The Sphere Packing Bound for DSPCs With Feedback à la Augustin

Barış Nakiboğlu, *Member, IEEE*

***Abstract*— Establishing the sphere packing bound for block codes on the discrete stationary product channels with feedback—which are commonly called the discrete memoryless channels with feedback—was considered to be an open problem until recently, notwithstanding the proof sketch provided by Augustin in 1978. A complete proof following Augustin's proof sketch is presented to demonstrate its adequacy and to draw attention to two novel ideas that it employs. These novel ideas (i.e., the Augustin's averaging and the use of subblocks) are likely to be applicable in other communication problems for establishing impossibility results.**

***Index Terms*— Feedback communications, reliability function, error exponent, sphere packing bound/exponent, error analysis.**

I. Introduction

AFTER the founding paper of Shannon [1], establishing the channel capacity as the threshold rate for reliable communication, one of the first challenges of the mathematical theory of communications was determining the behavior of the optimum error probability as a function of the block length at rates below the channel capacity. The optimum error probability was shown to decay exponentially with the block length and the exponent of this decay (i.e., the error exponent, or the reliability function) was determined at all rates between the critical rate and the capacity of the channel in [2]–[4] for various channel models. Although it was not always discussed in these terms, [2]–[4] proved the following two distinct results in order determine the error exponent at rates between the critical rate and the capacity of the channel.

(i) *The Random Coding Bound (RCB):* At all rates less than the capacity, the random coding exponent (RCE) is achievable, i.e., the error exponent is bounded from below by the RCE.

(ii) *The Sphere Packing Bound (SPB):* At any rate less than the capacity the error exponent is bounded from above by the sphere packing exponent (SPE).

The RCE and the SPE are equal to one another for all rates between the critical rate and the channel capacity. Thus (i) and (ii) determine the error exponent exactly for all rates between the critical rate and the capacity on any channel that they are established.

In [5], Gallager proved (i) not only for all of the models considered in [2]–[4], but also for essentially all memoryless channel models of interest, including the non-stationary ones. The elegance and the simplicity of Gallager's derivation and the generality of his result make his seminal paper [5] of interest to the contemporary researchers after decades [6], [7].

For the SPB—i.e., for (ii)—the progress did not happen all at once as it did for (i). The first two complete proofs of the SPB for arbitrary discrete stationary product channels (DSPCs)[1] by Shannon, Gallager, and Berlekamp in [8] and by Haroutunian in [9] both relied on expurgations based on the composition (i.e., the empirical distribution, or the type) of the input codewords. Thus the proofs in [8] and [9] hold only for codes on stationary channels with finite input sets. In [10], Augustin provided the first proof of the SPB on the product channels that does not assume either the stationarity of the channel or the finiteness of its input set. In [11], we have improved the approximation error term of the upper bound on the error exponent given in [10] from $O\left(n^{-0.5}\right)$ to $O\left(n^{-1}\ln n\right)$ for the block length n, using the Rényi capacity and center analyzed in [12].

Unlike the proofs in [8] and [10], Haroutunian's proof in [9] establishes the SPB not only for codes on the product channels but also for codes on the stationary memoryless channels with either composition or cost constraints. However, the finite input set hypothesis of [9] curbs its usefulness for models other than the discrete ones, e.g., [9] does not imply the SPB for the Poisson channels, derived for the first time in [13]. Building upon the techniques he developed in [10] and [14, §31] and employing the information measures he analyzed in [14, §34], Augustin proved the SPB on (possibly non-stationary) cost constrained memoryless channels with bounded cost functions in [14, §36]. Augustin's SPB given in [14, Thm. 36.6] applies to the Poisson channels, but not to various Gaussian channels analyzed in [2], [15], [16] because the quadratic cost function is not bounded. In [17], we have proved the SPB for codes on the cost constrained memoryless channels —without assuming the cost function to be bounded—using the constrained Augustin capacity and center analyzed in [14, §34] and [18].

Manuscript received June 29, 2018; revised January 2, 2019 and June 8, 2019; accepted July 16, 2019. Date of publication July 25, 2019; date of current version November 19, 2019. This work was partially supported by the NSF under Grant CCF-0917212 and Grant CNS-0932410. The associate editor coordinating the review of this article and approving it for publication was R. F. Schaefer.

The author is with the Department of Electrical and Electronics Engineering, Middle East Technical University, 06800 Ankara, Turkey (e-mail: bnakib@metu.edu.tr).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2019.2931302

[1]The channels that we call DSPCs are usually called discrete memoryless channels (i.e., DMCs). We use the name DSPC to underline the stationarity of these channels and the non-existence of constraints on their input sets; see §I-B for a more detailed discussion.

The SPB given in [17, Thm. 2] implies the SPB not only for the Poisson channels, but also for various Gaussian channels considered in [2], [15], [16].

Despite their generality, Augustin proofs in [10] and [14] did not have nearly as much impact as the proofs in [8] and [9]. This is partly due to the considerable simplification provided by the application of the composition based expurgations in [8] and [9]. This reliance on the composition based expurgations, however, were making the derivation of the SPB with the techniques in [8] and [9] rather convoluted and tedious—if at all possible—for codes on channels other than the stationary memoryless ones with finite input sets. For codes on DSPCs with feedback, for example, there is no evident generalization for the concept of composition of an input codeword that can be used in a derivation of the SPB similar to [8] or [9]. Thus establishing the SPB for arbitrary DSPCs with feedback has been a significant challenge. Nevertheless, several partial results have been reported over the years.

For DSPCs with feedback that have certain symmetries, Dobrushin established the SPB in [19]. For arbitrary DSPCs with feedback, Haroutunian [20] derived an upper bound on the error exponent, which is usually called Haroutunian's bound/exponent. Haroutunian's exponent is equal to the SPE only for DSPCs with certain symmetries; Haroutunian's exponent is strictly greater than the SPE even for non-symmetric binary input binary output channels. Sheverdyaev proposed a derivation of the SPB for codes on DSPCs with feedback using Taylor's expansion in [21]. Sheverdyaev's proof was, however, supported rather weakly on several critical steps, see [22, A7] for a more detailed discussion. Curtailing the ways feedback link can be used by appropriate assumptions, [22]–[24] derived the SPB for certain families of codes on arbitrary DSPCs with feedback.

Augustin presented a proof sketch establishing the SPB on arbitrary DSPCs with feedback in [14, §41]. Despite the novelty of Augustin's approach and the importance of his result, Augustin's proof sketch is not widely known. In fact, until very recently, establishing the SPB on DSPCs with feedback has been considered to be an open problem. In the following, we present a complete proof that is following Augustin's proof sketch without any significant modification. Our main aim is to make the two main ideas of Augustin's proof —the averaging and the use of subblocks—widely accessible via this relatively short article. We believe both ideas are likely to be useful in establishing impossibility results in other communications problems. We assume the channel to be discrete for simplicity and employ concepts that are not present, at least explicitly, in [14] —such as Rényi's information measures and stochastic sequences—whenever we think their use simplifies the discussion for the contemporary researcher.

Elsewhere, in [11, §V], we have proved the SPB for codes on DSPCs with feedback using the averaging and the subblock ideas of Augustin [14] together with the Taylor's expansion idea of Sheverdyaev [21] and the auxiliary channel idea of Haroutunian [9], [20]. In addition, we have shown in [11, §V-E] that Haroutunian's bound implies the SPB when considered together with the averaging and the subblock ideas of Augustin. Although proofs in [11, §V] do employ ideas from Augustin's proof sketch, both proofs also employ other fundamental observations which makes them substantially different from the proof we present in the following.

In the rest of the current section, we first describe our notation and model, then state the main asymptotic result, i.e., Theorem 1. In §II, we recall certain properties of Rényi's information measures and SPE, derive preliminary results on tilting and stochastic sequences, and state a sufficient condition for constructing a probability measure with a given set of conditional probabilities on a product space. In §III, we prove a non-asymptotic SPB for codes on DSPCs with feedback, which implies Theorem 1. In §IV, we discuss possible generalizations and alternative proofs for the main result of the paper, establishing the sphere packing exponent as an upper bound to the reliability function for channel with feedback.

A. Notation

We denote the set of all real numbers by $\mathbb{R}$, positive real numbers by $\mathbb{R}_{+}$, non-negative real numbers by $\mathbb{R}_{\geq 0}$, and integers by $\mathbb{Z}$. For any real number z, $\lfloor z \rfloor$ is the greatest integer less than or equal to z, $\lceil z \rceil$ is the least integer greater than or equal to z, and $|z|$ is the absolute value of z. For any set $\mathcal{A}$ the indicator function $\mathbb{1}_{\mathcal{A}}(\cdot)$ is defined as follows:

$$\mathbb{1}_{\mathcal{A}}(x) = \begin{cases} 1 & x \in \mathcal{A} \\ 0 & x \notin \mathcal{A}. \end{cases}$$

For any finite set $\mathcal{Y}$, we denote the set of all subsets of $\mathcal{Y}$ (i.e., the power set of $\mathcal{Y}$) by $2^{\mathcal{Y}}$ and the set of all probability mass functions (p.m.f.'s) on $\mathcal{Y}$ by $\mathcal{P}(\mathcal{Y})$. For any q and w in $\mathcal{P}(\mathcal{Y})$ the total variation distance between them is defined as

$$\|q - w\| \triangleq \sum\nolimits_{y \in \mathcal{Y}} |q(y) - w(y)|. \tag{1}$$

While discussing the continuity of functions, we will assume that the set of real numbers is equipped with its natural topology and the set of all p.m.f.'s is equipped with the total variation topology.

For any two finite sets $\mathcal{X}$ and $\mathcal{Y}$, we denote the Cartesian product of $\mathcal{X}$ and $\mathcal{Y}$ by $\mathcal{X} \times \mathcal{Y}$, the set of all functions from $\mathcal{X}$ to $\mathcal{Y}$ by $\mathcal{Y}^{\mathcal{X}}$, and the set of all stochastic matrices from $\mathcal{X}$ to $\mathcal{Y}$ by $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. We interpret stochastic matrices from $\mathcal{X}$ to $\mathcal{Y}$ as functions from $\mathcal{X}$ to $\mathcal{P}(\mathcal{Y})$, as well. Thus we use $W(x)$ and $W(\cdot|x)$ interchangeably for W's in $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. For any p in $\mathcal{P}(\mathcal{X})$ and W in $\mathcal{P}(\mathcal{Y}|\mathcal{X})$, $p \circledast W$ is the p.m.f. on $\mathcal{X} \times \mathcal{Y}$ whose marginal distribution on $\mathcal{X}$ is p and conditional distribution given x is $W(x)$. For any p in $\mathcal{P}(\mathcal{X})$ and q in $\mathcal{P}(\mathcal{Y})$, we denote their product, which is a p.m.f. on $\mathcal{X} \times \mathcal{Y}$, by $p \otimes q$. We use the symbol $\otimes$ to denote the product of σ-algebras, as well.

For any interval $\mathcal{A}$ on $\mathbb{R}$ the Borel σ-algebra of $\mathcal{A}$, denoted by $\mathcal{B}(\mathcal{A})$, is the minimum σ-algebra on the subsets of $\mathcal{A}$ that includes all the open subintervals of $\mathcal{A}$, [25, p. 143]. A pair $(\Omega, \mathcal{F})$ is a measurable space iff $\mathcal{F}$ is a σ-algebra of subsets of Ω. If in addition $\mathbf{P}$ is a probability on $\mathcal{F}$, then the triple $(\Omega, \mathcal{F}, \mathbf{P})$ form a probability space. A real valued function X on Ω is a random variable in the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ iff X is $\mathcal{F}$-measurable (i.e., the inverse image of every set in $\mathcal{B}(\mathbb{R})$ is in $\mathcal{F}$), [25, p. 170]. A sequence of pairs

$(\mathsf{X}_1,\mathcal{F}_1),\ldots,(\mathsf{X}_n,\mathcal{F}_n)$ is a stochastic sequence in $(\Omega,\mathcal{F},\mathbf{P})$ iff $\mathcal{F}_1,\ldots,\mathcal{F}_n$ are σ-algebras satisfying $\mathcal{F}_1\subset\cdots\mathcal{F}_n\subset\mathcal{F}$ and X_t's are $\mathcal{F}_t$-measurable random variables, [25, p. 476]. See [25, Ch. II], for an accessible introduction to the mathematical foundations of the probability theory.

