by noise, signal reconstruction is achieved by solving an l_1 norm minimization problem: $\min \|\underline{\theta}\|_1$ subject to $P\Psi\underline{\theta} = \underline{y}$. Minimizing the l_1 norm yields solutions that are zero except at a small number of isolated values and can be solved by efficient optimization algorithms which include Basis Pursuit, Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP), and methods based on the Expectation-Maximization (EM) algorithm. In our framework, however, measurements will be contaminated by noise (directly by the AWGN channel in Fig. 1(b) or implicitly when the kM measurements are imperfectly recovered from the M samples by the analog channel decoder in Fig. 1(a)), so that $\underline{y}_n = \underline{y} + \underline{n}$, and the direct application of the aforementioned methods results in poor performance. Rather, we will apply the Basis Pursuit Denoising algorithm (BPDN) [18] which takes into account the noise by solving the following problem

$$\min \|P\Psi\underline{\theta} - \underline{y}_n\|_2^2 + \lambda\|\underline{\theta}\|_1 \quad \text{s. t.} \quad \underline{y}_n = P\Psi\underline{\theta} + \underline{n}, \qquad (1)$$

where $\lambda > 0$ depends on the noise level and has to be carefully chosen to optimize performance.

B. Space-Filling Curves as Analog Joint-Source Channel Encoders

It is well known that point-to-point analog communications are optimal in some circumstances. For instance, when no bandwidth expansion/reduction is performed (i.e., the number of samples transmitted through the channel is the same as the number of source samples) optimality is achieved for Gaussian sources by just transmitting the source output directly through the channel [19], [20]. On the other hand, when bandwidth expansion/reduction is performed (i.e., the number of channel uses is greater/smaller than the number of source samples), the optimal communications system is more complex. Previous work in the literature has indeed investigated possible schemes based on analog transformations aiming at perfectly "matching" sources with channels [19], [21], [22], [23], [24]. However, research in this area is still in its infancy.

Among the few practical analog coding schemes that have appeared in the literature, those based on the use of space-filling curves (such as spirals, Hilbert curves, etc), already proposed more than 50 years ago by Shannon [16] and Kotel'nikov [15], have recently acquired a renewed importance due to the work of Fuldseth [25], Chung [26], Ramstad [27], and Hekland [28], among others. The encoding idea to perform bandwidth reduction (i.e., to reduce the number of samples to be transmitted) is to represent a tuple of n source samples as a point in an n-dimensional space where a space-filling surface of dimension k lives. Then, the n-tuple is projected onto the curve and the corresponding k-tuple is transmitted through the noisy channel. For instance, consider a 2:1 bandwidth reduction scenario where out of 1000 source samples only 500 samples are transmitted through the channel. We can think of a plane in which we draw a spiral (or another of the aforementioned space-filling curves) starting at the origin. Then, each two source samples (x_{2k}, x_{2k+1}) define a point in the plane, and the sample to be transmitted corresponds to the length of the fragment of the curve located between the origin and the point in the curve that is closest to (x_{2k}, x_{2k+1}). Maximum Likelihood (ML) or Minimum Mean Squared Error (MMSE) decoding is performed to recover the original data.

Most of the work on space-filling curves has focused on ML decoding and high SNR. In these conditions, it is possible to analyze the system performance, and use this analysis to optimize the curve parameters [26], [28]. For diverse sources and high SNR, ML decoding results in a performance very close to the theoretical limits. This is extremely interesting, since the system complexity is lower than that of similar performing digital systems based on separation between source and channel coding. Moreover, since no long blocks are used the delay is very small. We have recently shown that, by properly optimizing the curve parameters, the use of MMSE decoding for Gaussian and Laplacian sources transmitted over AWGN channels results in a performance that is very close to the theoretical limits in the whole SNR region [17].

