2022 IEEE Ninth International Conference on Communications and Electronics (ICCE) | 978-1-6654-9745-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCE55644.2022.9852067

Non-Asymptotic Bounds of Point-to-Point Communication With or Without Perfect Feedback Using α-Information Theory

Olivier Rioul
LTCI, Télécom Paris, Institut Polytechnique de Paris, France
0000-0002-8681-8916

Hang Nguyen
Télécom SudParis, Institut Polytechnique de Paris, France
0000-0002-9841-3772

***Abstract*—Some results of alpha-information theory are presented in order to derive simple non-asymptotic lower bounds on the probability of error for any binary block code used on symmetric channels with or without feedback. In particular, we obtain lower bounds on the signal-to-noise ratio for given code parameters and probability of error.**

***Index Terms*—Sibson's alpha information, Rényi's alpha divergence, converse theorems, perfect feedback, symmetric binary-input channel**

I. Introduction

Consider the classical point-to-point communication channel model with (or without) perfect feedback as depicted in Fig. 1. We assume that a (n, M) block code is used to transmit an M-ary information source W as a n-symbol codeword $\underline{X} = (X_1, \dots, X_n)$ through a memoryless channel. The channel output $\underline{Y} = (Y_1, \dots, Y_n)$ is decoded to retrieve the information source $\widehat{W}$ with probability of decoding error $\mathbb{P}_e = \mathbb{P}(\widehat{W} \neq W)$, that is, probability of successful decoding equal to $\mathbb{P}_s = 1 - \mathbb{P}_e = \mathbb{P}(\widehat{W} = W)$.

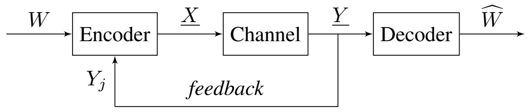

Fig. 1. The archetype of a point-to-point communication channel with perfect feedback such that $X_j = f(W, Y_1, \dots, Y_j)$ for each time instant j.

In a sense, this problem is well known and entirely solved by Shannon: The optimal coding rate for arbitrarily reliable communication is the channel capacity $C = \max_{p_X} I(X; Y)$ [13], which (since the channel is memoryless) is not increased by feedback [14]. In the sequel, we focus the determination of *converse theorems*, deriving lower bounds on $\mathbb{P}_e$ for any (n, M) code. Such bounds are established in a non-asymptotic setting (without requiring $n \to +\infty$) for practical communication models based on the recent development of *α-information theory*, with the hope to attract attention of the communication society community on such recent developments.

While most of the theoretical results used here were already established, we have put an effort to include self-contained and simplified proofs so as to convince the reader about the powerfulness of the approach. Following the seminal work of [10], several known results such as zero-error capacity in the presence of feedback [14], Arimoto's converse bound (strong converse theorem) [2] are easily recovered as special cases.

As "practical communication models" we consider binary (n, M) codes over symmetric channels arising from the additive white Gaussian noise (AWGN) channel with or without quantization at the output; in particular the binary symmetric channel (BSC), the binary erasure channel (BEC) or more generally the binary erasure and error channel, and the binary-input AWGN channel.

For each type of channel, not only do the converse theorems establish infeasible regions in terms of coding rate $R = \frac{\log M}{n}$ (in bits/bit[1]) and probability of error $\mathbb{P}_e$ for fixed dimension n, but also lower bounds on the required signal-to-noise ratio (SNR) depending of R and $\mathbb{P}_e$.

Notations: To simplify notations in the sequel, we follow [16] and consider random variables whose distributions are dominated by some σ-finite measure μ. This is always possible since we only consider a finite number of random variables at a time. Also, the quantities (α-divergence, α-information, α-capacity) defined below can be shown to be independent of the choice of the dominating measure μ.

From this assumption it follows that every considered random variable X admits a probability density p_X w.r.t. μ. In the sequel we shall identity X's distribution with p_X. When μ is the Lebesgue measure, p_X is the genuine probability density function (p.d.f.); when μ is a counting measure, p_X is the discrete probability mass function (p.m.f.).

The integral w.r.t. μ is accordingly denoted by the special symbol $\sum\!\!\!\!\!\!\int$ (classical integral for p.d.f.'s, discrete sum for p.m.f.'s). Thus, e.g., one always has $\sum\!\!\!\!\!\!\int p_X = \sum\!\!\!\!\!\!\int_x p_X(x) = 1$.

Outline: The remainder of the paper is organized as follows. Section II presents some basic definitions and properties of α-information theoretic quantities. Sections III and IV derive the main ingredients, data processing and Fano inequalities, respectively. The α-capacities are studied in Section V and the main result is presented in Section VI along with some theoretical and numerical applications. Section VII concludes.

[1] All logarithms log are taken to base 2, hence informational units are bits.

978-1-6654-9745-9/22/$31.00 ©2022 IEEE

II. α-Information

While the classical Shannon's information theory can be seen as based on the properties of the Kullback-Leibler divergence, the generalized α-information theory is based on the properties of the Rényi divergence [11] of order α.

Definition 1 (α-Divergence): *Given $\alpha \in (0,1) \cup (1,+\infty)$ and two probability distributions p, q,*

$$D_\alpha(p\|q) \triangleq \frac{1}{\alpha-1} \log \sum\!\!\!\!\!\!\int p^\alpha q^{1-\alpha} = \frac{\alpha}{\alpha-1} \log(p\|q)_\alpha \tag{1}$$

*where the α-*product

$$(p\|q)_\alpha \triangleq \bigl(\sum\!\!\!\!\!\!\int p^\alpha q^{1-\alpha}\bigr)^{1/\alpha} \tag{2}$$

is a.k.a. Hellinger integral or Bhattacharyya coefficient.

The α-divergence was extensively studied in [16]. In particular one has the following properties:

[D1] By Hölder's inequality for $\alpha < 1$ (or reverse Hölder inequality for $\alpha > 1$)

$$(p\|q)_\alpha \underset{\alpha>1}{\overset{\alpha<1}{\lesseqgtr}} \bigl((\sum\!\!\!\!\!\!\int p)^\alpha (\sum\!\!\!\!\!\!\int q)^{1-\alpha}\bigr)^{1/\alpha} = 1, \tag{3}$$

hence $D_\alpha(p\|q) \geq 0$ with equality $D_\alpha(p\|q) = 0$ if and only if $p \equiv q$ are the same distribution.

