Coding and Modulation for the Additive Exponential Noise Channel

Alfonso Martinez
Technische Universiteit Eindhoven
Eindhoven, The Netherlands
Email: alfonso.martinez@ieee.org

***Abstract*—Communication across an additive exponential noise (AEN) channel is studied. Constellations are designed for low signal-to-noise ratio, where the minimum energy per bit is not universally attained by all modulations, and for high signal-to-noise ratio, where an equiprobable non-uniform constellation 0.76 dB away from capacity is described. The pairwise error probability for binary codes is similar to that in an equivalent discrete-time Gaussian channel of identical signal-to-noise ratio.**

I. Introduction

As noticed by Verdú [1], the exponential distribution shares a number of special traits with the Gaussian distribution: it has the largest entropy of all non-negative random variables of a given mean and represents the worst possible non-negative additive noise. It is also the limiting form for continuous energy of the geometric distribution, the (discrete) energy distribution of the thermal noise present in radio receivers. Further, in the additive exponential noise (AEN) channel the capacity is given by $\log(1+\mathrm{SNR})$ nats per channel use, when SNR is the signal-to-noise ratio [1]; coded modulation used over complex-valued Gaussian channel (AWGN) admit a natural extension to the AEN channel, where they attain a similar performance.

II. Channel Model

We consider the following discrete-time channel model

$$y_k = x_k + z_k, \qquad k = 1, \dots, n, \tag{1}$$

where y_k is the k-th output, x_k the k-th signal component, and z_k the k-th sample of additive noise. These quantities are non-negative real numbers. The noise z_k is a sample of exponential noise of mean E_n, with density $p_Z(z) = \frac{1}{E_n} e^{-\frac{z}{E_n}} u(z)$, where $u(z)$ is a unit step function. The channel is used under a constraint on the total energy, of the form $\sum_{k=1}^{n} x_k \leq nE_s$, where E_s denotes the (maximum) energy per channel use. The average signal-to-noise ratio SNR is given by $\mathrm{SNR} = E_s/E_n$.

It is easy to establish a correspondence with an equivalent discrete-time AWGN model. Let us denote the AWGN channel variables by appending a prime. We can identify the signal and noise energies, that is $x_k = |x'_k|^2$ and $z_k = |z'_k|^2$. Also E_s remains the maximum permitted energy per channel use. Finally, the exponential density of the noise is the density of the squared amplitude of circularly-symmetric complex Gaussian noise with variance σ^2 when $E_n = \sigma^2$.

The channel variables can be re-normalized by a factor E_n, so that the input has unit average energy and the output conditional density is given by

$$p_{Y|X}(y|x) = e^{-(y-\mathrm{SNR}\,x)} u(y - \mathrm{SNR}\,x). \tag{2}$$

III. Channel Capacity

The capacity of the AEN channel was determined by Verdú [1]. The capacity $\mathrm{C(SNR)}$ of the AEN channel with signal-to-noise ratio $\mathrm{SNR} = E_s/E_n$ is given by

$$\mathrm{C} = \log(1+\mathrm{SNR}). \tag{3}$$

Throughout the paper the logarithms are in natural base and capacities (entropies) in nats, unless otherwise specified. The proof depends on the fact that an exponential density maximizes the differential entropy among the distributions of positive-valued random variables [1]. Recall that the entropy of an exponential random variable with mean ε is $\log(e\varepsilon)$ [1]. Observe that the capacity coincides with the capacity of the equivalent AWGN channel.

Verdú also determined the form of the input distribution which maximizes the output entropy and the capacity of the AEN channel [1]. The input density is given by

$$p_X(x) = \tfrac{E_s}{(E_s+E_n)^2} e^{-\frac{x}{E_s+E_n}} + \tfrac{E_n}{E_s+E_n}\delta(x), \quad x \geq 0. \tag{4}$$

Remark that in the Gaussian case, the capacity-achieving input is Gaussian, the same distribution that the noise has. However, the input here is not purely exponential.

