Using 2:1 Shannon Mapping for Joint Source-Channel Coding[1]

Fredrik Hekland   Geir E. Øien   Tor A. Ramstad

Department of Electronics and Telecommunications
Norwegian University of Science and Technology
O.S. Bragstadsplass 2B, 7491 Trondheim, Norway.
Tel./Fax: +47-735-94496/92640
E-mail: {hekland,oien,ramstad}@iet.ntnu.no

Abstract

The Archimedes' spiral can be used as a 2:1 bandwidth reducing mapping in a joint source-channel coding (JSCC) system. The combined point of two *iid* Gaussian sources (the source space) is mapped, or approximated, onto a double Archimedes' spiral (the codebook), and the squared angle from the origin to the mapped point is transmitted as an analogue channel symbol (the channel space), e.g. PAM. It is shown that the total distortion of this JSCC system is minimised when the distortion contributions from the approximation noise and channel noise are equal. The given system produces a channel input distribution close to a Laplacian probability density function (pdf) instead of the optimal Gaussian pdf. The loss when using this mismatched pdf is shown to be approximately equal to the relative entropy of the two pdf's.

1 Introduction

Shannon's separation theorem [7, Theorem 21] is a very powerful tool for designing communication system. This theorem states that for a stationary source, the source and channel coding can be separately designed without any loss of performance. However, the two coders must have infinite delay and complexity to achieve this. Joint source-channel coding (JSCC) may reduce these problems and achieve optimal performance in the case when having delay and complexity constraints. There are several ways to design a JSCC system, but most JSCC approaches utilise the same components as the traditional tandem systems, and optimise them jointly with respect to the current channel conditions. We have chosen a more radical approach by completely merging the source coder, channel coder into one joint processing operation. It is worth noting that this also corresponds to the classical block diagram, Fig. 1 in Shannon's 1948 paper [7] where no source-channel coder separation is implied. The joint source-channel coding is achieved by mapping source points directly onto the channel space, with no explicit channel coding. This mapping implies that continuous amplitude symbols are used on the channel. This way one might obtain higher spectral efficiency, whereas the properties of the mappings could also be designed to ensure robustness to channel errors. Moreover, the delay is low since the channel symbols are memoryless. A discussion on desirable properties of mappings can be found in Ramstad's paper [6]. Since Shannon proposed mappings like these in his 1949 paper [8], we have chosen to name these mappings Shannon Mappings [6].

[1]This work was supported by the Research Council of Norway under the project BEATS (URL: http://www.tele.ntnu.no/projects/beats).

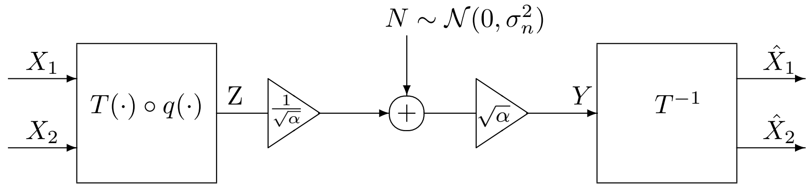

Figure 1: A generic 2 : 1 dimension (or bandwidth) reducing Shannon mapping. The $q(\cdot)$ operator projects the two-dimensional source tuple onto a point on a one dimensional curve in $\mathbb{R}^2$. This operation is not invertible when the curve has finite length. The $T(\cdot)$ operator is an invertible one-to-one mapping between the curve and the channel space. The scaling factor ensures that the power-constraint is fulfilled.

In this article, one particular mapping is analysed; a 2:1 dimension-reducing mapping where the channel space is restricted to a doubly intertwined Archimedes' spiral [9]. The reason for using this particular function is its similarity with the power-constrained channel-optimised vector quantiser codebook (PCCOVQ) developed by Fuldseth [4]. When a 2:1 PCCOVQ is trained for Gaussian sources over AWGN channels, spiral-like codebooks emerge. Having an analytical approximation of this PCCOVQ component enables us to analyse its performance, and compare it to an ideal mapping which attains OPTA[2]. Some results on the Archimedes' spiral have been given by Chung [1] and agree with the analytical results to be developed in Section 3. The remainder of this paper is organised as follows: First the optimal spiral parameters are determined, given source and channel statistics and a power-constraint. Having these parameters, we can determine the optimal balance between the distortion contributions from approximation noise and channel noise. Finally, the loss incurred by having a non-Gaussian channel input distribution on the AWGN channel is analysed.