[D2] The α-divergence between two binary distributions $(1-p, p)$ and $(1-q, q)$ is

$$d_\alpha(p\|q) \triangleq \tfrac{1}{\alpha-1} \log\bigl((1-p)^\alpha (1-q)^{1-\alpha} + p^\alpha q^{1-\alpha}\bigr). \tag{4}$$

It is easily checked directly that $q \mapsto d_\alpha(p\|q)$ decreases for $q \leq p$ and increases for $q \geq p$. (In fact, it is convex and vanishes at $q = p$.) In particular,

$$p \geq q \geq r \implies d_\alpha(p\|r) \geq d_\alpha(p\|q). \tag{5}$$

[D3] $D_\alpha(p,q)$ is nondecreasing in α [16, Thm. 3].

[D4] Limits [16, § II.C]:

- for $\alpha \nearrow 1$ one recovers the classical Kullback-Leibler divergence (a.k.a. relative entropy) $D_1(p\|q) = D(p\|q) = \displaystyle\sum\!\!\!\!\!\!\int p \log \frac{p}{q}$;
- for $\alpha \searrow 0$ one obtains $D_0(p\|q) = -\log \displaystyle\sum\!\!\!\!\!\!\int_{p>0} q$;
- for $\alpha \nearrow +\infty$ one obtains $D_\infty(p\|q) = \log \sup_q \frac{p}{q}$.

[D5] $D_\alpha(p,q)$ is lower semi-continuous in (p,q) w.r.t. the convergence in distribution [16, Thm. 19].

Now consider a *channel* $X \to Y$ with input X and output Y. The channel is defined by the transition probabilities $p_{Y|X}$ and we note $p_X \to \boxed{p_{Y|X}} \to p_Y$.

Knowing a particular observed output $Y = y$ modifies the distribution of X. The α-divergence between $p_{X|y}$ (X's distribution knowing $Y = y$) and p_X (X's distribution *not* knowing $Y = y$) is

$$D_\alpha(p_{X|y}\|p_X) = \frac{\alpha}{\alpha-1} \log(p_{X|y}\|p_X)_\alpha.$$

Sibson's α-information [15] is then obtained by averaging over Y inside the logarithm:

Definition 2 (α-Information and α-Response):

$$I_\alpha(X;Y) \triangleq \frac{\alpha}{\alpha-1} \log \mathbb{E}_Y (p_{X|Y}\|p_X)_\alpha. \tag{6}$$

For any p_X, define its α-response of the channel $X \to Y$ by

$$q_{Y,p_X} \triangleq \frac{(p_{X|Y}\|p_X)_\alpha \, p_Y}{\mathbb{E}_Y (p_{X|Y}\|p_X)_\alpha} = \frac{\bigl(\sum\!\!\!\!\!\!\int_x p_X p^\alpha_{Y|X}\bigr)^{1/\alpha}}{\sum\!\!\!\!\!\!\int_y \bigl(\sum\!\!\!\!\!\!\int_x p_X p^\alpha_{Y|X}\bigr)^{1/\alpha}}. \tag{7}$$

The normalization in the denominator shows that the α-response $q_{Y,p_X}(y)$ is a probability distribution in y.

Lemma 1 (Chain Rule for α-Product):

$$(p_{XY}\|q_{XY})_\alpha = \bigl((p_{X|Y}\|q_{X|Y})_\alpha \, p_Y \| q_Y\bigr)_\alpha. \tag{8}$$

Proof: Obvious by direct calculation. ■

Theorem 1 (Sibson's Identity [15]): *For any q_Y,*

$$D_\alpha(p_{XY}\|p_X q_Y) = D_\alpha(q_{Y,p_X}\|q_Y) + I_\alpha(X;Y). \tag{9}$$

In particular

$$I_\alpha(X;Y) = \min_{q_Y} D_\alpha(p_{XY}\|p_X q_Y) = D_\alpha(p_{XY}\|p_X q_{Y,p_X}) \tag{10}$$

where the α-response q_{Y,p_X} is the unique distribution achieving the minimum.

Proof: From the chain rule (8), $(p_{XY}\|p_X q_Y)_\alpha = \bigl((p_{X|Y}\|p_X)_\alpha \, p_Y \| q_Y\bigr)_\alpha = (q_{Y,p_X}\|q_Y)_\alpha \cdot \mathbb{E}_Y (p_{X|Y}\|p_X)_\alpha$. Taking the logarithm gives (9); (10) follows from **[D1]**. ■

The α-information was extensively studied in [6], [10], [15], [17]. In particular on has the following properties:

[I1] By (10), $I_\alpha(X;Y) \geq 0$ with equality $I_\alpha(X;Y) = 0$ if and only if X and Y are independent.

[I2] An easy calculation shows that

$$I_\alpha(X;Y) = \frac{\alpha}{\alpha-1} \log \sum\!\!\!\!\!\!\int_y \bigl(\sum\!\!\!\!\!\!\int_x p_X p^\alpha_{Y|X}\bigr)^{1/\alpha}. \tag{11}$$

In particular, letting $\phi(x) = \frac{1}{\alpha-1} \exp\bigl(\frac{\alpha-1}{\alpha} x\bigr)$, ϕ is an increasing function for all α and $\phi(I_\alpha(X;Y)) = \frac{1}{\alpha-1} \sum\!\!\!\!\!\!\int_y \bigl(\sum\!\!\!\!\!\!\int_x p_X p^\alpha_{Y|X}\bigr)^{1/\alpha}$ is concave in p_X for fixed channel $p_{Y|X}$ [6].

[I3] By (9) and **[D3]**, it is easily seen that $I_\alpha(X;Y)$ is non decreasing in α.

