Quantifying Performance Losses in Source-Channel Coding

Fredrik Hekland, Geir E. Øien and Tor A. Ramstad

***Abstract*— In this paper, we identify and quantify loss factors causing sub-optimal performance in joint source-channel coding. We show that both the loss due to non-Gaussian distributed channel symbols and the loss due to non-Gaussian quantization error equals the relative entropy between the actual distribution and the optimal Gaussian distribution, given an average power constraint and an mean-squared error (mse) distortion measure, respectively. Furthermore, losses occur whenever each channel symbol has an information bit rate lower than the channel capacity, and due to the presence of inter-channel correlation, source coder redundancy (intra-channel correlation), suboptimal receiver structures, and decoding errors. We quantify the loss associated with some of these factors and provide an example of communication scheme, showing that calculations and simulations agree.**

I. Introduction

In certain settings, joint source-channel coding (JSCC) can provide improved performance and robustness relative to the traditional separate source and channel coding approach, especially in cases where there are delay and complexity constraints [1]. After all, optimality in separate source and channel coding for a single user is achieved only when we allow infinite delay and complexity in both coders [2]. Joint source-channel coding, on the other hand, can come very close to the theoretical optimum with low system complexity. The most extreme example is when transmitting a Gaussian source over an additive white Gaussian noise (AWGN) channel of the same bandwidth, given an average power constraint and a squared-error distortion measure. By plugging the source directly onto the channel with a certain optimized amplification factor, optimum performance is achieved for all channel qualities, without any coding at all. The reason for this is that the source's probability density function (pdf) equals the channel signal pdf required to achieve the channel capacity, and the Gaussian noise from the channel equals the error signal pdf required to achieve the rate-distortion bound [3].

Earlier, we have proposed some schemes for joint source-channel coding [4], [5], [6], which we termed *Shannon mappings*. These schemes can achieve both bandwidth reduction (compression) and bandwidth expansion (error-control), but none of them achieve actual optimality. What is apparent from past work, however, is that systems which do compression come much closer to the theoretical bounds than when performing expansion. Intrigued by this, we try to identify more in detail which loss factors come into play when designing a JSCC system, in order to obtain some hints on what to do to improve the performance. The loss factors will be described in terms of information theoretic expressions. Thus, they are not specific to Shannon mappings or other direct source-channel mappings. Indeed, traditional schemes with separable quantizers and channel coders can also be analyzed, although our main emphasis in the example will be on JSCC systems.

The work leading to this paper was sponsored by The Research Council of Norway (NFR), in part within the project BEATS and in part within the NORDITE project CROPS. http://www.iet.ntnu.no/projects/beats

F. Hekland was with the Dept. Electronics & Telecommunications. Norwegian University of Science and Technology (NTNU), N-7491 Trondheim, Norway. He is now with ABB AS Corporate Research Center, N-1375 Billingstad, Norway. E-mail: hekland@gmail.com

G. E. Øien and T. A. Ramstad are with the Dept. Electronics & Telecommunications. Norwegian University of Science and Technology (NTNU), N-7491 Trondheim, Norway. E-mail: {oien,ramstad}@iet.ntnu.no

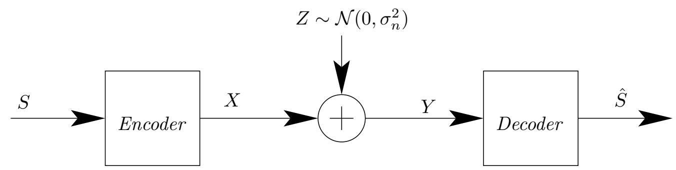

Fig. 1. A generic communication system

A good discussion about lossy source-channel coding system can be found in [3]. This paper covers the information theoretical aspects of optimal source-channel coding, and through Theorem 6, provides a criterion which can be used to check whether a system performs optimally or not. However, no clues are given on how to determine the actual loss whenever a system is not optimal. As most systems (especially those with delay or complexity constraints) do not achieve optimality, it is interesting to quantify these loss factors. This would enable us to know how much we can hope to improve a system, which parts of the system are responsible for the biggest losses, and thus where it pays off the most to make improvements.

The paper is organized as follows. First we identify the different loss factors; mismatched channel symbol distribution, under-utilization of the channel, inter-channel correlation, source coding redundancy, mismatched source distribution, suboptimal receiver structures, and decoding errors. Then we exemplify some of these with a 1:2 bandwidth-expanding JSCC system.

II. Preliminaries

Based to some extent on intuition, we will initially make some conjectures about what causes suboptimal performance in joint source-channel coding systems under a squared-error distortion measure. The conceptual communication chain is seen in Fig. 1 where S is the source symbol, X is the transmitted channel symbol, Z is the channel noise, Y is the received channel symbol, and $\hat{S}$ is the reconstructed source symbol. The encoder and decoder might be any JSCC system, or in fact even a traditional separate source and channel coding system. Moreover, the losses are stated in bits, but this does not limit us to looking at bit-based systems. Shannon mappings, which use direct source-channel mappings, can be analyzed as well.