2 Preliminaries

We are given two independent identically distributed (*iid*) Gaussian sources X_1 and X_2 with zero mean. This can be represented as a source vector $\mathbf{x} = (x_1, x_2)$. We want to perform a bandwidth reduction by transmitting a combination of two source samples as one channel sample. To achieve this we perform a mapping operation from the two dimensional source space onto a subspace which must consist of line segments. This is represented by the $q(\cdot)$ in Fig. 2, where any point in $\mathbb{R}^2$ is mapped to the closest point on any of the two spirals. In our case, the subspace is given as the double Archimedes' spiral, defined by the radii

$$r_+(\theta) = \frac{\Delta}{\pi}\theta \quad \text{and} \quad r_-(\theta) = \frac{\Delta}{\pi}(\theta + \pi), \tag{1}$$

for the solid and dashed spiral arms in Fig. 2 respectively. Here, Δ is the distance to neighbouring spiral arms in Fig. 2 and $\theta \in \mathbb{R}$ is the angle from the origin to a point on the double spiral. First, a source vector is approximated, or projected, onto the closest point on the spiral. In Fig. 2 this can be seen as the star being

[2]Optimal Performance Theoretically Attainable. Results from equating the expression for channel capacity and the rate-distortion function, and solving for signal to noise ratio.

approximated down to the circle on the spiral. The approximated source vector is still two-dimensional, but is restricted to the Archimedes' spirals. Then we perform a mapping using the invertible operator $T(\cdot)$ in Fig. 2 from the $2D$ subspace to a $1D$ channel representation. In this case, we us the square of the angle,

$$T_{+}(\theta) = +a\theta^2 \quad \text{and} \quad T_{-}(\theta) = -a\theta^2 \tag{2}$$

for the solid and dashed spiral arms respectively in Fig. 2. The parameter $a = 0.16\Delta$ makes this T-operator an approximation of the length along the spiral. This is then scaled to satisfy the power constraint and sent as an analogue symbol on the channel.

The optimal Δ depends on the channel noise and the transmit power constraint, and is the parameter to be determined. In the receiver, the square-root of the received symbol is taken. This indicates signal dependency of the channel noise. However, simulations suggests that in the high-CSNR region this effect is minimal, so we disregard this factor in the following.

When optimising the spiral mapping, the goal is to find the Δ that minimises the total distortion, given an average power-constraint P_{avg}. That is,

$$\Delta_{opt} = \underset{\Delta : E[z^2] \leq P_{avg}}{\arg\min} \left[D(\Delta) \right], \tag{3}$$

where $D(\Delta)$ is the resulting distortion after reception when using Δ as the spiral arm distance, and $E\left[z^2\right]$ is the average channel symbol power. The power-constraint will limit the range of θ, so we have $\theta \in \mathcal{S} \subset \mathbb{R}$.

We define the resulting distortion per source symbol after decoding at the receiver side as

$$D_{total} \overset{def}{=} \frac{E\left[\|\mathbf{x} - \hat{\mathbf{x}}\|^2\right]}{2},$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the original and decoded source vectors respectively. The denominator accounts for the fact that the distortion is distributed on two source symbols. As in the thesis by Chung [1] we decompose the error into a radial component resulting from the approximation operation $q(\cdot)$ in Fig. 2 and an angular component from channel noise N:

$$D_{total} = D_r + D_\theta = \frac{1}{2} E\left[\left(\|\mathbf{x}\| - \|\hat{\mathbf{x}}\|\right)^2\right] + \frac{1}{2} E\left[\|\hat{\mathbf{x}}\|^2 \left(\angle\mathbf{x} - \angle\hat{\mathbf{x}}\right)^2\right]. \tag{4}$$

This decomposition is possible under the assumption of high CSNR when the spiral is sufficiently dense at the origin, so that the average error is approximately equal in the two source samples after projecting the combined source point onto the spiral. The radial distortion will increase with increasing Δ, which is obvious since the distance to the closest spiral arm increases. The angular distortion will, however, increase with decreasing Δ. This is because the source will have to be down-scaled as a result of the power-constraint. The corresponding up-scaling at the receiver end will then amplify both signal and noise.