The capacity per unit cost is equal to that of the AWGN channel, namely E_n^{-1}. The minimum bit-energy-to-noise ratio $\mathrm{BNR}_{\min}$ is also given by $\mathrm{BNR}_{\min} = \log 2$, or -1.59 dB.

IV. Coded Modulation in the AEN Channel

A. Motivation

In the AEN channel, the input density which achieves the capacity, Eq. (4), depends on the value of the average signal E_s and the noise E_n energy levels. We next analyze an alternative input, scaled constellations $\mathrm{SNR}\,\mathcal{X}$, where $\mathcal{X}$ has unit energy.

The constellation points for amplitude modulation on the Gaussian channel have the form

$$\pm\beta_{\mathrm{PAM}}\left\{\frac{1}{2} + (i-1)\right\}, \quad i = 1, \dots, 2^{m-1}, \tag{5}$$

where β_{PAM} is a normalization factor. A straightforward extension to the AEN channel would be to consider pulse energy modulation (PEM), a set of the form, $\beta\{i-1\}$, for $i = 1, \ldots, 2^m$, where β is another normalization factor. We consider a more general constellation $\mathcal{X}_\lambda$ by choosing points of the form

$$\beta_{\text{PEM}}\big\{(i-1)^\lambda\big\}, \quad i = 1, \ldots, 2^m, \tag{6}$$

where $\lambda > 0$ is a positive real number and $\beta_{\text{PEM}}^{-1} = \frac{1}{2^m}\sum_{i=1}^{2^m}(i-1)^\lambda$ is a normalization factor. The free parameter λ allows us to optimize its value for a given channel property, such as the pairwise error probability or the mutual information, as we will see later.

As the number of points 2^m increases, even though the points are used with identical probabilities, the discrete constellations $\mathcal{X}_\lambda$ approach a continuous distribution $\mathcal{X}_\lambda^\infty$.

Proposition 1. *The constellation $\mathcal{X}_\lambda$ has bounded support in the interval $[0, \lambda+1]$. As $m \to \infty$, $\mathcal{X}_\lambda$ approaches a continuous constellation $\mathcal{X}_\lambda^\infty$ with non-uniform density*

$$p_X(x) = \frac{x^{\frac{1}{\lambda}-1}}{\lambda(1+\lambda)^{\frac{1}{\lambda}}} \tag{7}$$

in the interval $[0, \lambda+1]$ and zero outside. The differential entropy of a scaled constellation $\alpha\mathcal{X}_\lambda^\infty$, $\alpha > 0$, is given by $h(\alpha\mathcal{X}_\lambda^\infty) = 1 - \lambda + \log\big(\lambda(\lambda+1)\big) + \log\alpha$.

Here, and in the remainder of the paper, we omit the proof. In this case, the result is derived from the formulas for the density of a random variable as a function of the density of another.

For $\lambda > 1$ the density, not being flat as in the uniform density, is somewhat closer to the optimum density in Eq. (4).

B. Constrained Capacity

We now move on to study the constrained capacity for pulse energy modulation (PEM) with a constellation set $\mathcal{X}$. Symbols are used with probabilities $P(x)$, and have arbitrary first- and second-order moments, $\mu_1(\mathcal{X})$ and $\mu_2(\mathcal{X})$. For later use, we add a point at infinity, defined as $x_{|\mathcal{X}|+1} = \infty$, and sort the symbols in increasing order, i. e. $x_1 \leq x_2 \leq \ldots \leq x_{|\mathcal{X}|}$.

From the definition of mutual information we derive

Proposition 2. *The constrained capacity $C_\mathcal{X}(\text{SNR})$ for signalling with average signal-to-noise ratio* SNR *over a modulation set $\mathcal{X}$ with probabilities $P(x)$ is given by*

$$C_\mathcal{X}(\text{SNR}) = -\sum_x P(x) \sum_{x_j \geq x}\big(e^{\text{SNR}(x-x_j)} - e^{\text{SNR}(x-x_{j+1})}\big)\times \\ \times \log\left(\sum_{x' \leq x_j} P(x')e^{-\text{SNR}(x-x')}\right). \tag{8}$$

Here we have explicitly carried out the integration and rearranged the resulting terms.