In short, we want to determine the loss due to the inequalities in the proof of [3, Lemma 2], for convenience repeated here:

$$\begin{aligned} R(D) &= \min_{f_{\hat{s}|s}:\mathsf{E}d(S,\hat{S})\leq D} I(S;\hat{S}) \overset{a)}{\leq} I(S;\hat{S}) \\ &\overset{b)}{\leq} I(X;Y) \overset{c)}{\leq} \max_{f_x:\mathsf{E}\sigma_x^2\leq P} I(X;Y) = C(P), \end{aligned} \tag{1}$$

where E denotes the expectation operator with respect to the source distribution, $d(\cdot,\cdot)$ is the distortion measure, $R(D)$ denotes the rate-distortion bound for a distortion D, and $C(P)$ denotes the channel capacity for a given average power constraint P. A system performs optimally whenever the rate is equal to the channel capacity, i.e. when all the relations in (1) are equality signs. Equality in $a)$ is achieved for the rate-distortion bound achieving error distribution (given the distortion measure), equality in $b)$ is achieved for an information lossless encoder-decoder pair (when going to the channel space and back), and finally equality in $c)$ is achieved for the capacity-achieving channel symbol distribution (given the capacity-cost function).

It is worth noting that for an optimal pair $R(D) = C(P)$, one cannot lower D (or P) without changing $R(D)$ (or $C(P)$). This means that if the source rate is higher than the channel capacity, the information content must be reduced (i.e. D must be increased such that the equality is satisfied).

A. Loss from Mismatched Channel Distributions

We now show how much we lose by using non-Gaussian distributed channel symbols on an AWGN channel, causing an inequality in $c)$ of (1). This result was shown for a single channel case in [5], but is repeated here for the parallel channel case for completeness. To the authors' knowledge, this simple result has not been reported in the literature earlier than [5].

We know from information theory that Gaussian distributed channel symbols maximize the mutual information between a transmitted and received sequence over any memoryless channel with an average power constraint [7]. For the specific example of a n-dimensional AWGN channel with an average power constraint, the channel capacity is defined as [7]:

$$C = \max_{f(x_1,\ldots,x_n):\sum \sigma_{x_i}^2 \leq P} I(X_1,\ldots,X_n;Y_1,\ldots,Y_n), \tag{2}$$

where $I(\cdot;\cdot)$ denotes mutual information. Whenever we have a marginal channel symbol distribution $f(x_i) \nsim \mathcal{N}$ and the same power constraint, the achievable transmission rate will be less than the capacity of this specific channel, i.e.

$$C' = I(X_1,\ldots,X_n;Y_1,\ldots,Y_n)\big|_{f(x_i)\nsim\mathcal{N}} = C - \eta(f),$$

where $\eta(f)$ denotes the associated capacity loss for the distribution f. Expanding the mutual information in (2), and assuming that the signal and noise are uncorrelated, one can show that [7]

$$I(X_1,\ldots,X_n;Y_1,\ldots,Y_n) \leq \sum_{i=1}^{n} \left\{h(Y_i) - h(Z_i)\right\}, \tag{3}$$

where Z_i is the channel noise in channel i and $h(\cdot)$ denotes the differential entropy. Assuming that the covariance matrix of $(X_1,\ldots,X_n)$ is diagonal so that there is no correlation between the channels, the only loss we would experience is when the X_i's are not Gaussian distributed. To determine this loss, we expand the first term on the right hand side of (3). Since we always integrate over the same variable y, we write f_i instead of $f_{Y_i}(y)$ to increase the readability, although this in some sense is abuse of notation. Expanding the differential entropy, we obtain

$$\begin{aligned} h(Y_i) &= -\int f_i \log f_i dy = -\int f_i \log\left(f_i \frac{f_i^*}{f_i^*}\right) dy \\ &= -\int f_i \log f_i^* dy - \int f_i \log\left(\frac{f_i}{f_i^*}\right) dy \\ &\overset{(a)}{=} -\int f_i^* \log f_i^* dy - \int f_i \log\left(\frac{f_i}{f_i^*}\right) dy \qquad (4) \\ &= h^*(Y) - I(f_i \| f_i^*), \qquad (5) \end{aligned}$$

where $I(\cdot\|\cdot)$ is the relative entropy (or I-divergence) between two distributions[1], and (a) is valid when f_i^* is the Gaussian distribution and f_i is any continuous distribution with zero mean and same variance as f_i^*. To see why this is true consider

$$\begin{aligned} -\int f \ln f^* dy &= -\int f \ln\left(\frac{1}{\sqrt{2\pi}\sigma_G} e^{\frac{-y^2}{2\sigma_G^2}}\right) dy \\ &= \frac{1}{2\sigma_G^2}\int y^2 f dy + \ln\left(\sqrt{2\pi}\sigma_G\right)\int f dy \\ &= \frac{1}{2\sigma_G^2}\int y^2 f^* dy + \ln\left(\sqrt{2\pi}\sigma_G\right)\int f^* dy \\ &= -\int f^* \ln f^* dy, \end{aligned} \tag{6}$$

