The Sphere Packing Bound via Augustin's Method

Barış Nakiboğlu

***Abstract*— A sphere packing bound (SPB) with a prefactor that is polynomial in the block length n is established for codes on a length n product channel $W_{[1,n]}$, assuming that the maximum order 1/2 Rényi capacity among the component channels, i.e. $\max_{t\in[1,n]} C_{1/2,W_t}$, is $O(\ln n)$. The reliability function of the discrete stationary product channels with feedback is bounded from above by the sphere packing exponent. Both results are proved by first establishing a non-asymptotic SPB. The latter result continues to hold under a milder stationarity hypothesis.**

***Index Terms*— Channel coding, error analysis, error probability, feedback communications, block codes.**

I. Introduction

MOST proofs establishing the infeasibility of certain performance for the channel coding problem under fixed rate, fixed error probability, or slowly vanishing error probability hypotheses rely on either a type based expurgation [2], [31], [54] or a distinction of cases based on types [35], [49], [58], [59]. Although similar bounds can, usually, be obtained using the information spectrum approach [34] with greater generality, one has to give up the initial non-asymptotic bound in order to do so. This relative advantage of the method of types [14], [16] over the information spectrum approach [30] emerges from four distinct assumptions: the product structure of the sample space, the product structure of the probability measures, finiteness of the input set, and the stationarity of the channel. The finite input set assumption and the product structure assumptions can be removed and the stationarity assumption can be relaxed if one gives up the concept of type for the concept of typicality. The typicality arguments are, usually, employed for deriving asymptotic results, but they can also be used to obtain non-asymptotic bounds.

Augustin's proof [7] of the sphere packing bound (SPB) stands out in this high level classification of the techniques for deriving infeasibility results for the channel coding problem. It establishes a non-asymptotic bound without assuming the finiteness of the input set or the stationarity of the channel. The main aim of this article is to build an understanding of Augustin's method around the concepts of capacity and center. We believe such an understanding can guide us when we apply Augustin's method to the other information transmission problems. To build such an understanding, we derive the SPBs using Augustin's method in a way that makes the role of the Rényi capacity and center more explicit.

Shannon *et al.* [54, Th. 2] published the first rigorous proof of the SPB for arbitrary discrete stationary product channels[1] (DSPCs) in 1967. Haroutunian [31, Th. 2] published an alternative proof that holds for arbitrary stationary product channels (SPCs) with finite input sets in 1968. Haroutunian [31] expressed the sphere packing exponent in an alternative form, which he proved to be equal to the one in [54]. Augustin [6, Th. 4.7] published yet another proof of the SPB that holds for —possibly non-stationary— product channels with arbitrary (i.e. possibly infinite) input sets, in 1969. Augustin's SPB [6, Th. 4.7a] holds even for product channels with infinite channel capacity. In the same article, Augustin [6, Th. 4.8] also established a SPB with a polynomial prefactor, under a hypothesis that is satisfied by all DSPCs.

The first two proofs of the SPB for product channels, presented in [31] and [54], rely on expurgations based on the empirical distribution, i.e. the type or the composition, of the input codewords; as a result, they are valid only for the SPCs with finite input sets. Hence, even the non-stationary discrete product channels are beyond the reach of the results presented in [31] and [54], unless the channel has certain symmetries or the channel is —at least approximately— periodic. In order to see how the periodicity can be used to overcome non-stationarity, consider the product of a sequence of channels that alternates between two distinct channels at odd and even time instances. The resulting product channel is formally non-stationary; yet it can also be interpreted as a stationary product channel with larger components. Thus results of [31] and [54] are applicable to non-stationary but periodic DPCs, as well. Furthermore, if the channels in the sequence are from a finite set $\mathcal{W}$ of possible component channels and the frequencies of elements of $\mathcal{W}$ are asymptotically stable, then the results of [31] or [54] can be applied through larger component channels and appropriate worst case assumptions.[2] One can obtain the same result by making minor changes in the proofs presented in [31], or in [54], see [19, Sec. V.A] for one such modification for a related problem. In fact, with such changes one can handle infinite $\mathcal{W}$'s under appropriate finite approximability and asymptotic stability assumptions, albeit with crude approximation error terms and through a rather complicated proof.

