Capacity of Power Constrained Memoryless AWGN Channels with Fixed Input Constellations*

Nedeljko Varnica, Xiao Ma and Aleksandar Kavčić
Division of Engineering and Applied Sciences, Harvard University
33 Oxford Street, Cambridge, MA 02138
[varnica,xiaoma,kavcic]@hrl.harvard.edu

Abstract— **We propose a numerical method to compute the capacity of a power constrained memoryless additive white Gaussian noise (AWGN) channel with finite and fixed input alphabets. The method is based on a two-part algorithm. The first part is a modified version of the Blahut-Arimoto algorithm and the second part is a simple maximization algorithm over a single parameter. The optimal input distribution we obtain can be utilized to construct *probabilistic* codes for this channel. These codes promise to bridge the shaping gap between the uniform-input information rate and the capacity of the channel.**

I. Introduction

The channel capacity was introduced by Shannon in [1]. With the exception of some special cases (such as the memoryless power constrained additive white Gaussian noise (AWGN) channel and the binary symmetric channel) there is no closed form expression for the channel capacity [2] [3]. Blahut and Arimoto independently introduced a numerical method to compute the capacities of discrete input memoryless channels in [4] and [5]. For the capacities of channels with memory we refer the reader to [6].

In this contribution we are interested in computing the capacity of the memoryless AWGN channel with two constraints: 1) finite-size input signal constellation constraint and 2) a maximum power constraint. We also consider nearly optimal codes derived from the optimal input distribution, although the code construction method itself is beyond the scope of this paper.

We begin with a review of the memoryless AWGN channel and some well-known capacity results in Section II. In Section III we first formulate the problem and then present a two-part algorithmic solution with the proof of the validity of our algorithm. In Section IV we show the numerical results of our capacity computation method. Section V discusses constructions of (nearly) optimal codes based on a probabilistic construction method.

II. Background

The ideal discrete-time memoryless AWGN channel [3][7] is characterized by

$$Y_k = X_k + W_k, \quad k = 1, 2, \dots \tag{1}$$

At time k, the channel input X_k is a 2-dimensional (2D) random variable whose realization is $x \in \mathbb{R}^2$ (generalizations to multi-dimensional cases are straightforward). W_k is a 2D Gaussian noise random variable with zero mean and covariance matrix $\sigma^2 \mathbf{I}_2$, where $\mathbf{I}_2$ is the 2×2 identity matrix. The random variables $W_k,\ k = 1, 2, \cdots$ are assumed to be independent identically distributed (i.i.d.). From (1), the probabilistic law of the channel is characterized by the conditional probability density function

$$f_{Y_k|X_k}(y|x) = f_{Y|X}(y|x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|y-x\|^2}{2\sigma^2}\right) , \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm. The average mutual information rate $I(X;Y)$ is given by [2]

$$I(X;Y) = \mathrm{E}_{X,Y}\left[\log \frac{f_{Y|X}(y|x)}{f_Y(y)}\right], \tag{3}$$

where $\mathrm{E}_{X,Y}$ denotes the expectation taken with respect to the variables X and Y. The channel capacity is given by

$$C = \sup I(X;Y) \quad [\text{bits/2D}], \tag{4}$$

where the supremum is taken over all (possibly constrained) probability distributions of the input random variable X. The average input signal power is denoted by $\mathrm{E}_X(\|X\|^2)$. Using this notation, we distinguish several constrained scenarios:

Problem 1. (no input constraint) The input X can be any 2D random variable. In this case, $C = \infty$, and there is no optimal input (the supremum is not achievable).

Problem 2. (power constraint) The input X satisfies $\mathrm{E}_X(\|X\|^2) \le P_{av}$, where P_{av} is given. In this case, $C = \log(1 + P_{av}/(2\sigma^2))$, and the optimal input X is a zero-mean Gaussian random variable with the covariance matrix $(P_{av}/2) \cdot \mathbf{I}_2$, see [1].

Problem 3. (input signal constellation constraint) The realizations of the input X are constrained to a finite alphabet $\mathcal{X}$. For some special cases, for example, if $\mathcal{X}$ consists of QPSK signal points, it can be shown that the optimal input distribution is uniform, and that the capacity can be expressed as an integral. In general, neither the input distribution nor the capacity is known in closed form. However, the problem can be solved numerically using the Blahut-Arimoto algorithm [4][5].