which is valid for any well-behaved distribution f with zero mean. Inserting (5) and (3) into the definition of the channel capacity in (2) we obtain the maximum achievable transmission rate for input distributions f_i as

$$\begin{aligned} C' &= \sum_{i=1}^{n} h^*(Y_i) - h(N_i) - I(f_i \| f_i^*) \\ &= \sum_{i=1}^{n} \{C_i - I(f_i \| f_i^*)\}, \end{aligned} \tag{7}$$

which is the sum of the channel capacity of each channel minus the relative entropy between the mismatched distribution of each channel and the Gaussian. I.e., $\eta(f) = I(f_i \| f_i^*)$.

[1] We are using the less common form $I(\cdot\|\cdot)$ instead of $D(\cdot\|\cdot)$ in order to avoid confusion with the distortion D used later.

B. Loss from information rate less than the operational channel capacity $(R < C')$

Obviously, we cannot attain the theoretical optimum if the actual information rate of the symbols transmitted over the channel is strictly below the channel capacity (i.e., if relation b) in (1) is a strict inequality). This is for instance the case for channel codes with insufficient puncturing using more parity bits than necessary to achieve a certain bit-error rate (BER) [8], or for uncoded modulation with a rate lower than the channel capacity. This loss corresponds to [3, Eq.(8)], where $I(X;Y) - I(S;\hat{S})$ is the loss induced by too many parity bits or a too small modulation constellation. More generally we can say that the encoder-decoder pair in Fig. 1 is not information lossless. When communicating a continuous source over an AWGN channel, the Gaussian channel noise contaminating the source ensures that the first inequality sign in (1) is satisfied with equality. Furthermore, by plugging the source directly onto the channel and using the optimal Wiener filter as the receiver, the second inequality sign is satisfied with equality too. This means that we have $R = C'$ and the only loss will be caused by the non-Gaussian channel symbol distribution as described in Sec. II-A.

C. Loss from Correlated Channels

In a system with parallel communication channels, all the channels should be mutually uncorrelated in order to maximize the total achievable capacity. For independent noise Z, we have the following inequality for the mutual information

$$\begin{aligned} &I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) \\ &\quad = h(Y_1, Y_2, \ldots, Y_n) - \sum_i h(Z_i) \\ &\quad \leq \sum_i \{h(Y_i) - h(Z_i)\}, \end{aligned} \tag{8}$$

where equality is only achieved by independent Y's. Expanding the differential entropy of the Y's we have [7]

$$h(Y_1, Y_2, \ldots, Y_n) = \sum_{i=1}^{n} h(Y_i | Y_1, Y_2, \ldots, Y_{i-1}), \tag{9}$$

and consequently we have $h(Y_1, Y_2, \ldots, Y_n) \leq \sum h(Y_i)$. The resulting rate loss would be the difference $\sum h(Y_i) - \sum h(Y_j | Y_1, Y_2, \ldots, Y_{j-1})$. For examples of systems suffering from correlated channels, one could think of spatial correlation in multiple-input, multiple-output (MIMO) systems, or orthogonal frequency-division multiplex (OFDM) systems with correlated sub-channel gains. However, we will not discuss MIMO here, and we refer the reader to [9].

We could also classify the case of encoder-induced correlation into this section. By this we mean inter-channel correlation introduced by the encoder, as opposed to e.g. the spatial correlation leading to reduced rank MIMO channels. For example, the worst-case example would be block pulse amplitude modulation (BPAM) [10] for bandwidth expansion, where replicas of the source is transmitted in parallel ("repetition code"). Provided that the total power is doubled when the number of channels are doubled, the signal-to-noise ratio (SNR) is only increased by 3 decibel (dB) per doubling of channels. This is because the channels carry fully correlated information, and thus the available bandwidth is poorly utilized.

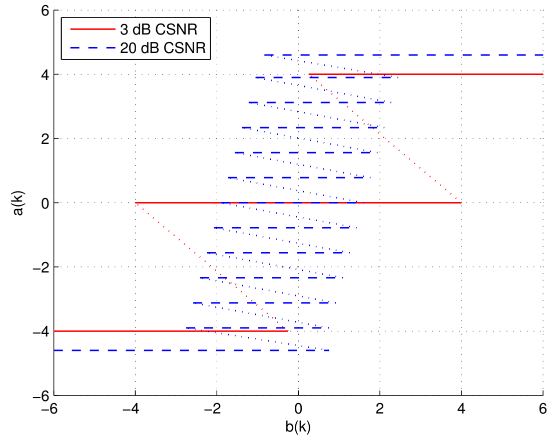

Fig. 2. The hybrid scalar quantizer, linear coder (HSQLC).