Figure 1 shows the constrained capacity for uniform 2^m-PEM, m being an integer and $\lambda = \frac{1}{2}(1+\sqrt{5})$. The reason for this specific choice will become apparent later. In Fig. 1a, capacity is plotted as a function of SNR, whereas in Fig. 1b, the bit-energy-to-noise ratio BNR, defined as $\text{BNR} = \frac{\text{SNR}}{\text{C}_\mathcal{X}}\log 2$, is given as a function of the constrained capacity.

(a) Capacity as a function of SNR.

(b) BNR as a function of the capacity.

Fig. 1: CM Capacity for PEM with $\lambda = \frac{1}{2}(1+\sqrt{5})$.

As clearly seen in Fig. 1a, the minimum bit-energy-to-noise ratio BNR_{min} is not universally attained by arbitrary constellation sets. For instance, the asymptotic value of BNR as $\text{SNR} \to 0$, denoted by BNR_0, is 0 dB for 2-PEM. In Section IV, we compute the first two terms of the Taylor expansion of the capacity around $\text{SNR} = 0$. It would be straightforward to exploit them to determine the value of BNR_0 as a function of m and λ, and provide constellation sets for which BNR_0 is arbitrarily close to BNR_{min}.

At high SNR, the capacity curves approach a common envelope as the number of constellation points increases. To emphasize this fact, a dotted line is shown; its form will be related later to the differential entropy of the input $\mathcal{X}_\lambda^\infty$.

The behaviour at low SNR of $\text{C}_\mathcal{X}$ is described in

Theorem 3. *The constrained capacity $C_\mathcal{X}(\text{SNR})$ of a set $\mathcal{X}$ at signal-to-noise ratio* SNR *has a Taylor series expansion $C_\mathcal{X}(\text{SNR}) = c_1\,\text{SNR} + c_2\,\text{SNR}^2 + O(\text{SNR}^3)$, as $\text{SNR} \to 0$, with c_1 and c_2 given by*

$$c_1 = -\sum_{j=1}^{|\mathcal{X}|-1}(x_{j+1}-x_j)q_j\log q_j \tag{9}$$

$$c_2 = \tfrac{1}{2}\sigma^2(\mathcal{X}) + \sum_{j=1}^{|\mathcal{X}|-1}(x_{j+1}-x_j)q'_j\log q_j, \tag{10}$$

where $\sigma^2(\mathcal{X}) = \mu_2(\mathcal{X}) - \mu_1^2(\mathcal{X})$, $q_j = \sum_{x' \leq x_j}P(x')$ and $q'_j = \frac{1}{2}(x_j + x_{j+1})q_j - \sum_{x' \leq x_j}x'P(x')$.

We characterize the behaviour at high SNR by assuming that the input can be described by the limiting continuous density in Eq. (7) and that the output (differential) entropy $h(Y)$ can be approximated by the differential entropy of the input $h(X)$. The differential entropy of the scaled input

SNR $\mathcal{X}_\lambda^\infty$ was determined in Proposition 1,

$$\log \mathrm{SNR} + 1 - \lambda + \log\big(\lambda(\lambda+1)\big). \quad (11)$$

Since noise has mean 1, its entropy is $h(Z) = \log(e)$, and the asymptotic expansion of the capacity at high SNR is

$$\mathrm{C}_{\text{asymp}}(\mathrm{SNR}) \simeq \log \mathrm{SNR} - \lambda + \log\big(\lambda(\lambda+1)\big). \quad (12)$$

The capacity is largest if λ is the solution of

$$-1 + \frac{1}{\lambda} + \frac{1}{\lambda+1} = 0 \Longrightarrow \lambda^2 - \lambda - 1 = 0, \quad (13)$$

that is for $\lambda = \frac{1+\sqrt{5}}{2}$, the golden number.