Problem 4. (power and input signal constellation constraints) The realizations of the input X are constrained to a finite alphabet $\mathcal{X}$. Furthermore, it is required that $\mathrm{E}_X(\|X\|^2) \le P_{av}$, where P_{av} is given. This is a special

* This work was supported by the National Science Foundation under Grant No. CCR-0118701

case of the problem given in [4] and [8], so it can be solved using the general methods presented there.

In the next section, we describe a generalization of **Problem 4**. We assume that we are given a channel and a power constraint. We also assume that the signaling constellation is finite and fixed (for example 16-QAM, see Figure 1A). As designers, we have the freedom of choosing the amplifier. Our task is to choose the amplification constant and the probability mass function (pmf) defined over the input alphabet in order to maximize the information rate.

III. Capacity Computation

A. Problem Formulation

We consider a memoryless AWGN channel. The channel input is constrained to a finite size alphabet $\mathcal{X} = \{x_1, x_2, ..., x_M\} \subset \mathbb{R}^2$. We denote the input pmf by $\underline{p} = [p_1, p_2, ..., p_M]$, where $p_i = Pr(X = x_i)$. The transmitted power for the *i-th* symbol in the constellation is $\|x_i\|^2$. We allow the amplitude levels of the signal points to change to model an amplifier in the transmitter. This is done by multiplying the constellation by a scaling parameter $\alpha > 0$, as shown in Figure 1 with the 16-QAM input constellation. We assume that the per-dimension noise variance σ^2 and the maximum average power P_{av} are given. At the receiving end we observe a noisy continuous-valued random variable Y. We are interested in computing

$$C = \max_{\underline{p}, \alpha} I(X;Y) \tag{5}$$

under the constraints

$$\sum_{i=1}^{M} p_i = 1 \tag{6}$$

$$\sum_{i=1}^{M} p_i \alpha^2 \|x_i\|^2 \leq P_{av} \tag{7}$$

and $$\alpha > 0.$$

B. The Blahut-Csiszár Solution

For a fixed α, this problem is reduced to **Problem 4**. Therefore, to solve the problem given in (5)-(7), we can embed the solution to **Problem 4** (given in [4] [8]) into an outer loop that varies α:

Solution 1 (Blahut-Csiszár)
step 1 For a given $\alpha > 0$, compute $C_\alpha = \max_{\underline{p}} I(X;Y)$ under the constraints (6) and (7).
step 2 Find $C = max_{\alpha>0} C_\alpha$.

Note that to compute C_α in step 1, the method described in [8] first computes a parametrized function $F(\gamma)$, and then finds the solution C_α as $C_\alpha = \min_{\gamma \geq 0}(F(\gamma) + \gamma P_{av})$ (see [8], page 138). The obvious problem with this approach is the minimization over γ because the function $F(\gamma)$ cannot be expressed in closed form, and γ can take

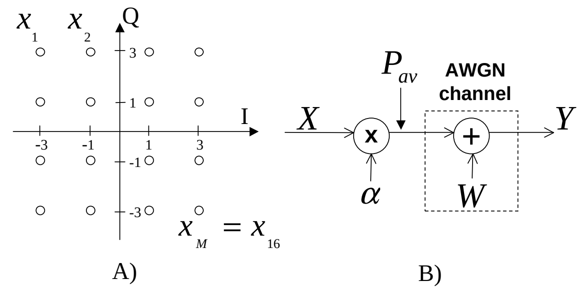

Fig. 1. A) 16-QAM example of the input signal constellation and B) the channel model. The constellation symbols are denoted by $x_1, x_2, ..., x_M$. The distribution vector of the symbols is $\underline{p} = [p_1, p_2, ..., p_M] = [Pr(X = x_1), Pr(X = x_2), ..., Pr(X = x_M)]$. The transmitted power for the *i-th* symbol in the constellation is $\|x_i\|^2$. The constellation is allowed to "breathe" - the scaling factor is denoted by α. The channel is an AWGN channel with the independent noise components.

any value in the interval $[0, \infty)$, which makes the exhaustive search computationally prohibitive.