Fig. 2 shows a specific 1:2 bandwidth expanding system (to be described more in detail in Sec. III-A). The line segments in the plane constitute the source space, and the plane is the channel space. BPAM with the same expansion factor would be a represented by a diagonal line in the plane. This means that the encoder introduces full correlation between the two channels and the channel space (the plane) is poorly utilized. The system shown in the figure has much less correlation and fills the channel space better.

D. Loss from Source Coder Redundancy

In order for a communication system to perform optimally, it also has to operate at the rate-distortion (R-D) bound (relation a) in (1)), and the source coder rate must thus be equal to the channel capacity [3]. The loss experienced from a non-optimal source coder is given as the excess rate required to achieve the same distortion as the R-D function evaluated at the channel capacity. First of all, ideal entropy coding must be applied in order to reach the source's differential entropy rate. Failing to do so will give a rate increase, with the amount of increase depending on the source pdf [11] and on the source redundancy.

Entropy coding alone, however, cannot achieve the R-D bound. For instance, a scalar quantizer can at best be 0.255 bits away from the R-D bound [11], regardless of the source pdf. This is due to the fact that the quantization noise cannot be shaped properly (to be a Gaussian) when we have only one dimension in the quantizer. vector quantization (VQ) can, on the other hand, reach the R-D bound if given infinite dimensions. This is referred to as the *space-filling advantage* in [12]. We will thus refer to the gap relative to an ideal VQ as the *space-filling loss*, and it applies to any source coder failing to produce Gaussian reconstruction-error noise.

Usually, the space filling loss is described in terms of high-resolution theory [12]. However, an alternative way of looking at the space filling loss is to examine the reconstruction (or quantization) noise which appears in the derivation of the R-D function for a Gaussian source. Given the squared error distortion measure, Gaussian-distributed quantization noise is needed to reach the rate-distortion bound [7], regardless of the input source distribution. This is because for an allowed distortion D, a Gaussian-distributed reconstruction error will minimize the necessary mutual information between the original and reconstructed sequences. We now determine the loss of a non-Gaussian reconstruction noise. The R-D function is defined as [7]:

$$R(D) = \min_{f(\hat{s}^m|s^m):\mathsf{E}d\left(S^m,\hat{S}^m\right)\leq D} I(S^m;\hat{S}^m), \qquad (10)$$

where $d(\cdot,\cdot)$ is the chosen distortion metric, mse in our case.

Assuming that the coder produces reconstruction noise which is not Gaussian distributed, we have an operational R-D function, $R'(D)$, which is higher than the R-D bound $R(D)$ given in (10):

$$R'(D) = I(S^m;\hat{S}^m)\big|_{f(\hat{s}^m|s^m)\nsim\mathcal{N}} = R(D) + \lambda(f), \qquad (11)$$

where $\lambda(f)$ is the penalty term due to the non-Gaussian reconstruction noise distributions. Looking at the mutual information, we can expand it as follows:

$$\begin{aligned} I(S^m;\hat{S}^m) &= \sum_{i=1}^{m} h(S_i) - \sum_{i=1}^{m} h(S_i|S^{i-1},\hat{S}^m) \\ &\geq \sum_{i=1}^{m} h(S_i) - \sum_{i=1}^{m} h(S_i|\hat{S}_i) \qquad (12) \\ &= \sum_{i=1}^{m} h(S_i) - \sum_{i=1}^{m} h(S_i-\hat{S}_i|\hat{S}_i) \\ &\geq \sum_{i=1}^{m} h(S_i) - \sum_{i=1}^{m} h(S_i-\hat{S}_i), \qquad (13) \end{aligned}$$

where (12) and (13) are due to the fact that conditioning reduces entropy. Using the definition of the differential entropy we can furthermore expand the last addend of (13) as follows (again slightly abusing notation):

$$\begin{aligned} h(S_i-\hat{S}_i) &= -\int f_i \log f_i dy = -\int f_i \log\left(f_i \frac{f_i^*}{f_i^*}\right) dy \\ &= -\int f_i \log f_i^* dy - \int f_i \log\left(\frac{f_i}{f_i*}\right) dy \\ &= -\int f_i^* \log f_i^* dy - \int f_i \log\left(\frac{f_i}{f_i^*}\right) dy \qquad (14) \\ &= h^*(S_i-\hat{S}_i) - I(f_i\|f_i^*), \qquad (15) \end{aligned}$$

where (14) is valid when f_i^* is the Gaussian distribution and f_i is any continuous distribution with zero mean[2] and the same variance as f_i^*, due to the same arguments as in (6). Inserting (15) into (13) we obtain

$$I(S^m;\hat{S}^m) \geq \sum_{i=1}^{m}\left\{h(S_i) - h^*(S_i-\hat{S}_i) + I(f_i\|f_i^*)\right\}, \quad (16)$$

where the two first terms on the right hand side correspond to $R(D)$ in (11) and the last term corresponds to the penalty $\lambda(f)$ in (11). This gives us the operational R-D function

$$R'(D) = \sum_{i=1}^{m}\left\{R(D_i) + I(f_i\|f_i^*)\right\}, \qquad (17)$$

i.e. the loss factor in (11) is $\lambda(f_i) = I(f_i\|f_i^*)$. This result is similar to the discrepancy shown in Sec. II-A, in the sense that the penalty for having a non-optimal distribution is the "distance" between the actual and the optimal distribution[3]. To the authors' knowledge, this result has not been previously reported in the literature.