Our notation will be overloaded for certain symbols, but the relations represented by these symbols will be clear from the context. We use the short hand $\mathcal{G}_t^n$ for the product of σ-algebras $\mathcal{G}_t,\ldots,\mathcal{G}_n$, $\mathcal{X}_t^n$ for the Cartesian product of sets $\mathcal{X}_t,\ldots,\mathcal{X}_n$, X_t^n for the random vector $(\mathsf{X}_t,\ldots,\mathsf{X}_n)$, and x_t^n for the vector $(x_t,\ldots,x_n)$.

B. The DSPCs With Feedback and the Channel Codes

A discrete channel with a finite input set $\mathcal{X}$ and a finite output set $\mathcal{Y}$, is represented by a stochastic matrix W. The product of a sequence of discrete channels $W_1,\ldots,W_n$ with the input sets $\mathcal{X}_1,\ldots,\mathcal{X}_n$ and the output sets $\mathcal{Y}_1,\ldots,\mathcal{Y}_n$ is a discrete channel from $\mathcal{X}_1^n$ to $\mathcal{Y}_1^n$, denoted by $W_{[1,n]}$, satisfying

$$W_{[1,n]}(y_1^n|x_1^n)=\prod\nolimits_{t=1}^{n} W_t(y_t|x_t)$$

for all x_1^n in $\mathcal{X}_1^n$ and y_1^n in $\mathcal{Y}_1^n$. A length n product channel $W_{[1,n]}$ is stationary iff all W_t's are identical. A discrete channel U from $\mathcal{Z}$ to $\mathcal{Y}_1^n$ is a length n memoryless channel if there exits a product channel $W_{[1,n]}$ with the input set $\mathcal{X}_1^n$ satisfying both $\mathcal{Z}\subset\mathcal{X}_1^n$ and $U(z)=W_{[1,n]}(z)$ for all $z\in\mathcal{Z}$.

The preceding definition of the memorylessness is wholly consistent with the one used in standard texts [26, p. 185], [27, (4.2.1)], [28, p. 84]. Nevertheless, the discrete product channels that are also stationary are customarily called discrete memoryless channels. Although the conventional name is not wrong, we prefer a more descriptive and accurate name: the discrete stationary product channels (DSPCs).

In discrete product channels (DPCs) probabilistic behavior of the channel outputs depend on the channel inputs, but the channel inputs do not depend on the channel outputs in any way. In DPCs with feedback, on the other hand, the channel input at any time instance may depend on the previous channel outputs, i.e., the channel input at time t can be a function from $\mathcal{Y}_1^{t-1}$ to $\mathcal{X}_t$ rather than an element of $\mathcal{X}_t$. We define the DPCs with feedback formally as follows.

Definition 1: For any positive integer n and $W_t:\mathcal{X}_t\to\mathcal{P}(\mathcal{Y}_t)$ for t in $\{1,\ldots,n\}$, the *length n discrete product channel with feedback* $W_{\overrightarrow{[1,n]}}:\overrightarrow{\mathcal{X}}_1^n\to\mathcal{P}(\mathcal{Y}_1^n)$ is defined via the following relation:

$$W_{\overrightarrow{[1,n]}}(y_1^n|\overrightarrow{x}_1^n)=W_1(y_1|\overrightarrow{x}_1)\prod\nolimits_{t=2}^{n} W_t(y_t|\overrightarrow{x}_t(y_1^{t-1})) \quad (2)$$

for all $\overrightarrow{x}_1^n\in\overrightarrow{\mathcal{X}}_1^n$ and $y_1^n\in\mathcal{Y}_1^n$ where $\overrightarrow{\mathcal{X}}_t=\mathcal{X}_t^{\mathcal{Y}_1^{t-1}}$ for $t\geq 2$ and $\overrightarrow{\mathcal{X}}_1=\mathcal{X}_1$. A DPC with feedback $W_{\overrightarrow{[1,n]}}$ is stationary, i.e., it is a DSPC with feedback, iff all W_t's are identical.

Broadly speaking, a channel code is a strategy to convey from the transmitter at the input of the channel to the receiver at the output of the channel, a random choice from a finite message set. The channel codes are usually described in terms of the amount of information they convey per channel use, i.e., in terms of their rate. In particular, a rate R *channel code* on a length n DPC with feedback $W_{\overrightarrow{[1,n]}}$ is an ordered pair (Ψ,Θ) composed of the *encoding function* Ψ that maps the message set $\mathcal{M}\triangleq\{1,2,\ldots,\lceil e^{nR}\rceil\}$ to the input set $\overrightarrow{\mathcal{X}}_1^n$ and the *decoding function* Θ that maps the output set $\mathcal{Y}_1^n$ to the message set $\mathcal{M}$.

The average error probability $P_{\mathbf{e}}^{av}$ of a rate R channel code (Ψ,Θ) on a length n DPC with feedback $W_{\overrightarrow{[1,n]}}$ is

$$P_{\mathbf{e}}^{av}\triangleq\frac{1}{\lceil e^{nR}\rceil}\sum\nolimits_{m\in\mathcal{M}} P_{\mathbf{e}}^{m}, \quad (3)$$

where $P_{\mathbf{e}}^{m}$ is *the conditional error probability* of the message m given by

$$P_{\mathbf{e}}^{m}\triangleq 1-\sum\nolimits_{y_1^n\in\mathcal{Y}_1^n}\mathbb{1}_{\{\Theta(y_1^n)\}}(m)\,W_{\overrightarrow{[1,n]}}(y_1^n|\Psi(m)). \quad (4)$$

C. Main Result

Definition 2: For any $\alpha\in(0,1]$, $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, and $p\in\mathcal{P}(\mathcal{X})$ *the order-α Rényi information for prior p* is

$$I_\alpha(p;W)\triangleq\begin{cases}\dfrac{\alpha}{\alpha-1}\ln\sum_y\left[\sum_x p(x)\left[W(y|x)\right]^\alpha\right]^{1/\alpha} & \alpha\in(0,1)\\ \sum_x p(x)\sum_y W(y|x)\ln\dfrac{W(y|x)}{q_{1,p}(y)} & \alpha=1,\end{cases}$$

where $q_{1,p}\in\mathcal{P}(\mathcal{Y})$ is defined as $q_{1,p}(y)\triangleq\sum_x p(x)W(y|x)$.

Definition 3: For any $\alpha\in(0,1]$ and $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$ *the order-α Rényi capacity of W* is

$$C_{\alpha,W}\triangleq\sup\nolimits_{p\in\mathcal{P}(\mathcal{X})} I_\alpha(p;W).$$

Both the Rényi information and the Rényi capacity are continuous non-decreasing functions of the order α on $(0,1]$, see [12, Lemmas 5 and 15]. We define the order-0 Rényi capacity as the continuous extension of the Rényi capacity at zero:[2]

$$C_{0,W}\triangleq\lim\nolimits_{\alpha\downarrow 0} C_{\alpha,W}. \quad (5)$$

Definition 4: For any stochastic matrix $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$ and rate $R\in\mathbb{R}_{\geq 0}$, *the sphere packing exponent* (SPE) is

$$E_{sp}(R,W)\triangleq\sup\nolimits_{\alpha\in(0,1)}\frac{1-\alpha}{\alpha}\left(C_{\alpha,W}-R\right).$$

Note that if $C_{0,W}=C_{1,W}$, then $E_{sp}(R,W)$ is infinite for R's in $[0,C_{1,W})$ and zero for R's in $[C_{1,W},\infty)$. For most stochastic matrices of interest, however, $C_{1,W}>C_{0,W}$ and consequently $E_{sp}(R,W)$ is a convex function of R that is infinite on $[0,C_{0,W})$, monotonically decreasing and continuous in R on $(C_{0,W},C_{1,W}]$, and zero on $[C_{1,W},\infty)$, see [11, Lemma 13].

Remark 1: For orders in $(0,1)$ the Rényi information is just a scaled and reparameterized version of the Gallager's function $E_0(\rho,p)$ introduced in [5]; in particular

$$I_\alpha(p;W)=\left.\frac{E_0(\rho,p)}{\rho}\right|_{\rho=\frac{1-\alpha}{\alpha}} \qquad \forall\alpha\in(0,1).$$

[2]The order-0 Rényi information is defined in a similar way and the supremum $I_0(p;W)$ over p's in $\mathcal{P}(\mathcal{X})$ is equal to $C_{0,W}$, as defined in (5), see [12, Lemma 16-(f)].

In [8], the function $E_0(\rho)$ is defined as the maximum of Gallager's function $E_0(\rho, p)$ over p's in $\mathcal{P}(\mathcal{X})$. Thus

$$C_{\alpha,W} = \left.\frac{E_0(\rho)}{\rho}\right|_{\rho=\frac{1-\alpha}{\alpha}} \qquad \forall\alpha\in(0,1).$$

Consequently, Definition 4 is merely a reparameterization of the definition used by Shannon, Gallager, and Berlekamp in [8, Thm. 2]. In [9], Haroutunian employed another expression for the SPE, which he proved to be equal to the one in [8]. This expression is commonly known as Haroutunian's form.

Theorem 1: For any $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$ satisfying $C_{0,W}\neq C_{1,W}$, and R_0, R_1 satisfying $C_{0,W} < R_0 < R_1 < C_{1,W}$, for all n large enough

$$P_{\mathbf{e}}^{av} \geq \exp\left(-n\left[E_{sp}\left(R-\frac{2\ln n}{n^{1/3}}, W\right)+\frac{2\ln n}{n^{1/3}}\right]\right) \quad (6)$$

for any rate R channel code on the length n DSPC with feedback $W_{\overrightarrow{[1,n]}}$ satisfying $W_t = W$ provided R satisfies

$$R_1 > R > R_0 + \frac{2\ln n}{n^{1/3}}. \quad (7)$$

Note that $\frac{2\ln n}{n^{1/3}}$ terms in (6) and (7) vanish as n increases; thus Theorem 1 establishes the SPE as an upper bound on the error exponent of any DSPC with feedback at any rate in $(C_{0,W}, C_{1,W})$, provided that $W_t = W$ for all t. In fact this result holds with uniform approximation error terms on every closed interval of rates in $(C_{0,W}, C_{1,W})$, as a result of Theorem 1. For rates less than $C_{0,W}$, SPE is infinite; thus the upper bound holds trivially. For rates larger than $C_{1,W}$, we already know that the optimal error probability of the channel codes converges to one by [21], [29].

II. Preliminaries

A. *Rényi's Information Measures and SPE*

Rényi's information measures have been studied explicitly [30]–[32] or implicitly [5], [8] since the sixties. For the finite sample space case, the propositions about them that we borrow from [12] and [33] in the following are relatively easy to prove and well-known, except for Lemma 4 establishing the continuity of the Rényi center as a function of the order. Lemma 5 states an immediate corollary of the monotonicity properties of the Rényi capacity and the definition of the SPE.

Definition 5: For any $\alpha\in(0,1]$ and $w, q\in\mathcal{P}(\mathcal{Y})$, *the order-$\alpha$ Rényi divergence between w and q* is

$$D_{\alpha}(w\|\, q) \triangleq \begin{cases} \sum_{y} w(y)\ln\frac{w(y)}{q(y)} & \alpha=1 \\ \frac{1}{\alpha-1}\ln\sum_{y}\left[w(y)\right]^{\alpha}\left[q(y)\right]^{1-\alpha} & \alpha\neq 1. \end{cases}$$

Note that for all $\alpha\in(0,1)$ and $w, q\in\mathcal{P}(\mathcal{Y})$ we have

$$\frac{1-\alpha}{\alpha}D_{\alpha}(w\|\, q) = D_{1-\alpha}(q\|\, w) \quad (8)$$

by definition. Using the derivatives of $e^{(\alpha-1)D_{\alpha}(w\|q)}$ with respect to α, one can show that as a function of its order the Rényi divergence is nondecreasing on $(0,1)$ and continuous from the left at one. Thus, we get the following proposition.

Lemma 1 ([33, Thms. 3, 7]): For any $w, q\in\mathcal{P}(\mathcal{Y})$, the Rényi divergence $D_{\alpha}(w\|\, q)$ is nondecreasing and continuous in α on $(0,1]$.

The Rényi divergence is non-negative as a result of the Jensen's inequality. This observation has been strengthened by the following inequality relating the Rényi divergence to the total variation distance [34], [35], called the Pinsker's inequality:

$$D_{\alpha}(w\|\, q) \geq \frac{\alpha}{2}\|w-q\|^{2} \quad (9)$$

for all $\alpha\in(0,1]$ and $w, q\in\mathcal{P}(\mathcal{Y})$.