C. New Solution: The Two-Part Algorithm

We introduce a simpler and computationally feasible alternative to **Solution 1**. Our solution, **Solution 2**, rests on the following proposition.

Proposition 1: The constraint $\sum_i p_i \alpha^2 \|x_i\|^2 \leq P_{av}$ is equivalent to the constraint $\sum_i p_i \alpha^2 \|x_i\|^2 = P_{av}$.
Proof: The proof is given in the appendix. □

Solution 2 (the two-part algorithm)
Part 1: Given $\alpha > 0$, compute $I_\alpha = \max_{\underline{p}} I(X;Y)$ under the constraints

$$\sum_i p_i = 1 \tag{8}$$

and $$\sum_i p_i \alpha^2 \|x_i\|^2 = P_{av} \ . \tag{9}$$

Part 2: Find $C = max_{\alpha>0} I_\alpha$.

Note that we have replaced the original constraint $\sum_i p_i \alpha^2 \|x_i\|^2 \leq P_{av}$ with the constraint $\sum_i p_i \alpha^2 \|x_i\|^2 = P_{av}$, which is justified by Proposition 1. This implies that $I_\alpha \leq C_\alpha$ and that our strategy is different from **Solution 1**.

D. Details of the Two-Part Algorithm

In this subsection we provide the details of Parts 1 and 2 of the two-part-algorithm given in the Subsection III-C. We start by rewriting the mutual information rate as

$$I(X;Y) = H(X) - H(X|Y) = \sum_{i=1}^{M} p_i \left[\log_2 \left(\frac{1}{p_i} \right) + T_i \right].$$

Here
$$T_i = \int f_{Y|X}(y|x_i) \log_2 P_{X|Y}(x_i|y) dy$$

$$= \mathrm{E}_Y \left[\frac{Pr(X = x_i|Y) \log_2 (Pr(X = x_i|Y))}{p_i} \right],$$

and $P_{X|Y}(x_i|y) = Pr(X = x_i|Y = y)$. The vector $\underline{T} = [T_1, T_2, ..., T_M]$ can be determined from the a posteriori probabilities $P_{X|Y}(x_i|y)$. If a closed form expression for the vector $\underline{T}$ can not be derived, we can compute it via a Monte-Carlo simulation [6].

We now present in detail the two-part solution given in Subsection III-C.

Part 1 (Modified Blahut-Arimoto algorithm)

step 1: Fix α. Choose $p_i > 0$, for $i = 1, 2..., M$ where $\sum_{i=1}^{M} p_i = 1$.
step 2: (**Expectation**) For fixed p_i $(i = 1, 2..., M)$, compute

$$T_i = \mathrm{E}_Y \left[\frac{Pr(X = x_i|Y) \log_2(Pr(X = x_i|Y))}{p_i} \right].$$

step 3: (**Maximization**) For fixed T_i $(i = 1, 2..., M)$, find

$$\underline{p} = \arg\max_{\underline{r}} \left[\sum_{i=1}^{M} r_i \left[\log_2 \left(\frac{1}{r_i} \right) + T_i \right] \right]$$

under constraints $\sum_i r_i = 1$ and $\sum_i r_i \alpha^2 \|x_i\|^2 = P_{av}$.
Termination: Repeat steps 2-3 until convergence. At the end, store the result in $I_\alpha = \sum_i p_i \left[\log_2 \left(\frac{1}{p_i} \right) + T_i \right]$.

We have implicitly assumed that the algorithm converges. We state this as a theorem and give a proof in Subsection III-E.

To find the maximum in step 3 of **Part 1**, we use the method of Lagrange multipliers. Let

$$F(\underline{p}, \underline{T}, \lambda_1, \lambda_2) = \sum_{i=1}^{M} p_i \left[\log_2 \left(\frac{1}{p_i} \right) + T_i \right] + \lambda_1 \sum_{i=1}^{M} p_i + \lambda_2 \sum_{i=1}^{M} p_i \alpha^2 \|x_i\|^2.$$