Specifically, we utilize spiral-like curves, which can be defined in the following parametric form [28]

$$\begin{cases} x = \dfrac{\Delta}{\pi}\theta\sin\theta \\ y = \dfrac{\Delta}{\pi}\theta\cos\theta \end{cases} for\ \theta \geq 0, \qquad \begin{cases} x = -\dfrac{\Delta}{\pi}\theta\sin\theta \\ y = \dfrac{\Delta}{\pi}\theta\cos\theta \end{cases} for\ \theta < 0, \qquad (2)$$

where Δ is the distance between two neighboring spiral arms and θ is the angle from the origin to the point (x, y) on the curve.

Notice that in the spiral curve described above there is a one-to-one correspondence between parameter θ and the points (x, y) on the curve, so that the curve gradually fills in the whole x-y plane as the absolute value of θ grows. Therefore, if all the other parameters are fixed, we can use a mapping function, $M_\Delta(\cdot)$, to project a pair of source samples (x, y) onto the curve by finding the closest point on the curve, which we will denote as $\hat{\theta} = M_\Delta(x, y)$. Then, an invertible function of $\hat{\theta}$, normalized so that the average transmitted energy per sample is equal to 1, is transmitted through the AWGN channel. In order to do so, we introduce a simple transform function $T_\alpha(x) = x^\alpha$ where $\alpha \in (0, 2]$ is a parameter that has to be optimized for different channel SNR, and denote the symbols transmitted through the channel as $T_\alpha(\hat{\theta})/\sqrt{\gamma}$, where $\sqrt{\gamma}$ is a normalization factor so that the average energy per sample transmitted through the channel is equal to 1 and thus $SNR = 10log\frac{1}{\sigma_n^2}$, where σ_n^2 is the noise variance.

We saw in [17] that by properly optimizing the curve parameters the resulting performance is excellent for both Gaussian and Laplacian sources. Specifically, for the whole SNR region the gap to the theoretical limits is below 1 dB in the case of Gaussian sources, and below 1.4 dB for Laplacian sources.

III. Simulation Results

In order to illustrate the potential of the proposed system, we consider the 256×256 image "Boat", which is sparse

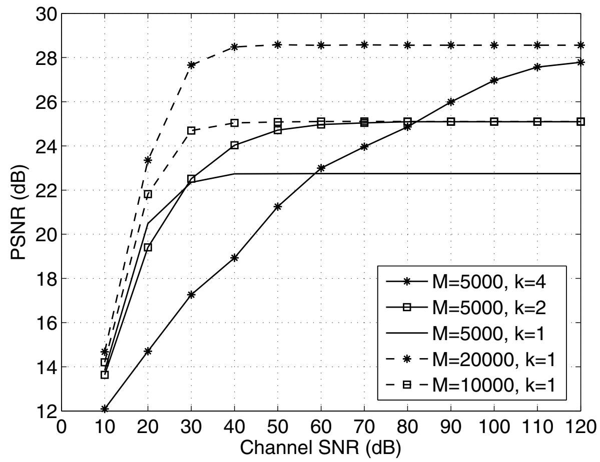


Fig. 2. For the 256×256 image "Boat", PSNR of the reconstructed image as a function of the channel SNR when $M = 5000$ samples are transmitted through an AWGN channel. Notice that for high SNR the proposed system with $k = 4$ and $k = 2$ clearly outperforms the standard CS scheme ($k = 1$). For comparison purposes, the performance of the standard stand-alone CS system ($k = 1$) with $M = 10000$ and $M = 20000$ measurements is also depicted.

in the Daubechies-8 wavelet basis. As explained before and illustrated in Fig. 1(a), we first perform CS to obtain kM measurements, and then we use a $k : 1$ non-linear mapping, so that the number of samples transmitted through the noisy channel, M, is the same as that of the system in Fig. 1(b) based on stand-alone CS ($k = 1$). Two values of k, $k = 2$ and $k = 4$, are considered. For $k = 2$, the applied non-linear mapping is the standard 2:1 spiral curve in (2), with the parameters optimized for each SNR. For $k = 4$, a suboptimal non-linear mapping consisting of the serial concatenation of two 2:1 spiral curves is used, and the curves parameters are optimized globally. In order to reduce the optimization complexity, in all results presented here ML rather than MMSE is considered, which leads to some performance loss.

