

MMSE DECODING FOR ANALOG JOINT SOURCE CHANNEL CODING USING MONTE CARLO IMPORTANCE SAMPLING

Yichuan Hu⁽¹⁾, Javier Garcia-Frias⁽¹⁾

Meritxell Lamarca^(1,2)

⁽¹⁾Dept. of Elec. and Comp. Engineering
University of Delaware
Newark, DE 19716
E-mail: {yhu, jgarcia}@ee.udel.edu

⁽²⁾Dept. of Signal Theory and Comm.
Technical University of Catalonia (UPC)
Barcelona, Spain
E-mail: xell@gps.tsc.upc.edu

ABSTRACT

We investigate the performance of a discrete-time all-analog-processing joint source-channel coding system for the transmission of i.i.d. Gaussian sources over additive white Gaussian noise (AWGN) channels. At the encoder, samples of an i.i.d. source are grouped and mapped into a channel symbol using a space-filling curve. Different from previous work in the literature, MMSE instead of ML decoding is considered, and we focus on both high and low channel SNR regions. In order to reduce complexity, Monte Carlo importance sampling is applied in the decoding process. The main contribution of this paper is to show that for a wide range of rates the proposed system presents a performance very close to the theoretical limits, even at low SNR, as long as the curve parameters are properly optimized.

1. INTRODUCTION

We consider analog transmission of discrete-time continuous-amplitude sources over AWGN channels. From a theoretical perspective, it is well known that a digital system based on separation between source and channel coding is optimal [1]. Thus, in order to approach the theoretical limit specified by $R_c R(D) < C$, where R_c is the code rate, $R(D)$ is the rate distortion function, and C is the channel capacity, in traditional communications systems the continuous source is first compressed (e.g., using very powerful vector quantization) up to the desired rate/distortion pair. Then, capacity approaching channel codes such as turbo codes or LDPC codes are applied.

Provided that both the source encoder and the channel encoder approach optimality, the separated scheme described above will achieve a performance close to the theoretical limits. However, the price to pay is a very high encoding/decoding complexity and significant delays, since any capacity approaching channel code requires long block lengths. Furthermore,

such separated system has 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.

Interestingly, it is well known that under some circumstances analog communications are optimal. For instance, direct transmission of uncoded Gaussian samples over AWGN channels is optimal [2]. In that sense, it is said that Gaussian sources are perfectly matched to Gaussian channels [3]. Therefore, it is not surprising that previous work in the literature has investigated possible schemes based on analog transformations aiming at perfectly “matching” sources with channels. However, although some promising schemes have been proposed (see [4, 5, 6, 7] among others), 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, already proposed more than 50 years ago by Shannon [8] and Kotel’nikov [9], have recently acquired a renewed importance due to the work of Fuldseth [4], Chung [5], Ramstad [6] and Hekland [7]. The encoding idea 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. Maximum Likelihood (ML) or Minimum Mean Square 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 signal to noise ratios (SNR). In these conditions, it is possible to analyze the system performance, and use this analysis to optimize the curve parameters [5, 7, 10, 11]. 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 much lower than that of a separated scheme. Moreover, since no long blocks are used the delay is very small. However, ML decoding does not perform well for low SNR, and even in the case of high SNR it is not optimal. The case of MMSE decoding has been mentioned in [5] for uniform sources and

This work was supported in part by NSF Award EECs-0725422. The work of M. Lamarca, while on sabbatical leave at the University of Delaware, was partially supported by the Spanish Government under project TEC2007-68094-C02-2.

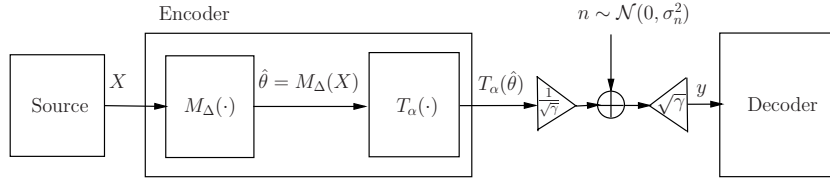


Fig. 1. System model.

Hilbert-like curves. However, in order to analyze/optimize performance in high SNR situations, [5] focuses on sub-optimal MMSE decoding, which results in performance degradation for low SNR. The common understanding about these mappings is that they perform quite well for high SNR, but experience significant degradation for low SNR.

In [12], we proposed a technique to perform quasi-optimal MMSE decoding of i.i.d. Gaussian and Laplacian sources transmitted over AWGN channels, optimizing the curve parameters to obtain a performance that is very close to the theoretical limits in the whole SNR region for 2:1 mappings. The performance improvement with respect to ML was very significant, and, contrary to ML decoding, the gap with respect to the theoretical limit gets smaller when SNR decreases.