3 Balancing the Distortion Contributions

Since we allow noise from both the approximation step and the channel to affect the source signals, we want to determine the optimal balance between these two noise

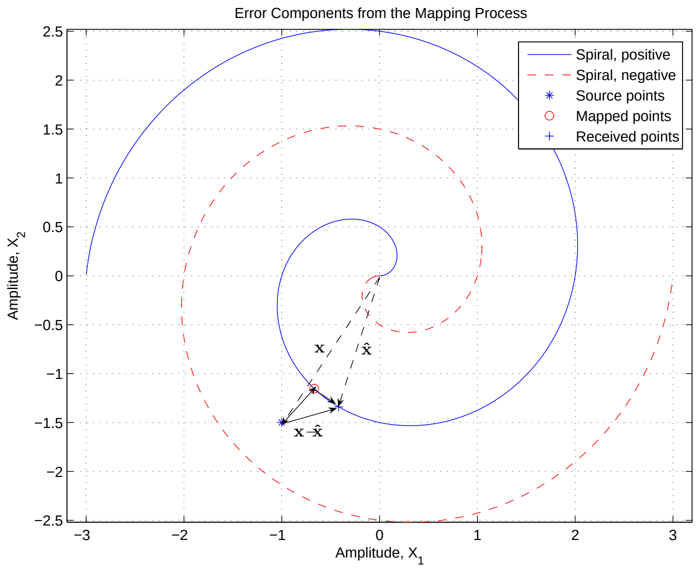

Figure 2: Indicates the error components from the mapping process. The total error $\|\mathbf{x} - \hat{\mathbf{x}}\|$ can be decomposed into two orthogonal components: A radial component, $\|\mathbf{x}\| - \|\hat{\mathbf{x}}\|$, from the approximation $q(\cdot)$ in Fig. 2. An angular component, $\|\hat{\mathbf{x}}\|\,(\angle\mathbf{x} - \angle\hat{\mathbf{x}})$, from the channel noise which moves the mapped point along the spiral.

contributions. That is, we want to determine the optimal Δ which minimises the total distortion. Then we will show that this implies $D_r = D_\theta$.

We need an expression for the total distortion, containing both distortion components and then we must minimise this under the given the power constraint. This expression is

$$D_{total}(\Delta, \sigma_n) = D_r(\Delta) + D_\theta(\Delta, \sigma_n), \tag{5}$$

The power constraint in (3) implies that we have to upscale the channel input signal with a factor $1/\sqrt{\alpha}$ in order to satisfy the average power constraint $E[(\frac{z}{\sqrt{\alpha}})^2] = \frac{1}{\alpha}E[z^2] = \frac{\sigma_z^2}{\alpha}$, where $\alpha = \sigma_z^2/P_{avg}$. The received channel signal must be up-scaled accordingly to obtain the original source power. Even though the inverse of the scaling parameter on the transmitter side is sub-optimal compared to a Wiener-type scaling factor, we use $\sqrt{\alpha}$ for simplicity. The scaling on the receiver side implies that the channel noise will also be scaled:

$$\text{Var}\left[\sqrt{\alpha}N\right] = \alpha\text{Var}\left[N\right] = \alpha\sigma_n^2. \tag{6}$$

For the two distortion terms in (5), we introduce some approximations valid in the high CSNR case (i.e. dense spiral). First, in the radial direction the approximation operation can be regarded as a standard scalar quantizer (except near the origin). Therefore we introduce the expression for the quantization error in a scalar quantizer, $\sigma_q^2 = \Delta^2/12$. With regard to the problems around the origin, it seems reasonable to assume that with a sufficiently dense spiral the approximation should still be valid. For the distortion in the angular direction, the scaled noise variance in (6) is used.

