entropy

MDPI

Article

Gaussian Multiple Access Channels with One-Bit Quantizer at the Receiver †,‡

Borzoo Rassouli [1,*], Morteza Varasteh [2,*] and Deniz Gündüz [2,*]

[1] School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK
[2] The Intelligent Systems and Networks group of Department of Electrical and Electronics, Imperial College London, London SW7 2AZ, UK
* Correspondence: b.rassouli@essex.ac.uk (B.R.); m.varasteh12@imperial.ac.uk (M.V.); d.gunduz@imperial.ac.uk (D.G.); Tel.: +44-(0)-207-594-6218 (D.G.)

† This paper is an extended version of our paper published in the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017.

‡ This work has been carried out when the first author was with the Information Processing and Communications Laboratory at Imperial College London.

Received: 21 May 2018; Accepted: 5 September 2018; Published: 7 September 2018



Abstract: The capacity region of a two-transmitter Gaussian multiple access channel (MAC) under average input power constraints is studied, when the receiver employs a zero-threshold one-bit analogue-to-digital converter (ADC). It is proven that the input distributions of the two transmitters that achieve the boundary points of the capacity region are discrete. Based on the position of a boundary point, upper bounds on the number of the mass points of the corresponding distributions are derived. Furthermore, a lower bound on the sum capacity is proposed that can be achieved by time division with power control. Finally, inspired by the numerical results, the proposed lower bound is conjectured to be tight.

Keywords: Gaussian multiple access channel; one-bit quantizer; capacity region

1. Introduction

The energy consumption of an analogue-to-digital converter (ADC) (measured in Joules/sample) grows exponentially with its resolution (in bits/sample) [1,2].When the available power is limited, for example, for mobile devices with limited battery capacity, or for wireless receivers that operate on limited energy harvested from ambient sources [3], the receiver circuitry may be constrained to operate with low resolution ADCs. The presence of a low-resolution ADC, in particular a one-bit ADC at the receiver, alters the channel characteristics significantly. Such a constraint not only limits the fundamental bounds on the achievable rate, but it also changes the nature of the communication and modulation schemes approaching these bounds. For example, in a real additive white Gaussian noise (AWGN) channel under an average power constraint on the input, if the receiver is equipped with a K-bin (i.e., $\log_2 K$-bit) ADC front end, it is shown in [4] that the capacity-achieving input distribution is discrete with at most $K+1$ mass points. This is in contrast with the optimality of the Gaussian input distribution when the receiver has infinite resolution.

Especially with the adoption of massive multiple-input multiple-output (MIMO) receivers and the millimetre wave (mmWave) technology enabling communication over large bandwidths, communication systems with limited-resolution receiver front ends are becoming of practical importance. Accordingly, there has been a growing research interest in understanding both the fundamental information theoretic limits and the design of practical communication protocols for systems with finite-resolution ADC front ends. In [5], the authors showed that for a Rayleigh fading

channel with a one-bit ADC and perfect channel state information at the receiver (CSIR), quadrature phase shift keying (QPSK) modulation is capacity-achieving. In case of no CSIR, [6] showed that QPSK modulation is optimal when the signal-to-noise ratio (SNR) is above a certain threshold, which depends on the coherence time of the channel, while for SNRs below this threshold, on-off QPSK achieves the capacity. For the point-to-point multiple-input multiple-output (MIMO) channel with a one-bit ADC front end at each receive antenna and perfect CSIR, [7] showed that QPSK is optimal at very low SNRs, while with perfect channel state information at the transmitter (CSIT), upper and lower bounds on the capacity are provided in [8].

To the best of our knowledge, the existing literature on communications with low-resolution ADCs focuses exclusively on point-to-point systems. Our goal in this paper is to understand the impact of low-resolution ADCs on the capacity region of a multiple access channel (MAC). In particular, we consider a two-transmitter Gaussian MAC with a one-bit quantizer at the receiver. The inputs to the channel are subject to average power constraints. We show that any point on the boundary of the capacity region is achieved by discrete input distributions. Based on the slope of the tangent line to the capacity region at a boundary point, we propose upper bounds on the cardinality of the support of these distributions. Finally, based on numerical analysis for the sum capacity, it is observed that we cannot obtain a sum rate higher than is achieved by time division with power control.

The paper is organized as follows. Section 2 introduces the system model. In Section 3, the capacity region of a general two-transmitter memoryless MAC under input average power constraints is investigated. The main result of the paper is presented in Section 3, and a detailed proof is given in Section 4. The proof has two parts: (1) it is shown that the support of the optimal distributions is bounded by contradiction; and (2) we make use of this boundedness to prove the finiteness of the optimal support by using Dubins' theorem [9]. Section 5 analyses the sum capacity, and finally, Section 6 concludes the paper.

Notations: Random variables are denoted by capital letters, while their realizations with lower case letters. $F_X(x)$ denotes the cumulative distribution function (CDF) of random variable X. The conditional probability mass function (pmf) $p_{Y|X_1,X_2}(y|x_1,x_2)$ will be written as $p(y|x_1,x_2)$. For integers $m \leq n$, we have $[m:n] = \{m, m+1, \ldots, n\}$. For $0 \leq t \leq 1$, $H_b(t) \triangleq -t\log_2 t - (1-t)\log_2(1-t)$ denotes the binary entropy function. The unit-step function is denoted by $s(\cdot)$.

2. System Model and Preliminaries

We consider a two-transmitter memoryless Gaussian MAC (as shown in Figure 1) with a one-bit quantizer Γ at the receiver front end. Transmitter $j = 1, 2$ encodes its message W_j into a codeword X_j^n and transmits it over the shared channel. The signal received by the decoder is given by:

$$Y = \Gamma(X_{1,i} + X_{2,i} + Z_i),\ i \in [1:n],$$

where $\{Z_i\}_{i=1}^n$ is an independent and identically distributed (i.i.d.) Gaussian noise process, also independent of the channel inputs X_1^n and X_2^n with $Z_i \sim \mathcal{N}(0,1), i \in [1:n]$. Γ represents the one-bit ADC operation given by:

$$\Gamma(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

This channel can be modelled by the triplet $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1,x_2), \mathcal{Y})$, where $\mathcal{X}_1, \mathcal{X}_2$ $(= \mathbb{R})$ and $\mathcal{Y}$ $(= \{0,1\})$, respectively, are the alphabets of the inputs and the output. The conditional pmf of the channel output Y conditioned on the channel inputs X_1 and X_2 (i.e., $p(y|x_1,x_2)$) is characterized by:

$$p(0|x_1,x_2) = 1 - p(1|x_1,x_2) = Q(x_1+x_2), \tag{1}$$

where $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-\frac{t^2}{2}} dt$.

We consider a two-transmitter stationary and memoryless MAC model $(\mathcal{X}_1 \times \mathcal{X}_2, p(y|x_1, x_2), \mathcal{Y})$, where $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{R}$, $\mathcal{Y} = \{0, 1\}$, $p(y|x_1, x_2)$ is given in (1).

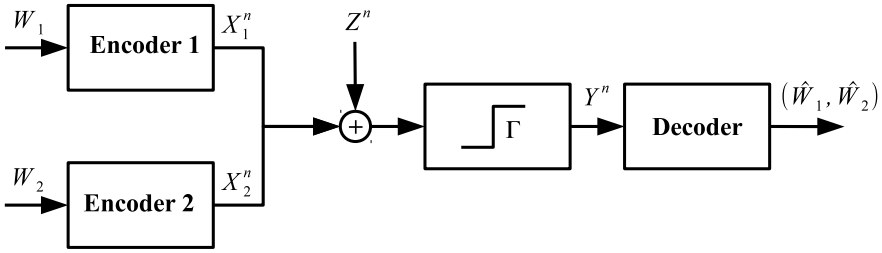


Figure 1. A two-transmitter Gaussian multiple access channel (MAC) with a one-bit analogue-to-digital converter (ADC) front end at the receiver.

A $(2^{nR_1}, 2^{nR_2}, n)$ code for this channel consists of (as in [10]):

- two message sets $[1 : 2^{nR_1}]$ and $[1 : 2^{nR_2}]$,
- two encoders, where encoder $j = 1, 2$ assigns a codeword $x_j^n(w_j)$ to each message $w_j \in [1 : 2^{nR_j}]$, and
- a decoder that assigns estimates $(\hat{w}_1, \hat{w}_2) \in [1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$ or an error message to each received sequence y^n.

The stationary property means that the channel does not change over time, while the memoryless property indicates that $p(y_i|x_1^i, x_2^i, y^{i-1}, w_1, w_2) = p(y_i|x_{1,i}, x_{2,i})$ for any message pair (w_1, w_2).

We assume that the message pair (W_1, W_2) is uniformly distributed over $[1 : 2^{nR_1}] \times [1 : 2^{nR_2}]$. The average probability of error is defined as:

$$P_e^{(n)} \triangleq \Pr\left\{(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\right\}. \tag{2}$$

Average power constraints are imposed on the channel inputs as:

$$\frac{1}{n}\sum_{i=1}^{n} x_{j,i}^2(w_j) \leq P_j\,,\ \forall w_j \in [1 : 2^{nR_j}],\ j \in \{1, 2\}, \tag{3}$$

where $x_{j,i}(w_j)$ denotes the i-th element of the codeword $x_j^n(w_j)$.

A rate pair (R_1, R_2) is said to be achievable for this channel if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes satisfying the average power constraints (3), such that $\lim_{n\to\infty} P_e^{(n)} = 0$. The capacity region $\mathscr{C}(P_1, P_2)$ of this channel is the closure of the set of achievable rate pairs (R_1, R_2).

3. Main Results

Proposition 1. *The capacity region $\mathscr{C}(P_1, P_2)$ of a two-transmitter stationary and memoryless MAC with average power constraints P_1 and P_2 is the set of non-negative rate pairs (R_1, R_2) that satisfy:*

$$\begin{aligned} R_1 &\leq I(X_1; Y|X_2, U), \\ R_2 &\leq I(X_2; Y|X_1, U), \\ R_1 + R_2 &\leq I(X_1, X_2; Y|U), \end{aligned} \tag{4}$$

for some $F_U(u)F_{X_1|U}(x_1|u)F_{X_2|U}(x_2|u)$, such that $\mathbb{E}[X_j^2] \leq P_j$, $j = 1, 2$. Furthermore, it is sufficient to consider $|\mathcal{U}| \leq 5$.

Proof of Proposition 1. The proof is provided in Appendix A. □

The main result of this paper is provided in the following theorem. It bounds the cardinality of the support set of the capacity-achieving distributions.

Theorem 1. *Let J be an arbitrary point on the boundary of the capacity region $\mathscr{C}(P_1, P_2)$ of the memoryless MAC with a one-bit ADC front end (as shown in Figure 1). J is achieved by a distribution in the form of $F_U^J(u)F_{X_1|U}^J(x_1|u)F_{X_2|U}^J(x_2|u)$. Furthermore, let l_J be the slope of the line tangent to the capacity region at this point. For any $u \in \mathcal{U}$, the conditional input distributions $F_{X_1|U}^J(x_1|u)$ and $F_{X_2|U}^J(x_2|u)$ have at most n_1 and n_2 points of increase (a point Z is said to be a point of increase of a distribution if for any open set Ω containing Z, we have $Pr\{\Omega\} > 0$), respectively, where:*

$$(n_1, n_2) = \begin{cases} (3,5) & l_J < -1 \\ (3,3) & l_J = -1 \\ (5,3) & l_J > -1 \end{cases}. \tag{5}$$

Furthermore, this result remains unchanged if the one-bit ADC has a non-zero threshold.