Here, the scalars λ_1 and λ_2 represent Lagrange multipliers. Differentiating $F(\underline{p}, \underline{T}, \lambda_1, \lambda_2)$ with respect to p_i and setting the derivatives to zero (for all $i = 1, 2, ..., M$) we obtain

$$p_i = \frac{2^{T_i + \lambda_2 \alpha^2 \|x_i\|^2}}{\sum_j 2^{T_j + \lambda_2 \alpha^2 \|x_j\|^2}}, \tag{10}$$

and
$$\sum_i (P_{av} - \alpha^2 \|x_i\|^2) 2^{T_i} \cdot 2^{\lambda_2 \alpha^2 \|x_i\|^2} = 0. \tag{11}$$

The solution of (10) and (11) achieves the global maximum of $G(\underline{p}, \underline{T}) = \sum_i p_i \left[\log_2 \left(\frac{1}{p_i} \right) + T_i \right]$ under the constraints (8) and (9) on the interval $0 \leq p_i \leq 1$ $(i = 1, 2, ..., M)$. To show this, we note that $G(\underline{p}, \underline{T})$ is concave in $\underline{p}$, since $\sum_i p_i \log_2(\frac{1}{p_i})$ is concave in $\underline{p}$, and $\sum_i p_i T_i$ is linear in $\underline{p}$, over the convex set given by the linear constraints (8) and (9). The solution λ_2 of (11) can be evaluated numerically (we use a simple interval halving procedure).

Note that in **Part 1** we have computed the value I_α by fixing the amplitude level α. Clearly, we now need to maximize I_α for all $\alpha > 0$. To do that, we use the quantization step δ with which we vary the amplitude α. Note that we can limit the search space to $\alpha_{min} < \alpha < \alpha_{max}$. Here α_{max} corresponds to the case when only the point(s) nearest to the origin (in the Euclidean distance sense) have non-zero probabilities, and α_{min} corresponds to the case when only the point(s) furthest from the origin have non-zero probabilities[1]. For the example shown in Figure 1, $\alpha_{max} = \sqrt{P_{av}/2}$ and $\alpha_{min} = \sqrt{P_{av}/18}$.

Part 2 (Maximizing over α)

step 1: Initialize $\alpha = \alpha_{min} + \delta$.
step 2: Perform **Part 1** (the modified Blahut-Arimoto algorithm) and compute I_α.
step 3: Increase α by δ $(\alpha \leftarrow \alpha + \delta)$.
step 4: Repeat steps 2-3 if $\alpha < \alpha_{max}$.
step 5: Find the maximum $C_\delta = \max_\alpha I_\alpha$ and the corresponding distribution $\underline{p}^*_\delta$ that achieves this maximum.

We now see that our strategy is simpler than **Solution 1**. **Part 1** of our solution is a convergent iterative algorithm (as opposed to the computationally prohibitive minimization over $\gamma \in [0, \infty)$ in **Solution 1**). **Part 2** is a maximization on the bounded interval $(\alpha_{min}, \alpha_{max})$, which is the same as in **Solution 1**. The maximum is obtained to within the precision of the quantization step δ.

E. Convergence Proof

Lemma 1: Assume $Q(x_i|y)$ is any given "a posteriori probability function" and define $g(\underline{p}, \underline{T}(Q)) = \sum_i p_i \left[\log_2(\frac{1}{p_i}) + T_i(Q) \right]$. Then, for a fixed parameter α

$$I_\alpha = \max_{\underline{p}} I(X; Y) = \max_{\underline{p}} \max_{Q} g(\underline{p}, \underline{T}(Q)).$$

Proof: The proof is given in the appendix. □

Theorem: The algorithm presented as **Part 1** (the modified Blahut-Arimoto algorithm) converges. As $\delta \to 0$ the probability vector obtained in **Part 2** becomes the optimal distribution vector, i.e., $\underline{p}^*_\delta \to \underline{p}^*$, where $\underline{p}^*$ denotes the optimal distribution.
Sketch of the proof: First we prove that the modified Blahut-Arimoto algorithm converges. Denote by $\underline{p}^{(i)}$ the probability distribution in the ith iteration. The corresponding expectation vector is denoted by $\underline{T}^{(i)}$. From Lemma 1 and the Maximization step (step 3) of **Part 1**, we have

$$g(\underline{p}^{(i)}, \underline{T}^{(i)}) \leq g(\underline{p}^{(i+1)}, \underline{T}^{(i)}) \leq g(\underline{p}^{(i+1)}, \underline{T}^{(i+1)}) \leq g(\underline{p}^{(i+2)}, \underline{T}^{(i+1)}) \leq g(\underline{p}^{(i+2)}, \underline{T}^{(i+2)}) \leq \cdots \leq \log_2 |\mathcal{X}|.$$