Inserting into (5) we obtain

$$D_{total}(\Delta,\sigma_n)=\frac{1}{2}\left(\frac{\Delta^2}{12}\right)+\frac{1}{2}\left(\alpha\sigma_n^2\right)=\frac{1}{2}\left(\frac{\Delta^2}{12}\right)+\frac{1}{2}\left(\frac{\sigma_z^2}{P_{avg}}\sigma_n^2\right) \tag{7}$$

where σ_z^2 is a function of Δ. To determine σ_z^2, we assume that in the high-CSNR region the spiral is sufficiently dense to disregard the approximation step. This means we can calculate the density on a circle around the origin. This is done by evaluating $Z=g(X,Y)=\pm a\theta^2=\pm a(X^2+Y^2)$, from the expression of the radius (1). For a $2D$ Gaussian circular symmetric source, this gives a Laplace distribution with variance $\sigma_z^2=2\lambda^2=2(2a\frac{\pi^2}{\Delta^2}\sigma_x^2)^2$, where σ_x^2 is the variance of the source, and a equals 0.16Δ as in (2). Inserting into (7) we have

$$D_{total}(\Delta,\sigma_n)=\frac{\Delta^2}{24}+\frac{8(0.16\pi^2\sigma_x^2\sigma_n)^2}{\Delta^2\cdot 2P_{avg}}. \tag{8}$$

We thus see, as stated above, that the radial distortion will decrease with smaller Δ whereas the angular distortion will increase with smaller Δ. Therefore we will have a distinct Δ which provides the minimal distortion for the given noise variance and power constraint. To find this minimum we differentiate D_{total} with respect to Δ and equate the resulting expression to zero.

$$\frac{\mathrm{d}D_{total}}{\mathrm{d}\Delta}=\frac{\Delta}{12}-\frac{8(0.16\pi^2\sigma_x^2\sigma_n)^2}{\Delta^3 P_{avg}}=0. \tag{9}$$

Solving this expression for Δ we obtain

$$\Delta=2\pi\sigma_x\sqrt[4]{\frac{6\cdot 0.16^2\sigma_n^2}{P_{avg}}}=2\pi\sigma_x\sqrt[4]{\frac{6\cdot 0.16^2}{CSNR}}. \tag{10}$$

Using the distortion expression in (8), and inserting (10) into (8), we find that minimal total distortion, both distortion contributions are

$$D_r=D_\theta=\frac{\Delta^2}{24}=\pi^2\sigma_x^2\sqrt{\frac{0.16^2}{6\cdot CSNR}}. \tag{11}$$

This implies that the total distortion is minimised when the approximation and channel noise are equal. The SNR is defined as

$$\begin{aligned} SNR=\frac{\sigma_x^2}{D_{total}}&=\frac{\sigma_x^2}{\Delta^2/12}=\frac{\sqrt{6}}{2\cdot 0.16\cdot\pi\cdot\pi}\sqrt{CSNR} \\ &\approx\frac{\sqrt{6}}{\pi}\sqrt{CSNR}, \end{aligned} \tag{12}$$

and comparing this to OPTA for the $2:1$ case:

$$SNR=\sqrt{1+CSNR}, \tag{13}$$

we see that the asymptotical gap from OPTA is $\sqrt{6}/\pi \approx 1.1$dB, which agrees to the findings of Chung [1]. In other words, using an Archimedes' spiral as mapping device, having a power constraint on the channel, and operating in the high-CSNR region, the total distortion will be minimised when the approximation noise and channel noise, D_r and D_θ, are equal. Each of the distortion contributions in (4) is $\Delta^2/24$, so the total distortion is $\Delta^2/12$ per source component, and the distance to the theoretical optimum is around 1.1 dB.

A comparison between this optimised spiral mapping, OPTA for 2 : 1 bandwidth reduction, and the ideal entropy-coded uniform scalar quantizer with perfect channel coder can be seen in Fig. 3. It is observed that the proposed system outperforms the scalar quantizer in the high-CSNR region where the assumptions hold, whereas below 20-25 dB CSNR, the assumptions in Section 3 are no longer valid, causing higher distortion. It is stressed that in contrast to the ideal entropy coded scalar quantizer with a perfect channel, the proposed system is memoryless and does not attain the theoretical channel capacity. Furthermore, it is very robust in the sense that incorrect channel-state information (CSI) in the receiver will not cause a breakdown in the decoder. The systems shows both graceful degradation and improvement, as can be seen in the two dashed curves in Fig. 3.