Proof of Theorem 1. The proof is provided in Section 4. □

Proposition 1 and Theorem 1 establish upper bounds on the number of mass points of the distributions that achieve a boundary point. The significance of this result is that once it is known that the optimal inputs are discrete with at most a certain number of mass points, the capacity region along with the optimal distributions can be obtained via computer programs.

4. Proof of Theorem 1

In order to show that the boundary points of the capacity region are achieved, it is sufficient to show that the capacity region is a closed set, i.e., it includes all of its limit points.

Let $\mathcal{U}$ be a set with $|\mathcal{U}| \leq 5$ and Ω be defined as:

$$\Omega \triangleq \left\{ F_{U,X_1,X_2} \middle|\ U \in \mathcal{U},\ X_1 - U - X_2,\ \mathbb{E}[X_j^2] \leq P_j,\ j = 1,2 \right\}, \tag{6}$$

which is the set of all CDFs on the triplet (U, X_1, X_2), where U is drawn from $\mathcal{U}$, and the Markov chain $X_1 - U - X_2$ and the corresponding average power constraints hold.

In Appendix B, it is proven that Ω is a compact set. Since a continuous mapping preserves compactness, the capacity region is compact. Since the capacity region is a subset of $\mathbb{R}^2$, it is closed and bounded (note that a subset of $\mathbb{R}^k$ is compact if and only if it is closed and bounded [11]). Therefore, any point P on the boundary of the capacity region is achieved by a distribution denoted by $F_U^J(u)F_{X_1|U}^J(x_1|u)F_{X_2|U}^J(x_2|u)$.

Since the capacity region is a convex space, it can be characterized by its supporting hyperplanes. In other words, any point on the boundary of the capacity region, denoted by (R_1^b, R_2^b), can be written as:

$$(R_1^b, R_2^b) = \arg \max_{(R_1,R_2)\in\mathscr{C}(P_1,P_2)} R_1 + \lambda R_2,$$

for some $\lambda \in (0, \infty)$. Here, we have excluded the cases $\lambda = 0$ and $\lambda = \infty$, where the channel is not a two-transmitter MAC any longer, and boils down to a point-to-point channel, whose capacity is already known.

Any rate pair $(R_1, R_2) \in \mathscr{C}(P_1, P_2)$ must lie within a pentagon defined by (4) for some $F_U F_{X_1|U} F_{X_2|U}$ that satisfies the power constraints. Therefore, due to the structure of the pentagon, the problem of finding the boundary points is equivalent to the following maximization problem.

$$\max_{(R_1,R_2)\in\mathscr{C}(P_1,P_2)} R_1 + \lambda R_2 = \begin{cases} \max I(X_1; Y|X_2, U) + \lambda I(X_2; Y|U) & 0 < \lambda \leq 1 \\ \max I(X_2; Y|X_1, U) + \lambda I(X_1; Y|U) & \lambda > 1 \end{cases}, \tag{7}$$

where on the right-hand side (RHS) of (7), the maximizations are over all $F_U F_{X_1|U} F_{X_2|U}$ that satisfy the power constraints. It is obvious that when $\lambda = 1$, the two lines in (7) are the same, which results in the sum capacity.

For any product of distributions $F_{X_1} F_{X_2}$ and the channel in (1), let I_λ be defined as:

$$I_\lambda(F_{X_1} F_{X_2}) \triangleq \begin{cases} I(X_1; Y|X_2) + \lambda I(X_2; Y) & 0 < \lambda \leq 1 \\ I(X_2; Y|X_1) + \lambda I(X_1; Y) & \lambda > 1 \end{cases}. \tag{8}$$

With this definition, (7) can be rewritten as:

$$\max_{(R_1, R_2) \in \mathscr{C}(P_1, P_2)} R_1 + \lambda R_2 = \max \sum_{i=1}^{5} p_U(u_i) I_\lambda(F_{X_1|U}(x_1|u_i) F_{X_2|U}(x_2|u_i)),$$

where the second maximization is over distributions of the form $p_U(u) F_{X_1|U}(x_1|u) F_{X_2|U}(x_2|u)$, such that:

$$\sum_{i=1}^{5} p_U(u_i) \mathbb{E}[X_j^2 | U = u_i] \leq P_j, \ j = 1, 2.$$

Proposition 2. *For a given F_{X_1} and any $\lambda > 0$, $I_\lambda(F_{X_1} F_{X_2})$ is a concave, continuous and weakly differentiable function of F_{X_2}. In the statement of this proposition, F_{X_1} and F_{X_2} could be interchanged.*

Proof of Proposition 2. The proof is provided in Appendix C. □

Proposition 3. *Let P_1', P_2' be two arbitrary non-negative real numbers. For the following problem:*

$$\max_{\substack{F_{X_1} F_{X_2}: \\ \mathbb{E}[X_j^2] \leq P_j', \ j=1,2}} I_\lambda(F_{X_1} F_{X_2}), \tag{9}$$

the optimal inputs $F_{X_1}^*$ and $F_{X_2}^*$, which are not unique in general, have the following properties,

(i) The support sets of $F_{X_1}^*$ and $F_{X_2}^*$ are bounded subsets of $\mathbb{R}$.
(ii) $F_{X_1}^*$ and $F_{X_2}^*$ are discrete distributions that have at most n_1 and n_2 points of increase, respectively, where:

$$(n_1, n_2) = \begin{cases} (5, 3) & 0 < \lambda < 1 \\ (3, 3) & \lambda = 1 \\ (3, 5) & \lambda > 1 \end{cases}.$$

Proof of Proposition 3. We start with the proof of the first claim. Assume that $0 < \lambda \leq 1$, and F_{X_2} is given. Consider the following optimization problem:

$$I_{F_{X_2}}^* \triangleq \sup_{\substack{F_{X_1}: \\ \mathbb{E}[X_1^2] \leq P_1'}} I_\lambda(F_{X_1} F_{X_2}). \tag{10}$$

Note that $I_{F_{X_2}}^* < +\infty$, since for any $\lambda > 0$, from (8),

$$I_\lambda \leq (\lambda + 1) H(Y) \leq (1 + \lambda) < +\infty.$$

From Proposition 2, I_λ is a continuous, concave function of F_{X_1}. Furthermore, the set of all CDFs with bounded second moment (here, P_1') is convex and compact. The compactness follows from Appendix I in [12], where the only difference is in using Chebyshev's inequality instead of Markov's inequality. Therefore, the supremum in (10) is achieved by a distribution $F_{X_1}^*$. Since for

any $F_{X_1}(x) = s(x - x_0)$ with $|x_0|^2 < P_1'$, we have $\mathbb{E}[X_1^2] < P_1'$, the Lagrangian theorem and the Karush–Kuhn–Tucker conditions state that there exists a $\theta_1 \geq 0$ such that:

$$I^*_{F_{X_2}} = \sup_{F_{X_1}} \left\{ I_\lambda(F_{X_1}F_{X_2}) - \theta_1 \left(\int x^2 dF_{X_1}(x) - P_1' \right) \right\}. \tag{11}$$

Furthermore, the supremum in (11) is achieved by $F^*_{X_1}$, and:

$$\theta_1 \left(\int x^2 dF^*_{X_1}(x) - P_1' \right) = 0. \tag{12}$$

Lemma 1. *The Lagrangian multiplier θ_1 is non-zero. From (12), this is equivalent to having $\mathbb{E}[X_1^2] = P_1'$, i.e., the first user transmits with its maximum allowable power (note that this is for $\lambda \leq 1$, as used in Appendix D).*

Proof of Lemma 1. In what follows, we prove that a zero Lagrangian multiplier is not possible. Having a zero Lagrangian multiplier means the power constraint is inactive. In other words, if $\theta_1 = 0$, (10) and (11) imply that:

$$\sup_{\substack{F_{X_1} \\ \mathbb{E}[X_1^2] \leq P_1'}} I_\lambda(F_{X_1}F_{X_2}) = \sup_{F_{X_1}} I_\lambda(F_{X_1}F_{X_2}). \tag{13}$$

We prove that (13) does not hold by showing that its left-hand side (LHS) is strictly less than one, while its RHS equals one. The details are provided in Appendix D. □

$I_\lambda(F_{X_1}F_{X_2})$ $(0 < \lambda \leq 1)$ can be written as:

$$\begin{aligned} I_\lambda(F_{X_1}F_{X_2}) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \sum_{y=0}^{1} p(y|x_1,x_2) \log \frac{p(y|x_1,x_2)}{[p(y;F_{X_1}F_{X_2})]^\lambda [p(y;F_{X_1}|x_2)]^{1-\lambda}} dF_{X_1}(x_1) dF_{X_2}(x_2) \\ &= \int_{-\infty}^{+\infty} \tilde{i}_\lambda(x_1;F_{X_1}|F_{X_2}) dF_{X_1}(x_1) \qquad (14) \\ &= \int_{-\infty}^{+\infty} i_\lambda(x_2;F_{X_2}|F_{X_1}) dF_{X_2}(x_2), \qquad (15) \end{aligned}$$

where we have defined:

$$\tilde{i}_\lambda(x_1;F_{X_1}|F_{X_2}) \triangleq \int_{-\infty}^{+\infty} \left(D\left(p(y|x_1,x_2)||p(y;F_{X_1}F_{X_2})\right) + (1-\lambda) \sum_{y=0}^{1} p(y|x_1,x_2) \log \frac{p(y;F_{X_1}F_{X_2})}{p(y;F_{X_1}|x_2)} \right) dF_{X_2}(x_2), \tag{16}$$

and:

$$i_\lambda(x_2;F_{X_2}|F_{X_1}) \triangleq \int_{-\infty}^{+\infty} D\left(p(y|x_1,x_2)||p(y;F_{X_1}F_{X_2})\right) dF_{X_1}(x_1) - (1-\lambda) D\left(p(y;F_{X_1}|x_2)||p(y;F_{X_1}F_{X_2})\right). \tag{17}$$

$p(y;F_{X_1}F_{X_2})$ is nothing but the pmf of Y with the emphasis that it has been induced by F_{X_1} and F_{X_2}. Likewise, $p(y;F_{X_1}|x_2)$ is the conditional pmf $p(y|x_2)$ when X_1 is drawn according to F_{X_1}. From (14), $\tilde{i}_\lambda(x_1;F_{X_1}|F_{X_2})$ can be considered as the density of I_λ over F_{X_1} when F_{X_2} is given. $i_\lambda(x_2;F_{X_2}|F_{X_1})$ can be interpreted in a similar way.

Note that (11) is an unconstrained optimization problem over the set of all CDFs. Since $\int x^2 dF_{X_1}(x)$ is linear and weakly differentiable in F_{X_1}, the objective function in (11) is concave and weakly differentiable. Hence, a necessary condition for the optimality of $F^*_{X_1}$ is:

$$\int \{\tilde{i}_\lambda(x_1;F^*_{X_1}|F_{X_2}) + \theta_1(P_1' - x_1^2)\} dF_{X_1}(x_1) \leq I^*_{F_{X_2}}, \quad \forall F_{X_1}. \tag{18}$$

Furthermore, (18) can be verified to be equivalent to:

$$\tilde{i}_\lambda(x_1; F^*_{X_1}|F_{X_2}) + \theta_1(P_1' - x_1^2) \leq I^*_{F_{X_2}}, \ \forall x_1 \in \mathbb{R}, \tag{19}$$

$$\tilde{i}_\lambda(x_1; F^*_{X_1}|F_{X_2}) + \theta_1(P_1' - x_1^2) = I^*_{F_{X_2}}, \ \text{if } x_1 \text{ is a point of increase of } F^*_{X_1}. \tag{20}$$

The justifications of (18)–(20) are provided in Appendix E.