In this paper, we extend the framework in [12] to general $n : k$ mappings. In order to reduce decoding complexity and make the proposed system practical, we resort to Monte Carlo techniques. We focus on 3:1 and 4:1 mappings, but the proposed technique can be easily extended to other mappings. The resulting performance is very close to the theoretical limits in the whole SNR region.

2. ENCODER

We consider the transmission of i.i.d. Gaussian sources over average power constrained AWGN channels. Without loss of generality, we assume the source produces Gaussian samples with zero mean and unit variance, and the average power for transmitting one channel symbol is constrained to 1. The key idea in the encoder is to group the source symbols into vectors X of dimension n , and to represent the vector X as a point in an n -dimensional space where a space-filling curve of dimension k lives. For instance, when $k = 1$, a single value, corresponding to the length of the fragment of the curve located between the origin and the point in the curve that is closest to X , is transmitted over the channel. As a result, $n : k$ bandwidth reduction is obtained by the above $n : k$ mapping.

In this paper, we focus on the 3:1 and 4:1 mappings proposed (for ML decoding) in [11, 13]:

$$\begin{cases} x_1 = \text{sign}(\theta) \frac{\Delta}{\pi} \theta \cos \frac{\theta}{2\pi} \sin \theta \\ x_2 = \frac{\Delta}{\pi} \theta \sin \frac{\theta}{2\pi} \sin \theta \\ x_3 = \text{sign}(\theta) \frac{\Delta}{\pi} \theta \cos \theta \end{cases} \quad \text{for } \theta \in \mathbb{R}. \quad (1)$$

$$\begin{cases} x_1 = \text{sign}(\theta) \frac{\Delta}{\pi} \theta \sin \frac{\theta}{3\pi} \cos \theta \\ x_2 = \text{sign}(\theta) \frac{\Delta}{\pi} \theta \cos \frac{\theta}{2\pi} \sin \theta \\ x_3 = \frac{\Delta}{\pi} \theta \sin \frac{\theta}{2\pi} \sin \theta \\ x_4 = \frac{\Delta}{\pi} \theta \cos \frac{\theta}{3\pi} \cos \theta \end{cases} \quad \text{for } \theta \in \mathbb{R}. \quad (2)$$

Notice that the curve gradually fills in the whole space as the absolute value of θ grows. Therefore, if Δ is known, we can use a mapping function, $M_\Delta(\cdot)$, to project X onto the curve by finding the closest point on the curve, which we will denote as $\hat{\theta} = M_\Delta(X)$. Then, an invertible function of $\hat{\theta}$, normalized so that the average transmitted power per sample is 1, is transmitted through the AWGN channel. The reason to introduce this function is to make the transmitted symbol follow a Gaussian distribution, which is necessary to approach channel capacity [14]. Specifically, the symbols transmitted through the channel are defined as $T_\alpha(\hat{\theta}) = \hat{\theta}^\alpha$ where $\alpha \in (0, 2]$ is a parameter that has to be optimized for the channel SNR (defined as $SNR = 10 \log \frac{1}{\sigma_n^2}$) of interest, and γ is a normalization factor such that $T_\alpha(M_\Delta(X))/\sqrt{\gamma}$ satisfies the average power constraint.

In sum, the received observation at the decoder can be expressed (see Figure 1) as $y = T_\alpha(M_\Delta(X)) + \sqrt{\gamma}n$, where $n \sim \mathcal{N}(0, \sigma_n^2)$ is additive white Gaussian noise.

3. DECODER

3.1. ML Decoding

Given a received symbol y , the ML estimate is obtained as the vector \hat{X}_{ML} belonging to the curve and satisfying

$$\begin{aligned} \hat{X}_{ML} &= \arg \max_{X \in \text{curve}} \{p(y|X)\} \\ &= \{X | X \in \text{curve and } T_\alpha(M_\Delta(X)) = y\}. \end{aligned} \quad (3)$$

ML decoding is equivalent to first applying the inverse function $T_\alpha^{-1}(\cdot)$ to received symbol y , $\hat{\theta}' = T_\alpha^{-1}(y) = y^{-\alpha}$, and then performing inverse mapping on $\hat{\theta}'$ according to (1) or (2).