Definition 6: For any $\alpha\in(0,1]$ and $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$ *the order-α Rényi radius of W* is

$$S_{\alpha,W} \triangleq \inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})}\max\nolimits_{x\in\mathcal{X}} D_{\alpha}(W(x)\|\, q).$$

The order-α Rényi capacity is defined as the supremum of the order-α Rényi information; however, it is also equal to the order-α Rényi radius, [32, Proposition 1]. In addition, there exists a unique order-α Rényi center corresponding to this radius. These observations are stated formally in Lemma 2.

Lemma 2 ([12, Thm. 1]): For any $\alpha\in(0,1]$ and $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$

$$C_{\alpha,W} = \inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})}\max\nolimits_{x\in\mathcal{X}} D_{\alpha}(W(x)\|\, q). \quad (10)$$

Furthermore, there exists a unique $q_{\alpha,W}$ in $\mathcal{P}(\mathcal{Y})$, called the order-α Rényi center of W, such that

$$C_{\alpha,W} = \max\nolimits_{x\in\mathcal{X}} D_{\alpha}(W(x)\|\, q_{\alpha,W}). \quad (11)$$

The Rényi capacity is nondecreasing in its order on $(0,1]$ as a result of Lemmas 1 and 2. Furthermore, $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ is nonincreasing in α on $(0,1)$, as a result of (8) and Lemmas 1 and 2. This implies the continuity of $C_{\alpha,W}$ in α on $(0,1)$, which can be extended to $(0,1]$.

Lemma 3 ([12, Lemma 15-(a,c)]): For any $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, $C_{\alpha,W}$ is nondecreasing and continuous in α on $(0,1]$ and $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ is nonincreasing in α on $(0,1)$.

As a result of Lemma 3, we have

$$C_{\alpha,W} \leq \frac{C_{1/2,W}}{1-\alpha} \qquad \forall\alpha\in(0,1). \quad (12)$$

The continuity of the Rényi capacity in the order implies the continuity of the Rényi center in the order.

Lemma 4 ([12, Lemma 20]): The Rényi center is a continuous function of its order on $(0,1]$, i.e., $\lim_{z\to\alpha}\|q_{z,W}-q_{\alpha,W}\| = 0$ for all $\alpha\in(0,1]$.

The continuity of the Rényi center in the order allows us to construct a probability measure that plays a crucial role in the proof of Theorem 1.

Proof of Lemma 4: The following identity, which is due to Sibson [31, p. 153], can be confirmed by substitution.

$$D_{\alpha}(p\circledast W\|p\otimes q) = I_{\alpha}(p;W) + D_{\alpha}(q_{\alpha,p}\|\, q) \quad (13)$$

where $q_{\alpha,p}$ is *the order-α Rényi mean* defined as follows

$$q_{\alpha,p}(y) \triangleq \frac{\left(\sum_{x} p(x)\left[W(y|x)\right]^{\alpha}\right)^{1/\alpha}}{\sum_{b}\left(\sum_{a} p(a)\left[W(b|a)\right]^{\alpha}\right)^{1/\alpha}}. \quad (14)$$

There exists a $p_\alpha^* \in \mathcal{P}(\mathcal{X})$ such that $I_\alpha(p_\alpha^*;W) = C_{\alpha,W}$ as a result of the extreme value theorem [36, 4.16] because $I_\alpha(p;W)$ is continuous in p on $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{X})$ is compact. Note that $q_{\alpha,p_\alpha^*} = q_{\alpha,W}$ by (9), (13), and Lemma 2. Applying (13) for $q = q_{\phi,W}$ and for $p = p_\alpha^*$ we get

$$\max_x D_\alpha(W(x)\| q_{\phi,W}) \geq C_{\alpha,W} + D_\alpha(q_{\alpha,W}\| q_{\phi,W}).$$

Then using the monotonicity of Rényi divergence in the order (i.e., Lemma 1) and Lemma 2 we get

$$C_{\phi,W} - C_{\alpha,W} \geq D_\alpha(q_{\alpha,W}\| q_{\phi,W}) \qquad \forall \phi\in[\alpha,1].$$

Then the lemma follows from (9) and Lemma 3. ∎

Lemma 5: For any stochastic matrix $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ satisfying $C_{0,W} \neq C_{1,W}$ and rate R in $(C_{0,W}, C_{1,W})$ there exists a $\phi \in (0,1)$ satisfying $C_{\phi,W} = R$ and an $\eta \in (\phi,1)$ satisfying $\frac{1-\eta}{\eta} C_{\eta,W} = E_{sp}(R,W)$.

Proof of Lemma 5: Since $C_{\alpha,W}$ is continuous in the order α by Lemma 3, the existence of the order ϕ follows from the intermediate value theorem [36, 4.23]. Then,

$$E_{sp}(R,W) = \sup_{\alpha\in(\phi,1)} \frac{1-\alpha}{\alpha}(C_{\alpha,W} - R),$$

because $C_{\alpha,W}$ is non-decreasing in the order α by Lemma 3. Thus $E_{sp}(R,W)$ is positive at all rates R in $(C_{0,W}, C_{1,W})$ because $C_{\beta,W} = \frac{R+C_{1,W}}{2}$ for some β in $(\phi,1)$ by the intermediate value theorem [36, 4.23]. Then there exists an order $z \in [\phi,1)$ satisfying

$$\begin{aligned} E_{sp}(R,W) &= \frac{1-z}{z}(C_{z,W} - R) \\ &< \frac{1-z}{z} C_{z,W}. \end{aligned}$$

Hence, $E_{sp}(R,W)$ is between the values of the function $\frac{1-\alpha}{\alpha} C_{\alpha,W}$ at $\alpha = z$ and at $\alpha = 1$. Then the continuity of $\frac{1-\alpha}{\alpha} C_{\alpha,W}$ in the order α—implied by Lemma 3—and the intermediate value theorem [36, 4.23] imply the existence of the order η in $(z,1)$, and hence in $(\phi,1)$. ∎

B. Tilting and the Selftilted Channel

Definition 7: For any $\alpha \in (0,1]$ and $w, q \in \mathcal{P}(\mathcal{Y})$ satisfying $D_\alpha(w\| q) < \infty$, *the order-α tilted p.m.f.* w_α^q is

$$w_\alpha^q(y) \triangleq e^{(1-\alpha)D_\alpha(w\|q)}[w(y)]^\alpha [q(y)]^{1-\alpha} \quad \forall y \in \mathcal{Y}.$$

One can confirm by substitution that

$$\alpha D_1(w_\alpha^q\| w) + (1-\alpha) D_1(w_\alpha^q\| q) = (1-\alpha) D_\alpha(w\| q), \quad (15)$$

provided that w_α^q is defined, i.e., $D_\alpha(w\| q) < \infty$.

The continuity of the tilted p.m.f. w_α^q in the order α on $(0,1)$ is an immediate consequence of its definition and Lemma 1. Interestingly, the continuity of w_α^q in the order α on $(0,1)$ holds even when q is changing continuously with α.

Lemma 6 ([11, Lemma 16]): Let q_α be a continuous function of the order α from $(0,1)$ to $\mathcal{P}(\mathcal{Y})$ and let $w \in \mathcal{P}(\mathcal{Y})$ satisfy $D_\alpha(w\| q_\alpha) < \infty$ for all $\alpha \in (0,1)$. Then

(a) $w_\alpha^{q_\alpha}$ is a continuous function of α from $(0,1)$ to $\mathcal{P}(\mathcal{Y})$, i.e., $\lim_{z\to\alpha} \|w_z^{q_z} - w_\alpha^{q_\alpha}\| = 0$ for all $\alpha \in (0,1)$.

(b) $D_\alpha(w\| q_\alpha)$, $D_1(w_\alpha^{q_\alpha}\| w)$, and $D_1(w_\alpha^{q_\alpha}\| q_\alpha)$ are continuous functions of α from $(0,1)$ to $\mathbb{R}_{\geq 0}$.

Since $\max_{x\in\mathcal{X}} D_\alpha(W(x)\| q_{\alpha,W})$ is finite by Lemma 2 and $q_{\alpha,W}$ changes continuously with α by Lemma 4, one can invoke Lemma 6 for $w = W(x)$ and $q_\alpha = q_{\alpha,W}$ for any $x \in \mathcal{X}$. In the proof of Theorem 1, this observation is used together with Lemma 8, given in the following, to construct a probability measure that is at the heart of the proof.

For establishing Theorem 1, we use two measure change arguments together with the Chebyshev's inequality. The bounds on the second moments, given in Lemma 7, are needed for applying the Chebyshev's inequality.

Lemma 7 ([14, Lemma 16.2-(a)]): If $D_\alpha(w\| q) < \infty$ for an $\alpha \in (0,1]$ and $w, q \in \mathcal{P}(\mathcal{Y})$, then

$$\sum_y w_\alpha^q(y) \ln^2 \frac{w_\alpha^q(y)}{w(y)} \leq 4e^{-2} + \frac{(1-\alpha)^2}{\alpha^2}[4 + [D_\alpha(w\| q)]^2], \quad (16)$$

$$\sum_y w_\alpha^q(y) \ln^2 \frac{w_\alpha^q(y)}{q(y)} \leq 4e^{-2} + \frac{4\alpha^2}{(1-\alpha)^2} + [D_\alpha(w\| q)]^2. \quad (17)$$

Proof of Lemma 7: Note that

$$\sum_y w_\alpha^q(y) \ln^2 \frac{w_\alpha^q(y)}{w(y)} \mathbb{1}_{[0,1]}\left(\frac{w_\alpha^q(y)}{w(y)}\right) \leq 4e^{-2}, \quad (18)$$

because $\sup_{\tau\in(0,1)} \tau \ln^2 \tau = \tau \ln^2 \tau\big|_{\tau=e^{-2}} \leq 4\, e^{-2}$.

Furthermore, let $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_+$ be

$$f(\tau) = 4e^{-2}\tau \mathbb{1}_{[0,e^2]}(\tau) + \ln^2 \tau \mathbb{1}_{(e^2,\infty)}(\tau).$$

Since f is a non-negative function satisfying $\ln^2 \tau \leq f(\tau)$ for all $\tau \geq 1$ we have

$$\begin{aligned} &\sum_y w_\alpha^q(y) \ln^2 \frac{w_\alpha^q(y)}{w(y)} \mathbb{1}_{(1,\infty)}\left(\frac{w_\alpha^q(y)}{w(y)}\right) \\ &= \left(\frac{1-\alpha}{\alpha}\right)^2 \sum_y w_\alpha^q(y) \ln^2 \left[\frac{w_\alpha^q(y)}{w(y)}\right]^{\frac{\alpha}{1-\alpha}} \mathbb{1}_{(1,\infty)}\left(\frac{w_\alpha^q(y)}{w(y)}\right) \\ &\leq \left(\frac{1-\alpha}{\alpha}\right)^2 \sum_y w_\alpha^q(y) f\left(\left[\frac{w_\alpha^q(y)}{w(y)}\right]^{\frac{\alpha}{1-\alpha}}\right). \end{aligned} \quad (19)$$

On the other hand the concavity of f, the Jensen's inequality, the definition of tilted p.m.f., and the monotonicity of f imply

$$\begin{aligned} &\sum_y w_\alpha^q(y) f\left(\left[\frac{w_\alpha^q(y)}{w(y)}\right]^{\frac{\alpha}{1-\alpha}}\right) \\ &\qquad \leq f\left(\sum_y w_\alpha^q(y) \left[\frac{w_\alpha^q(y)}{w(y)}\right]^{\frac{\alpha}{1-\alpha}}\right) \\ &\qquad \leq f\left(\sum_y q(y) e^{D_\alpha(w\|q)}\right) \\ &\qquad \leq (2 \vee D_\alpha(w\| q))^2. \end{aligned} \quad (20)$$

(16) follows from (18), (19), (20). One can prove (17), following a similar analysis and invoking (8). ∎

One can tilt the channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ with a q in $\mathcal{P}(\mathcal{Y})$, by tilting the individual $W(x)$'s; the resulting channel is called the tilted channel and denoted by W_α^q. If the Rényi center of the channel itself is used for tilting, then we call the resulting channel the selftilted channel.

Definition 8: For any $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $\alpha \in (0,1]$, *the order-α selftilted channel* $W_\alpha : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ is

$$W_\alpha(y|x) = [W(y|x)]^\alpha [q_{\alpha,W}(y)]^{1-\alpha} e^{(1-\alpha)D_\alpha(W(x)\|q_{\alpha,W})}$$

for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.