[I4] Limits [6]:

- for $\alpha \to 1$ one recovers the classical Shannon mutual information $I_1(X;Y) = I(X;Y)$;
- for $\alpha \searrow 0$ we get $I_0(X;Y) = -\log \sup_y \displaystyle\sum\!\!\!\!\!\!\int_{p_{y|X}>0} p_X$;
- for $\alpha \nearrow +\infty$ we get $I_\infty(X;Y) = \log \displaystyle\sum\!\!\!\!\!\!\int_y \sup_{p_X(x)>0} p_{Y|x}$.

In contrast to the case $\alpha = 1$, α-information is no longer "mutual" in the sense that $I_\alpha(X;Y) \neq I_\alpha(Y;X)$ for $\alpha \neq 1$.

III. Data Processing Inequalities (DPIs)

Theorem 2 (DPI for α-Divergence [16]): *When a given channel $p_{Y|X}$ responds to two different inputs: $p_X \to \boxed{p_{Y|X}} \to p_Y$ and $q_X \to \boxed{p_{Y|X}} \to q_Y$, one has $D_\alpha(p_Y\|q_Y) \leq D_\alpha(p_X\|q_X)$.*

In words, any random transformation reduces α-divergence. We give a short proof.

Proof: Since $p_{Y|X} = q_{Y|X}$, (8) with X and Y permuted gives $(p_{XY}\|q_{XY})_\alpha = (p_X\|q_X)_\alpha$. Now by (8) and (3), $(p_{XY}\|q_{XY})_\alpha = \bigl((p_{X|Y}\|q_{X|Y})_\alpha \, p_Y \| q_Y\bigr)_\alpha \underset{\alpha>1}{\overset{\alpha<1}{\lesseqgtr}} (p_Y\|q_Y)_\alpha$. Taking logarithms gives $D_\alpha(p_X\|q_X) \geq D_\alpha(p_Y\|q_Y)$. ■

Theorem 3 (DPI for α-Information [10]): *If $W-Y-\widehat{W}$ forms a Markov chain (as in Fig. 1), then $I_\alpha(W;Y) \geq I_\alpha(W;\widehat{W})$.*

In words, any processing (here the decoder) can only decrease α-information. We give the proof for completeness.

Proof: Consider the channel $p_{W\widehat{W}|WY} = p_{W|W}p_{\widehat{W}|WY}$ where $p_{\widehat{W}|WY} = p_{\widehat{W}|Y}$ by the Markov condition. Thus $p_{WY} \to \boxed{p_{W\widehat{W}|WY}} \to p_{W\widehat{W}}$, and if $q_Y \to \boxed{p_{\widehat{W}|Y}} \to q_{\widehat{W}}$, one has $p_W q_Y \to \boxed{p_{W\widehat{W}|WY} = p_{W|W}p_{\widehat{W}|Y}} \to p_W q_{\widehat{W}}$. Thus by Theorem 2, $D_\alpha(p_{WY}\|p_W q_Y) \geq D_\alpha(p_{W\widehat{W}}\|p_W q_{\widehat{W}}) \geq I_\alpha(W;\widehat{W})$ where the last inequality follows from (10). Again by (10), applying the resulting inequality to $q_Y = q_{Y,p_W}$ proves the theorem. ∎

It can be similarly shown [10] that α-information also satisfies the preprocessing inequality $I_\alpha(Y;\widehat{W}) \geq I_\alpha(W;\widehat{W})$.

IV. Fano-Type Inequalities

As is well known, the minimum probability of error $\mathbb{P}_e$ or the maximum probability of success $\mathbb{P}_s = \mathbb{P}(\widehat{W} = W)$ upon observation of the channel output $Y = y$ is achieved by the MAP rule:

$$\widehat{W}(y) = \arg\max_w p_{W|Y}(w|y) \tag{12}$$

In this case of optimal decoding we write

$$\mathbb{P}_s(W|Y) \triangleq \max_{W-Y-\widehat{W}} \mathbb{P}(\widehat{W} = W) = \mathbb{E}_Y\bigl(\max_w p_{W|Y}(w|Y)\bigr) \tag{13}$$

In particular, a blind estimation of W (without access to the channel output Y) gives

$$\mathbb{P}_s(W) \triangleq \max_w p_W(w) \tag{14}$$

It is obvious that $\mathbb{P}_s(W|Y) \geq \mathbb{P}_s(W)$ (observation increases success).

Theorem 4 (Fano's Inequality for α-Information [12]):

$$I_\alpha(W;Y) \geq d_\alpha\bigl(\mathbb{P}_s(W|Y)\big\|\mathbb{P}_s(W)\bigr) \tag{15}$$

where $d_\alpha(p\|q)$ is the binary α-divergence (4).

Proof: By Thm 3 & Eq. (10), $I_\alpha(W;Y) \geq I_\alpha(W;\widehat{W}) = D_\alpha(p_{W\widehat{W}}\|p_W q_{\widehat{W},p_W})$. Now apply Theorem 2 to the deterministic channel with input $(W,\widehat{W})$ and binary output $= 1$ if $\widehat{W} = W$, and $= 0$ otherwise. This gives $D_\alpha(p_{W\widehat{W}}\|p_W q_{\widehat{W},p_W}) \geq d_\alpha(\mathbb{P}(\widehat{W} = W)\|\mathbb{P}')$ where $\mathbb{P}' = \sum_w p_W(w) q_{\widehat{W},p_W}(w) \leq \max_w p_W(w) = \mathbb{P}_s(W)$. Assuming the MAP rule, $\mathbb{P}(\widehat{W} = W) = \mathbb{P}_s(W|Y) \geq \mathbb{P}_s(W) \geq \mathbb{P}'$, hence by (5), $d_\alpha(\mathbb{P}_s(W|Y)\|\mathbb{P}') \geq d_\alpha(\mathbb{P}_s(W|Y)\|\mathbb{P}_s(W))$. Combining yields the announced Fano inequality. ∎

Since our aim is to derive upper bounds on $\mathbb{P}_s$ (lower bounds on $\mathbb{P}_e$) we can always assume that the MAP rule holds. Notice, however, that any "reasonable" suboptimal decoding procedure should always give a probability of success $\mathbb{P}_s \geq \max_w p_W(w)$. In this case Fano's inequality still holds with $\mathbb{P}_s$ in place of the maximum $\mathbb{P}_s(W|Y)$.