In what follows, we prove that in order to satisfy (20), $F^*_{X_1}$ must have a bounded support by showing that the LHS of (20) goes to $-\infty$ with x_1. The following lemma is useful in the sequel for taking the limit processes inside the integrals.

Lemma 2. *Let X_1 and X_2 be two independent random variables satisfying $\mathbb{E}[X_1^2] \leq P_1'$ and $\mathbb{E}[X_2^2] \leq P_2'$, respectively ($P_1', P_2' \in [0, +\infty)$). Considering the conditional pmf in (1), the following inequalities hold.*

$$\left| D\left(p(y|x_1,x_2)||p(y;F_{X_1}F_{X_2})\right) \right| \leq 1 - 2\log Q(\sqrt{P_1'} + \sqrt{P_2'}) \tag{21}$$

$$p(y;F_{X_1}|x_2) \geq Q\left(\sqrt{P_1'} + |x_2|\right) \tag{22}$$

$$\left| \sum_{y=0}^{1} p(y|x_1,x_2) \log \frac{p(y;F_{X_1}F_{X_2})}{p(y;F_{X_1}|x_2)} \right| \leq -2\log Q\left(\sqrt{P_1'} + \sqrt{P_2'}\right) - 2\log Q\left(\sqrt{P_1'} + |x_2|\right) \tag{23}$$

Proof of Lemma 2. The proof is provided in Appendix F. □

Note that

$$\lim_{x_1\to+\infty} \int_{-\infty}^{+\infty} D\left(p(y|x_1,x_2)||p(y;F^*_{X_1}F_{X_2})\right) dF_{X_2}(x_2) = \int_{-\infty}^{+\infty} \lim_{x_1\to+\infty} D\left(p(y|x_1,x_2)||p(y;F^*_{X_1}F_{X_2})\right) dF_{X_2}(x_2) \tag{24}$$

$$= -\log p_Y(1;F^*_{X_1}F_{X_2}) \tag{25}$$

$$\leq -\log Q(\sqrt{P_1'} + \sqrt{P_2'}), \tag{26}$$

where (24) is due to the Lebesgue dominated convergence theorem [11] and (21), which permit the interchange of the limit and the integral; (25) is due to the following:

$$\lim_{x_1\to+\infty} D\left(p(y|x_1,x_2)||p(y;F^*_{X_1}F_{X_2})\right) = \lim_{x_1\to+\infty} \sum_{y=0}^{1} p(y|x_1,x_2) \log \frac{p(y|x_1,x_2)}{p(y;F^*_{X_1}F_{X_2})}$$
$$= -\log p_Y(1;F^*_{X_1}F_{X_2}),$$

since $p(0|x_1,x_2) = Q(x_1+x_2)$ goes to zero when $x_1 \to +\infty$ and $p_Y(y;F^*_{X_1}F_{X_2})$ $(y=0,1)$ is bounded away from zero by (A34) ; (26) is obtained from (A34) in Appendix F. Furthermore,

$$\lim_{x_1\to+\infty} \int_{-\infty}^{+\infty} \sum_{y=0}^{1} p(y|x_1,x_2) \log \frac{p(y;F^*_{X_1}F_{X_2})}{p(y;F^*_{X_1}|x_2)} dF_{X_2}(x_2) = \int_{-\infty}^{+\infty} \lim_{x_1\to+\infty} \sum_{y=0}^{1} p(y|x_1,x_2) \log \frac{p(y;F^*_{X_1}F_{X_2})}{p(y;F^*_{X_1}|x_2)} dF_{X_2}(x_2) \tag{27}$$

$$= \log p_Y(1;F^*_{X_1}F_{X_2}) - \int_{-\infty}^{+\infty} \log p(1;F^*_{X_1}|x_2) dF_{X_2}(x_2)$$
$$< -\log Q\left(\sqrt{P_1'} + \sqrt{P_2'}\right), \tag{28}$$

where (27) is due to the Lebesgue dominated convergence theorem along with (23) and (A39) in Appendix F; (28) is from (22) and the convexity of $\log Q(\alpha + \sqrt{t})$ in t when $\alpha \geq 0$ (see Appendix G).

Therefore, from (26) and (28),

$$\lim_{x_1 \to +\infty} \tilde{i}_\lambda(x_1; F^*_{X_1}|F_{X_2}) \leq -(2-\lambda)\log Q(\sqrt{P'_1} + \sqrt{P'_2}) < +\infty. \tag{29}$$

Using a similar approach, we can also obtain:

$$\lim_{x_1 \to -\infty} \tilde{i}_\lambda(x_1; F^*_{X_1}|F_{X_2}) \leq -(2-\lambda)\log Q(\sqrt{P'_1} + \sqrt{P'_2}) < +\infty. \tag{30}$$

From (29) and (30) and the fact that $\theta_1 > 0$ (see Lemma 1), the LHS of (19) goes to $-\infty$ when $|x_1| \to +\infty$. Since any point of increase of $F^*_{X_1}$ must satisfy (19) with equality and $I^*_{F_{X_2}} \geq 0$, it is proven that $F^*_{X_1}$ has a bounded support. Hence, from now on, we assume $X_1 \in [-A_1, A_2]$ for some $A_1, A_2 \in \mathbb{R}$ (note that A_1 and A_2 are determined by the choice of F_{X_2}).

Similarly, for a given F_{X_1}, the optimization problem:

$$I^*_{F_{X_1}} = \sup_{\substack{F_{X_2}: \\ \mathbb{E}[X_2^2] \leq P'_2}} I_\lambda(F_{X_1} F_{X_2}),$$

boils down to the following necessary condition:

$$i_\lambda(x_2; F^*_{X_2}|F_{X_1}) + \theta_2(P'_2 - x_2^2) \leq I^*_{F_{X_1}}, \quad \forall x_2 \in \mathbb{R}, \tag{31}$$

$$i_\lambda(x_2; F^*_{X_2}|F_{X_1}) + \theta_2(P'_2 - x_2^2) = I^*_{F_{X_1}}, \quad \text{if } x_2 \text{ is a point of increase of } F^*_{X_2}, \tag{32}$$

for the optimality of $F^*_{X_2}$. However, there are two main differences between (32) and (20). First is the difference between i_λ and $\tilde{i}_\lambda$. Second is the fact that we do not claim θ_2 to be nonzero, since the approach used in Lemma 1 cannot be readily applied to θ_2. Nonetheless, the boundedness of the support of $F^*_{X_2}$ can be proven by inspecting the behaviour of the LHS of (32) when $|x_2| \to +\infty$.

In what follows, i.e., from (33)–(38), we prove that the support of $F^*_{X_2}$ is bounded by showing that (32) does not hold when $|x_2|$ is above a certain threshold. The first term on the LHS of (32) is $i_\lambda(x_2; F^*_{X_2}|F_{X_1})$. From (17) and (21), it can be easily verified that:

$$\begin{aligned} \lim_{x_2 \to +\infty} i_\lambda(x_2; F^*_{X_2}|F_{X_1}) &= -\lambda \log p_Y(1; F_{X_1} F^*_{X_2}) \leq -\lambda \log Q(\sqrt{P'_1} + \sqrt{P'_2}), \\ \lim_{x_2 \to -\infty} i_\lambda(x_2; F^*_{X_2}|F_{X_1}) &= -\lambda \log p_Y(0; F_{X_1} F^*_{X_2}) \leq -\lambda \log Q(\sqrt{P'_1} + \sqrt{P'_2}) \end{aligned} \tag{33}$$

From (33), if $\theta_2 > 0$, the LHS of (32) goes to $-\infty$ with $|x_2|$, which proves that X^*_2 is bounded.

For the possible case of $\theta_2 = 0$, in order to show that (32) does not hold when $|x_2|$ is above a certain threshold, we rely on the boundedness of X_1, i.e., $X_1 \in [-A_1, A_2]$. Then, we prove that i_λ approaches its limit in (33) from below. In other words, there is a real number K such that $i_\lambda(x_2; F^*_{X_2}|F_{X_1}) < -\lambda \log p_Y(1; F_{X_1} F^*_{X_2})$ when $x_2 > K$, and $i_\lambda(x_2; F^*_{X_2}|F_{X_1}) < -\lambda \log p_Y(0; F_{X_1} F^*_{X_2})$ when $x_2 < -K$. This establishes the boundedness of X^*_2. In what follows, we only show the former, i.e., when $x_2 \to +\infty$. The latter, i.e., $x_2 \to -\infty$, follows similarly, and it is omitted for the sake of brevity.

By rewriting i_λ, we have:

$$\begin{aligned} i_\lambda(x_2; F^*_{X_2}|F_{X_1}) = &-\lambda p(1; F_{X_1}|x_2) \log p_Y(1; F_{X_1} F^*_{X_2}) \\ &- \int_{-A_1}^{A_2} H_b(Q(x_1+x_2)) dF_{X_1}(x_1) + (1-\lambda) \underbrace{H(Y|X_2=x_2)}_{H_b\left(\int Q(x_1+x_2) dF_{X_1}(x_1)\right)} \\ &- \lambda \underbrace{p(0; F_{X_1}|x_2)}_{\int Q(x_1+x_2) dF_{X_1}(x_1)} \log p_Y(0; F_{X_1} F^*_{X_2}). \end{aligned} \tag{34}$$

It is obvious that the first term on the RHS of (34) approaches $-\lambda \log p_Y(1; F_{X_1} F^*_{X_2})$ from below when $x_2 \to +\infty$, since $p(1; F_{X_1}|x_2) \leq 1$. It is also obvious that the remaining terms go to zero when $x_2 \to +\infty$. Hence, it is sufficient to show that they approach zero from below, which is proven by using the following lemma.

Lemma 3. *Let X_1 be distributed on $[-A_1, A_2]$ according to $F_{X_1}(x_1)$. We have:*

$$\lim_{x_2 \to +\infty} \frac{\int_{-A_1}^{A_2} H_b(Q(x_1+x_2)) dF_{X_1}(x_1)}{H_b\left(\int_{-A_1}^{A_2} Q(x_1+x_2) dF_{X_1}(x_1)\right)} = 1. \tag{35}$$

Proof of Lemma 3. The proof is provided in Appendix H. □

From (35), we can write:

$$\int_{-A_1}^{A_2} H_b(Q(x_1+x_2)) dF_{X_1}(x_1) = \gamma(x_2) H_b\left(\int_{-A_1}^{A_2} Q(x_1+x_2) dF_{X_1}(x_1)\right), \tag{36}$$

where $\gamma(x_2) \leq 1$ (due to the concavity of $H_b(\cdot)$), and $\gamma(x_2) \to 1$ when $x_2 \to +\infty$ (due to (35)). Furthermore, from the fact that $\lim_{x \to 0} \frac{H_b(x)}{cx} = +\infty$ $(c > 0)$, we have:

$$H_b\left(\int_{-A_1}^{A_2} Q(x_1+x_2) dF_{X_1}(x_1)\right) = -\eta(x_2) \log p_Y(0; F_{X_1} F^*_{X_2}) \int_{-A_1}^{A_2} Q(x_1+x_2) dF_{X_1}(x_1), \tag{37}$$

where $\eta(x_2) > 0$ and $\eta(x_2) \to +\infty$ when $x_2 \to +\infty$. From (36)–(37), the second and the third line of (34) become:

$$\left(1 - \gamma(x_2) + \frac{\lambda}{\eta(x_2)} - \lambda\right) \underbrace{\left(-\eta(x_2) \log p_Y(0; F_{X_1} F^*_{X_2}) \int_{-A_1}^{A_2} Q(x_1+x_2) dF_{X_1}(x_1)\right)}_{\geq 0}. \tag{38}$$

Since $\gamma(x_2) \to 1$ and $\eta(x_2) \to +\infty$ as $x_2 \to +\infty$, there exists a real number K such that $1 - \gamma(x_2) + \frac{\lambda}{\eta(x_2)} - \lambda < 0$ when $x_2 > K$. Therefore, the second and the third line of (34) approach zero from below, which proves that the support of X_2^* is bounded away from $+\infty$. As mentioned before, a similar argument holds when $x_2 \to -\infty$. This proves that X_2^* has a bounded support.