C. Construction of a Probability Measure With the Given Conditional Probabilities

In Definition 1, for describing the p.m.f. induced on the output set $\mathcal{Y}_1^n$ by an element $\overrightarrow{x}_1^n$ of the input set $\overrightarrow{\mathcal{X}}_1^n$, it was sufficient to specify the conditional p.m.f. given the past at each time instance. This is true for arbitrary finite sample spaces, as well. When constructing probability measures in a similar fashion for more general sample spaces, however, there are additional technical conditions one needs to ensure. If the conditional probability of events at each time are Borel functions of the past, then the existence of a unique probability measure is guaranteed, as demonstrated by the following lemma.

Lemma 8: Let $(\Omega_t, \mathcal{G}_t)$ be an arbitrary measurable space for each $t \in \{1, \ldots, n\}$ and $\Omega = \Omega_1^n$, $\mathcal{G} = \mathcal{G}_1^n$. Suppose that a probability measure $\mathbf{P}^{(1)}$ is given on $(\Omega_1, \mathcal{G}_1)$ and that, for every $\omega_1^t \in \Omega_1^t$ and $t \in \{1, \ldots, n-1\}$, probability measures $\mathbf{P}(\cdot|\omega_1^t)$ are given on $(\Omega_{t+1}, \mathcal{G}_{t+1})$. Suppose that for every $\mathcal{B} \in \mathcal{G}_{t+1}$ the functions $\mathbf{P}(\mathcal{B}|\omega_1^t)$ are Borel functions of ω_1^t and let

$$\mathbf{P}^{(t)}\left(\mathcal{A}_1^t\right) = \int_{\mathcal{A}_1} \mathbf{P}^{(1)}(\mathrm{d}\omega_1) \int_{\mathcal{A}_2} \mathbf{P}(\mathrm{d}\omega_2|\omega_1) \ldots \int_{\mathcal{A}_t} \mathbf{P}\left(\mathrm{d}\omega_t|\omega_1^{t-1}\right)$$

for all $\mathcal{A}_\imath \in \mathcal{G}_\imath$ and $t \in \{2, \ldots, n\}$. Then there is a unique probability measure $\mathbf{P}$ on $(\Omega, \mathcal{G})$ such that

$$\mathbf{P}(\{\omega : \omega_1 \in \mathcal{A}_1, \ldots, \omega_t \in \mathcal{A}_t\}) = \mathbf{P}^{(t)}\left(\mathcal{A}_1^t\right)$$

for every $t \in \{1, \ldots, n\}$.

Lemma 8 for $n=2$ case is [37, Thm. 2.6.2]. For arbitrary but finite n, Lemma 8 follows from a recursive application of [37, Thm. 2.6.2]. Lemma 8 is also implied by Ionescu Tulcea's theorem [25, Ch.II §9 Thm. 2], which establishes a more general result for the infinite horizon (i.e., n) case.

Remark 2: $\mathbf{P}(\mathcal{B}|\omega_1^t)$ is a Borel function iff the inverse image of every Borel set is in $\mathcal{G}_1^t$, i.e., if $\{\omega_1^t : \mathbf{P}(\mathcal{B}|\omega_1^t) \in \mathcal{C}\} \in \mathcal{G}_1^t$ for every $\mathcal{C} \in \mathcal{B}([0,1])$. If—for example—$(\Omega_t, \mathcal{G}_t) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ for all t, then $\mathbf{P}(\mathcal{B}|\omega_1^t)$'s are Borel functions whenever $\mathbf{P}(\mathcal{B}|\omega_1^t)$ are continuous in ω_1^t.

Remark 3: Lemma 8 requires $\mathbf{P}(\mathcal{B}|\omega_1^t)$'s to be Borel functions of ω_1^t. This is general enough for our purposes because we work with real valued random variables. More generally, this condition is stated as the measurability of $\mathbf{P}(\mathcal{B}|\omega_1^t)$ in $\mathcal{G}_1^t$, which makes $\mathbf{P}(\cdot|\cdot)$'s transition probabilities (i.e., Markov kernels or stochastic kernels), see [38, §10.7] for a more complete discussion. The same measurability condition makes $\mathbf{P}(\cdot|\cdot)$'s conditional distributions in the sense of [39, p. 343], as well.

The proof of Theorem 1 presented in the following section employs Lemma 8 in order to assert the existence of a probability with certain conditional probabilities. It is worth mentioning that we are not asserting that one needs to consider infinite sample spaces in order to calculate the average error probability of a channel code on a DSPC with feedback. The expressions in (3) and (4) determine the value of the average error probability relying solely on a finite sample space model. What we are saying is that Augustin's approach relies on a probability space with an infinite sample space in order to bound the minimum average error probability of channel codes on a given DSPC with feedback.

D. Chebyshev's Inequality

Lemma 9: Let $a_1, \ldots, a_n$ be a sequence of real numbers and $(\mathsf{X}_1, \mathcal{F}_1), \ldots, (\mathsf{X}_n, \mathcal{F}_n)$ be a stochastic sequence satisfying $\mathbf{E}[\mathsf{X}_t|\mathcal{F}_{t-1}] \le a_t$ and $\mathbf{E}\left[(\mathsf{X}_t)^2\right] < \infty$ for all t in $\{1, \ldots, n\}$, and σ satisfy $\sigma^2 = \sum_{t=1}^n \mathbf{E}\left[(\mathsf{X}_t)^2\right]$. Then

$$\mathbf{P}\left(\sum\nolimits_{t=1}^n \mathsf{X}_t < \gamma + \sum\nolimits_{t=1}^n a_t\right) \ge 1 - \frac{\sigma^2}{\gamma^2} \tag{21}$$

for all $\gamma \in \mathbb{R}_+$.

Lemma 9 is essentially a corollary of the Chebyshev's inequality, see [40, Appendix] for a proof. A similar lemma was stated for a particular stochastic sequence and probability space in [14, Lemma 41.4].

III. SPB FOR CODES ON DSPCs WITH FEEDBACK

The main aim of this section is to prove a non-asymptotic SPB, i.e., Lemma 10 given in the following. We use this non-asymptotic SPB to prove the asymptotic one given in Theorem 1 at the end of this section in §III-F. Let us start with stating the aforementioned non-asymptotic SPB.

Lemma 10: For any $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ satisfying $C_{0,W} \neq C_{1,W}$ and R_1, R_2 satisfying $C_{0,W} < R_0 < R_1 < C_{1,W}$, let $\phi \in (0,1)$ satisfy $C_{\phi,W} = R_0$, $\eta \in (\phi, 1)$ satisfy[3] $\frac{1-\eta}{\eta} C_{\eta,W} = E_{sp}(R_1, W)$, positive parameter ϵ satisfy $\epsilon \le \frac{\phi \wedge (1-\eta)}{2}$, and positive integers n, κ satisfy $\kappa \le n$. Then any rate R channel code on the length n DSPC with feedback $W_{\overrightarrow{[1,n]}}$ satisfying $W_t = W$ satisfies

$$P_{\mathbf{e}}^{av} \ge e^{-n[E_{sp}(R-\delta_1, W)+\delta_2]}, \tag{22}$$

provided that

$$R_1 \ge R \ge R_0 + \delta_1, \tag{23}$$

where

$$\delta_1 \triangleq \frac{\ln 4}{n} + 8\frac{2 + C_{1/2,W}}{(1-\eta)\sqrt{\kappa}} + \frac{\kappa}{n} \ln\left(n + \tfrac{1}{\epsilon}\right), \tag{24}$$

$$\delta_2 \triangleq \frac{\ln 4}{n} + 8\frac{2 + C_{1/2,W}}{\phi\sqrt{\kappa}} + \frac{\kappa \ln n}{n} + \frac{2R\epsilon}{\phi^2}. \tag{25}$$

The proof of Lemma 10 relies on a pigeon hole argument and a measure change argument. In this respect, it is similar to the standard proofs of the SPB. Its principle novelty is in the choice/construction of the probability spaces and measures to apply these arguments. We present this construction and the proof through self contained pieces in §III-A-§III-E.

- In §III-A, we divide the block length into κ subblocks of approximately equal length.

[3] Such a ϕ and η can always be found as a result of Lemmas 3 and 5.

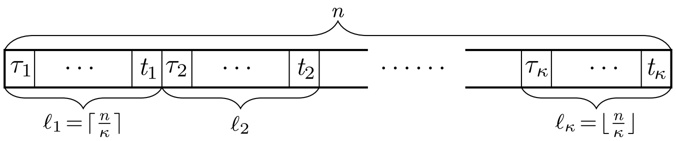

Fig. 1. A typical partitioning of the length n block into κ subblocks. The length of the first subblock is always $\lceil \frac{n}{\kappa}\rceil$ and the length of the last subblock is always $\lfloor \frac{n}{\kappa}\rfloor$.

- In §III-B, we extend the natural finite sample space that is used to describe the channel codes by introducing a positive valued random variable at beginning of each subblock and construct probability measures $\mathbf{P}$, $\mathbf{P}_v$, $\mathbf{P}_q$ for the extended sample space using a sequence of functions $g_1,\ldots,g_\kappa$ to be determined later. The probability of the error event under $\mathbf{P}$ will be equal to $P_{\mathbf{e}}^{av}$ by construction.
- In §III-C, we describe a choice of the functions $g_1,\ldots,g_\kappa$ that bounds the order-one Rényi divergence between the conditional p.m.f.'s of the outputs of the subblocks, i.e. $\mathsf{Y}_{1+t_{\imath-1}}^{t_\imath}$'s, under $\mathbf{P}_v$ and $\mathbf{P}_q$ —as well as under $\mathbf{P}_v$ and $\mathbf{P}$—$\mathbf{P}_v$-almost surely.
- In §III-D, we use Chebyshev's inequality to find an event $\mathcal{E}$ in the extended probability spaces satisfying $\mathbf{P}_v(\mathcal{E})\geq 0.5$ for which both $\mathbf{P}(\mathcal{E}\cap\mathcal{B})\gtrapprox e^{-nE_{sp}(R,W)}\mathbf{P}_v(\mathcal{E}\cap\mathcal{B})$ and $\mathbf{P}_q(\mathcal{E}\cap\mathcal{B})\gtrapprox e^{-nR}\mathbf{P}_v(\mathcal{E}\cap\mathcal{B})$ hold for any event $\mathcal{B}$ in the extended probability spaces.
- In §III-E, we apply a measure change argument together with a pigeon hole argument to prove Lemma 10.

In the following, we assume without loss of generality that the input and output sets are finite subsets of $\mathbb{R}$. This will allow us to call the channel input and output at time t random variables and to denote them by X_t and Y_t, respectively. Similarly, we assume that $\mathcal{M}$ is a subset of $\mathbb{R}$ and denote the random variables associated with the transmitted and decoded messages by M and $\widehat{\mathsf{M}}$, respectively. We denote the realizations of the random variables such as M, $\mathsf{Z}_\imath$, $\widehat{\mathsf{M}}$ or vectors such as X_τ^t, Y_τ^t by the corresponding lower case letters such as m, $z_\imath$, $\widehat{m}$ or x_τ^t, y_τ^t. We denote the expected value of a random variable Q under $\mathbf{P}_v$ by $\mathbf{E}_v[\mathsf{Q}]$. As it is customary, we denote the expected value of a random variable Q conditioned on the random variable Z (i.e., conditioned on the minimum σ-algebra generated by Z) by $\mathbf{E}[\mathsf{Q}|\,\mathsf{Z}]$. When we are working with $\mathbf{P}_v$ instead of $\mathbf{P}$, we use $\mathbf{E}_v[\mathsf{Q}|\,\mathsf{Z}]$ rather than $\mathbf{E}[\mathsf{Q}|\,\mathsf{Z}]$.

A. Division Into κ Subblocks

We divide the length n block into κ subblocks of length either $\lfloor \frac{n}{\kappa}\rfloor$ or $\lceil \frac{n}{\kappa}\rceil$. In particular, we set t_0 to zero and define $\ell_\imath$ and $t_\imath$ for $\imath\in\{1,\ldots,\kappa\}$ as follows

$$\ell_\imath \triangleq \lceil n/\kappa\rceil \mathbb{1}_{(0,n-\lfloor n/\kappa\rfloor\kappa]}(\imath)+\lfloor n/\kappa\rfloor \mathbb{1}_{(n-\lfloor n/\kappa\rfloor\kappa,\kappa]}(\imath),$$
$$t_\imath \triangleq t_{\imath-1}+\ell_\imath.$$

The last time instance of the $\imath^{th}$ subblock is $t_\imath$; for brevity, we denote the first time instance by $\tau_\imath$, i.e.,

$$\tau_\imath \triangleq t_{\imath-1}+1.$$

Figure 1 demonstrates a typical partitioning of the length n block into κ subblocks.