For *equiprobable* M-ary source W (a usual assumption in the communication model), Fano's inequality (15) reduces to $I_\alpha(W;Y) \geq d_\alpha\bigl(\mathbb{P}_s\big\|\frac{1}{M}\bigr)$. When $\alpha \to 1$ one recovers the classical Fano inequality [5].

V. α-Capacity

A. Definition and Characterization

By analogy with Shannon's formula $C = \max_{p_X} I(X;Y)$, we define the α-capacity as follows[2].

Definition 3:

$$C_\alpha \triangleq \max_{p_X} I_\alpha(X;Y). \tag{16}$$

Thus C_α depends only on the considered channel $p_{Y|X}$. The α-capacity was extensively studied in [9] and [3].

Theorem 5 (Characterization of α-Capacity [3], [4]): *For discrete X,*

$$C_\alpha = \min_{q_Y}\max_x D_\alpha(p_{Y|x}\|q_Y) = \max_x D_\alpha(p_{Y|x}\|q_{Y,p_X^*}) \tag{17}$$

where q_{Y,p_X^*} is the α-response of the distribution p_X^* achieving the maximum in (16).

We give a simple proof of this key result.

Proof: From the definition (16) and (10) we observe that $C_\alpha = \max_{p_X} D_\alpha(p_X p_{Y|X}\|p_X q_{Y,p_X})$ where q_{Y,p_X} is the unique distribution achieving $\min_{q_Y} D_\alpha(p_X p_{Y|X}\|p_X q_Y)$.

Now let $\psi(x) = \operatorname{sgn}(\alpha-1)\exp\bigl((\alpha-1)x\bigr)$; ψ is an increasing function for all α and $\psi(D_\alpha(p_X p_{Y|X}\|p_X q_Y)) = \operatorname{sgn}(\alpha-1)\sumint_x p_X(x)\sumint_y p_{Y|X}^\alpha q_Y^{1-\alpha}$ is linear in p_X for fixed q_Y. It is also lower semi-continuous in q_Y for fixed p_X by **[D5]**. Therefore, $f(p_X,q_Y) = \psi(D_\alpha(p_X p_{Y|X}\|p_X q_Y))$ satisfies the conditions of Lemma 3 in the Appendix, andf we have $C_\alpha = \min_{q_Y}\max_{p_X} D_\alpha(p_X p_{Y|X}\|p_X q_Y) = \max_{p_X} D_\alpha(p_X p_{Y|X}\|p_X q_{Y,p_X^*})$.

Finally, since $\psi(D_\alpha(p_X p_{Y|X}\|p_X q_Y))$ is linear in p_X, its maximum over p_X is necessarily achieved when p_X is a Dirac distribution at some $X = x$. This proves (17). ∎

The following Lemma is also useful.

Lemma 2: *One has*

$$D_\alpha(p_{XY}\|q_X q_Y) \leq D_\alpha(p_X\|q_X) + \max_x D_\alpha(p_{Y|x}\|q_Y). \tag{18}$$

Proof: Consider the difference $D_\alpha(p_{XY}\|q_X q_Y) - D_\alpha(p_X\|q_X)$. By (1) and (8), an easy calculation shows that $\frac{(p_{XY}\|q_X p_Y)_\alpha}{(p_X\|q_X)_\alpha} = \frac{((p_{Y|X}\|q_Y)_\alpha p_X\|q_X)_\alpha}{(p_X\|q_X)_\alpha}$ has the form $(r_X p_{Y|X}\|r_X q_Y)_\alpha$ where $r_X = \frac{p_X^\alpha q_X^{1-\alpha}}{(p_X\|q_X)_\alpha^\alpha}$ is a distribution. Thus $D_\alpha(p_{XY}\|q_X q_Y) - D_\alpha(p_X\|q_X) = D_\alpha(r_X p_{Y|X}\|r_X q_Y) \leq \max_x D_\alpha(p_{Y|x}\|q_Y)$ by the same argument as in the last part of the proof of Theorem 5. ∎

B. Memoryless Channel With (or Without) Perfect Feedback

Consider the memoryless channel illustrated in Fig. 1. Then it is easily seen that

$$p_{\underline{Y}|W} = \prod_{j=1}^n p_{Y_j|W,Y_1,\ldots,Y_{j-1}} = \prod_{j=1}^n p_{Y_j|X_j} \tag{19}$$

where $X_j = f(W,Y_1,\ldots,Y_{j-1})$ for $j = 1,\ldots,n$.

Theorem 6: [10] *With the above assumptions, one has*

$$I_\alpha(W,\underline{Y}) \leq n\cdot C_\alpha \tag{20}$$

where C_α is the α-capacity (16) *in bits per symbol (for the single-letter channel).*

In the absence of the feedback link in a memoryless channel, one has in fact [9], [10] $I_\alpha(W,\underline{Y}) \leq \max_{p_{\underline{X}}} I_\alpha(\underline{X};\underline{Y}) = n\cdot C_\alpha$. A simple proof of Theorem 6 is as follows.

[2] Notice, however, that this is not an operational definition.

Proof: Let $q_{Y_j} = q_{Y_j, p^*_{X_j}}$ be the distribution achieving (17) for the jth symbol, $j = 1, \ldots, n$. Then from (10), $I_\alpha(W, \underline{Y}) \leq D_\alpha(p_{W\underline{Y}} \| p_W q_{Y_1} \cdots q_{Y_n})$. Now by n applications of (18) using (19), $D_\alpha(p_{WY_1 \cdots Y_n} \| p_W q_{Y_1} \cdots q_{Y_n}) \leq D_\alpha(p_W \| p_W) + \sum_{j=1}^n \max_{x_j} D_\alpha(p_{Y_j|x_j} \| q_{Y_j})$. The first term $= 0$ and all terms in the sum are $= C_\alpha$ by (17). ■

C. Binary Input Symmetric Channel Model

We now derive the general expression of C_α for binary-input symmetric channels $X \to Y$. Without loss of generality we can assume that the input X has values ± 1 so that by symmetry, $p_{Y|X=1}(y) = p_{Y|X=-1}(-y)$, that is, $p_{Y|1} = p_{-Y|-1}$.