Fig. 2 presents the Peak Signal to Noise Ratio, PSNR, of the reconstructed image as a function of the channel SNR when $M = 5000$. As expected, the best performance at high SNR is obtained when $k = 4$ (i.e., when the non-linear mapping performs the greatest reduction in the number of samples), and the worst for $k = 1$ (i.e., standard stand-alone CS). On the other hand, the best performance at low SNR is obtained for stand-alone CS, and the use of non-linear mapping in this case results in performance degradation. The reason is that the applied 4:1 mapping, ML decoding, and also the proposed decoding technique (simple serial decoding of the space-filling curve and the CS block) are all suboptimal. The huge advantage of the proposed method in the high SNR region is obvious in Fig. 3, where the reconstructed images corresponding to $k = 1$, $k = 2$ and $k = 4$ are shown when the channel $SNR = 120$ dB.

IV. Conclusion

We have proposed a system based on the encoding of the compressive sensing output with non-linear mappings that act as analog joint source-channel encoders, producing continuous amplitude samples that are transmitted directly through the noisy channel. Simulation results have shown that the proposed framework is readily applicable in practical systems such as imaging, clearly outperforming systems based on stand-alone compressive sensing.

References

[1] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp.118-121, July 2007.

[2] E. Candes, 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. 489-509, February 2006.

[3] 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, August 2006.

[4] E. Candes and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?," *IEEE Trans. on Information Theory*, vol. 52, no. 12, pp. 5406-5425, December 2006.

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

[6] I. Esnaola and J. Garcia-Frias, "Exploiting prior knowledge in the recovery of non-sparse signals from noisy random projections," *Proc. CISS'07*, March 2007.

[7] I. Esnaola and J. Garcia-Frias, "Distributed compression of correlated signals using random projections," *Proc. DCC'08*, March 2008.

[8] J. Garcia-Frias and I. Esnaola, "Exploiting prior knowledge in the recovery of signals from noisy random projections," *Proc. DCC'07*, March 2007.

[9] I. Esnaola and J. Garcia-Frias, "MMSE estimation of distributely coded correlated Gaussian sources using random projections," *Proc. Globecom'08*, December 2008.

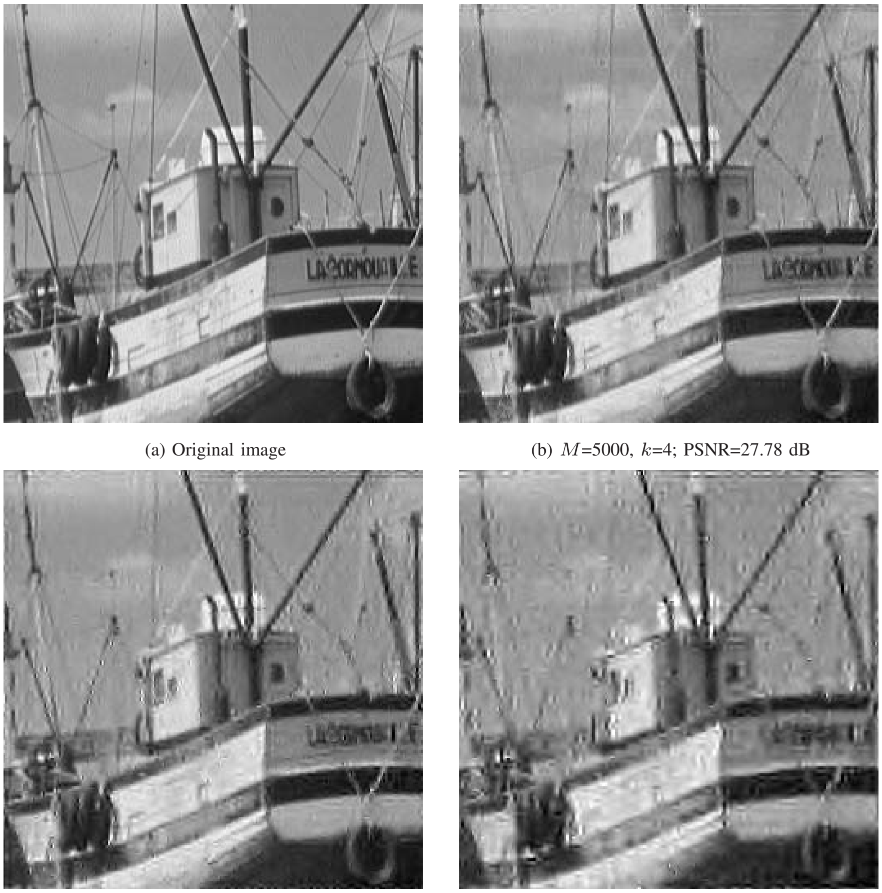
[10] Z. Wang, G. R. Arce, and J. L. Paredes, "Colored random projections for compressed sensing," *Proc. ICASSP'07*, April 2007.