E. Loss From Mismatched Source Distribution

When designing a source coder, knowing the source's input pdf is important to ensure optimal performance. However, exact knowledge about the source is usually not available and it is usually estimated from data, resulting in mismatch in the source coder. We will call the resulting loss *source mismatch loss*. Some work has been done on this for high-rate entropy-constrained vector quantizers [13], where they show that the loss incurred by designing the quantizer for a pdf f and applying a pdf g is equal to the relative entropy between the two, i.e. $I(f\|g)$. The restrictions on this result is that f/g must be bounded, which in many cases is not true.

F. Loss Due to Suboptimal Receiver Structures

The design of receiver is crucial for the system performance, since it is responsible for extracting the transmitted signal from the contaminating channel noise and possibly other impairments such as fading and interference. Moreover, depending on whether the source symbols are uniformly distributed or not, the optimal receiver would be either be a maximum likelihood (ML) or a maximum a-posteriori probability (MAP) receiver [14]. The former is basically a minimum-distance detector for the AWGN channel, but in some cases, even an ML detector might become prohibitively complex, and more simpler receiver structures are employed. Any sub-optimality in the receiver of course results in degraded performance. Often there is a certain correlation between channels, either because of non-optimal source coding or deliberately introduced to aid decoding. However, if the decoder is not decoding the channels jointly, but rather independently, performance degradations will occur. This typically has bigger impact at low than at high channel signal-to-noise ratio (CSNR).

It is, however, rather difficult to quantify this degradation in general terms, since it is very system dependent. Usually, this is a trade-off between the complexity of the receiver and performance.

[2]This is not a restricting assumption, since the mean can always be subtracted before encoding the source.

[3]The relative entropy, also known as the Kullback-Leibler distance, is not a true distance metric, but still it provides a measure of the rate difference of two distributions.

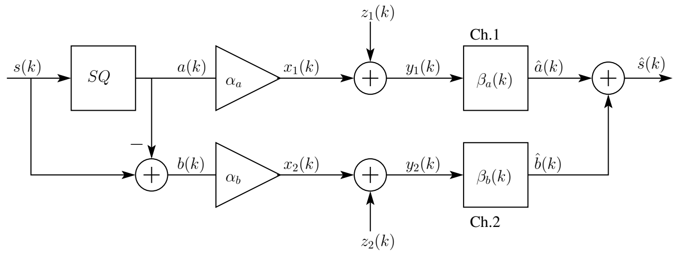

Fig. 3. The HSQLC system

G. Loss From Incorrectly Decoded Channel Symbols

Whenever the channel impairments induce incorrect decisions in the decoder, we have an information loss responsible for an increasing inequality in relation b) of (1). What this loss translates to, depends strongly on the actual system. For a communication system using entropy coding, bit errors can cause catastrophic breakdown ruining the entire transmission. Direct modulation of a scalar quantizer, however, only causes increased distortion when incorrectly decoding to neighboring intervals. In both cases, this effect is often referred to as the *threshold effect*, since the performance above a certain CSNR threshold is good, but rapidly decreases below the threshold.

III. Example

We now present a specific example of joint source-channel coding system, where the relevant loss factors described in Sec. II are shown to be able to completely explain the discrepancy between the systems' performance and the theoretical optimum. Even though these systems do not operate with bit representations, the losses are stated in bits and converted to decibels with the common 6.02 dB per bit rule [11]. This is valid for the case of scalar, memoryless sources with a squared-error distortion measure.

A. 1:2 HSQLC, Gaussian Source and AWGN Channel

The hybrid scalar quantizer, linear coder (HSQLC) [15] seen in Fig. 3 is a bandwidth expanding single-channel joint source-channel coding scheme with two channels, where the total channel bandwidth is twice that of the source bandwidth. The bandwidth expansion is a way of performing error control, which allows higher SNR of the reconstructed source than for equal source and channel bandwidths. Theoretically, the best we can achieve is optimal performance, theoretically attainable (OPTA) [16], which is found by equating the rate-distortion function with the channel capacity. For a Gaussian source and an AWGN channel we have

$$2B_s \cdot \frac{1}{2}\log_2(\text{SNR}) = 2B_c \cdot \frac{1}{2}\log_2(1+\text{CSNR})$$
$$\text{SNR} = (1+\text{CSNR})^{\frac{B_c}{B_s}} = (1+\text{CSNR})^2. \quad (18)$$

The HSQLC system's performance is, however, far away from this bound. At around 30 dB CSNR when the two channels are allocated the same amount of power by the optimization algorithm, the system performance is a little less than 8.5 dB away from OPTA.