The capacity of the AEN channel asymptotically behaves as $\log \mathrm{SNR}$ for large SNR. We compute the energy loss incurred by using a non-optimal constellation by setting $\log \mathrm{SNR} = \log \mathrm{SNR}' - \lambda + \log\big(\lambda(\lambda+1)\big)$, or equivalently

$$\frac{\mathrm{SNR}'}{\mathrm{SNR}} = \frac{e^\lambda}{\lambda(\lambda+1)}. \quad (14)$$

The lowest loss, achieved for $\lambda = \frac{1}{2}(1+\sqrt{5})$, is about 0.76 dB, lower than the loss of 1.53 dB for uniform square QAM constellations in the Gaussian channel. Uniform PEM, with $\lambda = 1$, suffers from an energy loss $2/e$, or approximately 1.33 dB, the same value as a uniform distribution in a circle for the Gaussian channel [2]. The approximations used to derive this energy loss are very accurate. At high SNR, PEM constellations in the AEN channel are more energy-efficient than QAM or PAM modulations in the AWGN channel. This is opposite to the situation at low SNR, where typical AWGN constellations require a lower SNR to achieve the same capacity.

V. Error Probability in the AEN Channel

A. Introduction

In previous sections, we have studied the constrained capacity for pulse-energy modulation. We complement our analysis by estimating the pairwise error probability for binary transmission. Since the pairwise error probability is a key element in the analysis of the code performance, the tools developed here may prove of use in future research on code design.

As we did in [3], we estimate the pairwise error probability at Hamming distance d by the tail probability

$$\mathrm{pep}(d) = \Pr\left(\sum_{j=1}^{d} \Lambda_j > 0\right) + \frac{1}{2}\Pr\left(\sum_{j=1}^{d} \Lambda_j = 0\right), \quad (15)$$

where the variables Λ_j is independent and identically distributed log-likelihood ratios with sample value

$$\lambda = \log \frac{P_{B|Y}(\bar{b}|y)}{P_{B|Y}(b|y)} = \log \frac{\sum_{x \in \mathcal{X}_i^{\bar{b}}} p_{Y|X}(y|x)}{\sum_{x \in \mathcal{X}_i^{b}} p_{Y|X}(y|x)}. \quad (16)$$

The density of λ is determined as follows. An input bit '0' (the all-zero codeword is transmitted) is mapped to either $b = 0$ or $b = 1$ with probability 1/2, then an index i, for $i = 1, \dots, m$ (the modulation has 2^m points), is chosen, and finally one of the $\frac{|\mathcal{X}|}{2}$ symbols in $\mathcal{X}_i^b$, the set of constellation symbols with bit b in the i-th position of the binary label, selected and sent over the channel to generate the output y. The density is thus a function of all possible input choices and channel realizations.

We prove that the pairwise error probability is very close to that of an AWGN channel with identical signal-to-noise ratio. This closeness strongly suggests that binary codes will achieve essentially the same performance in these two channels.

B. Binary Modulation

We consider first binary modulation (2-PEM). At the transmitter, each bit b_k is mapped onto a binary symbol x_k, with $x_k = \{0, +2\}$; the mapping rule μ is $x_k = 0$ if $b_k = 0$ and $x_k = +2$ if $b_k = 1$, used with probability 1/2, and its complement $\bar{\mu}$, used with probability 1/2. The choice between μ and $\bar{\mu}$ is known at the receiver. This step renders the channel output-symmetric. The output y_k is given by the sum

$$y_k = \mathrm{SNR}\, x_k + z_k, \quad (17)$$

where z_k is a sample of exponential noise of mean 1. The inclusion of fading, especially of Nakagami/gamma fading would be straightforward. Most of the results reported here admit a natural extension to this case.