Manuscript received November 21, 2016; revised November 7, 2017; accepted November 17, 2018. Date of publication November 28, 2018; date of current version January 18, 2019.

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

Communicated by I. Sason, Associate Editor at Large.

Digital Object Identifier 10.1109/TIT.2018.2882547

[1] Recently, Dalai gave an account of the earlier results in [17, Appendix B].

[2] The asymptotic stability of frequencies of elements of $\mathcal{W}$, rather than the periodicity of the channel, suffices because the lack of contiguity for the subcomponents of the component channels is inconsequential for the performance of the codes on product channels.

The stationarity of the channel has been assumed even in the proofs of the SPB tailored for specific noise models, such as the ones for the Poisson channels in [10] and [62]. In their current form, without major changes, neither the approach of Burnashev and Kutoyants in [10] nor Wyner's approach in [62] —relying on discretization — can establish the SPB for a zero dark current Poisson channel whose inputs are intensity functions, i.e. f's, that are bounded as follows:

$$0 \leq f(t) \leq g(t) \quad \forall t \in \Re_{+}$$

where g is a non-periodic function that is integrable on all bounded intervals. On the other hand, this channel satisfies the hypothesis of [6, Th. 4.7b] by [40, eq. (92)] and the SPB for this channel follows from Augustin's general proof for the product channels, provided that g satisfies rather mild conditions.

The Shannon, Gallager, Berlekamp proof and Haroutunian's proof had greater impact on the field than Augustin's proof. Variants of Haroutunian's proof can be found in [14], [16], [18], [33], and [45]. For the DSPCs, Haroutunian's method leads to a SPB with a polynomial prefactor, i.e. a prefactor of the form $e^{-O(\ln n)}$. The prefactor of the Shannon, Gallager, Berlekamp proof in [54] is $e^{-O(\sqrt{n})}$, which is considerably worse. Valembois and Fossorier [60] have improved the prefactor of [54] for moderate block lengths; the asymptotic behavior of the prefactor, however, is still $e^{-O(\sqrt{n})}$. Wiechman and Sason [61] improved the prefactor of [60] for channels with certain symmetries by eliminating the type based expurgation step of the derivation and the resulting contribution to the rate back-off term in the bound. This improvement, however, is inconsequential for the asymptotic behavior of the prefactor in [61], which is $e^{-O(\sqrt{n})}$, as well. Augustin derived a SPB with a polynomial prefactor for certain product channels in [6, Th. 4.8]; however, the prefactors of his general results [6, Th. 4.7], [7, Th. 31.4] are $e^{-O(\sqrt{n})}$. One of our main contributions is establishing the SPB with a polynomial prefactor for a large class of product channels.

Using the list decoding variant of Gallager's bound [23], [28, ex 5.20], one can see that the exponential decay rate of the SPB, i.e. the sphere packing exponent, is tight. But determining the right prefactor for the SPB is still an open problem even for the DSPCs. Altuğ and Wagner [1] considered the DSPCs with positive transition probabilities satisfying certain symmetry conditions [28, p. 94] and established a SBP with a prefactor of the form $n^{-\frac{1+\epsilon}{2\alpha}}$ for any $\epsilon > 0$ for certain α in $(0, 1)$. Their result is tight because later they have proved in [3] that Gallager's bound [27] can be improved to have a prefactor $n^{-\frac{1}{2\alpha}}$, for the aforementioned α, for arbitrary DSPCs. For arbitrary DSPCs, we only have bounds for the constant composition codes that are also due to Altuğ and Wagner [4].