4 Mismatch of the Channel Input Distribution

It is well known that in order to achieve capacity on an AWGN channel, the channel input must follow a Gaussian distribution [2]. As soon as the channel input differs from the Gaussian, we are not fully utilising the channel. Thus a certain loss in the transmission is to be expected. We know that Gaussian input to a linear system yields Gaussian output. This is, however, not in general the case for non-linear systems, such as the Archimedes' spiral and quadratic channel representation. For the case of a circular symmetric Gaussian source, and using the squared angle along the spiral arm as channel representation, the channel signal in fact follows a Laplace distribution. In the following section, the loss incurred from using a non-Gaussian channel input distribution is calculated. The channel capacity is given by the well-known formula [2]

$$C = \max_{f(z):E[z^2]\leq P} I(Z;Y), \tag{14}$$

where $I(X;Y)$ denotes mutual information. For a Gaussian channel, the capacity is achieved if the channel input distribution also is Gaussian. In our case, however, we will not attain the theoretical channel capacity since the channel input distribution is approximately Laplacian:

$$f_{\mathbf{y}}(y) = \frac{1}{2\lambda}\mathrm{e}^{-|y|/\lambda}. \tag{15}$$

The best we can achieve in this case is the channel capacity *minus* a constant, which is determined by the misfit of the Laplace distribution.

$$C_z = I(X;Y)\big|_{f_{\mathbf{x}}(x)=f_{\mathbf{z}}(z)} = C_{AWGN} - k. \tag{16}$$

Our goal is to quantify the loss k resulting from using a Laplace distributed channel input on a Gaussian channel.

First, the definition of mutual information is

$$I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(N), \tag{17}$$

where the last equality comes from the fact that signal and channel noise are independent [2]. Further, we expand the differential entropy of the received signal as (shortening the notation since we always integrate over y)

$$\begin{aligned} h(Y) &= -\int f \ln f = -\int f \ln\left(f\frac{f^*}{f^*}\right) = -\int f \ln f^* - \int f \ln\left(\frac{f}{f^*}\right) \\ &\overset{(a)}{=} -\int f^* \ln f^* - \int f \ln\left(\frac{f}{f^*}\right) \\ &= h(Y^*) - D\left(f\|f^*\right), \end{aligned} \tag{18}$$

where f^* is the Gaussian pdf and f is given by (15), and (a) is valid when both distributions have zero mean and equal variance, that is $\mu_{Gauss} = \mu_{Laplace} = 0$, and $\sigma^2_{Gauss} = \sigma^2_{Laplace}$, since then

$$\begin{aligned} -\int f \ln f^* &= -\int f \ln\left(\frac{1}{\sqrt{2\pi}\sigma_G}\mathrm{e}^{\frac{-(y^2)}{2\sigma_G^2}}\right) = \frac{1}{2\sigma_G^2}\int y^2 f + \ln\left(\sqrt{2\pi}\sigma_G\right)\int f \\ &= \frac{1}{2\sigma_G^2}\int y^2 f^* + \ln\left(\sqrt{2\pi}\sigma_G\right)\int f^* = -\int f^* \ln f^*. \end{aligned}$$

We have now come to the second result in this paper. Inserting (17) and (18) into (16), we obtain

$$C_z = h(Y^*) - h(N) - D\left(f\|f^*\right) = C_{AWGN} - D\left(f\|f^*\right). \tag{19}$$

We see that the price we have to pay for using the incorrect distribution is the relative entropy between the distribution in use and the optimal distribution. Interestingly, this is similar to what Gray and Linder [5] found when designing a variable-rate vector quantizer for a source probability density function (pdf) g and applying a source with a different pdf f. There, the loss in performance was equal to the relative entropy $D(f\|g)$. The loss can be found explicitly by inserting the pdf for a Gaussian distribution and (15) into the last term in (19), and after some algebraic manipulation we obtain

$$\begin{aligned} D\left(f\|f^*\right) &= \int f \ln\frac{f}{f^*} = -h(f) - \int f \ln f^* \\ &= -h(f) + \int f \ln\left(\sqrt{2\pi}\sigma_G\right) + \int f \frac{y^2}{2\sigma_G^2} \\ &= -h(f) + \ln\left(\sqrt{2\pi}\sigma_G\right) + \frac{1}{2\sigma_G^2}\int y^2 f \\ &\overset{(b)}{=} -\ln 2\lambda - 1 + \ln\left(\sqrt{2\pi}\sigma_G\right) + 1/2 \\ &= \ln\left(\frac{\sqrt{2\pi}\sigma_G}{2\lambda\sqrt{\mathrm{e}}}\right). \end{aligned} \tag{20}$$