B. Construction of Auxiliary Probability Measures for a Given Sequence of Functions $g_1,\ldots,g_\kappa$

Let the sample space Ω and σ-algebra of its subsets $\mathcal{F}$ be

$$\Omega \triangleq \mathcal{M}\times\mathcal{Z}_1\times\mathcal{Y}_{\tau_1}^{t_1}\times\cdots\times\mathcal{Z}_\kappa\times\mathcal{Y}_{\tau_\kappa}^{t_\kappa},$$
$$\mathcal{F} \triangleq 2^{\mathcal{M}}\otimes\mathcal{B}(\mathcal{Z}_1)\otimes 2^{\mathcal{Y}_{\tau_1}^{t_1}}\otimes\cdots\otimes\mathcal{B}(\mathcal{Z}_\kappa)\otimes 2^{\mathcal{Y}_{\tau_\kappa}^{t_\kappa}},$$

where $\mathcal{Z}_\imath$ is the open interval $(0,1)$ and $\mathcal{B}(\mathcal{Z}_\imath)$ is the associated Borel σ-algebra for each $\imath$ in $\{1,\ldots,\kappa\}$.

Let the σ-algebras $\mathcal{F}_0,\ldots,\mathcal{F}_\kappa$ be

$$\mathcal{F}_0 \triangleq 2^{\mathcal{M}},$$
$$\mathcal{F}_\imath \triangleq \mathcal{F}_{\imath-1}\otimes\mathcal{B}(\mathcal{Z}_\imath)\otimes 2^{\mathcal{Y}_{\tau_\imath}^{t_\imath}} \quad \forall\imath\in\{1,\ldots,\kappa\}.$$

In the following, we construct three probability measures on $(\Omega,\mathcal{F})$—i.e., $\mathbf{P}$, $\mathbf{P}_v$, and $\mathbf{P}_q$—through their marginal distributions on $\mathcal{M}$ and their conditional distributions using Lemma 8. The marginal distributions of $\mathbf{P}$, $\mathbf{P}_v$, and $\mathbf{P}_q$ on the message set $\mathcal{M}$ are all equal to the uniform distribution. We specify the conditional distributions of $\mathsf{Z}_\imath$'s individually and the conditional distributions of Y_t's jointly through the conditional distributions of the vectors of the form $\mathsf{Y}_{\tau_\imath}^{t_\imath}$. In both cases, however, we demonstrate the conditional distributions to be Borel functions. This allows us to invoke the existence of unique probability measures $\mathbf{P}$, $\mathbf{P}_v$, and $\mathbf{P}_q$ on $(\Omega,\mathcal{F})$ with the given conditional distributions[4] via Lemma 8.

Let us first describe the conditional distributions of Z's. Let g_1 be a function from $\mathcal{M}$ to $(0,1)$ to be determined later. Similarly, for each $\imath$ in $\{2,\ldots,n\}$, let $g_\imath:\mathcal{M}\times\mathcal{Y}_1^{t_{\imath-1}}\to(0,1)$ be a function that is to be determined later. The conditional distribution of $\mathsf{Z}_\imath$ is the same for $\mathbf{P}$, $\mathbf{P}_v$, and $\mathbf{P}_q$ and it is determined by the function $g_\imath$ as follows:

$$\mathbf{P}\Big(\mathcal{A}\big|\, m,z_1^{\imath-1},y_1^{t_{\imath-1}}\Big)=\frac{1}{\epsilon}\int_{(1-\epsilon)\alpha}^{\alpha+\epsilon(1-\alpha)}\mathbb{1}_{\mathcal{A}}(z)\mathrm{d}z \tag{26}$$

for all $\mathcal{A}\in\mathcal{B}(\mathcal{Z}_\imath)$, where $\alpha=g_\imath(m,y_1^{t_{\imath-1}})$. Since $\mathcal{M}\times\mathcal{Y}_1^{t_{\imath-1}}$ is a finite set, all of the elements of its power set are Borel sets and $\mathbf{P}\Big(\mathcal{A}\big|\, m,z_1^{\imath-1},y_1^{t_{\imath-1}}\Big)$ is a Borel function for any $\mathcal{A}\in\mathcal{B}(\mathcal{Z}_\imath)$.

Let us proceed with the description of the conditional probability distributions of Y's. For $\mathbf{P}$ we have

$$\mathbf{P}\Big(y_{\tau_\imath}^{t_\imath}\big|\, m,z_1^\imath,y_1^{t_{\imath-1}}\Big)=\prod\nolimits_{t=\tau_\imath}^{t_\imath}W(y_t|x_t) \tag{27}$$

for all $y_{\tau_\imath}^{t_\imath}\in\mathcal{Y}_{\tau_\imath}^{t_\imath}$ where x_t is the channel input at time t, which is nothing but $\overrightarrow{x}_t(y_1^{t-1})$ for $\overrightarrow{x}_1^n$ satisfying $\Psi(m)=\overrightarrow{x}_1^n$. Note that $\mathbf{P}\Big(\mathcal{A}\big|\, m,z_1^\imath,y_1^{t_{\imath-1}}\Big)$ does not depend on $z_1^\imath$. Thus $\mathbf{P}\Big(\mathcal{A}\big|\, m,z_1^\imath,y_1^{t_{\imath-1}}\Big)$ is a Borel function for all $\mathcal{A}\subset\mathcal{Y}_{\tau_\imath}^{t_\imath}$ as a consequence of the finiteness of $\mathcal{M}\times\mathcal{Y}_1^{t_{\imath-1}}$.

For $\mathbf{P}_q$ we have

$$\mathbf{P}_q\Big(y_{\tau_\imath}^{t_\imath}\big|\, m,z_1^\imath,y_1^{t_{\imath-1}}\Big)=\prod\nolimits_{t=\tau_\imath}^{t_\imath}q_{z_\imath,W}(y_t) \tag{28}$$

for all $y_{\tau_\imath}^{t_\imath}\in\mathcal{Y}_{\tau_\imath}^{t_\imath}$. Since Rényi center is continuous in its order by Lemma 4, $\mathbf{P}_q\Big(\mathcal{A}\big|\, m,z_1^\imath,y_1^{t_{\imath-1}}\Big)$ is a continuous and hence a Borel function of $z_\imath$ for all $\mathcal{A}\subset\mathcal{Y}_{\tau_\imath}^{t_\imath}$.

[4]Those readers who are not already familiar with the technical subtleties about the conditional probabilities might benefit from taking the existence of $\mathbf{P}$, $\mathbf{P}_v$, and $\mathbf{P}_q$ on $(\Omega,\mathcal{F})$ with the conditional distributions given in (26), (27), (28), and (29) granted, at least in their initial reading.

For $\mathbf{P}_v$ we have

$$\mathbf{P}_v\left(y_{\tau_\imath}^{t_\imath}\middle| m, z_1^\imath, y_1^{t_\imath-1}\right)=\prod\nolimits_{t=\tau_\imath}^{t_\imath} W_{z_\imath}(y_t|\Psi_t(m, y_1^{t-1})) \quad (29)$$

for all $y_{\tau_\imath}^{t_\imath} \in \mathcal{Y}_{\tau_\imath}^{t_\imath}$ where $W_{z_\imath}$ is the order-$z_\imath$ selftilted channel described in Definition 8 and x_t is the channel input at time t. Since $W_\alpha(\cdot|x)$ is continuous in α for any x by Lemmas 4 and 6, $\mathbf{P}_v\left(\mathcal{A}\middle| m, z_1^\imath, y_1^{t_\imath-1}\right)$ is a continuous function of $z_\imath$ for any $y_1^{t_\imath-1}$, which does not depend on $z_1^{\imath-1}$. Since $\mathcal{Y}_1^{t_\imath-1}$ is a finite set, this will ensure $\mathbf{P}_v\left(\mathcal{A}\middle| m, z_1^\imath, y_1^{t_\imath-1}\right)$ to be a Borel function for any $\mathcal{A}\subset \mathcal{Y}_{\tau_\imath}^{t_\imath}$.

C. A Choice of $g_1, \ldots, g_\kappa$

The preceding construction works for any choice of the functions $g_1, \ldots, g_\kappa$. However, only some of the choices are appropriate for our purposes. In the following, we choose $g_1, \ldots, g_\kappa$ by determining the value of $g_\imath(m, y_1^{t_\imath-1})$ for each $\imath$, m, and $y_1^{t_\imath-1}$ individually and commit to the resulting $g_1, \ldots, g_\kappa$'s for the rest of the paper. In order to find the aforementioned appropriate choice we analyze the value of certain conditional expectation—i.e., $\mathbf{E}_v\left[\mathsf{H}_\imath\middle| m, y_1^{t_\imath-1}\right]$—as a function of the value of $g_\imath$ at $(m, y_1^{t_\imath-1})$—i.e., as a function of $g_\imath(m, y_1^{t_\imath-1})$—at each $(m, y_1^{t_\imath-1})$ individually.

Note that $D_1(W_{\mathsf{Z}_\imath}(\mathsf{X}_t)\| q_{\mathsf{Z}_\imath,W})$ is a random variable that is measurable in the σ-algebra generated by X_t and $\mathsf{Z}_\imath$ because $D_1(W_z(x)\| q_{z,W})$ is continuous in z by Lemmas 4 and 6. For any $\imath\in\{1, \ldots, \kappa\}$, let the random variable $\mathsf{H}_\imath$ be

$$\mathsf{H}_\imath \triangleq \sum\nolimits_{t=\tau_\imath}^{t_\imath} \mathbf{E}_v[D_1(W_{\mathsf{Z}_\imath}(\mathsf{X}_t)\| q_{\mathsf{Z}_\imath,W})| \mathcal{F}_{\imath-1},\mathsf{Z}_\imath]. \quad (30)$$

Note that $\mathsf{H}_\imath$ is a non-negative random variable by (9). Furthermore $D_1(W_{\mathsf{Z}_\imath}(\mathsf{X}_t)\| q_{\mathsf{Z}_\imath,W}) \leq D_{\mathsf{Z}_\imath}(W(\mathsf{X}_t)\| q_{\mathsf{Z}_\imath,W})$ by (9) and (15) and $D_{\mathsf{Z}_\imath}(W(\mathsf{X}_t)\| q_{\mathsf{Z}_\imath,W})\leq C_{\mathsf{Z}_\imath,W}$ by Lemma 2. Thus for any $\imath\in\{1, \ldots, \kappa\}$, the random variables $\mathsf{H}_\imath$ and $C_{\mathsf{Z}_\imath,W}$ satisfy

$$0 \leq \mathsf{H}_\imath \leq \ell_\imath C_{\mathsf{Z}_\imath,W} \quad (31)$$

for all realizations of $\mathcal{F}_{\imath-1}$ and $\mathsf{Z}_\imath$. Then for all realizations of $\mathcal{F}_{\imath-1}$, the conditional expectation $\mathbf{E}_v[\mathsf{H}_\imath| \mathcal{F}_{\imath-1}]$ is a continuous function of the value of $g_\imath$ at $(m, y_1^{t_\imath-1})$ —i.e., $g_\imath(m, y_1^{t_\imath-1})$— as a result of (26) defining the conditional distribution of $\mathsf{Z}_\imath$ for $\mathbf{P}$, $\mathbf{P}_v$, and $\mathbf{P}_q$, because $C_{\alpha,W}$ is nondecreasing in α and finite on $(0, 1)$ by Lemma 3. Thus we can tune the value of $\mathbf{E}_v[\mathsf{H}_\imath| \mathcal{F}_{\imath-1}]$ by changing the value of the function $g_\imath$ for different realizations of M and $\mathsf{Y}_1^{t_\imath-1}$.

On the other hand as a result of the construction, we have

$$\mathbf{E}_v[\mathsf{H}_\imath| \mathcal{F}_{\imath-1}] = \mathbf{E}_v\left[\mathsf{H}_\imath\middle| \mathsf{M}, \mathsf{Y}_1^{t_\imath-1}\right]. \quad (32)$$

We use the following rule to choose the value $g_\imath$ at each $(m, y_1^{t_\imath-1})$ depending on the rate of the code R and the positive constant δ_1 defined in (24).