Hence there exists a constant $\hat{I}_\alpha$ such that

$$\lim_{i \to \infty} g(\underline{p}^{(i)}, \underline{T}^{(i)}) = \hat{I}_\alpha. \tag{12}$$

[1]In M-QAM we have four such points due to the symmetry of the constellation.

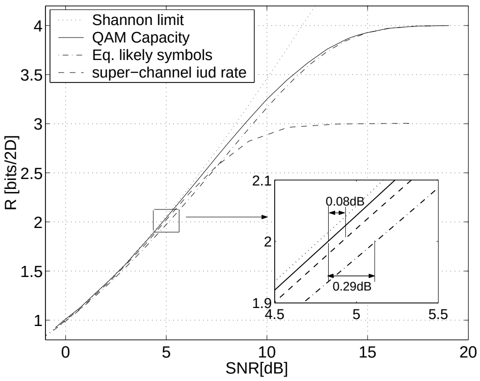

Fig. 2. 16-QAM capacity, together with the Shannon limit and the 16-QAM mutual information rate when the input symbols are equiprobable. Also shown is the i.u.d. super-channel information rate (Section V). The i.u.d. super-channel information rate is within 0.08dB of the QAM capacity at rate $r = 2$bits/2D.

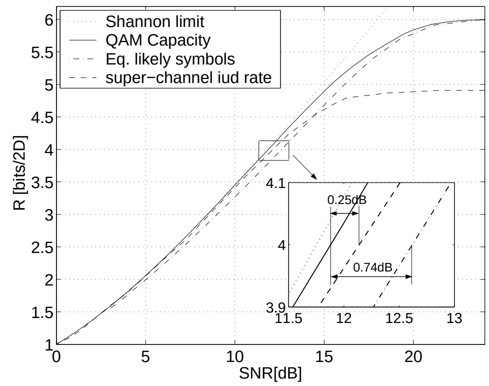

Fig. 3. 64-QAM capacity, together with the Shannon limit and the 64-QAM mutual information rate when the input symbols are equiprobable. Also shown is the i.u.d. super-channel information rate (Section V). The i.u.d. super-channel information rate is within 0.25dB of the QAM capacity at rate $r = 4$bits/2D.

To prove that $\hat{I}_\alpha = I_\alpha = \max_{\underline{p}} I(X;Y)$ for fixed α, we refer to [9], chapter 10. Since both constraints (8) and (9) are linear in $\underline{p}$, it can be verified that $g(\underline{p}, \underline{T})$ satisfies the conditions required in [9].

The second part of the theorem is obvious, since I_α is a continuous function of α. $\square$

IV. Numerical Results

We use the algorithm proposed in the previous section to compute the capacities of the power constrained AWGN channels with 16-QAM and 64-QAM input signal constellations. More precisely, we compute the capacities of these channels to within the precision of the quantization step δ. In our simulations we set $\delta^2/P_{av} = 0.01$. The results are shown in Figures 2 and 3. Here, the signal-to-noise ratio is defined as $SNR = 10\log_{10}(P_{av}/(2\sigma^2))$.

We can see from Figure 2 that for small-size signal constellations, only very little can be gained by using non-uniform input distributions. For example, in the 16-QAM constellation, the difference between the uniform-input information rate and the capacity (this difference is also known as the shaping gap [7]) is about 0.29dB at rate $r = 2$ bits/2D. However, the shaping gap for large-size constellations is more significant. For the 64-QAM constellation this gap is about 0.74dB at rate $r = 4$ bits/2D. In the limit, for very large QAM constellations, the shaping gap becomes 1.53dB [7]. A simple code construction method to bridge this gap is proposed in Section V.