3.2. MMSE Decoding

Although ML decoding is simple, it is not optimal for the mean square error distortion criterion (MSE). In this case, MMSE decoding is optimal, and can be expressed as

$$\begin{aligned}\hat{X}_{MMSE} &= E\{X|y\} = \int Xp(X|y)dX \\ &= \frac{1}{p(y)} \int Xp(y|X)p(X)dX.\end{aligned}\quad (4)$$

Since the conditional probability $p(y|X)$ involves the mapping function $M_{\Delta}(\cdot)$, which is discontinuous and highly non-linear, (4) can not be expressed in a closed form. In [12], for the case of 2:1 mapping, calculation of the above integral was simplified to only multiplication and addition operations by discretizing the vector X with a uniform step. However, the computational complexity of this method increases exponentially with the dimension of X , and, therefore, it cannot be applied to 3:1 and 4:1 mappings.

Monte Carlo (MC) techniques [15, 16] provide an alternative way to calculate (4). Specifically, we can write (4) as

$$\hat{X}_{MMSE} = \frac{\int Xp(y|X)p(X)dX}{\int p(y|X)p(X)dX} = \frac{E_{p(X)}[Xp(y|X)]}{E_{p(X)}[p(y|X)]}.\quad (5)$$

The basic idea is that by the Strong Law of Large Numbers the sample mean converges almost surely to the true mean value. Therefore, by generating M samples $X_1 \dots X_M$, where $X_i \sim \mathcal{N}(0, I)$, (5) can be approximated by

$$\hat{X}'_{MMSE} = \frac{\sum_{i=1}^M X_i p(y|X_i)}{\sum_{i=1}^M p(y|X_i)}.\quad (6)$$

Obviously, the numerator and denominator of (6) are unbiased estimators of $E_{p(X)}[Xp(y|X)]$ and $E_{p(X)}[p(y|X)]$, respectively. Another remarkable property of these estimators is that due to the Central Limit Theorem their variances are independent of the dimension of vector X [15, 16]. In other words, their variances only depend on the number of samples M and on the distribution chosen to generate these samples. Notice, however, that sampling from $p(X)$ is not a good choice. For example, if \hat{X}_{MMSE} is a non-zero vector and $p(X|y)$ only has non-zero values in a small region around \hat{X}_{MMSE} , then most of the samples X_i generated according to $p(X)$ ($\mathcal{N}(0, I)$) will contribute very little to (6), since $p(y|X_i)$ will be close to 0. Thus, the convergence speed of (6) may be very slow if a bad sample distribution is used.

The alternative to directly sampling from $p(X)$ for the evaluation of (5) is to use importance sampling. By sampling $X_1 \dots X_M$ from another distribution $q(X)$, called importance distribution [15], (5) can be estimated by

$$\hat{X}''_{MMSE} = \frac{\sum_{i=1}^M \frac{X_i p(y|X_i) p(X_i)}{q(X_i)}}{\sum_{i=1}^M \frac{p(y|X_i) p(X_i)}{q(X_i)}},\quad (7)$$

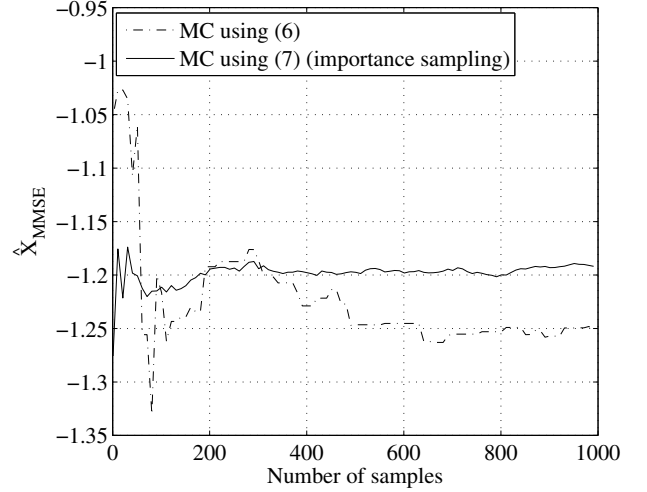


Fig. 2. Convergence of the estimators obtained using Monte Carlo techniques as described in (6) and (7). The first component of vector \hat{X}_{MMSE} is shown.

since (5) can be written as

$$\hat{X}_{MMSE} = \frac{\int \frac{Xp(y|X)p(X)}{q(X)} q(X)dX}{\int \frac{p(y|X)p(X)}{q(X)} q(X)dX} = \frac{E_{q(X)}[\frac{Xp(y|X)p(X)}{q(X)}]}{E_{q(X)}[\frac{p(y|X)p(X)}{q(X)}]}.\quad (8)$$