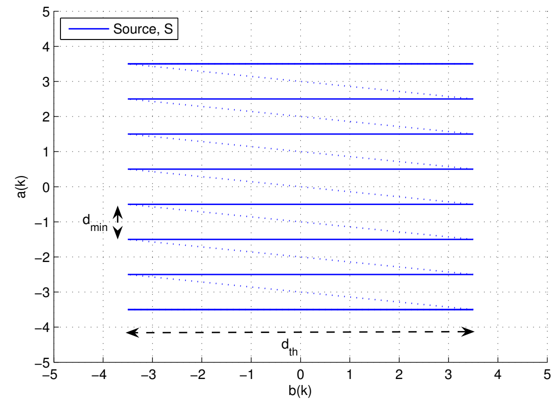

Fig. 4. Simplified HSQLC system.

We will now scrutinize this simplified system in order to try to identify the main causes of this big loss. We will concentrate on the high CSNR region, since in this region the optimization algorithm distributes the power equally on the two channels (hence we have the same CSNR on the two channels). Moreover, in this region, the quantizer has sufficiently high rate to assume that the quantizer output and quantization error are uncorrelated (given that the usual required conditions for this approximation to be valid are fulfilled). Also, in the high CSNR case, the effect of non-ideal receivers is more or less negligible. In order to facilitate the analysis, we abandon the optimized quantizer characteristics (Fig. 2), and instead stack the quantizer levels directly above each other (Fig. 4). This degrades the performance for lower CSNR, since incorrectly decoded symbols $\hat{a}(k)$ will induce higher distortion, and furthermore the error-signal distribution will be uniform instead of something resembling a Gaussian. However, at 30 dB and above, the difference between the simplified and the optimal HSQLC system is negligible.

The schematic diagram of the HSQLC system is shown in Fig. 3. The input signal $s(k)$ is first quantized by a uniform scalar quantizer, and the quantized signal $a(k)$ is allocated a certain amount of power and transmitted over the channel as a multilevel pulse-amplitude modulation (PAM) symbol $x_1(k)$. The quantization error $b(k)$ is allocated the rest of the power and transmitted with direct PAM over the other channel as $x_2(k)$. For high CSNR ($>$ 30 dB) the optimization described in [15] divides the available transmit power equally between the two channels. We assume that the two channels have equal characteristics. The receivers are not optimal, as the two channels are decoded independently. However, simulations confirm that the difference in performance between the optimal and simplified receivers is negligible at high CSNR [15].

Loss from mismatched channel symbol distributions: The quantized signal transmitted over channel 1 will have a probability mass function (pmf) approximating the continuous Gaussian distribution, and there will be almost no loss due to incorrect channel symbol distribution as long as the number of levels is sufficiently high.

As the signal on channel 2 is transmitted as direct PAM, it will have the same distribution as the quantization error $b(k)$. The scalar quantizer is of high rate (at high CSNR) and thus the quantization error is approximately following a uniform distribution. According to Sec. II-A we should thus experience a loss equal to the relative entropy between the actual uniform distribution and the optimal Gaussian channel symbol distribution of the same variance. Denoting the uniform distribution as f and the Gaussian as f^* we have

$$\begin{aligned} I(f\|f^*) &= \int f \ln \frac{f}{f^*} dy = -h(f) - \int f \ln f^* dy \\ &= -h(f) + \ln\left(\sqrt{2\pi}\sigma_G\right) + \int f \frac{y^2}{2\sigma_G^2} dy \\ &= -\ln(b_{max} - b_{min}) + \ln\left(\sqrt{2\pi}\sigma_G\right) + 1/2 \\ &\overset{(c)}{=} -\ln\sqrt{12\sigma_U^2} + \ln\left(\sqrt{2\pi\sigma_G^2}\right) + \ln\sqrt{e} \\ &= \frac{1}{2}\ln\left(\frac{2\pi e \sigma_G^2}{12\sigma_U^2}\right) = \frac{1}{2}\ln\left(\frac{\pi e}{6}\right), \end{aligned} \tag{19}$$

where we use the fact that $\sigma_U^2 \equiv \sigma_G^2$ (comparing distortion at the same power level), and in (c) we use the fact that $\sigma_U^2 = 1/12(b_{max} - b_{min})^2$. In bits, this loss is $0.5\log_2(\pi e/6) = 0.255$. This value is confirmed by estimating the relative entropy of the simulated quantization error and a Gaussian distribution of the same variance, using the method described in [17]. In decibel, this loss is 1.53 dB which is equal to the shaping gain of 1.53 dB seen in modulation [18].