V. Code Construction Proposal

We are concerned with the construction of nearly optimal codes for the power constrained AWGN channels with finite input alphabets. One possible construction is the multilevel coding method in [10], where the authors used a Gaussian approximation to obtain a good lower bound on the QAM-capacity and the nearly optimal input distribution. Another method, which is explained in [11], is to use the optimal input distribution (here obtained by the two-part algorithm) to construct a trellis code which we call a *matched information rate* code. The strategy proposed in [11] (shown in Figure 4) is to use this trellis code as the inner code and an LDPC (low density parity check) code [12] as the outer linear code.

In this paper we are only concerned with the inner code, which is constructed to maximize its i.u.d. (independent uniformly distributed) super-channel information rate[2]. We speculate that a properly constructed outer LDPC code can nearly achieve this information rate. We constructed a 16-state inner trellis code for 16-QAM and a 64-state inner trellis code for 64-QAM. The trellis construction method is beyond the scope of this paper, and is based on the probabilistic method similar to the one in [13] combined with Ungerboeck's trellis construction rules [14]. The trellises are not given in this paper because of their excessive sizes, but can be found in [15]. In Figures 2 and 3 we plot the i.u.d. information rates of the super-channels for the constructed nearly optimal 16-QAM and 64-QAM trellis codes, respectively.

The rate of the 16-state trellis code constructed for 16-QAM is $r_{in} = 3$ bits/symbol, whereas the rate of the 64-state trellis code constructed for 64-QAM is $r_{in} = 5$ bits/symbol. The trellis codes were optimized at rates $r = 2$ bits/symbol and $r = 4$ bits/symbol for the 16-QAM and 64-QAM signal constellations, respectively. The i.u.d. information rates of these super-channels are very close to the capacities of their respective channels. The distances from the capacities are 0.08dB (at $r = 2$ bits/symbol) for 16-QAM and 0.25dB (at $r = 4$ bits/symbol) for 64-QAM.

In [11] we have shown that one can construct nearly optimal outer LDPC codes for a binary input Gaussian channel with intersymbol interference. By analogy, we believe that nearly optimal concatenated codes for the *memory-*

[2]The i.u.d. information rate is defined as the information rate when the input sequence is i.u.d.

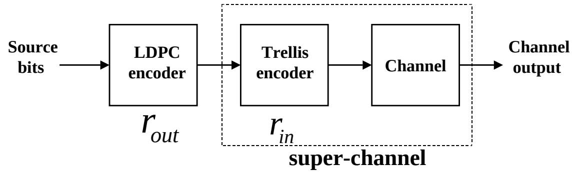

Fig. 4. Concatenated coding scheme. The inner code is a trellis code with rate r_{in}. The outer code is an LDPC code with rate r_{out}. The total rate equals $r = r_{in} \cdot r_{out}$.

less AWGN channels considered in this paper can be constructed by an appropriate optimization of LDPC codes on the super-channel. More about the optimization of the outer LDPC code can be found in [11] and the references therein.

VI. Conclusion

We have presented a numerical method to compute the capacity of the power constrained memoryless AWGN channel with a finite input constellation. A two-part algorithm was proposed to compute the optimal input distribution and the optimal scaling parameter of the amplifier in the transmitter. Our numerical results show that for the 64-QAM constellation, the information rate gain over the uniform-input information rate is about 0.74dB. A proposed probabilistic code construction strategy promises to very closely approach the capacity of the channel and effectively close the shaping gap.

APPENDIX

Proof of Proposition 1: Let α_1 and α_2 be two scaling (amplification) factors, and $Y^{(1)}$ and $Y^{(2)}$ be the corresponding output random variables, respectively. Assume that $\alpha_1^2 \mathrm{E}_X(\|X\|^2) < P_{av}$ and $\alpha_2^2 \mathrm{E}_X(\|X\|^2) = P_{av}$. It suffices to prove that $I(X;Y^{(2)}) \geq I(X;Y^{(1)})$. Intuitively, it must be true since $\alpha_2 X$ has more energy than $\alpha_1 X$. This can also be verified by considering the two equivalent channels in Fig. 5. By using the data-processing theorem [3], we have

$$I(X;Y^{(2)}) \geq I(X;Z^{(2)}) = I(X;Z^{(1)}) = I(X;Y^{(1)}).$$

The last equality holds because $Z^{(1)}$ and $Y^{(1)}$ can be determined from each other. □