Remark 1. *We remark here that the order of showing the boundedness of the supports is important. First, for a given F_{X_2} (not necessarily bounded), it is proven that $F^*_{X_1}$ is bounded. Then, for a given bounded F_{X_1}, it is shown that $F^*_{X_2}$ is also bounded. Hence, the boundedness of the supports of the optimal input distributions is proven by contradiction. The order is reversed when $\lambda > 1$, and it follows the same steps as in the case of $\lambda \leq 1$. Therefore, it is omitted.*

We next prove the second claim in Proposition 3. We assume that $0 < \lambda < 1$, and a bounded F_{X_1} is given. We already know that for a given bounded F_{X_1}, $F^*_{X_2}$ has a bounded support denoted by $[-B_1, B_2]$. Therefore,

$$\begin{aligned} I^*_{F_{X_1}} &= \sup_{\substack{F_{X_2}: \\ \mathbb{E}[X_2^2] \le P_2'}} I_\lambda(F_{X_1}F_{X_2}) \\ I^*_{F_{X_1}} &= \sup_{\substack{F_{X_2} \in \mathscr{S}_2: \\ \mathbb{E}[X_2^2] \le P_2'}} I_\lambda(F_{X_1}F_{X_2}), \end{aligned} \tag{39}$$

where $\mathscr{S}_2$ denotes the set of all probability distributions on the Borel sets of $[-B_1, B_2]$. Let $p_0^* = p_Y(0; F_{X_1}F^*_{X_2})$ denote the probability of the event $Y = 0$, induced by $F^*_{X_2}$ and the given F_{X_1}. Furthermore, let P_2^* denote the second moment of X_2 under $F^*_{X_2}$. The set:

$$\mathscr{F}_2 = \left\{ F_{X_2} \in \mathscr{S}_2 \middle| \int_{-B_1}^{B_2} p(0|x_2) dF_{X_2}(x_2) = p_0^*, \int_{-B_1}^{B_2} x_2^2 dF_{X_2}(x_2) = P_2^* \right\} \tag{40}$$

is the intersection of $\mathscr{S}_2$ with two hyperplanes (note that $\mathscr{S}_2$ is convex and compact). We can write:

$$I^*_{F_{X_1}} = \sup_{F_{X_2} \in \mathscr{F}_2} I_\lambda(F_{X_1}F_{X_2}). \tag{41}$$

Note that having $F_{X_2} \in \mathscr{F}_2$, the objective function in (41) becomes:

$$\underbrace{\lambda H(Y)}_{\text{constant}} + \underbrace{(1-\lambda)H(Y|X_2) - H(Y|X_1, X_2)}_{\text{linear in } F_{X_2}}. \tag{42}$$

Since the linear part is continuous and $\mathscr{F}_2$ is compact (The continuity of the linear part follows similarly the continuity arguments in Appendix C. Note that this compactness is due to the closedness of the intersecting hyperplanes in $\mathscr{F}_2$, since a closed subset of a compact set is compact [11]. The hyperplanes are closed due to the continuity of x_2^2 and $p(0|x_2)$ (see (A16)).), the objective function in (41) attains its maximum at an extreme point of $\mathscr{F}_2$, which, by Dubins' theorem, is a convex combination of at most three extreme points of $\mathscr{S}_2$. Since the extreme points of $\mathscr{S}_2$ are the CDFs having only one point of increase in $[-B_1, B_2]$, we conclude that given any bounded F_{X_1}, $F^*_{X_2}$ has at most three mass points.

Now, assume that an arbitrary F_{X_2} is given with at most three mass points denoted by $\{x_{2,i}\}_{i=1}^3$. It is already known that the support of $F^*_{X_1}$ is bounded, which is denoted by $[-A_1, A_2]$. Let $\mathscr{S}_1$ denote the set of all probability distributions on the Borel sets of $[-A_1, A_2]$. The set:

$$\mathscr{F}_1 = \left\{ F_{X_1} \in \mathscr{S}_1 \middle| \int_{-A_1}^{A_2} p(0|x_1, x_{2,j}) dF_{X_1}(x_1) = p(0; F^*_{X_1}|x_{2,j}),\ j \in [1:3],\ \int_{-A_1}^{A_2} x_1^2 dF_{X_1}(x_1) = P_1' \right\}, \tag{43}$$

is the intersection of $\mathscr{S}_1$ with four hyperplanes. Note that here, since we know $\theta_1 \neq 0$, the optimal input attains its maximum power of P_1'. In a similar way,

$$I^*_{F_{X_2}} = \sup_{F_{X_1} \in \mathscr{F}_1} \left\{ I_\lambda(F_{X_1}F_{X_2}) \right\}, \tag{44}$$

and having $F_{X_1} \in \mathscr{F}_1$, the objective function in (44) becomes:

$$I_\lambda = \underbrace{\lambda H(Y) + (1-\lambda) \sum_{i=1}^{3} p_{X_2}(x_{2,i}) H(Y|X_2 = x_{2,i})}_{\text{constant}} - \underbrace{H(Y|X_1, X_2)}_{\text{linear in } F_{X_1}} \tag{45}$$

Therefore, given any F_{X_2} with at most three points of increase, $F^*_{X_1}$ has at most five mass points.

When $\lambda = 1$, the second term on the RHS of (45) disappears, which means that $\mathscr{F}_1$ could be replaced by:

$$\left\{F_{X_1} \in \mathscr{S}_1 | \int_{-A_1}^{A_2} p(0|x_1) dF_{X_1}(x_1) = \tilde{p}_0^*, \int_{-A_1}^{A_2} x_1^2 dF_{X_1}(x_1) = P_1'\right\},$$

where $\tilde{p}_0^* = p_Y(0; F^*_{X_1} F_{X_2})$ is the probability of the event $Y = 0$, which is induced by $F^*_{X_1}$ and the given F_{X_2}. Since the number of intersecting hyperplanes has been reduced to two, it is concluded that $F^*_{X_1}$ has at most three points of increase. □

Remark 2. *Note that, the order of showing the discreteness of the support sets is also important. First, for a given bounded F_{X_1} (not necessarily discrete), it is proven that $F^*_{X_2}$ is discrete with at most three mass points. Then, for a given discrete F_{X_2} with at most three mass points, it is shown that $F^*_{X_1}$ is also discrete with at most five mass points when $\lambda < 1$ and at most three mass points when $\lambda = 1$. When $\lambda > 1$, the order is reversed, and it follows the same steps as in the case of $\lambda < 1$. Therefore, it is omitted.*

Remark 3. *If $\mathcal{X}_1, \mathcal{X}_2$ are assumed finite initially, similar results can be obtained by using the iterative optimization in the previous proof and the approach in Chapter 4, Corollary 3 of [13].*

5. Sum Rate Analysis

In this section, we propose a lower bound on the sum capacity of a MAC in the presence of a one-bit ADC front end at the receiver, which we conjecture to be tight. The sum capacity is given by:

$$C_{\text{sum}} = \sup\ I(X_1, X_2; Y|U), \tag{46}$$

where the supremum is over $F_U F_{X_1|U} F_{X_2|U}$ ($|\mathcal{U}| \leq 5$), such that $\mathbb{E}[X_j^2] \leq P_j$, $j = 1, 2$. We obtain a lower bound for the above by considering only those input distributions that are zero-mean per any realization of the auxiliary random variable U, i.e., $\mathbb{E}[X_j|U = u] = 0, \forall u \in \mathcal{U}, j = 1, 2$. Let P_1' and P_2' be two arbitrary non-negative real numbers. We have:

$$\sup_{\substack{F_{X_1} F_{X_2}: \\ \mathbb{E}[X_j^2] \leq P_j' \\ \mathbb{E}[X_j] = 0,\ j=1,2}} I(X_1, X_2; Y) \leq \sup_{\substack{F_{\tilde{X}}: \\ \mathbb{E}[\tilde{X}^2] \leq P_1' + P_2'}} I(\tilde{X}; Y) \tag{47}$$

$$= 1 - H_b\left(Q\left(\sqrt{P_1' + P_2'}\right)\right) \tag{48}$$

where in (47), $\tilde{X} \triangleq X_1 + X_2$, $p_{Y|\tilde{X}}(0|\tilde{x}) = Q(\tilde{x})$; (48) follows from [4] for the point-to-point channel. Therefore, when $\mathbb{E}[X_j|U = u] = 0, \forall u \in \mathcal{U}, j = 1, 2$, we can write:

$$\begin{aligned} I(X_1, X_2; Y|U) &= \sum_{i=1}^{5} p_U(u_i) I(X_1, X_2; Y|U = u_i) \\ &\leq 1 - \sum_{i=1}^{5} p_U(u_i) H_b\left(Q\left(\sqrt{\mathbb{E}[X_1^2|U = u_i] + \mathbb{E}[X_2^2|U = u_i]}\right)\right) \\ &\leq 1 - H_b\left(Q\left(\sqrt{\mathbb{E}[X_1^2] + \mathbb{E}[X_2^2]}\right)\right) \qquad (49) \\ &\leq 1 - H_b\left(Q\left(\sqrt{P_1 + P_2}\right)\right), \qquad (50) \end{aligned}$$

where (49) is due to the fact that $H_b\left(Q(\sqrt{x + y})\right)$ is a convex function of (x, y), and (50) follows from $\mathbb{E}[X_j^2] \leq P_j, j = 1, 2$.