In (b), $h(f)$ is given from [2], and to have equal power (or variance, since the mean, $\mu = 0$) in the two distributions, we set $\int y^2 f = \sigma_L^2 = 2\lambda^2 = \sigma_G^2 \Rightarrow \sigma_G = \sqrt{2}\lambda$. Inserting this into (20) we finally obtain

$$D\left(f \| f^*\right) = \ln\left(\sqrt{\frac{\pi}{\mathrm{e}}}\right) = \frac{1}{2}\ln\left(\frac{\pi}{e}\right) \approx 0.072 \text{ nats} = 0.104 \text{ bits}. \tag{21}$$

This indicates that the loss induced by the misfitted channel input distribution is independent of the CSNR, and is comparable to the shape gain [3] obtainable for digital non-binary signal alphabets. This can also be seen by looking at the high-CSNR region in Fig. 3 where the loss is constant for higher CSNR. In the lower CSNR region, the assumptions for the calculated spiral arm distance Δ are no longer exact, and we no longer have optimal system parameters. Still, the curve of the received signal is still almost parallel to the optimum performance theoretically attainable (OPTA). Thus the loss is independent of the CSNR in the regions where the assumptions hold.

If we utilise the scalar quantizer measure $SNR = 6.02R + Constant$ and insert the value from (21), the loss in terms of SNR for the one-dimensional channel symbol is approximately 0.6 dB. However, since we have a 2 : 1 compression, i.e. each channel symbol carries two source symbols, this number must be multiplied by two. Then the total loss for the two sources is around 1.2 dB which is more or less the distance to OPTA in the high-CSNR region in Fig. 3.

5 Conclusion

In this paper we have examined a joint source-channel coding system which provides a 2 : 1 bandwidth reduction by transmitting two source samples on one channel symbol. Since we have a power-constraint, this dimension reduction is bound to introduce some approximation noise into the sources. Moreover, the channel symbols are discrete time, but continuous amplitude. This implies that we have added channel noise. It was shown that the total distortion is minimised when the approximation noise and channel noise are equal. For high CSNR, the spiral mapping outperforms an ideal entropy-coded uniform scalar quantizer even though the latter has an ideal channel, and asymptotically the distance to OPTA is around 1.1dB. It could be mentioned that numerically optimised vector quantizers like the PCCOVQ [4] performs slightly better. Furthermore, the spiral mapping proves to be very robust in the sense that it shows both graceful degradation and graceful improvement in the case of non-perfect channel-state information (CSI). Finally, it was shown that due to a mismatched channel input distribution, the system experiences a performance loss slightly larger than the relative entropy between the actual and the optimal channel input distribution.