Proof of Lemma 1: Note that α is given. It suffices to show that

$$\max_Q g(\underline{p}, \underline{T}(Q)) = I(X;Y).$$

In fact,

$$\begin{aligned} & I(X;Y) - g(\underline{p}, \underline{T}(Q)) \\ &= \sum_{i=1}^{M} p_i \int f_{Y|X}(y|x_i) \log \frac{P_{X|Y}(x_i|y)}{Q(x_i|y)}\, dy \\ &= \int f_Y(y) \sum_{i=1}^{M} P_{X|Y}(x_i|y) \log \frac{P_{X|Y}(x_i|y)}{Q(x_i|y)}\, dy \geq 0. \end{aligned}$$

The last inequality holds since, for given y,

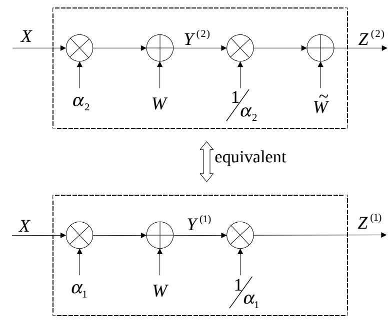

Fig. 5. Two equivalent channel models. Here $\tilde{W}$ is a 2-dimensional Gaussian noise random variable with zero mean and variance $\tilde{\sigma}^2 \mathbf{I_2}$, where $\tilde{\sigma}^2 = \sigma^2/\alpha_1 - \sigma^2/\alpha_2$ and $\alpha_2 > \alpha_1 > 0$.

$$\sum_{i=1}^{M} P_{X|Y}(x_i|y) \log \frac{P_{X|Y}(x_i|y)}{Q(x_i|y)}$$

is the Kullback-Leibler distance [3]. □

References

[1] C. E. Shannon, “A mathematical theory of communications,” *Bell Systems Technical Journal*, vol. 27, pp. 379–423 (part I) and 623–656 (part II), 1948.
[2] R. G. Gallager, *Information Theory and Reliable Communication.* New York: John Wiley and Sons, Inc., 1968.
[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: John Wiley and Sons, 1991.
[4] R. E. Blahut, “Computation of channel capacity and rate-distortion functions,” *IEEE Transactions on Information Theory*, vol. 18, pp. 460–473, July 1972.
[5] S. Arimoto, “An algorithm for computing the capacity of arbitrary discrete memoryless channels,” *IEEE Transactions on Information Theory*, vol. 18, pp. 14–20, January 1972.
[6] A. Kavčić, “On the capacity of Markov sources over noisy channels,” in *Proceedings IEEE Global Communications Conference 2001*, (San Antonio, Texas), Nov 2001. available at http://hrl.harvard.edu/˜kavcic/publications.html.
[7] G. D. Forney, Jr. and G. Ungerboeck, “Modulation and coding for linear Gaussian channels,” *IEEE Transactions on Information Theory*, vol. 44, pp. 2384–2415, October 1998.
[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York-San Francisco-London: Academic Press, Inc., 1981.
[9] R. W. Yeung, *A First Course in Information Theory.* Kluwer Academic/Plenum Publishers, 2002.
[10] U. Wachsmann, R. F. Fischer, and J. B. Huber, “Multilevel codes: theoretical concepts and practical design rules,” *IEEE Transactions on Information Theory*, vol. 45, pp. 1361–1391, July 1999.
[11] X. Ma, N. Varnica, and A. Kavčić, “Matched information rate codes for binary ISI channels,” in *Proc. 2002 IEEE International Symposium on Information Theory*, (Lausanne, Switzerland), July 2002.
[12] R. G. Gallager, *Low-Density Parity-Check Codes.* Cambridge, MA: MIT Press, 1962.
[13] A. Kavčić, X. Ma, M. Mitzenmacher, and N. Varnica, “Capacity approaching signal constellations for channels with memory.” in Allerton’01 Conference, October 2001.
[14] G. Ungerboeck, “Channel coding with multilevel/phase signals,” *IEEE Transactions on Information Theory*, vol. 28, pp. 55–67, January 1982.
[15] Nearly optimal inner trellis codes - available at http://hrl.harvard.edu/˜varnica/publications/trellises.