Theorem 7 (α-Capacity of a binary-input symmetric channel):

$$C_\alpha = 1 - \frac{\alpha}{1-\alpha} \log \sum\!\!\!\!\!\!\int \tfrac{1}{2} \bigl(p^\alpha_{Y|1} + p^\alpha_{-Y|1}\bigr)^{1/\alpha}. \qquad (21)$$

Proof: By concavity **[I2]** of $I_\alpha(X;Y)$ in the binary distribution p_X, the α-capacity is achieved for equiprobable inputs $p^*_X = (\frac{1}{2}, \frac{1}{2})$. Then the corresponding α-response (7) is $q_{Y,p^*_X} \propto (\sum_x p^*_X p^\alpha_{Y|X})^{1/\alpha} = \bigl(\frac{1}{2}(p^\alpha_{Y|1} + p^\alpha_{Y|-1})\bigr)^{1/\alpha}$. Thus, by symmetry, one has $D_\alpha(p_{Y|1} \| q_{Y,p^*_X}) = D_\alpha(p_{Y|-1} \| q_{Y,p^*_X}) = \max_x D_\alpha(p_{Y|x} \| q_{Y,p^*_X})$. Therefore, by (17), $C_\alpha = D_\alpha(p_{Y|1} \| q_{Y,p^*_X}) = \frac{1}{\alpha-1} \log \dfrac{\sum\!\!\!\!\!\!\int p^\alpha_{Y|1} \bigl(\frac{1}{2}(p^\alpha_{Y|1} + p^\alpha_{Y|-1})\bigr)^{\frac{1-\alpha}{\alpha}}}{\bigl(\sum\!\!\!\!\!\!\int \bigl(\frac{1}{2}(p^\alpha_{Y|1} + p^\alpha_{Y|-1})\bigr)^{\frac{1}{\alpha}}\bigr)^{1-\alpha}}$ where the two sums/integrals are equal by channel symmetry. Simplifying gives (21). ■

One has the following properties:

[C1] For a binary-input channel,

$$C_\alpha \leq 1 \text{ bit}. \qquad (22)$$

For a symmetric channel, this can be checked directly using Minkowski/reverse Minkowski inequality to show that the second term in (21) is nonnegative. More generally, this can easily seen by noting that from Theorem 3, since $X - X - Y$ forms a Markov chain, $I_\alpha(X;Y) \leq I_\alpha(X;X) \leq C_\alpha(0)$, the α-capacity of the binary noiseless channel, which by (21) equals $1 - \frac{\alpha}{1-\alpha} \log 1 = 1$ bit.

[C2] By **[I3]**, C_α is nondecreasing in α.

[C3] Limits [9, Lemma 15]: $\alpha \mapsto C_\alpha$ is continuous in α for finitely many inputs (in particular for binary inputs). Thus, one recovers the following particular cases:

- For $\alpha \searrow 0$ one obtains the *feedback zero-error capacity* C_0 (see § VI-A below). In the binary-input symmetric case, the limit of (21) is easily computed as $C_0 = 1 - \log \sup_y (p^0_{y|1} + p^0_{y|-1})$, i.e., $C_0 = 1 - \log 2 = 0$ bit for a noisy channel and $C_0 = 1 - \log 1 = 1$ for a noiseless channel.
- For $\alpha = \frac{1}{2}$, $C_{1/2} = R_0$ is known as the *cut-off rate* as defined in [18, § 5.4, 6.2] for binary inputs (see also [4]). Following Massey [8], this had been long adopted as an important criterion for coding systems design. The value of (21) at $\alpha = \frac{1}{2}$ is indeed
$$C_{1/2} = 1 - \log\bigl(1 + \sum\!\!\!\!\!\!\int \sqrt{p_{Y|1} \, p_{Y|-1}}\bigr). \qquad (23)$$
- For $\alpha \to 1$, $C_1 = C$ is the genuine Shannon capacity (which is equal to the feedback capacity for a memoryless channel [14]). It can be checked directly by l'Hospital rule that (21) when $\alpha \to 1$ gives the expression of $I(X;Y)$ for equiprobable X.
- For $\alpha \to +\infty$, $C_\infty = 1 + \log \sum\!\!\!\!\!\!\int \dfrac{1}{2} \max(p_{Y|1}, p_{Y|-1})$. A usual assumption (for small noise) is $p_{Y|1}(y) \geq p_{Y|-1}(y)$ when $y \geq 0$ and the opposite inequality for $y \leq 0$. Using symmetry, the expression then easily simplifies to
$$C_\infty = 1 - \log \frac{1}{1 - p_e} = \log \frac{p_s}{1/2} \qquad (24)$$
where $p_e = 1 - p_s = \sum\!\!\!\!\!\!\int_{y \leq 0} p_{Y|1}(y)$ is the bit-error probability obtained by the usual threshold detector at the channel output (possibly with 50% chance in case of a tie $y = 0$). This is equivalent to using the maximum likelihood rule (here equivalent to the MAP rule for equiprobable input), and with the notations of § IV one obtains $C_\infty = \log \frac{\mathbb{P}_s(X|Y)}{\mathbb{P}_s(X)}$.

D. α-Capacities of Some Known Channels

We consider binary-input symmetric channels arising from the additive white Gaussian noise (AWGN) channel with or without quantization at the output. With quantization the considered channel models are depicted in Fig. 2.

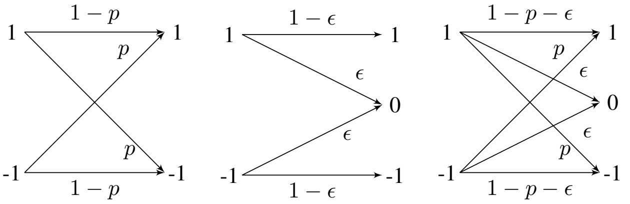

Fig. 2. Flow graphs of symmetric channels. Left to right: BSC; BEC; BSEC.