- If $\mathbf{E}_v\left[\mathsf{H}_\imath\middle| m, y_1^{t_\imath-1}\right] \leq \ell_\imath(R-\delta_1)$ for $g_\imath(m, y_1^{t_\imath-1})=\frac{\eta}{1-\epsilon}$, then $g_\imath(m, y_1^{t_\imath-1})=\frac{\eta}{1-\epsilon}$.
- If $\mathbf{E}_v\left[\mathsf{H}_\imath\middle| m, y_1^{t_\imath-1}\right] > \ell_\imath(R-\delta_1)$ for $g_\imath(m, y_1^{t_\imath-1})=\frac{\eta}{1-\epsilon}$, then $g_\imath(m, y_1^{t_\imath-1})=\alpha$ for an α in $[\frac{\phi-\epsilon}{1-\epsilon}, \frac{\eta}{1-\epsilon})$ satisfying $\mathbf{E}_v\left[\mathsf{H}_\imath\middle| m, y_1^{t_\imath-1}\right] = \ell_\imath(R - \delta_1)$. The existence of such an α follows from the continuity of $\mathbf{E}_v\left[\mathsf{H}_\imath\middle| m, y_1^{t_\imath-1}\right]$ in the value of $g_\imath(m, y_1^{t_\imath-1})$, the intermediate value theorem [36, 4.23], and the inequality $\mathbf{E}_v\left[\mathsf{H}_\imath\middle| m, y_1^{t_\imath-1}\right] \leq \ell_\imath(R - \delta_1)$ for $g_\imath(m, y_1^{t_\imath-1}) = \frac{\phi-\epsilon}{1-\epsilon}$. In order to see why the inequality at $\frac{\phi-\epsilon}{1-\epsilon}$ holds, first note that (26) and (31) imply

$$\mathbf{E}_v\left[\mathsf{H}_\imath\middle| m, y_1^{t_\imath-1}\right] \leq \frac{\ell_\imath}{\epsilon}\int_{\phi-\epsilon}^{\phi} C_{z,W}\,\mathrm{d}z.$$

Then the inequality follows from (23), $R_0 = C_{\phi,W}$, and the monotonicity of the Rényi capacity in its order.

The choice of $g_\imath$'s described above ensures not only

$$0 \leq \mathbf{E}_v[\mathsf{H}_\imath| \mathcal{F}_{\imath-1}] \leq \ell_\imath(R - \delta_1) \quad (33)$$

for all $\imath \in \{1, \ldots, \kappa\}$, but also

$$\frac{\phi-\epsilon}{1-\epsilon} \leq \mathsf{G}_\imath \leq \frac{\eta}{1-\epsilon} \quad (34)$$

$$\phi-\epsilon \leq \mathsf{Z}_\imath \leq \eta+\epsilon \quad (35)$$

for all $\imath \in \{1, \ldots, \kappa\}$, where $\mathsf{G}_\imath$ is the random variable defined as $\mathsf{G}_\imath \triangleq g_\imath(\mathsf{M}, \mathsf{Y}_1^{t_\imath-1})$.

D. Application of Chebyshev's Inequality to Find an Event With Substantial Probability Under the Auxiliary Measure

The preceding choice of the functions $g_1, \ldots, g_\kappa$, bounds the expected value of random variables that are used in the measure change argument. In order to apply the measure change argument, we first prove —using Lemma 9—that these random variables take values that are close to their means with substantial probability under $\mathbf{P}_v$.

For any $\imath\in\{1, \ldots, \kappa\}$, let $\mathcal{F}_\imath$-measurable random variable $\mathsf{Q}_\imath$ be

$$\mathsf{Q}_\imath \triangleq \ln\frac{\mathbf{P}_v\left(\mathsf{Y}_{\tau_\imath}^{t_\imath}\middle| \mathsf{M}, \mathsf{Z}_1^\imath, \mathsf{Y}_1^{t_\imath-1}\right)}{\mathbf{P}_q\left(\mathsf{Y}_{\tau_\imath}^{t_\imath}\middle| \mathsf{M}, \mathsf{Z}_1^\imath, \mathsf{Y}_1^{t_\imath-1}\right)}. \quad (36)$$

Note that (28), (29), (30), and the definition of order-one Rényi divergence imply

$$\mathbf{E}_v[\mathsf{Q}_\imath| \mathcal{F}_{\imath-1},\mathsf{Z}_\imath]=\mathsf{H}_\imath.$$

Then (33) implies

$$0 \leq \mathbf{E}_v[\mathsf{Q}_\imath| \mathcal{F}_{\imath-1}] \leq \ell_\imath(R - \delta_1) \quad (37)$$

for all $\imath \in \{1, \ldots, \kappa\}$.

Let us proceed with bounding the second moments of $\mathrm{Q}_\imath$'s from above. The Cauchy-Schwarz inequality implies

$$\begin{aligned}\mathbf{E}_v[(\mathsf{Q}_\imath)^2] &= \sum\nolimits_{t=\tau_\imath}^{t_\imath} \sum\nolimits_{\jmath=\tau_\imath}^{t_\imath} \mathbf{E}_v[\mathrm{D}_t\mathrm{D}_\jmath] \\ &\leq \sum\nolimits_{t=\tau_\imath}^{t_\imath} \sum\nolimits_{\jmath=\tau_\imath}^{t_\imath} \sqrt{\mathbf{E}_v[(\mathrm{D}_t)^2]\mathbf{E}_v[(\mathrm{D}_\jmath)^2]}\end{aligned}$$

where $\mathrm{D}_t \triangleq \ln\frac{\mathbf{P}_v(\mathrm{Y}_t|\mathrm{M},\mathrm{Z}_1^\imath,\mathrm{Y}_1^{t-1})}{\mathbf{P}_q(\mathrm{Y}_t|\mathrm{M},\mathrm{Z}_1^\imath,\mathrm{Y}_1^{t-1})}$ for all $t \in \{\tau_\imath, \ldots, t_\imath\}$.

On the other hand using the definition of the order-one Rényi divergence and (17) of Lemma 7 we get

$$\begin{aligned}\mathbf{E}_v[(\mathrm{D}_t)^2] &= \mathbf{E}_v[\mathbf{E}_v[(\mathrm{D}_t)^2 \mid \mathcal{F}_{\imath-1}, \mathrm{Z}_\imath, \mathrm{Y}_{\tau_\imath}^{t-1}]] \\ &\leq \mathbf{E}_v\left[\frac{4}{(1-\mathrm{Z}_\imath)^2} + (D_{Z_\imath}(W(\mathrm{X}_t)\|q_{Z_\imath,W}))^2\right]\end{aligned}$$

First invoking (11) and (12) to bound $D_{Z\imath}(W(X_t)\|q_{Z_\imath,W})$, and then using the identity $1-\mathsf{Z}_\imath \geq \frac{1-\eta}{2}$, which follows from (35) and the hypothesis $\epsilon \leq \frac{\phi\wedge(1-\eta)}{2}$, we get

$$\mathbf{E}_v[(\mathsf{D}_t)^2] \leq \mathbf{E}_v\left[\frac{4+(C_{1/2,W})^2}{(1-\mathsf{Z}_\imath)^2}\right] \leq 4\frac{4+(C_{1/2,W})^2}{(1-\eta)^2}.$$

Thus using $\ell_\imath \leq 2\frac{n}{k}$ we get

$$\mathbf{E}_v[(\mathsf{Q}_\imath)^2] \leq \ell_1^2 4\frac{4+(C_{1/2,W})^2}{(1-\eta)^2} \leq 16\frac{4+(C_{1/2,W})^2}{(1-\eta)^2}\frac{n^2}{k^2}. \tag{38}$$

Applying Lemma 9, for $a_\imath = \ell_\imath(R-\delta_1)$ to the stochastic sequence[5] $(\mathsf{Q}_1,\mathcal{F}_1),\ldots,(\mathsf{Q}_\kappa,\mathcal{F}_\kappa)$ via (37) we get

$$\mathbf{P}_v(\mathsf{Q}\leq n(R-\delta_1)+\gamma) \geq 1-\frac{\sum_{\imath=1}^{\kappa}\mathbf{E}_v\left[(\mathsf{Q}_\imath)^2\right]}{\gamma^2},$$

where Q is defined as

$$\mathsf{Q} \triangleq \sum\nolimits_{\imath=1}^{\kappa}\mathsf{Q}_\imath. \tag{39}$$

Setting $\gamma = 8\frac{(2+C_{1/2,W})n}{(1-\eta)\sqrt{\kappa}}$ and invoking (24) and (38) we get

$$\mathbf{P}_v(\mathcal{E}_q) \geq \frac{3}{4}, \tag{40}$$

where $\mathcal{E}_q$ is defined as

$$\mathcal{E}_q \triangleq \left\{\omega\in\Omega : \mathsf{Q}(\omega)\leq nR-\ln 4-\kappa\ln\left(n+\frac{1}{\epsilon}\right)\right\}. \tag{41}$$

Recall that for all $\imath\in\{1,\ldots,\kappa\}$ the conditional distributions of $\mathbf{P}_v$ and $\mathbf{P}_q$ for $\mathsf{Z}_\imath$'s given $\mathcal{F}_{\imath-1}$ are identical because of (26). Thus $\mathsf{Q}(\omega) = \ln\frac{\mathrm{d}\mathbf{P}_v}{\mathrm{d}\mathbf{P}_q}(\omega)$ and consequently

$$\mathbf{P}_q(\mathcal{B}\cap\{\mathsf{Q}\leq\lambda\}) \geq e^{-\lambda}\mathbf{P}_v(\mathcal{B}\cap\{\mathsf{Q}\leq\lambda\}) \tag{42}$$

for any $\mathcal{B}\in\mathcal{F}$ and $\lambda\in\mathbb{R}$.

We need identities analogous to (40) and (42) for $\mathbf{P}$ and $\mathbf{P}_v$, as well. The random variables $\mathsf{V}_1,\ldots,\mathsf{V}_\kappa$ are used to obtain those identities. For any $\imath\in\{1,\ldots,\kappa\}$, let $\mathcal{F}_\imath$-measurable random variable $\mathsf{V}_\imath$ be

$$\mathsf{V}_\imath \triangleq \ln\frac{\mathbf{P}_v\left(\mathsf{Y}_{\tau_\imath}^{t_\imath}\middle|\,\mathsf{M},\mathsf{Z}_1^\imath,\mathsf{Y}_1^{t_{\imath-1}}\right)}{\mathbf{P}\left(\mathsf{Y}_{\tau_\imath}^{t_\imath}\middle|\,\mathsf{M},\mathsf{Z}_1^\imath,\mathsf{Y}_1^{t_{\imath-1}}\right)}. \tag{43}$$

Then as a result of (27), (29), and the definition of order-one Rényi divergence

$$\mathbf{E}_v[\mathsf{V}_\imath|\,\mathcal{F}_{\imath-1},\mathsf{Z}_\imath] = \sum\nolimits_{t=\tau_\imath}^{t_\imath}\mathbf{E}_v[D_1(W_{\mathsf{Z}_\imath}(\mathsf{X}_t)\|\,W(\mathsf{X}_t))|\,\mathcal{F}_{\imath-1},\mathsf{Z}_\imath].$$

On the other hand as a result of (15) and Lemma 2, we have

$$D_1(W_{\mathsf{Z}_\imath}(\mathsf{X}_t)\|\,W(\mathsf{X}_t)) \leq \frac{1-\mathsf{Z}_\imath}{\mathsf{Z}_\imath}\left(C_{\mathsf{Z}_\imath,W}-D_1(W_{\mathsf{Z}_\imath}(\mathsf{X}_t)\|\,q_{\mathsf{Z}_\imath,W})\right)$$

for all $t\in\{\tau_\imath,\ldots,t_\imath\}$.

Then the non-negativity of the Rényi divergence and the definition of $\mathsf{H}_\imath$ given in (30) imply

$$0 \leq \mathbf{E}_v[\mathsf{V}_\imath|\,\mathcal{F}_{\imath-1}] \leq \mathbf{E}_v\left[\frac{1-\mathsf{Z}_\imath}{\mathsf{Z}_\imath}\left(\ell_\imath C_{\mathsf{Z}_\imath,W}-\mathsf{H}_\imath\right)\middle|\,\mathcal{F}_{\imath-1}\right]. \tag{44}$$

We bound the expression on the right hand side of (44) through a case by case analysis based on the value of $\mathsf{G}_\imath$.