An interesting problem arises here: how to choose a proper $q(X)$ to improve the convergence speed? In theory, the optimal $q(X)$ (in the sense of minimizing the variance of the estimator) should be proportional to $|X|p(X|y)$ [16]. However, this is not applicable in practice because $p(X|y)$ can not be easily calculated. Instead, since the results from ML decoding provide some information about the location of \hat{X}_{MMSE} , we can generate samples around \hat{X}_{ML} by letting $q(X)$ be a normal distribution with mean vector \hat{X}_{ML} and covariance matrix $(\frac{\Delta}{2})^2 \cdot I$. Figure 2 shows the convergence of estimators \hat{X}'_{MMSE} (6) and \hat{X}''_{MMSE} (7). By using importance sampling from $q(X)$, the estimator \hat{X}''_{MMSE} converges after 300 samples. On the contrary, for \hat{X}'_{MMSE} , direct sampling from $p(X)$ leads to very unstable results (i.e., oscillations) even if more than 1000 samples are used. Therefore, in all our simulation results only importance sampling (with 300 samples) will be considered.

4. SIMULATION RESULTS

In our first set of simulations, we evaluate the performance of the 3:1 non-linear mapping defined in (1) for the transmission of Gaussian sources over AWGN channels. The performance is measured in terms of output SDR (defined as

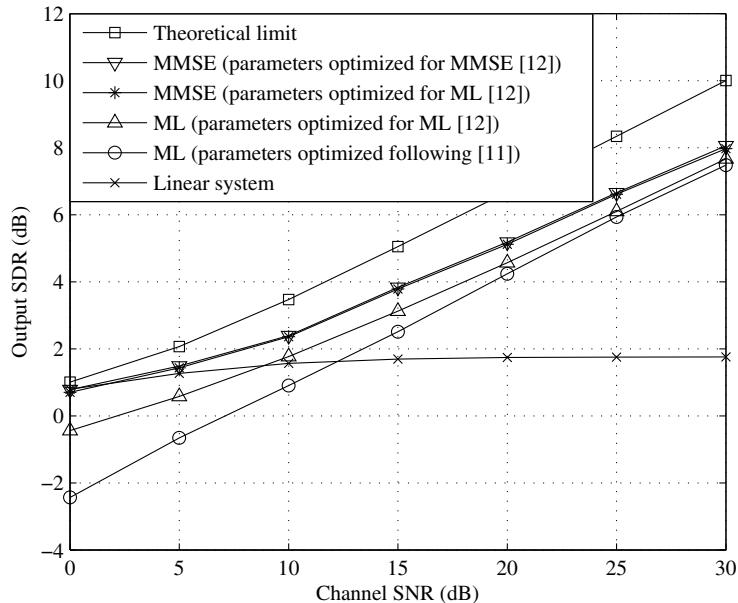


Fig. 3. Performance of the 3:1 non-linear mapping for the transmission of i.i.d. Gaussian sources over AWGN channels. For comparison purposes, the performance of a linear system is also depicted.

$SDR = 10 \log \frac{1}{MSE}$) versus channel SNR, and is compared with the theoretical limit, obtained by equating the channel capacity and the rate distortion function for i.i.d. Gaussian sources ($SDR = (1 + SNR)^{\frac{k}{n}}$). Figure 3 shows the results when MMSE and ML decoding are applied and the curve parameters, α and Δ , are optimized using two criteria: i) α always equals to 2 and Δ calculated based on the high SNR analysis for ML decoding developed in [11], ii) α and Δ jointly optimized by the technique proposed in [12]. The worst performance is obtained when ML decoding is applied and the curve parameters are optimized following i). When parameters are optimized using ii) (still with ML decoding), we can observe a performance gain in output SDR of around 0.9 - 2 dB for low SNR ($SNR \in [0, 10]$ dB). In this low SNR region, additional gains of around 0.6 - 1.1 dB are obtained if MMSE is used instead of ML decoding. In this case, Figure 3 shows that parameter optimization is less critical than in ML decoding, and the best performance is obtained by applying MMSE decoding and optimizing the parameters for this case. It is remarkable that in this situation the gap with respect to the theoretical limit is less than 1.9 dB for all values of SNR. Moreover, the gap gets smaller as the channel SNR decreases. This occurs because parameter optimization for low SNR makes the space-filling curve approximate a linear system, which is optimal when the channel SNR goes to $-\infty$ [17].

We next consider the transmission of Gaussian sources over AWGN channels using the 4:1 non-linear mapping defined in (2). Figure 4 presents the resulting performance when

either ML or MMSE decoding is applied. Similar to the case of 3:1 mapping, for low SNR 1.2 - 1.7 dB improvement in output SDR is obtained by parameter optimization when ML decoding is performed. When MMSE decoding is applied, additional gains of 0.5 - 0.7 dB and 0.3 - 0.5 dB can be obtained in low and high SNR regions, respectively, so that the gap to the theoretical limit is below 1.9 dB for the whole SNR region.