The natural counterpart of the results for the Gaussian channel, where the pairwise error probability [4] is

$$\mathrm{pep}(d) = Q(\sqrt{2d\,\mathrm{SNR}}) \simeq \frac{1}{2\sqrt{\pi d\,\mathrm{SNR}}} e^{-d\,\mathrm{SNR}}, \quad (18)$$

where $Q(x)$ is the Gaussian tail function, is the following

Theorem 4. *For 2-PEM, the saddlepoint approximation to the pairwise error probability for Hamming distance d is*

$$\mathrm{pep}(d) \simeq \begin{cases} \frac{2}{\sqrt{2\pi d}}\left(\sum_{j=0}^{\frac{d-1}{2}} e^{-2j\,\mathrm{SNR}}\right) e^{-(d+1)\,\mathrm{SNR}}, & d \text{ odd}, \\ \frac{2}{\sqrt{2\pi d}}\left(\frac{1}{2} + \sum_{j=1}^{\frac{d}{2}} e^{-2j\,\mathrm{SNR}}\right) e^{-d\,\mathrm{SNR}}, & d \text{ even} \end{cases}. \quad (19)$$

The Chernoff bound to the pairwise error probability is

$$\mathrm{pep}(d) \le e^{-d\,\mathrm{SNR}}. \quad (20)$$

A number of conclusions follow from this theorem. First, the Chernoff bound coincides with the value of BPSK in Gaussian channels [4]. The error performance of both modulation formats should be similar for a given SNR.

Second, the error probability decays as $e^{-(d+1)\,\mathrm{SNR}}$ when d is odd, a slightly faster decay than that given by the Chernoff bound. This effect may create some room to design efficient codes for this channel, since the error probability for d odd is similar to the error probability with $d+1$, "buying", so to speak, some Hamming distance from the channel itself.

A special case of d odd is uncoded transmission, for which $d = 1$, and the exact bit error rate is $\Pr_b = \frac{1}{2} e^{-2\,\mathrm{SNR}}$. This value is very close to the approximation in Eq. (19),

$$\Pr_b \simeq \frac{2}{\sqrt{2\pi}} e^{-2\,\mathrm{SNR}} \simeq 0.79788 e^{-2\,\mathrm{SNR}}. \quad (21)$$

Figure 2 depicts the word error probabilities for several values of d. The simulated values match well with the approximation in Theorem 4, except that the Chernoff bound does not give the correct dependence with d for odd d.

Fig. 2: Comparison of simulation and saddlepoint approximation for pairwise error probability of 2-PEM, $d = 1, \dots, 6$.

The main element of the proof of Theorem 4 is the fact that binary PEM in the AEN channel is a Z-channel with inputs used with probability 1/2. This leads to a following general statement for the Z-channel:

Proposition 5. *In the Z-channel with transition probability ϵ and inputs used with probability 1/2, the pairwise error probability at Hamming distance d is given by Eq.* (19) *(saddlepoint approximation) or Eq.* (20) *(Chernoff bound), by setting $\epsilon = e^{-2\,\mathrm{SNR}}$.*

C. Bit-Interleaved Coded Modulation

Bit-interleaved coded modulation (BICM) is an efficient way of coupling good binary codes for use with non-binary modulations [5]. In BICM, binary codewords are mapped onto an array of channel symbols $\big(x_1, \dots, x_n\big)$ by bit-interleaving the binary codeword and mapping it on the signal constellation $\mathcal{X}$ with a binary labelling rule μ. With no loss of generality, the constellation set is assumed to have 2^m elements, so that m bits are necessary to index one symbol. Bit j-th is interleaved onto position $\pi(j)$ and accordingly to symbol index $k = \lceil \frac{\pi(j)}{m} \rceil$, and $k = 1, \dots, n$.

It is possible to approximate the error probability by using a saddlepoint approximation, as we previously did for binary modulation. Of special interest is the following:

Proposition 6. *For large* SNR*, BICM in the AEN channel behaves as a binary modulation with distance $d_{\min} = \min_{x,x' \in \mathcal{X}} |x - x'|$, in the sense that*

$$\lim_{\mathrm{SNR}\to\infty} \frac{\log \mathrm{pep}(d)}{\mathrm{SNR}} = -\frac{d\, d_{min}}{2}. \tag{22}$$

For uniform 2^m-PEM and large SNR, the pairwise error probability is then approximately $\mathrm{pep}(d) \simeq e^{-d\frac{\mathrm{SNR}}{2^m-1}}$.