- If $\mathsf{G}_\imath = \frac{\eta}{1-\epsilon}$, then $\mathsf{Z}_\imath \geq \eta$ by construction. On the other hand $\frac{1-\eta}{\eta}C_{\eta,W} = E_{sp}(R_1,W)$ by the hypothesis and $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ is nonincreasing in α by Lemma 3. Thus $\mathbf{E}_v[\mathsf{V}_\imath|\,\mathcal{F}_{\imath-1}] \leq E_{sp}(R_1,W)$ as a result of the non-negativity of $\mathsf{H}_\imath$ established by (31). Since $E_{sp}(R,W)$ is nonincreasing in R by definition we get
$$\mathbf{E}_v\left[\frac{1-\mathsf{Z}_\imath}{\mathsf{Z}_\imath}\left(\ell_\imath C_{\mathsf{Z}_\imath,W}-\mathsf{H}_\imath\right)\middle|\,\mathcal{F}_{\imath-1}\right] \leq \ell_\imath E_{sp}(R-\delta_1,\,W). \tag{45}$$
- If $\mathsf{G}_\imath \neq \frac{\eta}{1-\epsilon}$, then $\mathbf{E}_v[\mathsf{H}_\imath|\,\mathcal{F}_{\imath-1}] = \ell_\imath(R-\delta_1)$ by construction. Thus $\mathsf{H}_\imath \geq 0$—established in (31)—and (26) imply
$$\begin{aligned}\mathbf{E}_v&\left[\frac{1-\mathsf{Z}_\imath}{\mathsf{Z}_\imath}\left(\ell_\imath(R-\delta_1)-\mathsf{H}_\imath\right)\middle|\,\mathcal{F}_{\imath-1}\right]\\ &\leq \frac{1-(1-\epsilon)\mathsf{G}_\imath}{(1-\epsilon)\mathsf{G}_\imath}\ell_\imath(R-\delta_1)\\ &\qquad-\frac{(1-\epsilon)(1-\mathsf{G}_\imath)}{\mathsf{G}_\imath+\epsilon(1-\mathsf{G}_\imath)}\mathbf{E}_v[\mathsf{H}_\imath|\,\mathcal{F}_{\imath-1}]\\ &= \frac{\ell_\imath(R-\delta_1)\epsilon}{(\mathsf{G}_\imath-\epsilon\mathsf{G}_\imath)(\mathsf{G}_\imath+\epsilon(1-\mathsf{G}_\imath))}.\end{aligned}$$
On the other hand $\mathsf{G}_\imath \geq \frac{\phi-\epsilon}{1-\epsilon}$ by (34), $\epsilon\leq\frac{\phi}{2}$ by hypothesis and $\frac{1-\mathsf{Z}_\imath}{\mathsf{Z}_\imath}(C_{\mathsf{Z}_\imath,W}-(R-\delta_1)) \leq E_{sp}(R-\delta_1,\,W)$ by the definition of $E_{sp}(R,\,W)$ given in Definition 4. Thus
$$\mathbf{E}_v\left[\frac{1-\mathsf{Z}_\imath}{\mathsf{Z}_\imath}\left(\ell_\imath C_{\mathsf{Z}_\imath,W}-\mathsf{H}_\imath\right)\middle|\,\mathcal{F}_{\imath-1}\right] \leq \ell_\imath E_{sp}(R-\delta_1,\,W) + \ell_\imath\frac{2R\epsilon}{\phi^2}. \tag{46}$$

Using (44), (45), and (46) we get

$$0\leq\mathbf{E}_v[\mathsf{V}_\imath|\,\mathcal{F}_{\imath-1}]\leq\ell_\imath E_{sp}(R-\delta_1,\,W)+\ell_\imath\frac{2R\epsilon}{\phi^2} \tag{47}$$

for all $\imath\in\{1,\ldots,\kappa\}$

The analysis for bounding the conditional second moments of $\mathsf{V}_\imath$'s is analogous to the one for bounding the conditional second moments of $\mathsf{Q}_\imath$'s. We invoke $\mathsf{Z}_\imath \geq \phi-\epsilon$ instead of $\mathsf{Z}_\imath \leq \eta+\epsilon$.

$$\mathbf{E}_v\left[(\mathsf{V}_\imath)^2\middle|\,\mathcal{F}_{\imath-1}\right] \leq 16\frac{(2+C_{1/2,W})^2n^2}{(\phi)^2\kappa^2}. \tag{48}$$

Applying Lemma 9, for $a_\imath = \ell_\imath(E_{sp}(R-\delta_1,\,W)+\frac{2R\epsilon}{\phi^2})$ to the stochastic sequence $(\mathsf{V}_1,\mathcal{F}_1),\ldots,(\mathsf{V}_\kappa,\mathcal{F}_\kappa)$ via (47) we get

$$\mathbf{P}_v\left(\mathsf{V}\leq n(E_{sp}(R-\delta_1,\,W)+\frac{2R\epsilon}{\phi^2})+\gamma\right) \geq 1-\frac{\sum_{\imath=1}^{\kappa}\mathbf{E}_v\left[(\mathsf{V}_\imath)^2\right]}{\gamma^2},$$

[5]Note that $\mathcal{F}_\imath$'s are not defined as σ-algebras on Ω and hence they are not sub-σ-algebras of $\mathcal{F}$. Nevertheless, for each $\mathcal{F}_\imath$ there is a corresponding $\widetilde{\mathcal{F}}_\imath \subset \mathcal{F}$ that uniquely determines $\mathcal{F}_\imath$ and that is uniquely determined by $\mathcal{F}_\imath$. When applying Lemma 9 we are in fact considering $(\mathsf{Q}_1,\widetilde{\mathcal{F}}_1),\ldots,(\mathsf{Q}_\kappa,\widetilde{\mathcal{F}}_\kappa)$ rather than $(\mathsf{Q}_1,\mathcal{F}_1),\ldots,(\mathsf{Q}_\kappa,\mathcal{F}_\kappa)$.

where V is defined as

$$\mathsf{V} \triangleq \sum\nolimits_{\imath=1}^{\kappa} \mathsf{V}_{\imath}. \tag{49}$$

Setting $\gamma = 8\frac{(2+C_{1/2,W})n}{\phi\sqrt{\kappa}}$ and invoking (25) and (48) we get

$$\mathbf{P}_v(\mathcal{E}_v) \geq \frac{3}{4}, \tag{50}$$

where $\mathcal{E}_v$ is defined as

$$\mathcal{E}_v \triangleq \left\{\omega \in \Omega : \mathsf{V}(\omega) \leq n(E_{sp}(R-\delta_1, W)+\delta_2) + \ln\frac{1}{4n^{\kappa}}\right\}. \tag{51}$$

The conditional distribution of $\mathbf{P}_v$, and $\mathbf{P}$ for $\mathsf{Z}_\imath$'s given $\mathcal{F}_{\imath-1}$ are identical for all $\imath \in \{1, \ldots, \kappa\}$ because of (26). Thus $\mathsf{V}(\omega) = \ln\frac{\mathrm{d}\mathbf{P}_v}{\mathrm{d}\mathbf{P}}(\omega)$ and consequently

$$\mathbf{P}(\mathcal{B} \cap \{\mathsf{V} \leq \lambda\}) \geq e^{-\lambda}\mathbf{P}_v(\mathcal{B} \cap \{\mathsf{V} \leq \lambda\}) \tag{52}$$

for any $\mathcal{B} \in \mathcal{F}$ and $\lambda \in \mathbb{R}$.

As a result of (40) and (50) we have

$$\mathbf{P}_v(\mathcal{E}_q \cap \mathcal{E}_v) \geq \frac{1}{2}, \tag{53}$$

where $\mathcal{E}_q$ and $\mathcal{E}_v$ are defined in (41) and (51), respectively.

Remark 4: If we could show $\mathbf{P}_q\left(\mathsf{M} \neq \widehat{\mathsf{M}}\right) \approx e^{-nR}$, then we would use (42), (52), and (53), to bound the error probability under $\mathbf{P}$—i.e., to bound $P_{\mathbf{e}}^{av}$—from below. However, the distribution of Y_1^n depends on M not only under $\mathbf{P}$ and $\mathbf{P}_v$ but also under $\mathbf{P}_q$ because of Z's. We cope with this issue using a pigeon hole argument.

E. A Change of Measure Argument Together With a Pigeon Hole Argument

Let us consider the random variables $\mathsf{G}_1, \ldots, \mathsf{G}_\kappa$. Let us divide the interval $(0, 1]$ into n intervals of length $1/n$. Thus for each $\imath$ in $\{1, \ldots, \kappa\}$, the value of the random variable $\mathsf{G}_\imath$ will be in only one of the n intervals for each sample point $\omega \in \Omega$. Thus we get n^κ disjoint κ-cubes whose union is $(0, 1]^\kappa$ for the vector G_1^κ. For each κ-cube ζ, let us define the event $\mathcal{E}_\zeta \in \mathcal{F}$ as

$$\mathcal{E}_\zeta \triangleq \{\omega \in \Omega : \mathsf{G}_1^\kappa(\omega) \in \zeta\}.$$

As a result of (53) there exists at least one κ-cube ζ^* satisfying

$$\mathbf{P}_v(\mathcal{E}_q \cap \mathcal{E}_v \cap \mathcal{E}_{\zeta^*}) \geq \frac{1}{2n^\kappa}. \tag{54}$$

Let us assume without loss of generality that

$$\zeta^* = \left(\beta_1 - \frac{\beta_1}{n}, \beta_1 + \frac{1-\beta_1}{n}\right] \times \cdots \times \left(\beta_\kappa - \frac{\beta_\kappa}{n}, \beta_\kappa + \frac{1-\beta_\kappa}{n}\right]$$

for some $\beta_1, \ldots, \beta_\kappa$. Let us define the probability measure $\mathbf{P}_u$ on $(\Omega, \mathcal{F})$ by setting its marginal on $\mathcal{M}$ to the uniform distribution and defining its conditional distributions as follows:

$$\mathbf{P}_u\left(\mathcal{A} \middle| m, z_1^{\imath-1}, y_1^{t_\imath-1}\right) = \frac{1}{\tilde{\epsilon}}\int_{(1-\tilde{\epsilon})\beta_\imath}^{\beta_\imath+\tilde{\epsilon}(1-\beta_\imath)} \mathbb{1}_{\mathcal{A}}(z)\mathrm{d}z \tag{55}$$

for all $\mathcal{A} \in \mathcal{B}(\mathcal{Z}_\imath)$, where $\tilde{\epsilon} = \epsilon + \frac{1-\epsilon}{n}$ and

$$\mathbf{P}_u\left(y_{\tau_\imath}^{t_\imath} \middle| m, z_1^\imath, y_1^{t_\imath-1}\right) = \prod\nolimits_{t=\tau_\imath}^{t_\imath} q_{z_\imath,W}(y_t) \tag{56}$$

for all $y_{\tau_\imath}^{t_\imath} \in \mathcal{Y}_{\tau_\imath}^{t_\imath}$.

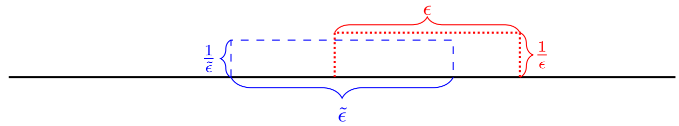

Fig. 2. A representation of the conditional probability density functions of $\mathsf{Z}_\imath$ given $\mathcal{F}_{\imath-1}$ under $\mathbf{P}_u$ and $\mathbf{P}_q$, which are described in (55) and (26). For all realizations of $\mathcal{F}_{\imath-1}$, the conditional probability density function of $\mathsf{Z}_\imath$ under $\mathbf{P}_u$ is the same: it is equal to $1/\tilde{\epsilon}$ on an interval of length $\tilde{\epsilon}$ and zero elsewhere. We represent it with a dashed line in the above figure. For all realizations of $\mathcal{F}_{\imath-1}$, the conditional probability density function of $\mathsf{Z}_\imath$ under $\mathbf{P}_q$ is equal to $1/\epsilon$ on some interval of length ϵ and zero elsewhere, as well. However, the starting point of this interval, i.e. $(1-\epsilon)\mathsf{G}_\imath$, —and hence the conditional density function $\mathsf{Z}_\imath$ under $\mathbf{P}_q$ itself—depends on the realization of $\mathcal{F}_{\imath-1}$. We represent it with a dotted line in the above figure.