5. CONCLUSION

We have presented a discrete-time all-analog-processing joint source-channel coding scheme, based on the use of spiral-like curves, achieving 3:1 and 4:1 bandwidth reduction for the transmission of i.i.d. Gaussian sources over AWGN channels. In order to reduce the computational complexity, Monte Carlo importance sampling is applied to perform MMSE decoding. The proposed system outperforms previous existing schemes in the literature, with a gap with respect to the theoretical limits below 1.9 dB for the whole SNR range. Although we have presented these techniques for the cases of 3:1 and 4:1 bandwidth reduction, they can easily be applied for other reduction factors.

6. REFERENCES

- [1] C. E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, pp. 379-423, 1948.

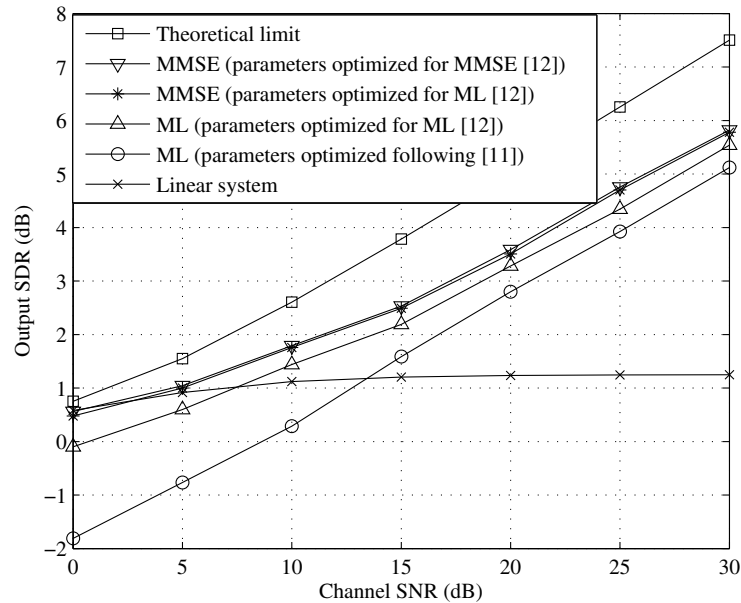


Fig. 4. Performance of the 4:1 non-linear mapping for the transmission of i.i.d. Gaussian sources over AWGN channels. For comparison purposes, the performance of a linear system is also depicted.

- [2] T. J. 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.
- [3] 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.
- [4] 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.
- [5] S. -Y. Chung, "On The Construction of Some Capacity-Approaching Coding Schemes," *Ph.D. dissertation, Massachusetts Institute of Technology*, 2000.
- [6] T. A. Ramstad, "Shannon Mappings for Robust Communication," *Teletronikk*, vol. 98, no. 1, pp. 114-128, 2002.
- [7] F. Hekland, G. E. Oien, and T. A. Ramstad, "Using 2:1 Shannon Mapping for Joint Source-Channel Coding," *Proc. DCC'05*, March 2005.
- [8] C. E. Shannon, "Communication in The Presence of Noise," *Proc. IRE*, vol. 37, pp. 10-21, January 1949.
- [9] V. A. Kotel'nikov, *The Theory of Optimum Noise Immunity*, New York: McGraw-Hill, 1959.
- [10] F. Hekland, "On The Design and Analysis of Shannon-Kotel'nikov Mappings for Joint Source-Channel Coding," *Ph.D. dissertation, Norwegian University of Science and Technology*, 2007.
- [11] P. A. Floor, "On The Theory of Shannon-Kotel'nikov Mappings in Joint Source-Channel Coding," *Ph.D. dissertation, Norwegian University of Science and Technology*, 2008.
- [12] Y. Hu, J. Garcia-Frias, and M. Lamarca, "Analog Joint Source-Channel Coding Using Space-Filling Curves and MMSE Decoding," *Proc. DCC'09*, March 2009.
- [13] P. A. Floor and T. A. Ramstad, "Dimension Reducing Mappings in Joint Source-Channel Coding," *Proc. NORSIG*, June 2006.
- [14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 2006.
- [15] A. Doucet and X. Wang, "Monte Carlo Methods for Signal Processing: A Review in The Statistical Signal Processing Context," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 152-170, November 2005.
- [16] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, New York: Springer-Verlag, 2000.
- [17] 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.