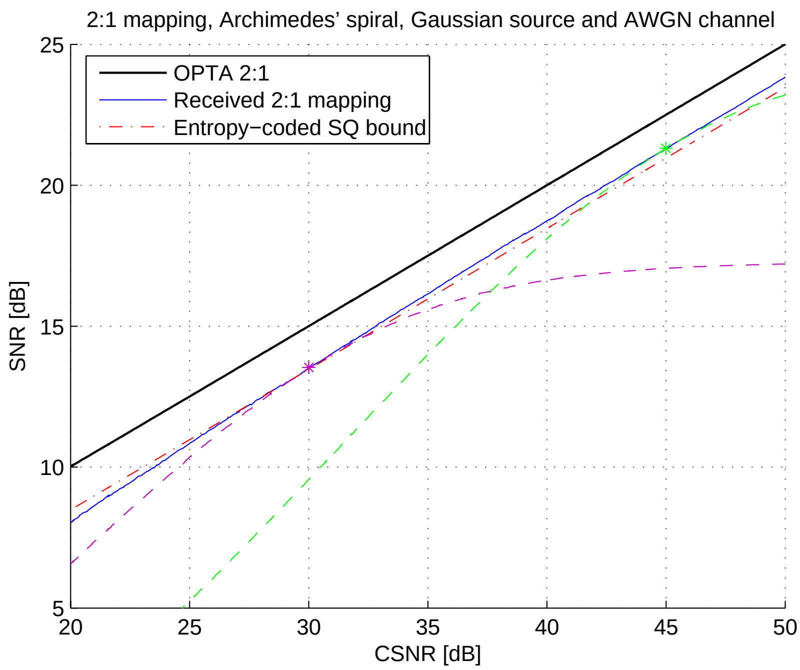

Figure 3: Performance of the proposed system. As a reference, the theoretical bound for the best scalar quantizer operating at the channel capacity is shown in the dash-dotted line. The robustness of the mapping can be seen in the dashed curves, where the spiral is optimised for 30 and 45 dB CSNR and still perform within 5 dB of the curve with perfect channel-state information (CSI), over a 20 dB CSNR range.

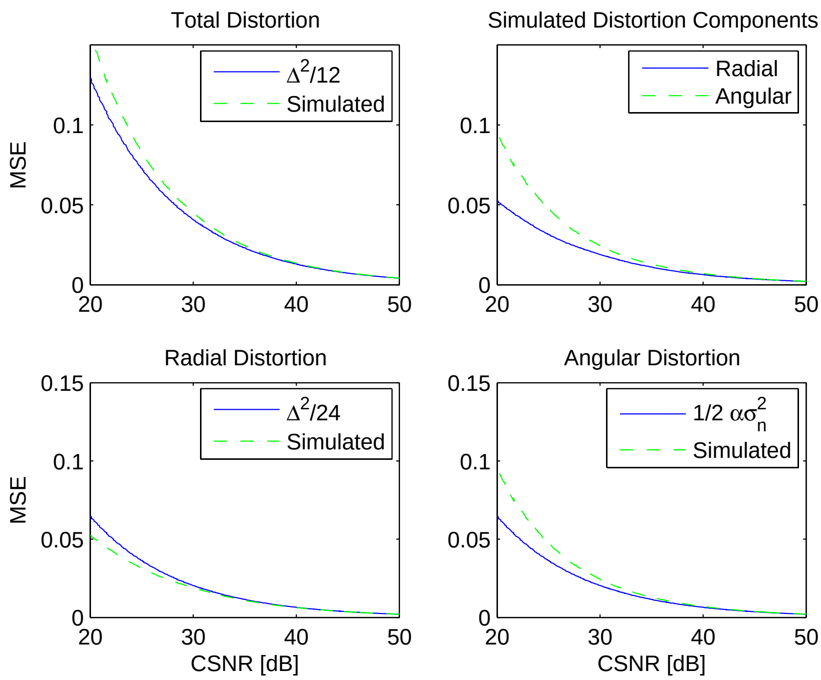

Figure 4: Theoretical distortion expressions compared to simulated distortion. It can be seen that the models are good above 35dB CSNR.

References

[1] Sae-Young Chung. *On the Construction of Some Capacity-Approaching Coding Schemes.* PhD thesis, Massachusetts Institute of Technology, September 2000. http://lids.mit.edu/∼sychung/thesis/index.html.

[2] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

[3] G.David Jr. Forney and Gottfried Ungerboeck. Modulation and coding for linear gaussian channels. *IEEE Trans. Information Theory*, 44(6):2384–2415, Oct. 1998.

[4] Arild Fuldseth. *Robust subband video compression for noisy channels with multilevel signaling.* PhD thesis, Norwegian University of Science and Engineering (NTNU), 1997.

[5] Robert M. Gray and Tamás Linder. Mismatch in high-rate entropy-constrained vector quantization. *IEEE Trans. Information Theory*, 49(5):1204–1217, May 2003.

[6] Tor A. Ramstad. Shannon mappings for robust communication. *Telektronikk*, 98(1):114–128, 2002. Information Theory and its Applications.

[7] C. E. Shannon. A mathematical theory of communication. *The Bell System technical journal*, 27:379–423, 1948.

[8] Claude E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37:10–21, jan 1949.

[9] Daniel Zwillinger. *CRC Standard Mathematical Tables and Formulae.* CRC Press, Boca Raton, Fla., 31st. edition, 2003.