Comparing (55) and (56) describing the conditional distributions of $\mathbf{P}_u$ with (26) and (28) describing the conditional distributions of $\mathbf{P}_q$, we can conclude that

$$\mathbf{P}_q(\mathcal{B} \cap \mathcal{E}_{\zeta^*}) \leq \left(\frac{\tilde{\epsilon}}{\epsilon}\right)^\kappa \mathbf{P}_u(\mathcal{B}) \tag{57}$$

for any $\mathcal{B} \in \mathcal{F}$.

Since the distribution of Y_1^n does not depend on M under $\mathbf{P}_u$, we have

$$\mathbf{P}_u\left(\mathsf{M} = \widehat{\mathsf{M}}\right) \leq \frac{1}{\lceil e^{nR} \rceil}.$$

Invoking (57) for $\mathcal{B} = \mathcal{E}_q \cap \mathcal{E}_v \cap \{\mathsf{M} = \widehat{\mathsf{M}}\}$ we get

$$\mathbf{P}_q\left(\mathcal{E}_q \cap \mathcal{E}_v \cap \mathcal{E}_{\zeta*} \cap \{\mathsf{M} = \widehat{\mathsf{M}}\}\right) \leq \left(\frac{\tilde{\epsilon}}{\epsilon}\right)^\kappa e^{-nR}.$$

If we use (41) and (42) for $\lambda = nR - \ln 4 - \kappa\ln(n + \frac{1}{\epsilon})$ and recall $\tilde{\epsilon} = \epsilon + \frac{1-\epsilon}{n}$ we get

$$\begin{aligned}
&\mathbf{P}_v\left(\mathcal{E}_q \cap \mathcal{E}_v \cap \mathcal{E}_{\zeta*} \cap \{\mathsf{M} = \widehat{\mathsf{M}}\}\right) \\
&\qquad \leq \frac{e^{nR}}{4}\left(n + \frac{1}{\epsilon}\right)^{-\kappa}\left(\frac{\tilde{\epsilon}}{\epsilon}\right)^\kappa e^{-nR} \\
&\qquad = \frac{1}{4}\left(\frac{1}{\epsilon n + 1}\right)^\kappa\left(\frac{\epsilon n + (1-\epsilon)}{n}\right)^\kappa \\
&\qquad \leq \frac{1}{4n^\kappa}.
\end{aligned}$$

Then as a result of (54),

$$\mathbf{P}_v\left(\mathcal{E}_q \cap \mathcal{E}_v \cap \mathcal{E}_{\zeta*} \cap \{\mathsf{M} \neq \widehat{\mathsf{M}}\}\right) \geq \frac{1}{4n^\kappa}.$$

If we use (51) and (52) for $\lambda = n(E_{sp}(R-\delta_1, W)+\delta_2) + \ln\frac{1}{4n^\kappa}$, then we get

$$\mathbf{P}\left(\mathcal{E}_q \cap \mathcal{E}_v \cap \mathcal{E}_{\zeta*} \cap \{\mathsf{M} \neq \widehat{\mathsf{M}}\}\right) \geq e^{-n[E_{sp}(R-\delta_1, W)+\delta_2]}.$$

Then (22) holds because $P_{\mathbf{e}}^{av} = \mathbf{P}\left(\mathsf{M} \neq \widehat{\mathsf{M}}\right)$.

F. Proof of Theorem 1

If $\epsilon_n = \frac{\phi\wedge(1-\eta)}{n}$ and $\kappa_n = \lfloor n^{2/3} \rfloor$, then there exists an n_0 for which δ_1 defined in (24) and δ_2 defined in (25) satisfy $\delta_1 \vee \delta_2 \leq \frac{2\ln n}{n^{1/3}}$ for all $n \geq n_0$. Then for any $n \geq n_0$, the hypotheses of Lemma 10 is satisfied by any code satisfying the hypotheses of Theorem 1 and Theorem 1 follows from Lemma 10.

IV. Discussion

We have proved both the non-asymptotic SPB given in Lemma 10 and the asymptotic SPB given in Theorem 1 for codes on DSPC with feedback in order keep the analysis as simple as possible. Nevertheless, the proofs work, essentially, as is for codes on finite output set stationary product channels with feedback, as well. Augustin, on the other hand, stated his asymptotic result [14, Thm. 41.7] for codes on finite input set stationary product channels with feedback.

In a general stationary product channel with feedback, the stochastic matrix $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ is replaced by a transition probability $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, see [11, Definition 8]. In order to generalize Lemma 10 to stationary product channels with feedback, we first need to prove Lemma 5. That can be done rather easily by assuming

$$\lim\nolimits_{\alpha\uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha,W} = 0. \tag{58}$$

The challenge lies in the construction of probability measures $\mathbf{P}$, $\mathbf{P}_v$, and $\mathbf{P}_q$ and in determining the functions $g_1, \ldots, g_\kappa$: we need to show that expressions given in (26), (27), (28), (29) define Borel functions for an appropriate choice of the functions $g_1, \ldots, g_\kappa$ and that the same choice ensures (33), (34), and (35). The other parts of the proof of Lemma 10 and the proof of Theorem 1 work as is. Augustin has asserted in [14, Corollary 41.9] that his proof sketch works for codes on stationary product channels with feedback whose component channel W satisfies[6] (58).

The SPBs are customarily stated for the list decoding, e.g. [8, (1.4) and Thm. 2]; however, we have confined our discussion to the case without the list decoding for the sake of simplicity. Nevertheless, both Lemma 10 and Theorem 1 can be extended to the list decoding case in a straightforward way.

Recently, we have proposed another proof for the SPB for codes on DSPCs with feedback [11, Thm. 3] and generalized it to codes on (possibly non-stationary) DPCs with feedback [11, Thm. 4]. It seems analogous generalizations are possible for Theorem 1 and Lemma 10 under similar hypotheses. A natural next step would be considering codes on the cost constrained stationary memoryless channels with feedback. Under certain hypothesis, it is possible to establish the SPB using the proof technique applied here, but we are not aware of a general proof that will work for all cost constrained stationary discrete memoryless channels with feedback.

Acknowledgment

The author would like to thank Fatma Nakiboğlu and Mehmet Nakiboğlu for their hospitality; this paper simply would not have been possible without it. He completed his initial study confirming the soundness of Augustin's method at UC Berkeley in 2012 fall while he was working as a Post-Doctoral Researcher. He would like to thank Imre Csiszár for pointing out Agustin's work at Austin in 2010 ISIT, Harikrishna R. Palaiyanur for sending him Augustin's manuscript [14], and the reviewers for their comments and suggestions on the manuscript.

[6] [14, Corollary 41.9] assumes the conditional weak compactness, which is just another way of assuming (58). The condition given in (58) and being conditionally weakly compact —i.e., having compact closure in the weak topology—are equivalent by [12, Lemma 24-(d)].

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[2] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, 1959.

[3] P. Elias, "Coding for two noisy channels," in *Proc. 3rd London Symp. Inf. Theory*. London, U.K.: Butterworth Scientific, 1955, pp. 61–74.

[4] R. L. Dobrushin, "Asymptotic estimates of the probability of error for transmission of messages over a discrete memoryless communication channel with a symmetric transition probability matrix," *Theory Probab. Appl.*, vol. 7, no. 3, pp. 270–300, 1962.

[5] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. IT-11, no. 1, pp. 3–18, Jan. 1965.

[6] H. Shin and M. Z. Win, "Gallager's exponent for MIMO channels: A reliability-rate tradeoff," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 972–985, Apr. 2009.

[7] J. Zhang, M. Matthaiou, G. K. Karagiannidis, H. Wang, and Z. Tan, "Gallager's exponent analysis of STBC MIMO systems over η-μ and κ-μ fading channels," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1028–1039, Mar. 2013.

[8] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Inf. Control*, vol. 10, no. 1, pp. 65–103, Feb. 1967.

[9] E. A. Haroutunian, "Estimates of the error probability exponent for a semicontinuous memoryless channel," *Problems Inf. Transmiss.*, vol. 4, no. 4, pp. 37–48, 1968.

[10] U. Augustin, "Error estimates for low rate codes," *Zeitschrift Wahrscheinlichkeitstheorie Verwandte Gebiete*, vol. 14, no. 1, pp. 61–88, 1969.

[11] B. Nakiboğlu, "The sphere packing bound via Augustin's method," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 816–840, Feb. 2019. [Online]. Available: https://arxiv.org/abs/1611.06924

[12] B. Nakiboğlu, "The Renyi capacity and center," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 841–860, Feb. 2019. [Online]. Available: https://arxiv.org/abs/1608.02424

[13] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel. II," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 6, pp. 1462–1471, Nov. 1988.

[14] U. Augustin, "Noisy channels," Habilitation thesis, Univ. Erlangen-Nürnberg, Erlangen, Germany, 1978. [Online]. Available: http://bit.ly/2ID8h7m

[15] P. M. Ebert, "Error bounds for parallel communication channels," Res. Lab. Electron. Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 448, 1966. [Online]. Available: http://hdl.handle.net/1721.1/4295

[16] J. S. Richters, "Communication over fading dispersive channels," Res. Lab. Electron., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 464, 1967. [Online]. Available: http://hdl.handle.net/1721.1/4279

[17] B. Nakiboğlu, "The sphere packing bound for memoryless channels," 2018. *arXiv:1804.06372*. [Online]. Available: http://arxiv.org/abs/1804.06372v2

[18] B. Nakiboğlu, "The Augustin capacity and center," 2018, *arXiv:1803.07937*. [Online]. Available: http://arxiv.org/abs/1803.07937v3

[19] R. L. Dobrushin, "An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback," *Problemy Kibernetiki*, no. 8, pp. 161–168, 1962.

[20] E. A. Haroutunian, "Lower bound for error probability in channels with feedback," *Problems Inf. Transmiss.*, vol. 13, no. 2, pp. 36–44, 1977.

[21] A. Y. Sheverdyaev, "Lower bound for error probability in a discrete memoryless channel with feedback," *Problems Inf. Transmiss.*, vol. 18, no. 4, pp. 5–15, 1982.

[22] H. R. Palaiyanur, "The impact of causality on information-theoretic source and channel coding problems," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, Berkeley, CA, USA, May 2011. [Online]. Available: http://escholarship.org/uc/item/1m29g0bp

[23] H. Palaiyanur and A. Sahai, "An upper bound for the block coding error exponent with delayed feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 246–250.
[24] G. Como and B. Nakiboğlu, "Sphere-packing bound for block-codes with feedback and finite memory," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 251–255.
[25] A. N. Shiryaev, *Probability*. New York, NY, USA: Springer-Verlag, 1995.
[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
[27] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
[28] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[29] I. Csiszár and J. Körner, "Feedback does not affect the reliability function of a DMC at rates above capacity (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 28, pp. 92–93, Jan. 1982.
[30] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math., Statist. Probab.*, 1961, pp. 547–561.
[31] R. Sibson, "Information radius," *Z. Wahrscheinlichkeitstheorie Verwandte Gebiete*, vol. 14, no. 2, pp. 149–160, 1969.
[32] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 26–34, Jan. 1995.
[33] T. van Erven and P. Harremoés, "Rényi divergence and Kullback–Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.
[34] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungarica*, vol. 2, nos. 3–4,pp. 299–318, 1967.
[35] G. L. Gilardoni, "On Pinsker's and Vajda's type inequalities for Csiszár's f-Divergences," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5377–5386, Nov. 2010.
[36] W. Rudin, *Principles of Mathematical Analysis*. New York, NY, USA: McGraw-Hill, 1976.
[37] R. B. Ash and C. Doleans-Dade, *Probability and Measure Theory*. New York, NY, USA: Academic, 2000.
[38] V. I. Bogachev, *Measure Theory*. Berlin, Germany: Springer-Verlag, 2007.
[39] R. M. Dudley, *Real Analysis and Probability*, vol. 74. New York, NY, USA: Cambridge Univ. Press, 2002.
[40] B. Nakiboğlu, "The sphere packing bound for DSPCs with feedback a la Augustin," 2018, *arXiv:1806.11531*. [Online]. Available: https://arxiv.org/abs/1806.11531

Barış Nakiboğlu received the B.S. degrees in electrical and electronics engineering and in physics from Middle East Technical University (METU), Ankara, Turkey, in 2002, and the M.S. degree in electrical engineering and computer science and the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2005 and 2011, respectively. He was a Post-Doctoral Scholar with the University of California at Berkeley from 2012 to 2013. He is currently with the Department of Electrical and Electronics Engineering, METU.