As depicted in Fig. 3, which represents the simulation and approximations to the pairwise error probability for Gray mapping and several PEM modulations, the asymptotic losses with respect to 2-PEM ($d_{\min} = 2$) are 4.77 dB for 4-PEM, 11.76 dB for 16-PEM, and 17.99 dB for 64-PEM, in good agreement with the results for 2-PEM presented in Fig. 2. Note that the asymptotic loss with respect to an equivalent 2^m modulation in the AWGN case is $\frac{3}{2}$, or 1.76 dB, since in the Gaussian channel the pairwise error probability behaves as $\mathrm{pep}_{\mathrm{AWGN}}(d) \simeq e^{-d\frac{3\,\mathrm{SNR}}{2(2^m-1)}}$ for 2^m-QAM modulation. Here we combined the result in [3] on the asymptotics of $\mathrm{pep}(d)$ for BICM in the Gaussian channel with a trite computation of the minimum distance of 2^m-QAM.

Fig. 3: Comparison of simulation, Chernoff bound, and saddlepoint approximation of the $\mathrm{pep}(5)$ for 4-, 16-, and 64-PEM.

VI. Conclusions

In this paper, we have derived a number of results for the discrete-time additive exponential noise (AEN) channel. First, we have considered the design of pulse-energy modulation, more specifically a modulation whose symbols are used with equal probabilities and placed at non equally spaced positions,

$$\beta_{\mathrm{PEM}}\big\{(i-1)^{\lambda}\big\}, \quad i = 1, \dots, 2^m, \tag{23}$$

where β_{PEM} is an energy normalization factor. The choice $\lambda = \frac{1}{2}(1+\sqrt{5})$ gives the highest mutual information at high SNR, namely about 0.76 dB away from capacity. We have also determined the first two terms of the Taylor expansion of the coded modulation capacity at zero SNR. Generic modulations do not attain the minimum energy per bit, differently from the situation in the Gaussian channel.

Additionally, we have determined the Chernoff bound and the saddlepoint approximation to the pairwise error probability, and found that they have similar forms to those in the AWGN channel. Finally, we have seen that the performance of bit-interleaved coded modulation (BICM) is close to that of coded modulation and to its Gaussian channel counterpart.

Appendix A
CM Capacity Expansion at Low SNR

For the sake of compactness, we define $\gamma = \mathrm{SNR}$. Also, we respectively denote the first- and second-order moments of the constellation by μ_1 and μ_2. For each of the summands in the $\log(\cdot)$ in Eq. (8) we use $e^t = 1 + t + \frac{1}{2}t^2 + \mathrm{O}(t^3)$ to obtain

$$e^{-\,\mathrm{SNR}(x_l - x')} \doteq 1 + \gamma(x' - x_l) + \tfrac{1}{2}\gamma^2\big(x'^2 + x_l^2 - 2x'x_l\big). \tag{24}$$

We use the symbol $\doteq$ to represent equality neglecting higher order terms in the relevant variable, in this case t.

Let us define the variables q_j, q'_j, and q''_j respectively as

$$q_j = \sum_{x' \le x_j} P(x'), q'_j = \sum_{x' \le x_j} x'P(x'), q''_j = \sum_{x' \le x_j} x'^2 P(x'). \tag{25}$$

Rearranging, the sum in the logarithm over x' gives

$$q_j\left(1 + \gamma\frac{q'_j}{q_j} - \gamma x_l + \tfrac{1}{2}\gamma^2\frac{q''_j}{q_j} - \gamma^2 x_l\frac{q'_j}{q_j} + \tfrac{1}{2}\gamma^2 x_l^2\right). \tag{26}$$