The SPB has been conjectured to hold for the channel codes on DSPCs with feedback. Assuming certain symmetries, Dobrushin [20] proved it to be the case. However, it was challenging to prove the conjecture for arbitrary DSPCs with feedback because of the reliance of the standard proofs on the type based expurgations. In 1977, Haroutunian [32] established a lower bound on the error probability of codes on arbitrary DSPCs with feedback; but the exponent of Haroutunian's bound is equal to the sphere packing exponent only for DSPCs with certain symmetries. Haroutunian points out in [32] that his exponent is strictly larger than the sphere packing exponent even for the stationary binary input binary output channel with the following transition probability matrix

$$\begin{bmatrix} 1/2 & 1/2 \\ 0 & 1 \end{bmatrix}.$$

There are other partial results [11], [46], [48] establishing the SPB for certain families of codes —rather than all codes— on the DSPCs with feedback.

In 1978, Augustin [7, Th. 41.7] presented a proof sketch establishing the SPB for codes on arbitrary DSPCs with feedback. A complete proof following Augustin's sketch can be found in [41]. One of our main contributions is the new derivation of the SPB for codes on DSPCs with feedback. Furthermore, our result holds for non-stationary and non-periodic DPCs with feedback under an appropriate stationarity hypothesis, see Assumption 4 and Theorem 4.

In 1982, Sheverdyaev [55] suggested another proof. Sheverdyaev's proof, however, is supported rather weakly at certain critical points. Palaiyanur's thesis [48, A7] includes an in depth discussion of the subtleties of [55]. It is worth mentioning that Sheverdyaev [55] has two major claims about DSPCs with feedback. Our reservations are for the claim about the SPB. Sheverdyaev proves the claim about the strong converse satisfactorily and demonstrates that the exponential decay rate of the probability of successful transmission is not changed with the availability of the feedback for rates above capacity. Earlier that year, Csiszár and Körner [15] established the same result. The result in question was also reported by Augustin [7, Th. 41.3], as Csiszár and Körner pointed out in [15].

In the rest of this section, we describe our notation, model, and contributions. In §I-A, we describe the notion we use throughout the article. In §I-B, we define the channel coding problem, product channels, stationarity, memorylessness, and product channels with feedback. In §I-C, we provide an overview of the article and our main contributions.

A. The Notation

We denote the set of all reals by $\Re$, positive reals by $\Re_{+}$, non-negative reals by $\Re_{\geq 0}$, and integers by $\mathbb{Z}$. For any $x \in \Re$, $\lfloor x \rfloor$ is the greatest integer less than or equal to x and $\lceil x \rceil$ is the least integer greater than or equal to x. We call $(-\infty, \infty]$ valued functions continuous if they satisfy the topological definition of continuity for the order topology on $(-\infty, \infty]$.

For any 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 that are non-zero only on finitely many members of $\mathcal{Y}$ by $\mathcal{P}(\mathcal{Y})$. We call the set of all y's for which $p(y) > 0$ the support of p and denote it by $\mathtt{supp}(p)$. Let $\mathcal{X}$ be another set; then we denote the set of all functions from $\mathcal{X}$ to $\mathcal{Y}$ by $\mathcal{Y}^{\mathcal{X}}$.

For any measurable space $(\mathcal{Y}, \mathscr{Y})$, we denote the set of all finite signed measures by $\mathcal{M}(\mathscr{Y})$, the set of all non-zero finite measures by $\mathcal{M}^{+}(\mathscr{Y})$, and the set of all probability measures by $\mathcal{P}(\mathscr{Y})$. For any pair of measurable spaces $(\mathcal{X}, \mathscr{X})$ and

$(\mathcal{Y},\mathscr{Y})$, we denote the set of all $(\mathscr{X},\mathscr{Y})$-measurable functions from $\mathcal{X}$ to $\mathcal{Y}$ by $\mathscr{Y}^{\mathscr{X}}$ and the set of all transition probabilities from $(\mathcal{X},\mathscr{X})$ to $(\mathcal{Y},\mathscr{Y})$ by $\mathcal{P}(\mathscr{Y}|\mathscr{X})$. The formal definition of a transition probability is as follows.