As is well known, for *binary* output, we obtain the binary symmetric channel (BSC) of parameter $p = Q(\sqrt{\frac{2E_b}{N_0}})$ where $Q(\cdot)$ is the Q-function and E_b/N_0 is the SNR per bit. To simplify the obtained expressions, we assume $E_b = 1$ for input $X \in \{\pm 1\}$ and write $\sigma^2 = N_0/2$ (noise sample variance) so that $p = Q(\frac{1}{\sigma})$. One may always assume $p \leq 1/2$.

For a *ternary* output, we obtain the binary symmetric error & erasure channel (BSEC) with raw error probability p and erasure probability ϵ. Typical expressions (for a uniform output quantization) are $p = Q(\frac{3}{2\sigma})$ and $p + \epsilon = Q(\frac{1}{2\sigma})$. For large SNR (small noise) p is negligible w.r.t. ϵ in which case the channel model reduces to a binary erasure channel (BEC) of parametrer $\epsilon = Q(\frac{1}{2\sigma})$.

Finally, without output quantization we have the binary-input AWGN model with $X \in \{\pm 1\}$ and $X = Y + Z$ where $Z \sim \mathcal{N}(0, \sigma^2)$.

Applying (21) readily gives the expressions in Table I. We recover in particular the results of [3] for BSC and BEC. Some α-capacities are given in Fig. 3. Notice that the α-capacity of the AWGN is always larger then that of the BSC or BSEC by the DPI (Theorem 3) applied to the output quantizer.

TABLE I

SOME α-CAPACITIES OF BINARY-INPUT MEMORYLESS CHANNELS. $h(p) = -p\log p - (1-p)\log(1-p)$ IS THE BINARY ENTROPY FUNCTION.

	C_α	cut-off $C_{1/2}$	usual capacity $C = C_1$	C_∞
BSC	$1 - \frac{1}{1-\alpha}\log(p^\alpha + (1-p)^\alpha)$	$1 - \log(1 + 2\sqrt{p(1-p)})$	$1 - h(p)$	$1 - \log\frac{1}{1-p}$
BEC	$1 - \frac{\alpha}{1-\alpha}\log(1 - \epsilon + 2^{\frac{1-\alpha}{\alpha}}\epsilon)$	$1 - \log(1+\epsilon)$	$1 - \epsilon$	$1 - \log\frac{1}{1-\epsilon/2}$
BSEC	$1 - \frac{\alpha}{1-\alpha}\log\big((p^\alpha + (1-p-\epsilon)^\alpha)^{\frac{1}{\alpha}} + 2^{\frac{1-\alpha}{\alpha}}\epsilon\big)$	$1 - \log(1 + \epsilon + 2\sqrt{p(1-p-\epsilon)})$	$(1-\epsilon)(1 - h(\frac{p}{1-\epsilon}))$	$1 - \log\frac{1}{1-p-\epsilon/2}$
AWGN	$1 - \frac{\alpha}{1-\alpha}\log\int_{-\infty}^{\infty}\frac{e^{-(y^2+1)/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \times\frac{1}{2}(e^{y\alpha/\sigma^2} + e^{-y\alpha/\sigma^2})^{1/\alpha}\,\mathrm{d}y$	$1 - \log(1 + e^{-1/2\sigma^2})$	$1 - \int_{-\infty}^{\infty}\frac{e^{-(y-1)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \times\log(1 + e^{-2y/\sigma^2})\,\mathrm{d}y$	$1 - \log\frac{1}{1-Q(1/\sigma)}$

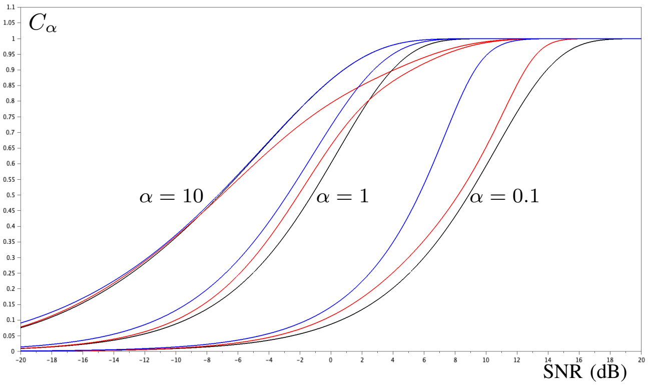

Fig. 3. α-capacities of binary-input BSC (black), BSEC (red) and AWGN channel (blue) as a function of SNR$= 1/(2\sigma^2)$ per transmitted bit.

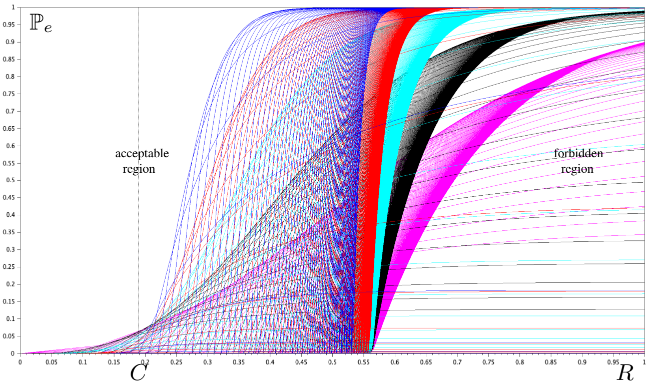

Fig. 4. Lower bounds on error probability $\mathbb{P}_e$ vs. coding rate R on a BSC(.25) for $n = 8$ (magenta), 16 (black), 32 (cyan), 64 (red), 128 (blue) for $\alpha \in [0, +\infty]$ with stepsize $= 0.1$. The dashed vertical line shows the capacity C.

VI. α-CONVERSE THEOREM

The following Theorem provides infinitely many non-asymptotic upper bounds on $\mathbb{P}_s$ (lower bounds on $\mathbb{P}_e$).