Taking logarithms, and using the expansion $\log(1+t) \doteq t - \frac{1}{2}t^2$ around $\gamma = 0$, we obtain

$$\log q_j + \gamma\frac{q'_j}{q_j} - \gamma x_l + \tfrac{1}{2}\gamma^2\frac{q''_j}{q_j} - \gamma^2 x_l\frac{q'_j}{q_j} + \tfrac{1}{2}\gamma^2 x_l^2 - \tfrac{1}{2}\gamma^2\left(\frac{q'_j}{q_j} - x_l\right)^2 \tag{27}$$

$$\doteq \log q_j + \gamma\frac{q'_j}{q_j} - \gamma x_l + \tfrac{1}{2}\gamma^2\frac{q''_j}{q_j} - \tfrac{1}{2}\gamma^2\left(\frac{q'_j}{q_j}\right)^2. \tag{28}$$

We now move on to the summation over j in Eq. (8). We use the Taylor expansion of the exponential function in the summation over j, separating the last term as special. Starting at it, $j = |\mathcal{X}|$, we note that the sum over $x' \leq x_{|\mathcal{X}|}$ includes all symbols, and its contribution to the sum is

$$\Big(1 - \gamma(x_{|\mathcal{X}|} - x_l) + \tfrac{1}{2}\gamma^2(x_{|\mathcal{X}|} - x_l)^2\Big)\times \\ \times \left(\log q_j + \gamma\frac{q_j'}{q_j} - \gamma x_l + \tfrac{1}{2}\gamma^2\frac{q_j''}{q_j} - \tfrac{1}{2}\gamma^2\left(\frac{q_j'}{q_j}\right)^2\right) \tag{29}$$

$$= \Big(1 - \gamma(x_{|\mathcal{X}|} - x_l) + \tfrac{1}{2}\gamma^2(x_{|\mathcal{X}|} - x_l)^2\Big)\times \\ \times \Big(\gamma(\mu_1 - x_l) + \tfrac{1}{2}\gamma^2\big(\mu_2 - \mu_1^2\big)\Big), \tag{30}$$

since $\sum_{x'} P(x') = 1$. Carrying out the expectation over x_l, and discarding terms of order $\mathrm{O}(\gamma^3)$, this term contributes with

$$-\tfrac{1}{2}\gamma^2\big(\mu_2 - \mu_1^2\big). \tag{31}$$

As for the terms $j < |\mathcal{X}|$, the following terms contribute

$$\sum_{j=l}^{|\mathcal{X}|-1}\Bigg\{\gamma(x_{j+1} - x_j)\log q_j + \tfrac{1}{2}\gamma^2(x_j^2 - x_{j+1}^2)\log q_j \\ - \gamma^2(x_j - x_{j+1})x_l\log q_j + \gamma^2(x_{j+1} - x_j)\left(\frac{q_j'}{q_j} - x_l\right)\Bigg\}. \tag{32}$$

The last step is the averaging over x_l. The order of the double summation over l and j can be reversed, with the summation limits becoming

$$\sum_{j=1}^{|\mathcal{X}|-1}\sum_{l\leq j} P(x_l)\Bigg\{\gamma(x_{j+1} - x_j)\log q_j \\ + \gamma^2(x_j - x_{j+1})\big(\tfrac{1}{2}(x_j + x_{j+1}) - x_l\big)\log q_j \\ + \gamma^2(x_{j+1} - x_j)\left(\frac{\sum_{x'\leq x_j} x'P(x')}{q_j} - x_l\right)\Bigg\} \tag{33}$$

$$= \sum_{j=1}^{|\mathcal{X}|-1}\Bigg(\gamma(x_{j+1} - x_j)q_j\log q_j \\ + \gamma^2(x_j - x_{j+1})\big(\tfrac{1}{2}(x_j + x_{j+1})q_j - q_j'\big)\log q_j\Bigg). \tag{34}$$

This gives the desired expression for the CM capacity.

References

[1] S. Verdú, "The exponential distribution in information theory," *Prob. Per. Inf.*, vol. 32, no. 1, pp. 86–95, Jan-Mar 1996.

[2] G. D. Forney, Jr., R. G. Gallager, G. R. Lang, F. M. Longstaff, and S. U. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 632–647, September 1984.

[3] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Error probability of bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 262–271, January 2006.

[4] J. G. Proakis, *Digital Communications*. McGraw-Hill, 1995.

[5] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.