[11] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. on Signal Processing*, vol. 55, no. 12, pp. 5695-5702, December 2007.

[12] K.-H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. on Communications*, vol. 24, no. 12, pp. 1283-1290, December 1976.

[13] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, 27:379-423, 1948.

[14] Z. Wang and G. R. Arce, "Variable density compressed image sampling," *Submitted to IEEE Trans. on Image Processing*.



(a) Original image

(b) M=5000, k=4; PSNR=27.78 dB

(c) M=5000, k=2; PSNR=25.09 dB

(d) M=5000, k=1; PSNR=22.74 dB

Fig. 3. For the original 256×256 image "Boat" represented in (a), reconstructed images corresponding to (b) $k = 4$; (c) $k = 2$; and (d) $k = 1$ (stand-alone CS) when $M = 5000$ samples are transmitted through an AWGN channel and $SNR = 120$ dB.

[15] V. A. Kotel'nikov, "The theory of optimum noise immunity," *New York: McGraw-Hill*, 1959.

[16] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10-21, January 1949.

[17] Y. Hu, J. Garcia-Frias, and M. Lamarca, "Analog joint source-channel coding using space-filling curves and MMSE decoding," *To appear in DCC'09*.

[18] S. S. Chen, D. L. Donoho, and M. A. Saunders,"Atomic decomposition by basis pursuit," *SIAM SCI. Comput.*, vol 20, no. 1, pp. 33-61, August 1998.

[19] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Trans. on Information Theory*, vol. 49, no. 5, pp. 1147-1158, May 2003.

[20] T. Goblick, "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. on Information Theory*, vol. 11, no. 4, pp. 558-567, October 1965.

[21] M. Gastpar, "To code or not to code," *Ph.D. Dissertation, Ecole Polytechnique Federale (EPFL)*, Lausanne, Switzerland, 2002.

[22] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code or not to code," *Proc. ISIT'00*, June 2000.

[23] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," *Information Processing in Sensor Networks*, Springer Berlin, 2003.

[24] Y. Kochman and R. Zamir, "Analog matching of colored sources to colored channels," *Proc. ISIT'06*, July 2006.

[25] A. Fuldseth and T. A. Ramstad, "Bandwidth compression for continuous-amplitude channels based on vector approximation to a continuous subset of the source signal space," *Proc. ICASSP'97*, April 1997.

[26] S. -Y. Chung, "On the construction of some capacity-approaching coding schemes," *Ph.D. Thesis, Massachusetts Institute of Technology*, 2000.

[27] T. A. Ramstad, "Shannon mappings for robust communication," *Telektronikk*, vol. 98, no. 1, pp. 114-128, 2002.

[28] F. Hekland, G. E. Oien, and T. A. Ramstad, "Using 2:1 Shannon mapping for joint source-channel coding," *Proc. DCC'05*, March 2005.