Loss from threshold effect: Channel 1 has multi-level symbols with no explicit error protection. Channel noise will therefore cause transition to neighboring symbol levels, and if we try to increase the rate on this channel, the threshold effect increases. This implies that in order to obtain an acceptable symbol error rate, each channel symbol must carry strictly less information than the channel capacity allows, and we will have a loss like in Sec. II-B. Since we in the HSQLC use direct modulation of the quantized values, without entropy coding or bit-representation, and the channel symbols are memoryless, a symbol error will not cause complete breakdown in the coder. Only one source sample is affected, and the result is higher source distortion (Sec. II-G).

Calculation of the mse caused by the threshold effect in channel 1 is done as follows. The source length of each segment in Fig. 4 is $d_{th} = (s_{max} - s_{min})/L$, where L is the number of levels. Since the simplified system has the line segments stacked on top of each other, the distortion caused by a jump to neighboring levels is simply $\|k \cdot d_{th}\|^2$, where k is the number of jumps caused by the channel noise. The total mse from the threshold effect is the distortion from all jumps to neighboring levels, multiplied by the probability of transitions, integrated over the entire source S:

$$\varepsilon_{th}^2 = \sum_{i=1}^{L}\sum_{j=1}^{L}\int_{s_i}^{s_i+d_{th}} ((i-j)d_{th})^2 Q\left(\frac{|i-j|d_{min}}{2\sigma_n}\right) f_S(s) ds, \tag{20}$$

where $Q(\cdot)$ is the Q-function which returns the area under the tail of the standardized normal distribution, d_{min} is the distance between quantizer levels on the channel, $s_1 = s_{min}$, $s_L = s_{max}$ and $s_i = s_{i-1} + d_{th}$ (assuming that the overload distortion from not letting s go to $\pm\infty$ is negligible).

Translating the mse in (20) into its associated SNR loss in an exact manner is not trivial. Starting with the expression for the total SNR

$$\text{SNR} = 10\log_{10}\left(\frac{\sigma_x^2}{\varepsilon_{ch.2}^2 + \varepsilon_{th}^2}\right), \tag{21}$$

we can try to calculate the SNR loss in dB by taking the difference of the SNR from only channel 2 and the total SNR as follows: $\text{SNR}_{thr} = 10\log_{10}\left(\frac{\sigma_x^2}{\varepsilon_{ch.2}^2}\right) - \text{SNR}$. This is not exact, however, since $\log(\gamma + \delta) \neq \log\gamma + \log\delta$, but when the mse from the threshold effect is small it gives a reasonably accurate result.

Loss from rate lower than C: Since the channel symbols are memoryless and no entropy coding is involved, the HSQLC system can allow a certain amount of symbol errors. Still, threshold effect calculated above restricts the rate to a number below the channel capacity. At 30 dB CSNR the number of PAM-levels is 28 (for the simplified system). Using [11, Eq. 4.108] for a Gaussian distribution with the optimal step size, we find the resulting entropy to be 3.89 information bits. This is 1.09 bits below the channel capacity. Channel 2, carrying the quantization error is of continuous-amplitude and has a rate equal to the channel capacity. The only loss in this channel is from the incorrect channel symbol distribution as explained above.

Loss from channel symbol correlation: Since the quantizer has relatively high rate and the input distribution is Gaussian, we can treat the quantization error and the input as uncorrelated random variables. Since the quantizer output is a deterministic function of the input, we can also treat the quantization error and the output as uncorrelated random variables. This means that the correlation between the two channels is close to zero, and thus there is no loss as explained in Sec. II-C. This, however, is a simplification which does not hold for low CSNR, where the rate on channel 1 drops. Looking at Fig. 2, we see the quantizer levels of the original HSQLC system [15]. The shift of the quantized levels reduces the effect of erroneously decoded symbols and shapes the quantization error into a more Gaussian-like distribution. Whether the latter has any effect is, however, not clear, since this also increases the power of channel 2 and thus provide less channel noise suppression. The shift introduces a dependency between the two channels which should be accounted for in the receiver, as discussed in Sec. II-F.

Table I sums up all the redundancy factors of the simplified HSQLC-system and shows the estimated loss in decibels, calculated from the 6.02 dB per bit rule. The total estimated loss of 8.56 dB is very close to the loss we see from simulations, with simulation results varying a bit due to the threshold effect contribution. When the probability of the threshold effect is

Loss (30 dB CSNR)	Sec.	**bits**	**dB**
Mismatched pmf (Ch.1)	II-A	∼0	∼0
$R_{Ch.1} < C$	II-B	1.09	6.56
Mismatched pdf (Ch.2)	II-A	0.255	1.53
$R_{Ch.2} < C$	II-B	∼0	∼0
Inter-channel correlation	II-C	∼0	∼0
Source Coder Redundancy	II-D	∼0	∼0
Mismatced Source distribution	II-E	∼0	∼0
Suboptimal receiver	II-F	∼0	∼0
Threshold effect (Ch.1)	II-G	-	0.47
Sum		**1.345**	**8.56**
Simulation			∼8.58

TABLE I
LOSS ACCOUNT FOR SIMPLIFIED HSQLC, GAUSSIAN SOURCE AND AWGN CHANNEL.

estimated correctly in (20), the calculated and simulated loss agree within 0.02 dB.