Theorem 8 (α-Converse Theorem): *For any $\alpha \in [0, +\infty]$ and any block code (n, M) with rate $R = \frac{\log M}{n}$ and decoding error probability $\mathbb{P}_e = 1 - \mathbb{P}_s$ on a memoryless channel (with or without perfect feedback) of α-capacity C_α,*

$$d_\alpha(\mathbb{P}_s \| \mathbb{P}_s') \le n \cdot C_\alpha \tag{25}$$

where $\mathbb{P}_s' = \max_w p_W(w) \le \mathbb{P}_s$; in particular, $\mathbb{P}_s' = \frac{1}{M}$ for equiprobable messages W.

Proof: Combine Theorems 4 and 6. (The values $\alpha = 0$, 1, $+\infty$ are obtained by taking limits.) ■

For varying $\alpha \in [0, +\infty]$, (25) provides non-asymptotic lower bounds on $\mathbb{P}_e$ (upper bounds on $\mathbb{P}_s$) for any particular choice of block code parameters (n, M)—or for any choice of code length n with varying coding rate $R = \frac{\log M}{n}$. An illustration is given in Fig. 4 for increasing code lengths.

A. Application to the Zero-Error Problem

If one requires strictly zero error [14], that is, $\mathbb{P}_e = 0$ and $\mathbb{P}_s = 1$, then (25) applies with equiprobable messages, where $d_\alpha(1 \| \frac{1}{M}) = \log M$. Thus (25) takes the form of a coding rate bound $R = \frac{\log M}{n} \le C_\alpha$. By **[C2]** $\inf C_\alpha = C_0$, so this all boils down to the inequality

$$R \le C_0 = \max_{p_X} I_0(X;Y) = \max_{p_X}\inf_{y} \log\frac{1}{\sum\limits_{p_{y|X}>0} p_X} \tag{26}$$

(see **[I4]**). As noticed in [17], this is exactly Shannon's expression of the zero-error capacity with feedback in the case where this capacity is > 0 (when not all inputs pairs can cause the same output [14]).

B. Application: Strong Converse

For $\alpha > 1$, (25) readily implies the strong converse theorem (for equiprobable messages):

Theorem 9 (Strong Converse): *If $R > C$, then $\mathbb{P}_e$ tends exponentially to 1 as $n \to +\infty$.*

As shown in [10], Arimoto's converse bound [2] can be recovered from this result.

Proof: For $\alpha > 1$, $\frac{1}{\alpha-1}\log(\mathbb{P}_s^\alpha \frac{1}{M^{1-\alpha}}) < d_\alpha(\mathbb{P}_s \| \frac{1}{M}) \le nC_\alpha$. Simplifying gives $\mathbb{P}_s < 2^{-n(R-C_\alpha)\frac{\alpha-1}{\alpha}}$. If $R > C$, since $C_\alpha \searrow C$ as $\alpha \searrow 1$, one has $R > C_\alpha + \epsilon$ for some $\alpha > 1$ and $\epsilon > 0$, and $\mathbb{P}_s < 2^{-n\epsilon\frac{\alpha-1}{\alpha}} \to 0$ exponentially. ■

Fig. 4 illustrates Theorem 9 by showing lower bounds on $\mathbb{P}_e$ for increasing lengths. The "hard limit" at Shannon's capacity C is only attainted for immeasurably large n.

C. Application: Lower Bound on the SNR

In our channel models, letting E_b/N_0 be the SNR per (information) bit, C_α is expressed as functions of $\frac{1}{\sigma^2} = 2R\cdot$SNR per coded bit sent on the channel. Since C_α is increasing in SNR (as illustrated in Fig. 3), (25) gives a lower bound on the feasible SNR for a given performance level $(\mathbb{P}_e, R)$ over a given channel.

In particular for $n \to +\infty$ and $R \to 0$ we recover the well-known Shannon limits -1.59 dB and 0.37 dB for binary-input AWGN and BSC, respectively. What is more interesting, however, is the non-asymptotic regions for a given choice of code parameters as illustrated in Fig. 5, 6 and 7.

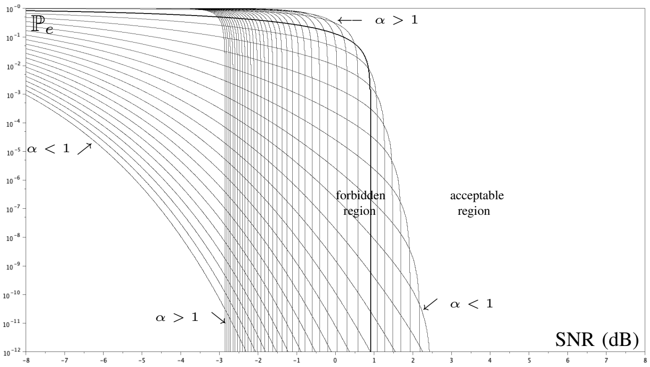

Fig. 5. Lower bounds on error probability $\mathbb{P}_e$ vs. SNR for a $[128, 64]$ code ($n = 128$, $R = 1/2$) on a BSEC. The thick curve is for $\alpha = 1$.

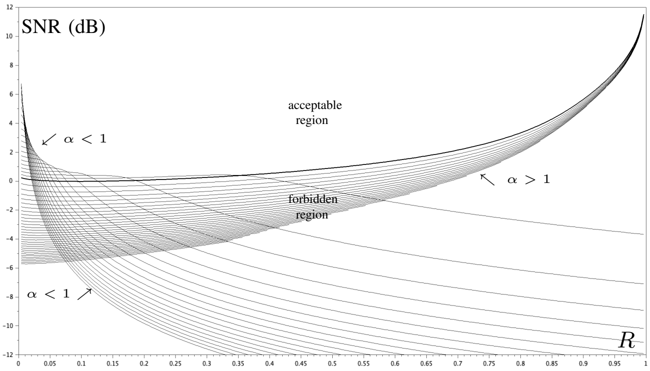

Fig. 6. Lower bounds on SNR vs. coding rate for $n = 1024$ on a BSEC. The thick curve is for $\alpha = 1$.