The upper bound in (50) can be achieved by time division with power control as follows. Let $\mathcal{U} = \{0,1\}$ and $p_U(0) = 1 - p_U(1) = \frac{P_1}{P_1+P_2}$. Furthermore, let $F_{X_1|U}(x|1) = F_{X_2|U}(x|0) = s(x)$, where $s(\cdot)$ is the unit step function, and:

$$F_{X_1|U}(x|0) = F_{X_2|U}(x|1) = \frac{1}{2}s(x + \sqrt{P_1 + P_2}) + \frac{1}{2}s(x - \sqrt{P_1 + P_2}).$$

With this choice of $F_U F_{X_1|U} F_{X_2|U}$, the upper bound in (50) is achieved. Therefore,

$$C_{\text{sum}} \geq 1 - H_b\left(Q\left(\sqrt{P_1 + P_2}\right)\right). \tag{51}$$

A numerical evaluation of (46) is carried out as follows (the codes that are used for the numerical simulations are available at https://www.dropbox.com/sh/ndxkjt6h5a0yktu/AAAmfHkuPxe8rMNV1KzFVRgNa?dl=0). Although $\mathbb{E}[X_j^2]$ is upper bounded by P_j ($j = 1,2$), the value of $\mathbb{E}[X_j^2|U = u]$ ($\forall u \in \mathcal{U}$) has no upper bound and could be any non-negative real number. However, in our numerical analysis, we further restrict our attention to the case $\mathbb{E}[X_j^2|U = u] \leq 20P_j, \forall u \in \mathcal{U}, j = 1,2$. Obviously, as this upper bound tends to infinity, the approximation becomes more accurate (This further bounding of the conditional second moments is justified by the fact that the sum capacity is not greater than one, which is due to the one-bit quantization at the receiver. As a result, $I(X_1, X_2; Y|U = u)$ increases at most sublinearly with $\mathbb{E}[X_j^2|U = u], j = 1,2$, while $p_U(u)$ needs to decrease at least linearly to satisfy the average power constraints. Hence, the product $p_U(u)I(X_1, X_2; Y|U = u)$ decreases with $\mathbb{E}[X_j^2|U = u]$ when $\mathbb{E}[X_j^2|U = u]$ is above a threshold.). Each of the intervals $[0, 20P_1]$ and $[0, 20P_2]$ are divided into 201 points uniformly, which results in the discrete intervals $\frac{P_1}{10}[0:200]$ and $\frac{P_2}{10}[0:200]$, respectively. Afterwards, for any pair $(\alpha, \beta) \in \frac{P_1}{10}[0:200] \times \frac{P_2}{10}[0:200]$, the following is carried out for input distributions with at most three mass points.

$$\max_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_1^2]\leq\alpha, \mathbb{E}[X_2^2]\leq\beta}} I(X_1, X_2; Y) \tag{52}$$

The results are stored in a 201×201 matrix accordingly. In the above optimization, the MATLAB function fmincon is used with three different initial values, and the maximum of these three experiments is chosen. Then, the problem boils down to finding proper gains, i.e., the mass probabilities of U, that maximize $I(X_1, X_2; Y|U)$ and satisfy the average power constraints $\mathbb{E}[X_j^2] \leq P_j$. This is done via a linear program, which can be efficiently solved by the linprog function in MATLAB. Several cases were considered, such as $(P_1, P_2) = (1,1), (P_1, P_2) = (1,2), (P_1, P_2) = (3,1)$, etc. In all these cases, the numerical evaluation of (46) leads to the same value as the lower bound in (51). Since the problem is not convex, it is not known whether the numerical results are the global optimum solutions; hence, we leave it as a conjecture that the sum capacity can be achieved by time division with power control.

6. Conclusions

We have studied the capacity region of a two-transmitter Gaussian MAC under average input power constraints and one-bit ADC front end at the receiver. We have derived an upper bound on the cardinality of this auxiliary variable, and proved that the distributions that achieve the boundary points of the capacity region are finite and discrete. Finally, a lower bound is proposed on the sum capacity of this MAC that is achieved by time division with power control. Through numerical analysis, this lower bound is shown to be tight.

Author Contributions: All three authors contributed to the paper. B.R. derived most of the claims with discussion with the M.V. and D.G.; The numerical results regarding the sum-rate analysis were obtained by the M.V.; Writing and editing of the paper has been done jointly, and all three authors have read and approved the final manuscript.

Funding: This research was supported in part by the European Research Council (ERC) through Starting Grant BEACON (Agreement No. 677854), by the U.K. Engineering and Physical Sciences Research Council (EPSRC) through the project COPES (EP/N021738/1) and by the British Council Institutional Link Program under Grant NO. 173605884.

Acknowledgments: The authors thank Professor Barbie and Professor Shirokov for their help in showing the preservation of the Markov chain in the weak convergence of the joint distributions in Appendix B.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The capacity region of the discrete memoryless (DM) MAC with input cost constraints has been addressed in Exercise 4.8 of [10]. If the input alphabets are not discrete, the capacity region is still the same because: (1) the converse remains the same if the inputs are from a continuous alphabet; (2) the region is achievable by coded time sharing and the discretization procedure (see Remark 3.8 in [10]). Therefore, it is sufficient to show the cardinality bound $|\mathcal{U}| \leq 5$.

Let $\mathscr{P}$ be the set of all product distributions (i.e., of the form $F_{X_1}(x_1)F_{X_2}(x_2)$) on $\mathbb{R}^2$. Let $\mathbf{g} : \mathscr{P} \to \mathbb{R}^5$ be a vector-valued mapping defined element-wise as:

$$\begin{aligned}
g_1(F_{X_1|U}(\cdot|u)F_{X_2|U}(\cdot|u)) &= I(X_1;Y|X_2,U=u),\\
g_2(F_{X_1|U}(\cdot|u)F_{X_2|U}(\cdot|u)) &= I(X_2;Y|X_1,U=u),\\
g_3(F_{X_1|U}(\cdot|u)F_{X_2|U}(\cdot|u)) &= I(X_1,X_2;Y|U=u),\\
g_4(F_{X_1|U}(\cdot|u)F_{X_2|U}(\cdot|u)) &= \mathbb{E}[X_1^2|U=u],\\
g_5(F_{X_1|U}(\cdot|u)F_{X_2|U}(\cdot|u)) &= \mathbb{E}[X_2^2|U=u].
\end{aligned} \tag{A1}$$

Let $\mathscr{G} \subset \mathbb{R}^5$ be the image of $\mathscr{P}$ under the mapping $\mathbf{g}$ (i.e., $\mathscr{G} = \mathbf{g}(\mathscr{P})$). Given an arbitrary $(U, X_1, X_2) \sim F_U F_{X_1|U} F_{X_2|U}$, we obtain the vector $\mathbf{r}$ as:

$$\begin{aligned}
r_1 &= I(X_1;Y|X_2,U) = \int_{\mathcal{U}} I(X_1;Y|X_2,U=u)dF_U(u),\\
r_2 &= I(X_2;Y|X_1,U) = \int_{\mathcal{U}} I(X_2;Y|X_1,U=u)dF_U(u),\\
r_3 &= I(X_1,X_2;Y|U) = \int_{\mathcal{U}} I(X_1,X_2;Y|U=u)dF_U(u),\\
r_4 &= \mathbb{E}[X_1^2] = \int_{\mathcal{U}} \mathbb{E}[X_1^2|U=u]dF_U(u),\\
r_5 &= \mathbb{E}[X_2^2] = \int_{\mathcal{U}} \mathbb{E}[X_2^2|U=u]dF_U(u).
\end{aligned}$$

Therefore, $\mathbf{r}$ is in the convex hull of $\mathscr{G} \subset \mathbb{R}^5$. By Carathéodory's theorem [9], $\mathbf{r}$ can be written as a convex combination of six ($= 5+1$) or fewer points in $\mathscr{G}$, which states that it is sufficient to consider $|\mathcal{U}| \leq 6$. Since $\mathscr{P}$ is a connected set ($\mathscr{P}$ is the product of two connected sets; therefore, it is connected. Each of the sets in this product is connected because of being a convex vector space.) and the mapping $\mathbf{g}$ is continuous (this is a direct result of the continuity of the channel transition probability), $\mathscr{G}$ is a connected subset of $\mathbb{R}^5$. Therefore, the connectedness of $\mathscr{G}$ refines the cardinality of U to $|\mathcal{U}| \leq 5$.

It is also important to note that for the boundary points of $\mathscr{C}(P_1, P_2)$ that are not sum-rate optimal, it is sufficient to have $|\mathcal{U}| \leq 4$. The proof is as follows. Any point on the boundary of the capacity region that does not maximize $R_1 + R_2$ is either of the form $(I(X_1;Y|X_2,U), I(X_2;Y|U))$ or $(I(X_1;Y|U), I(X_2;Y|X_1,U))$ for some $F_U F_{X_1|U} F_{X_2|U}$ that satisfies $\mathbb{E}[X_j^2] \leq P_j, j = 1,2$. In other words, it is one of the corner points of the corresponding pentagon in (4). As in the proof of Proposition 1, define the mapping $\mathbf{g} : \mathscr{P} \to \mathbb{R}^4$, where g_1 and g_2 are the coordinates of this boundary point

conditioned on $U = u$, and g_3, g_4 are the same as g_4 and g_5 in (A1), respectively. The sufficiency of $|\mathcal{U}| \leq 4$ in this case follows similarly.

Appendix B

Since $|\mathcal{U}| \leq 5$, we assume $\mathcal{U} = \{0, 1, 2, 3, 4\}$ without loss of generality, since what matters in the evaluation of the capacity region is the mass probability of the auxiliary random variable U, not its actual values.

In order to show the compactness of Ω, we adopt a general form of the approach in [12].

First, we show that Ω is tight (a set of probability distributions Θ defined on $\mathbb{R}^k$, i.e., the set of CDFs $F_{X_1,X_2,\ldots,X_k}$, is said to be tight, if for every $\epsilon > 0$, there is a compact set $K_\epsilon \subset \mathbb{R}^k$ such that [14]:

$$\Pr\{(X_1, X_2, \ldots, X_k) \in \mathbb{R}^k \backslash K_\epsilon\} < \epsilon, \ \forall F_{X_1,X_2,\ldots,X_k} \in \Theta.$$

Choose T_j, $j = 1, 2$, such that $T_j > \sqrt{\frac{2P_j}{\epsilon}}$. Then, from Chebyshev's inequality,

$$\Pr\{|X_j| > T_j\} \leq \frac{P_j}{T_j^2} < \frac{\epsilon}{2}, \ j = 1, 2. \tag{A2}$$

Let $K_\epsilon = [0, 4] \times [-T_1, T_1] \times [-T_2, T_2] \subset \mathbb{R}^3$. It is obvious that K_ϵ is a closed and bounded subset of $\mathbb{R}^3$ and, therefore, compact. With this choice of K_ϵ, we have:

$$\begin{aligned} \Pr\{(U, X_1, X_2) \in \mathbb{R}^3 \backslash K_\epsilon\} &\leq \Pr\{U \notin [0, 4]\} + \Pr\{X_1 \notin [-T_1, T_1]\} + \Pr\{X_2 \notin [-T_2, T_2]\} \\ &< 0 + \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \end{aligned} \tag{A3}$$

where (A3) is due to (A2). Hence, Ω is tight.

From Prokhorov's theorem [14] (p. 318), a set of probability distributions is tight if and only if it is relatively sequentially compact (a subset of topological space is relatively compact if its closure is compact). This means that for every sequence of CDFs $\{F_n\}$ in Ω, there exists a subsequence $\{F_{n_k}\}$ that is weakly convergent (the weak convergence of $\{F_n\}$ to F (also shown as $F_n(x) \xrightarrow{w} F(x)$) is equivalent to:

$$\lim_{n\to\infty} \int_{\mathbb{R}} \psi(x) dF_n(x) = \int_{\mathbb{R}} \psi(x) dF(x), \tag{A4}$$

for all continuous and bounded functions $\psi(\cdot)$ on $\mathbb{R}$. Note that $F_n(x) \xrightarrow{w} F(x)$ if and only if $d_L(F_n, F) \to 0$.) to a CDF F_0, which is not necessarily in Ω. If we can show that this F_0 is also an element of Ω, then the proof is complete, since we have shown that Ω is sequentially compact and, therefore, compact (Compactness and sequential compactness are equivalent in metric spaces. Note that Ω is a metric space with Lévy distance.).