As the reduced rate on the digital channel is responsible for 75% of the loss for the HSQLC, it would make sense to try to include a more efficient coding on that channel, like for instance a bandwidth-efficient low-density parity check (LDPC)-coded modulation [19][4]. Using this on channel 1, we could get as close as 3.88 dB to the Shannon limit with a BER of 10^{-7}. This translates to a capacity loss of 0.643 bits per symbol. Assuming no incorrectly decoded symbols, the calculated distance to OPTA would be $0.643 \cdot 6.02 + 1.53 = 5.39$ dB. For a capacity-achieving code, the distance to OPTA would be only 1.53 dB. Both these cases are verified by simulations.

To further improve the performance, a higher dimensional vector quantizer in place of the scalar quantizer must be employed, since that is the only way to produce a more Gaussian distributed quantization noise.

IV. CONCLUDING REMARKS

In this paper, we have tried to identify factors in (lossy) joint source-channel coding systems leading to sub-optimal performance. We only considered systems with a mean squared-error distortion measure and an average power constraint on the channel.

The loss factors identified are the following; Non-Gaussian channel symbol distribution (Sec.II-A), operational rate lower than the channel capacity (Sec. II-B), correlation between channels (Sec. II-C), source coder redundancy (Sec. II-D), mismatched source distribution (Sec. II-E),suboptimal receiver structures (Sec. II-F) and decoding errors (Sec. II-G). The list of loss factors is not exhaustive, but should cover the most important loss factors of source-channel coding systems. By means of an example, some of the identified factors were shown to be able to explain the loss experienced and to provide important hints on where to start when trying to improve a given system's performance. Due to lack of space, the extent of dependency between the different loss factors have not been investigated thoroughly. However, it seems to be system dependent, thus it might be hard to generalize this.

Future work will include verification of these findings on other joint source-channel coding systems, and research on how to utilize the (currently to a large degree non-constructive) knowledge of the losses to improve on current systems. Moreover, the topic of inter-channel correlation for direct source-channel coding systems needs more research in order to be able to quantify the losses, especially for low CSNR where this effect is more pronounced.

[4] This would, however, further reduce the system's robustness compared to direct modulation.

REFERENCES

[1] J. Lim and D. Neuhoff, "Joint and tandem source-channel coding with complexity and delay constraints," *IEEE Trans. Commun.*, vol. 51, no. 5, pp. 757–766, 2003.

[2] S. Vembu, S. Verdu, and Y. Steinberg, "The source-channel separation theorem revisited," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 44–54, 1995.

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

[4] T. A. Ramstad, "Shannon mappings for robust communication," *Telektronikk*, vol. 98, no. 1, pp. 114–128, 2002. [Online]. Available: http://www.telenor.com/telektronikk/volumes/pdf/1.2002/Page_114-128.pdf

[5] F. Hekland, G. Øien, and T. Ramstad, "Using 2:1 Shannon mapping for joint source-channel coding," in *Data Compression Conference, 2005. Proceedings. DCC 2005*, Mar. 2005, pp. 223–232.

[6] P. A. Floor and T. A. Ramstad, "Noise analysis for dimension expanding mappings in source-channel coding," in *7th Workshop on Signal Processing Advances in Wireless Communications*. Cannes, France: IEEE, July 2006.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[8] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, Apr. 1988.

[9] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, 2003.

[10] K.-H. Lee and D. P. Petersen, "Optimal linear coding for vector channels," *IEEE Trans. Commun.*, vol. COM-24, no. 12, pp. 1283–1290, Dec. 1976.

[11] N. S. Jayant and P. Noll, *Digital Coding of Waveforms, Principles and Applications to Speech and Video*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1984.

[12] T. Lookabaugh and R. Gray, "High-resolution quantization theory and the vector quantizer advantage," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1020–1033, Sept. 1989.

[13] R. M. Gray and T. Linder, "Mismatch in high-rate entropy-constrained vector quantization," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1204–1217, May 2003.

[14] J. R. Barry, E. A. Lee, and M. D. G., *Digital Communication*, 3rd ed. Kluwer Academic Publishers, 2004.

[15] H. Coward, "Joint source-channel coding: Development of methods and utilization in image communications," Ph.D. dissertation, NTNU, Trondheim, 2001.

[16] J. Goblick, T., "Theoretical limitations on the transmission of data from analog sources," *IEEE Trans. Inf. Theory*, vol. 11, no. 4, pp. 558–567, 1965.

[17] Q. Wang, S. Kulkarni, and S. Verdu, "Divergence estimation of continuous distributions based on data-dependent partitions," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3064–3074, Sept. 2005.

[18] G. J. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.

[19] E. Eleftheriou and S. Olcer, "Low-density parity-check codes for digital subscriber lines," in *Proc. Int. Conf. on Communications (ICC)*, vol. 3, 2002, pp. 1752–1757.