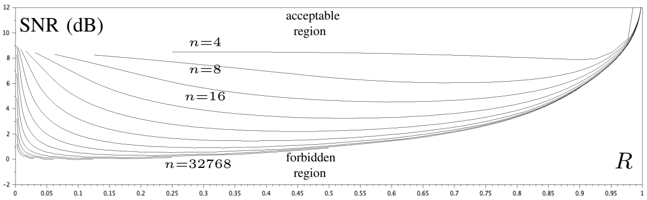

Fig. 7. Lower bounds (maximized over α) on SNR vs. coding rate for $n = 4, 8, 16, \ldots, 32768$ on a BSEC.

VII. Conclusion and Perspectives

We have presented some results of α-information theory in order to derive simple non-asymptotic lower bounds on the probability of error for any binary block code used on symmetric memoryless channels with or without feedback. Such bounds can be rewritten as lower bounds on the SNR for any given code parameters.

Since $I_\alpha(X;Y) \neq I_\alpha(Y,X)$, one can also define a "reverse" α-capacity $C'_\alpha = \max_{p_X} I_\alpha(Y;X)$. Since [1] $C_\alpha \leq C'_\alpha$, the resulting bounds (at least without feedback) cannot be tighter than the bounds in this paper.

As a perspective, the obtained converse bounds can be compared to other known finite-length bounds, applied to more general types of channels and perhaps other types of problems—in fact, data processing and Fano's inequalities were recently applied to side-channel analysis in [7].

Appendix: A Technical Lemma

Lemma 3: *Let $\mathcal{P}$ be a probability space and let $f(p,q)$ be defined for $p, q \in \mathcal{P}$ such that*

- *$f(p,q)$ is linear in p for fixed q;*
- *$f(p,q)$ is lower semi-continous in q for fixed p;*
- *$\forall\ p \in \mathcal{P}$, there exists a unique $q = q_p$ achieving $\min_q f(p,q) = f(p,q_p)$, where q_p in continuous in p.*

Further assume p^* achieves $\max_p f(p,q_p) = f(p^*,q_{p^*}) = C$. Then $C = \max_p f(p,q_{p^*}) = \min_q \max_p f(p,q)$.

Proof: Let $p \in \mathcal{P}$ and $\varepsilon > 0$ and consider the perturbation $p^*_\varepsilon = (1-\varepsilon)p^* + \varepsilon p$ so that $p = \varepsilon^{-1}p^*_\varepsilon + (1-\varepsilon^{-1})p^*$. Then by lower semi-continuity for small enough $\varepsilon > 0$, $f(p,q_{p^*}) \leq f(p,q_{p^*_\varepsilon})$; by linearity in p, $f(p,q_{p^*_\varepsilon}) = \varepsilon^{-1}f(p^*_\varepsilon,q_{p^*_\varepsilon}) + (1-\varepsilon^{-1})f(p^*,q_{p^*_\varepsilon})$, where $f(p^*,q_{p^*_\varepsilon}) \geq f(p^*,q_{p^*}) = C$ by definition of q_{p^*} and $f(p^*_\varepsilon,q_{p^*_\varepsilon}) \leq f(p^*,q_{p^*}) = C$ by definition of p^*. Combining we obtain $f(p,q_{p^*}) \leq \varepsilon^{-1}C + (1-\varepsilon^{-1})C = C$ $(\forall p)$ which proves $C = \max_p f(p,q_{p^*})$. Therefore, $C \geq \min_q \max_p f(p,q)$. Now $C \leq \max_p f(p,q)$ $(\forall q)$ hence $C \leq \min_q \max_p f(p,q)$, which proves the Lemma. ■

References

[1] G. Aishwarya and M. Madiman, "Conditional Rényi entropy and the relationships between Rényi capacities," *Entropy*, vol. 22, no. 526, pp. 1–17, May 2020.

[2] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Topics in Inf. Theory, Proc. 2nd Coll. Math. Societ. J. Bolyai*, no. 16, 1977, pp. 41–52.

[3] C. Cai and S. Verdú, "Conditional Rényi divergence saddlepoint and the maximization of α-mutual information," *Entropy*, vol. 21, no. 969, pp. 1–25, Oct. 2019.

[4] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 26–34, Jan. 1995.

[5] R. M. Fano, *Class notes for course 6.574: Transmission of Information*, MIT, Cambridge, MA, 1952.

[6] S.-W. Ho and S. Verdú, "Convexity/concavity of Renyi entropy and α-mutual information," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, China, June 14–19, 2015, pp. 745–749.

[7] Y. Liu, W. Cheng, S. Guilley, and O. Rioul, "On conditional alpha-information and its application in side-channel analysis," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Kanazawa, Japan, Oct. 17-21 2021.

[8] J. L. Massey, "Coding and modulation in digital communications," in *Proc. Int. Zurich Seminar Digit. Comm.*, Mar. 1974, pp. E2(1)–E2(4).

[9] B. Nakiboğlu, "The Rényi capacity and center," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 841–860, Feb. 2019.

[10] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *Proc. Forty-Eighth Annual Allerton Conference*, Allerton House, UIUC, IL, Oct. 2010, pp. 1327–1333.

[11] A. Rényi, "On measures of entropy and information," in *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1: Contributions to the Theory of Statistics. Berkeley, CA: University of California Press, 1961, pp. 547–561.

[12] O. Rioul, "A primer on alpha-information theory with application to leakage in secrecy systems," in *5th conference on Geometric Science of Information (GSI'21), Paris, France, 21-23 July 2021*, 2021.

[13] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3 & 4, pp. 379–423 & 623–656, July & Oct. 1948.

[14] ——, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, Sept. 1956.

[15] R. Sibson, "Information radius," *Zeitschrift Wahrscheinlichkeitstheorie Verwandte Gebiete*, vol. 14, pp. 149–160, Jun. 1969.

[16] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.

[17] S. Verdú, "α-mutual information," in *Proc. Information Theory and Applications Workshop*, La Jolla, CA, Feb. 2015, pp. 1–6.

[18] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. John Wiley & Sons, 1965.