Assume a sequence of distributions $\{F_n(\cdot, \cdot, \cdot)\}$ in Ω that converges weakly to $F_0(\cdot, \cdot, \cdot)$. In order to show that this limiting distribution is also in Ω, we need to show that both the average power constraints and the Markov chain $(X_1 - U - X_2)$ are preserved under F_0. The preservation of the second moment follows similarly to the argument in (Appendix I, [12]). In other words, since x^2 is continuous and bounded from below, from Theorem 4.4.4 in [15]:

$$\begin{aligned} \int x_j^2 d^3 F_0(u, x_1, x_2) &\leq \liminf_{n\to\infty} \int x_j^2 d^3 F_n(u, x_1, x_2) \\ &\leq P_j, \ j = 1, 2, \end{aligned} \tag{A5}$$

Therefore, the second moments are preserved under the limiting distribution F_0.

For the preservation of the Markov chain $X_1 - U - X_2$, we need the following proposition.

Proposition A1. *Assume a sequence of distributions* $\{F_n(\cdot,\cdot)\}$ *over the pair of random variables* (X,Y) *that converges weakly to* $F_0(\cdot,\cdot)$. *Furthermore, assume that* Y *has a finite support, i.e.,* $\mathcal{Y} = \{1,2,\ldots,|\mathcal{Y}|\}$. *Then, the sequence of conditional distributions (conditioned on* Y*) converges weakly to the limiting conditional distribution (conditioned on* Y*), i.e.,*

$$F_n(\cdot|y) \xrightarrow{w} F_0(\cdot|y),\ \forall y \in \mathcal{Y}, p_0(y) > 0. \tag{A6}$$

Proof of Proposition A1. The proof is by contradiction. If (A6) is not true, then there exists $y' \in \mathcal{Y}$, such that $p_0(y') > 0$ and $F_n(\cdot|y') \not\xrightarrow{w} F_0(\cdot|y')$. This means, from the definition of weak convergence, that there exists a bounded continuous function of x, denoted by $g_{y'}(x)$, such that:

$$\int g_{y'}(x) dF_n(x|y') \not\to \int g_{y'}(x) dF_0(x|y'). \tag{A7}$$

Let $f(x,y)$ be any bounded continuous function that satisfies:

$$f(x,y) = \begin{cases} 0 & y \in \mathcal{Y},\ y \neq y' \\ g_{y'}(x) & y = y' \end{cases}. \tag{A8}$$

With this choice of $f(x,y)$, we have:

$$\int f(x,y) d^2F_n(x,y) \not\to \int f(x,y) d^2F_0(x,y), \tag{A9}$$

which violates the assumption of the weak convergence of $F_n(\cdot,\cdot)$ to $F_0(\cdot,\cdot)$. Therefore, (A6) holds. □

Since $\{F_n(\cdot,\cdot,\cdot)\}$ in Ω converges weakly to $F_0(\cdot,\cdot,\cdot)$ and $\mathcal{U}$ is finite, from Proposition A1, we have:

$$F_n(\cdot,\cdot|u) \xrightarrow{w} F_0(\cdot,\cdot|u),\ \forall u \in \mathcal{U}, \tag{A10}$$

where it is obvious that the arguments are x_1 and x_2. Since $F_n \in \Omega$, we have $F_n(x_1,x_2|u) = F_n(x_1|u)F_n(x_2|u)\ \forall u \in \mathcal{U}$. Furthermore, since the convergence of the joint distribution implies the convergence of the marginals, we have (Theorem 2.7, [16,17]),

$$F_0(x_1,x_2|u) = F_0(x_1|u)F_0(x_2|u)\ \forall u \in \mathcal{U}, \tag{A11}$$

which states that under the limiting distribution F_0, the Markov chain $X_1 - U - X_2$ is preserved (Alternatively, this could be proven by the lower-semicontinuity of the mutual information as follow:

$$I_{F_0}(X_1;X_2|U=u) \leq \liminf_{n\to\infty} I_{F_n}(X_1;X_2|U=u) \tag{A12}$$

$$= 0,\ \forall u \in \mathcal{U}, \tag{A13}$$

where I_F denotes the mutual information under distribution F. The last equality is from the conditional independence of X_1 and X_2 given $U=u$ under F_n. Therefore, $I_{F_0}(X_1;X_2|U=u) = 0,\ \forall u \in \mathcal{U}$, which is equivalent to (A11).). This completes the proof of the compactness of Ω.

Appendix C

Appendix C.1. Concavity

When $0 < \lambda \leq 1$, we have:

$$I_\lambda(F_{X_1}F_{X_2}) = \lambda H(Y) + (1-\lambda)H(Y|X_2) - H(Y|X_1,X_2). \tag{A14}$$

For a given F_{X_1}, $H(Y)$ is a concave function of F_{X_2}, while $H(Y|X_2)$ and $H(Y|X_1, X_2)$ are linear in F_{X_2}. Therefore, I_λ is a concave function of F_{X_2}. For a given F_{X_2}, $H(Y)$ and $H(Y|X_2)$ are concave functions of F_{X_1}, while $H(Y|X_1, X_2)$ is linear in F_{X_1}. Since $(1-\lambda) \geq 0$, I_λ is a concave function of F_{X_1}. The same reasoning applies to the case $\lambda > 1$.

Appendix C.2. Continuity

When $\lambda \leq 1$, the continuity of the three terms on the RHS of (A14) is investigated. Let $\{F_{X_2,n}\}$ be a sequence of distributions, which is weakly convergent to F_{X_2}. For a given F_{X_1}, we have:

$$\begin{aligned}\lim_{x_2 \to x_2^0} p(y; F_{X_1}|x_2) &= \lim_{x_2 \to x_2^0} \int Q(x_1 + x_2) dF_{X_1}(x_1) \\ &= \int \lim_{x_2 \to x_2^0} Q(x_1 + x_2) dF_{X_1}(x_1) \qquad \text{(A15)} \\ &= p(y; F_{X_1}|x_2^0), \qquad \text{(A16)}\end{aligned}$$

where (A15) is due to the fact that the Q function can be dominated by one, which is an absolutely integrable function over F_{X_1}. Therefore, $p(y; F_{X_1}|x_2)$ is continuous in x_2, and combined with the weak convergence of $\{F_{X_2,n}\}$, we can write:

$$\begin{aligned}\lim_{n \to \infty} p(y; F_{X_1} F_{X_2,n}) &= \lim_{n \to \infty} \int p(y; F_{X_1}|x_2) dF_{X_2,n}(x_2) \\ &= \int p(y; F_{X_1}|x_2) dF_{X_2}(x_2) \\ &= p(y; F_{X_1} F_{X_2}).\end{aligned}$$

This allows us to write:

$$\lim_{n \to \infty} -\sum_{y=0}^{1} p(y; F_{X_1} F_{X_2,n}) \log p(y; F_{X_1} F_{X_2,n}) = -\sum_{y=0}^{1} p(y; F_{X_1} F_{X_2}) \log p(y; F_{X_1} F_{X_2}),$$

which proves the continuity of $H(Y)$ in F_{X_2}. $H(Y|X_2 = x_2)$ is a bounded ($\in [0,1]$) continuous function of x_2, since it is a continuous function of $p(y; F_{X_1}|x_2)$, and the latter is continuous in x_2 (see (A16)). Therefore,

$$\lim_{n \to \infty} \int H(Y|X_2 = x_2) dF_{X_2,n}(x_2) = \int H(Y|X_2 = x_2) dF_{X_2}(x_2),$$

which proves the continuity of $H(Y|X_2)$ in F_{X_2}. In a similar way, it can be verified that $\int H(Y|X_1 = x_1, X_2 = x_2) dF_{X_1}(x_1)$ is a bounded and continuous function of x_2, which guarantees the continuity of $H(Y|X_1, X_2)$ in F_{X_2}, since:

$$H(Y|X_1, X_2) = \int \left(\int H(Y|X_1 = x_1, X_2 = x_2) dF_{X_1}(x_1) \right) dF_{X_2}(x_2) \qquad \text{(A17)}$$

Therefore, for a given F_{X_1}, I_λ is a continuous function of F_{X_2}. Exchanging the roles of F_{X_1} and F_{X_2}, also the case $\lambda > 1$ can be addressed similarly, so they are omitted for the sake of brevity.

Appendix C.3. Weak Differentiability

For a given F_{X_1}, the weak derivative of I_λ at $F_{X_2}^0$ is given by:

$$I'_\lambda(F_{X_1} F_{X_2})|_{F_{X_2}^0} = \lim_{\beta \to 0^+} \frac{I_\lambda(F_{X_1}((1-\beta)F_{X_2}^0 + \beta F_{X_2})) - I_\lambda(F_{X_1} F_{X_2}^0)}{\beta}, \qquad \text{(A18)}$$

if the limit exists. It can be verified that:

$$\begin{aligned}
&I'_\lambda(F_{X_1}F_{X_2})|_{F^0_{X_2}} \\
&= \lim_{\beta\to 0^+} \frac{\int i_\lambda(x_2;(1-\beta)F^0_{X_2}+\beta F_{X_2}|F_{X_1})d((1-\beta)F^0_{X_2}(x_2)+\beta F_{X_2}(x_2)) - \int i_\lambda(x_2;F^0_{X_2}|F_{X_1})dF^0_{X_2}(x_2)}{\beta} \\
&= \int i_\lambda(x_2;F^0_{X_2}|F_{X_1})dF_{X_2}(x_2) - \int i_\lambda(x_2;F^0_{X_2}|F_{X_1})dF^0_{X_2}(x_2) \\
&= \int i_\lambda(x_2;F^0_{X_2}|F_{X_1})dF_{X_2}(x_2) - I_\lambda(F_{X_1}F^0_{X_2}),
\end{aligned}$$

where i_λ has been defined in (17). In a similar way, for a given F_{X_2}, the weak derivative of I_λ at $F^0_{X_1}$ is:

$$I'_\lambda(F_{X_1}F_{X_2})|_{F^0_{X_1}} = \int \tilde{i}_\lambda(x_1;F^0_{X_1}|F_{X_2})dF_{X_1}(x_1) - I_\lambda(F^0_{X_1}F_{X_2}), \tag{A19}$$

where $\tilde{i}_\lambda$ has been defined in (16). The case $\lambda > 1$ can be addressed similarly.

Appendix D

We have:

$$\begin{aligned}
\sup_{\substack{F_{X_1}:\\ \mathbb{E}[X_1^2]\le P'_1}} I_\lambda(F_{X_1}F_{X_2}) &\le \sup_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P'_j,\, j=1,2}} I_\lambda(F_{X_1}F_{X_2}) \\
&\le \sup_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P'_j,\, j=1,2}} I(X_1,X_2;Y) && \text{(A20)} \\
&\le \sup_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P'_j,\, j=1,2}} H(Y) - \inf_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P'_j,\, j=1,2}} H(Y|X_1,X_2) \\
&= 1 - \inf_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P'_j,\, j=1,2}} \iint H_b\left(Q(x_1+x_2)\right) dF_{X_1}(x_1)dF_{X_2}(x_2) \\
&= 1 - \inf_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P'_j,\, j=1,2}} \iint H_b\left(Q\left(\sqrt{x_1^2}+\sqrt{x_2^2}\right)\right) dF_{X_1}(x_1)dF_{X_2}(x_2) && \text{(A21)} \\
&\le 1 - \inf_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P'_j,\, j=1,2}} \iint Q\left(\sqrt{x_1^2}+\sqrt{x_2^2}\right) dF_{X_1}(x_1)dF_{X_2}(x_2) && \text{(A22)} \\
&= 1 - Q\left(\sqrt{P'_1}+\sqrt{P'_2}\right) && \text{(A23)} \\
&< 1, && \text{(A24)}
\end{aligned}$$

where (A20) is from the non-negativity of mutual information and the assumption that $0 < \lambda \le 1$; (A21) is justified since the Q function is monotonically decreasing and the sign of the inputs does not affect the average power constraints; X_1 and X_2 can be assumed non-negative (or alternatively non-positive) without loss of optimality; in (A22), we use the fact that $Q\left(\sqrt{x_1^2}+\sqrt{x_2^2}\right) \le \frac{1}{2}$, and for $t \in [0, \frac{1}{2}]$, $H_b(t) \ge t$; (A23) is based on the convexity and monotonicity of the function $Q(\sqrt{u}+\sqrt{v})$ in (u,v), which is shown in Appendix G. Therefore, the LHS of (13) is strictly less than one.

Since X_2 has a finite second moment ($\mathbb{E}[X_2^2] \leq P_2'$), from Chebyshev's inequality, we have:

$$P(|X_2| \geq M) \leq \frac{P_2'}{M^2}, \ \ \forall M > 0. \tag{A25}$$

Fix $M > 0$, and consider $X_1 \sim F_{X_1}(x_1) = \frac{1}{2}[s(x_1 + 2M) + s(x_1 - 2M)]$. By this choice of F_{X_1}, we get:

$$\begin{aligned} I_\lambda(F_{X_1}F_{X_2}) &= I(X_1;Y|X_2) + \lambda I(X_2;Y) \\ &\geq I(X_1;Y|X_2) \qquad\qquad (A26) \\ &= \int_{-\infty}^{+\infty} I(X_1;Y|X_2 = x_2) dF_{X_2}(x_2) \\ &\geq \int_{-M}^{+M} I(X_1;Y|X_2 = x_2) dF_{X_2}(x_2) \\ &\geq \inf_{F_{X_2}} \int_{-M}^{+M} H(Y|X_2 = x_2) dF_{X_2}(x_2) - \sup_{F_{X_2}} \int_{-M}^{+M} H(Y|X_1, X_2 = x_2) dF_{X_2}(x_2) \\ &\geq \left(1 - \frac{P_2'}{M^2}\right) H_b\left(\frac{1}{2} - \frac{1}{2}(Q(3M) + Q(M))\right) - H_b\left(Q(2M)\right), \qquad\qquad (A27) \end{aligned}$$

where (A27) is due to (A25) and the fact that $H(Y|X_2 = x_2) = H_b(\frac{1}{2}Q(2M + x_2) + \frac{1}{2}Q(-2M + x_2))$ is minimized over $[-M, M]$ at $x_2 = M$ (or, alternatively at $x_2 = -M$) and $H(Y|X_1, X_2 = x_2) = \frac{1}{2}H_b(Q(2M + x_2)) + \frac{1}{2}H_b(Q(-2M + x_2))$ is maximized at $x_2 = 0$. (A27) shows that I_λ (≤ 1) can become arbitrarily close to one given that M is large enough. Hence, its supremum over all distributions F_{X_1} is one. This means that (13) cannot hold, and $\theta_1 \neq 0$.

Appendix E. Justification of (18), (19) and (20)

Let X be a vector space and Z be a real-valued function defined on a convex domain $D \subset X$. Suppose that x^* maximizes Z on D and that Z is Gateaux differentiable (weakly differentiable) at x^*. Then, from (Theorem 2, p. 178, [18]),

$$Z'(x)|_{x^*} \leq 0, \tag{A28}$$

where $Z'(x)|_{x^*}$ is the weak derivative of Z at x^*.

From (A19), we have the weak derivative of I_λ at $F_{X_1}^*$ as:

$$I_\lambda'(F_{X_1}F_{X_2})|_{F_{X_1}^*} = \int \tilde{i}_\lambda(x_1; F_{X_1}^*|F_{X_2}) dF_{X_1}(x_1) - I_\lambda(F_{X_1}^*F_{X_2}). \tag{A29}$$

Now, the derivation of (18) is immediate by inspecting that the weak derivative of the objective of (11) at $F_{X_1}^*$ is given by:

$$\begin{aligned} I_\lambda'(F_{X_1}F_{X_2})|_{F_{X_1}^*} - \theta_1 \left(\int x_1^2 dF_{X_1}(x_1) - \int x_1^2 dF_{X_1}^*(x_1)\right) &= \int \tilde{i}_\lambda(x_1; F_{X_1}^*|F_{X_2}) dF_{X_1}(x_1) - I_\lambda(F_{X_1}^*F_{X_2}) \\ &\quad - \theta_1 \left(\int x_1^2 dF_{X_1}(x_1) - \int x_1^2 dF_{X_1}^*(x_1)\right). \qquad (A30) \end{aligned}$$

Letting (A30) be lower than or equal to zero (as in (A28)) results in (18).

The equivalence of (18) to (19) and (20) follows similarly to the proof of Corollary 1 in (p. 210, [19]).

Appendix F

Equation (21) is obtained as follows.

$$\begin{aligned}\left|D\left(p(y|x_1,x_2)||p(y;F_{X_1}F_{X_2})\right)\right| &= \left|\sum_{y=0}^{1} p(y|x_1,x_2)\log\frac{p(y|x_1,x_2)}{p(y;F_{X_1}F_{X_2})}\right| \\ &\le \left|H(Y|X_1=x_1,X_2=x_2)\right| + \left|\sum_{y=0}^{1} p(y|x_1,x_2)\log p(y;F_{X_1}F_{X_2})\right| \\ &\le 1+\left|\sum_{y=0}^{1}\log p(y;F_{X_1}F_{X_2})\right| \qquad \text{(A31)} \\ &= 1-\sum_{y=0}^{1}\log p(y;F_{X_1}F_{X_2}) \\ &\le 1-2\min\left\{\log p_Y(0;F_{X_1}F_{X_2}),\log p_Y(1;F_{X_1}F_{X_2})\right\} \\ &\le 1-2\log Q(\sqrt{P_1'}+\sqrt{P_2'}) \qquad \text{(A32)} \\ &< \infty,\end{aligned}$$

where (A31) is due to the fact that the binary entropy function is upper bounded by one. (A32) is justified as follows.

$$\begin{aligned}\min\left\{p_Y(0;F_{X_1}F_{X_2}),p_Y(1;F_{X_1}F_{X_2})\right\} &\ge \inf_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P_j'}} \min\left\{p_Y(0;F_{X_1}F_{X_2}),p_Y(1;F_{X_1}F_{X_2})\right\} \\ &= \inf_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P_j'}} p_Y(0;F_{X_1}F_{X_2}) \\ &= \inf_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P_j'}} \iint Q(x_1+x_2)dF_{X_1}(x_1)dF_{X_2}(x_2) \\ &= \inf_{\substack{F_{X_1}F_{X_2}:\\ \mathbb{E}[X_j^2]\le P_j'}} \iint Q\left(\sqrt{x_1^2}+\sqrt{x_2^2}\right)dF_{X_1}(x_1)dF_{X_2}(x_2) \qquad \text{(A33)} \\ &\ge Q\left(\sqrt{P_1'}+\sqrt{P_2'}\right), \qquad \text{(A34)}\end{aligned}$$

where (A34) is based on the convexity and monotonicity of the function $Q(\sqrt{u}+\sqrt{v})$, which is shown in Appendix G.

Equation (22) is obtained as follows.

$$\begin{aligned}p(y;F_{X_1}|x_2) &\ge \min\left\{p(0;F_{X_1}|x_2),p(1;F_{X_1}|x_2)\right\} \\ &\ge \int Q\left(|x_1|+|x_2|\right)dF_{X_1}(x_1) \\ &= \int Q\left(\sqrt{x_1^2}+|x_2|\right)dF_{X_1}(x_1) \\ &\ge Q\left(\sqrt{P_1'}+|x_2|\right), \qquad \text{(A35)}\end{aligned}$$

where (A35) is due to the convexity of $Q(\alpha+\sqrt{x})$ in x for $\alpha \ge 0$.

Equation (23) is obtained as follows.

$$\left|\sum_{y=0}^{1} p(y|x_1,x_2)\log\frac{p(y;F_{X_1}F_{X_2})}{p(y;F_{X_1}|x_2)}\right| \leq -\sum_{y=0}^{1} p(y|x_1,x_2)\log p(y;F_{X_1}|x_2) - \sum_{y=0}^{1} p(y|x_1,x_2)\log p(y;F_{X_1}F_{X_2})$$

$$\leq -\sum_{y=0}^{1}\log p(y;F_{X_1}|x_2) - \sum_{y=0}^{1}\log p(y;F_{X_1}F_{X_2}) \tag{A36}$$

$$\leq -2\log Q\left(\sqrt{P_1'}+\sqrt{P_2'}\right) - 2\log Q\left(\sqrt{P_1'}+|x_2|\right), \tag{A37}$$

where (A36) is from $p(y|x_1,x_2)\leq 1$; and (A37) is from (A35) and (A34).

Note that (A37) is integrable with respect to F_{X_2} due to the concavity of $-\log Q(\alpha+\sqrt{x})$ in x for $\alpha\geq 0$ as shown in Appendix G. In other words,

$$\int_{-\infty}^{+\infty}\left(-2\log Q\left(\sqrt{P_1'}+\sqrt{P_2'}\right) - 2\log Q\left(\sqrt{P_1'}+|x_2|\right)\right)dF_{X_2}(x_2) < -4\log Q\left(\sqrt{P_1'}+\sqrt{P_2'}\right) \tag{A38}$$

$$< +\infty. \tag{A39}$$

Appendix G. Two Convex Functions

Let $f(x)=\log Q(a+\sqrt{x})$ for $x,a\geq 0$. We have,

$$f'(x) = -\frac{e^{-\frac{(a+\sqrt{x})^2}{2}}}{2\sqrt{2\pi x}Q(a+\sqrt{x})},$$

and:

$$f''(x) = \frac{e^{-\frac{(a+\sqrt{x})^2}{2}}}{4x\sqrt{2\pi}Q^2(a+\sqrt{x})}\left((a+\sqrt{x}+\frac{1}{\sqrt{x}})Q(a+\sqrt{x}) - \phi(a+\sqrt{x})\right), \tag{A40}$$

where $\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Note that:

$$(1+at+t^2)Q(a+t)+a\phi(a+t) > \left(1+(a+t)^2\right)Q(a+t) \tag{A41}$$

$$> (a+t)\phi(a+t),\ \ \forall a,t>0, \tag{A42}$$

where (A41) and (A42) are, respectively, due to $\phi(x)>xQ(x)$ and $(1+x^2)Q(x)>x\phi(x)$ $(x>0)$. Therefore,

$$(a+\sqrt{x}+\frac{1}{\sqrt{x}})Q(a+\sqrt{x}) > \phi(a+\sqrt{x}),$$

which makes the second derivative in (A40) positive and proves the (strict) convexity of $f(x)$.

Let $f(u,v)=Q(\sqrt{u}+\sqrt{v})$ for $u,v\geq 0$. By simple differentiation, the Hessian matrix of f is:

$$\mathbf{H} = \frac{e^{-\frac{(\sqrt{u}+\sqrt{v})^2}{2}}}{\sqrt{2\pi}}\begin{bmatrix} \frac{1}{2u\sqrt{u}}+\frac{\sqrt{u}+\sqrt{v}}{4u} & \frac{\sqrt{u}+\sqrt{v}}{4\sqrt{u}\sqrt{v}} \\ \frac{\sqrt{u}+\sqrt{v}}{4\sqrt{u}\sqrt{v}} & \frac{1}{2v\sqrt{v}}+\frac{\sqrt{u}+\sqrt{v}}{4v}\end{bmatrix}. \tag{A43}$$

It can be verified that $\det(\mathbf{H})>0$ and $\mathrm{trace}(\mathbf{H})>0$. Therefore, both eigenvalues of $\mathbf{H}$ are positive, which makes the matrix positive definite. Hence, $Q(\sqrt{u}+\sqrt{v})$ is (strictly) convex in (u,v).

Appendix H

Let $A\triangleq\max\{A_1,A_2\}$.

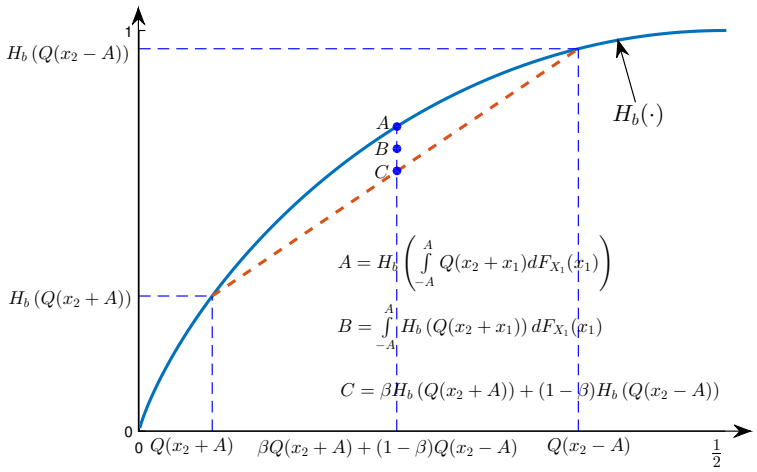


Figure A1. The figure depicting (A46) and (A48). Note that in the statement of Lemma 3, $x_2 \to +\infty$. Hence, we have assumed $x_2 > A$ in the figure.

It is obvious that:

$$Q(x_2+A) \leq \int_{-A}^{A} Q(x_1+x_2)dF_{X_1}(x_1) \leq Q(x_2-A). \tag{A44}$$

Therefore, we can write:

$$\int_{-A}^{A} Q(x_1+x_2)dF_{X_1}(x_1) = \beta Q(x_2+A) + (1-\beta)Q(x_2-A), \tag{A45}$$

for some $\beta \in [0,\ 1]$. Note that β is a function of x_2. Furthermore, due to the concavity of $H_b(\cdot)$, we have:

$$H_b\left(\int_{-A}^{A} Q(x_1+x_2)dF_{X_1}(x_1)\right) \geq \int_{-A}^{A} H_b(Q(x_1+x_2))dF_{X_1}(x_1). \tag{A46}$$

From the fact that:

$$H_b(x) \geq \frac{H_b(p)-H_b(a)}{p-a}(x-a) + H_b(a),\ \forall x \in [a,\ p],\ \forall a,p \in [0,\ 1](a<p), \tag{A47}$$

we can also write:

$$\begin{aligned}\int_{-A}^{A} H_b(Q(x_1+x_2))dF_{X_1}(x_1) &\geq \frac{H_b(Q(x_2-A))-H_b(Q(x_2+A))}{Q(x_2-A)-Q(x_2+A)}\left(\int_{-A}^{A} Q(x_1+x_2)dF_{X_1}(x_1) - Q(x_2+A)\right)\\ &\quad + H_b(Q(x_2+A))\\ &= \beta H_b(Q(x_2+A)) + (1-\beta)H_b(Q(x_2-A)),\end{aligned} \tag{A48}$$

where (A45) and (A47) have been used in (A48). (A46) and (A48) are depicted in Figure A1.

From (A45) and (A48), we have:

$$\frac{\beta H_b(Q(x_2+A)) + (1-\beta)H_b(Q(x_2-A))}{H_b\Big(\beta Q(x_2+A) + (1-\beta)Q(x_2-A)\Big)} \leq \frac{\int_{-A}^{A} H_b(Q(x_1+x_2))dF_{X_1}(x_1)}{H_b\left(\int_{-A}^{A} Q(x_1+x_2)dF_{X_1}(x_1)\right)} \leq 1. \tag{A49}$$

Let:

$$\beta^* \triangleq \arg\min_{\beta} \frac{\beta H_b(Q(x_2+A)) + (1-\beta)H_b(Q(x_2-A))}{H_b\Big(\beta Q(x_2+A) + (1-\beta)Q(x_2-A)\Big)}. \tag{A50}$$

This minimizer satisfies the following equality:

$$\frac{d}{d\beta}\left(\frac{\beta H_b(Q(x_2+A)) + (1-\beta)H_b(Q(x_2-A))}{H_b\Big(\beta Q(x_2+A) + (1-\beta)Q(x_2-A)\Big)}\right)\Bigg|_{\beta=\beta^*} = 0. \tag{A51}$$

Therefore, we can write:

$$\frac{\beta H_b(Q(x_2+A)) + (1-\beta)H_b(Q(x_2-A))}{H_b\Big(\beta Q(x_2+A) + (1-\beta)Q(x_2-A)\Big)} \geq \frac{\beta^* H_b(Q(x_2+A)) + (1-\beta^*)H_b(Q(x_2-A))}{H_b\Big(\beta^* Q(x_2+A) + (1-\beta^*)Q(x_2-A)\Big)} \tag{A52}$$

$$= \frac{\frac{H_b(Q(x_2-A))-H_b(Q(x_2+A))}{Q(x_2-A)-Q(x_2+A)}}{H_b'\Big(\beta^* Q(x_2+A) + (1-\beta^*)Q(x_2-A)\Big)} \tag{A53}$$

$$\geq \frac{\frac{H_b(Q(x_2-A))-H_b(Q(x_2+A))}{Q(x_2-A)-Q(x_2+A)}}{H_b'(Q(x_2+A))}, \tag{A54}$$

where (A52) is from the definition in (A50); (A53) is from the expansion of (A51), and $H_b'(t) = \log(\frac{1-t}{t})$ is the derivative of the binary entropy function; (A54) is due to the fact that $H_b'(t)$ is a decreasing function.

Applying L'Hôspital's rule multiple times, we obtain:

$$\begin{aligned}
\lim_{x_2\to+\infty} \frac{\frac{H_b(Q(x_2-A))-H_b(Q(x_2+A))}{Q(x_2-A)-Q(x_2+A)}}{H_b'(Q(x_2+A))} &= \lim_{x_2\to+\infty} \frac{H_b(Q(x_2-A))\left(1-\frac{H_b(Q(x_2+A))}{H_b(Q(x_2-A))}\right)}{Q(x_2-A)\left(1-\frac{Q(x_2+A)}{Q(x_2-A)}\right)\log(\frac{1-Q(x_2+A)}{Q(x_2+A)})} \\
&= \lim_{x_2\to+\infty} -\frac{H_b(Q(x_2-A))}{Q(x_2-A)\log(Q(x_2+A))} \\
&= \lim_{x_2\to+\infty} \frac{e^{-\frac{(x_2-A)^2}{2}}\log(Q(x_2-A))}{e^{-\frac{(x_2-A)^2}{2}}\log(Q(x_2+A)) + \frac{Q(x_2-A)}{Q(x_2+A)}e^{-\frac{(x_2+A)^2}{2}}} \\
&= \lim_{x_2\to+\infty} \frac{\log(Q(x_2-A))}{\log(Q(x_2+A))+1} \\
&= \lim_{x_2\to+\infty} \frac{Q(x_2+A)e^{Ax_2}}{Q(x_2-A)e^{-Ax_2}} \\
&= 1
\end{aligned} \tag{A55}$$

From (A49), (A54) and (A55), (35) is proven. Note that the boundedness of X_1 is crucial in the proof. In other words, the fact that $Q(x_2 - A) \to 0$ as $x_2 \to +\infty$ is the very result of $A < +\infty$.

References

1. Walden, R.H. Analog-to-digital converter survey and analysis. *IEEE J. Sel. Areas Commun.* **1999**, *17*, 539–550. [CrossRef]
2. Murmann, B. ADC Performance Survey. *CoRR* 2014. Available online: http://web.stanford.edu/~murmann/adcsurvey.html (accessed on 20 May 2018).
3. Gunduz, D.; Stamatiou, K.; Michelusi, N.; Zorzi, M. Designing intelligent energy harvesting communication systems. *IEEE Comm. Mag.* **2014**, *52*, 210–216. [CrossRef]

4. Singh, J.; Dabeer, O.; Madhow, U. On the limits of communication with low-precision analog-to-digital Conversion at the Receiver. *IEEE Trans. Commun.* **2009**, *57*, 3629–3639. [CrossRef]
5. Krone, S.; Fettweis, G. Fading channels with 1-bit output quantization: Optimal modulation, ergodic capacity and outage probability. In Proceedings of the 2010 IEEE Information Theory Workshop, Dublin, Ireland, 30 August–3 September 2010; pp. 1–5.
6. Mezghani, A.; Nossek, J.A. Analysis of Rayleigh-fading channels with 1-bit quantized output. In Proceedings of the 2008 IEEE International Symposium on Information Theory, Toronto, ON, Canada, 6–11 July 2008.
7. Mezghani, A.; Nossek, J.A. On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; pp. 1286–1289.
8. Mo, J.; Heath, R. Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information. *IEEE Trans. Signal Process.* **2015**, *63*, 5498–5512. [CrossRef]
9. Witsenhausen, H.S. Some aspects of convexity useful in information theory. *IEEE Trans. Inf. Theory* **1980**, *26*, 265–271. [CrossRef]
10. El Gamal, A.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2012.
11. Rudin, W. *Principles of Mathematical Analysis*; McGraw-Hill: New York, NY, USA, 1976.
12. Abou-Faycal, I.C.; Trott, M.D.; Shamai, S. The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE Trans. Inf. Theory* **2001**, *47*, 1290–1300. [CrossRef]
13. Gallager, R.G. *Information Theory and Reliable Communication*; Wiley: New York, NY, USA, 1968.
14. Shiryaev, A.N. *Probability*, 2nd ed.; Springer: Berlin, Germany, 1996.
15. Chung, K.L. *A Course in Probability Theory*, 2nd ed.; Academic Press: New York, NY, USA, 1974.
16. Billingsley, P. *Convergence of Probability Measures*, 2nd ed.; Wiley: New York, NY, USA, 1968.
17. Sagitov, S. Lecture Notes: Weak Convergence of Probability Measures. Available online: http://www.math.chalmers.se/~serik/C-space.pdf (accessed on 20 May 2018).
18. Luenberger, D. *Optimization by Vector Space Methods*; Wiley: New York, NY, USA, 1969.
19. Smith, J.G. The information capacity of amplitude and variance constrained scalar Gaussian channels. *Inf. Control* **1971**, *18*, 203–219. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).