Definition 1: Let $(\mathcal{X},\mathscr{X})$ and $(\mathcal{Y},\mathscr{Y})$ be measurable spaces. Then a function $W:\mathscr{Y}\times\mathcal{X}\to[0,1]$ is called a transition probability (stochastic kernel, Markov kernel) from $(\mathcal{X},\mathscr{X})$ to $(\mathcal{Y},\mathscr{Y})$ if it satisfies the following two constraints:

(i) For all $x\in\mathcal{X}$, the function $W(\cdot|x):\mathscr{Y}\to[0,1]$ is a probability measure on $(\mathcal{Y},\mathscr{Y})$.

(ii) For all $\mathcal{E}\in\mathscr{Y}$, the function $W(\mathcal{E}|\cdot):\mathcal{X}\to[0,1]$ is a $(\mathscr{X},\mathcal{B}([0,1]))$-measurable function.

If $\mathscr{X}$ is the power set of $\mathcal{X}$, then the second constraint, i.e. the measurability constraint, is void because it is always satisfied. Hence, the above definition is consistent with the customary use of the term transition probability in information theory, in which $\mathcal{X}$ and $\mathcal{Y}$ are finite sets and $\mathscr{X}$ and $\mathscr{Y}$ are their power sets. Recall that for any transition probability W, $p\circledast W$ defines a joint probability measure with desired properties for all probability measures p on $(\mathcal{X},\mathscr{X})$ by [9, Th. 10.7.2].

A measure μ on $(\mathcal{Y},\mathscr{Y})$ is absolutely continuous with respect to another measures ν on $(\mathcal{Y},\mathscr{Y})$, i.e. $\mu\prec\nu$, iff $\mu(\mathcal{E})=0$ for any $\mathcal{E}\in\mathscr{Y}$ such that $\nu(\mathcal{E})=0$.

We denote the integral of a measurable function f on $(\mathcal{Y},\mathscr{Y})$ with respect to a probability measure $\nu\in\mathcal{P}(\mathscr{Y})$, i.e. the expected value of f under ν, by $\mathbf{E}_{\nu}\left[f\right]$ or $\mathbf{E}_{\nu}\left[f(\mathsf{Y})\right]$. If the integral is on the real line and with respect to the Lebesgue measure, we denote it by $\int f\mathrm{d}y$ or $\int f(y)\mathrm{d}y$, as well.

When discussing the convergence of sequences of functions, we denote the $\nu-$almost everywhere convergence by $\xrightarrow{\nu-a.e.}$, the convergence in measure for ν by $\xrightarrow{\nu}$ and the convergence in variation, i.e. $\mathcal{L}^1(\nu)$ convergence, by $\xrightarrow{\mathcal{L}^1(\nu)}$.

Our notation will be overloaded for certain symbols; but the relations represented by these symbols will be clear from the context. We denote the product of topologies, σ-algebras, and measures by $\otimes$. We denote the Cartesian product of sets by $\times$. We use the short hand $\mathcal{X}_t^n$ for the Cartesian product of sets $\mathcal{X}_t,\ldots,\mathcal{X}_n$ and $\mathscr{Y}_t^n$ for the product of the σ-algebras $\mathscr{Y}_t,\ldots,\mathscr{Y}_n$. We use $|\cdot|$ to denote the absolute value of reals and the size of sets.

The sign $\leq$ stands for the usual less than or equal to relation for reals and the corresponding pointwise inequality for functions. For μ and ν in $\mathcal{M}(\mathscr{Y})$, $\mu\leq\nu$ iff $\mu(\mathcal{E})\leq\nu(\mathcal{E})$ for all $\mathcal{E}\in\mathscr{Y}$.

The minimum of reals x and y is denoted by $x\wedge y$. For the real valued functions f and g, $f\wedge g$ stands for their pointwise minimum. We use the symbol $\vee$ analogously to $\wedge$; but we represent maxima and suprema with it, rather than minima and infima.

B. The Channel Model and Channel Coding Problem

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. Once the transmitter and receiver agree on a strategy, the transmitter is given an element of the message set, i.e. the message. Then the transmitter chooses the channel input, according to the strategy, using the message. The channel input determines the probabilistic behavior of the channel output. The receiver observes the realization of the channel output and then chooses the decoded list based on the channel output, according to the strategy. If the message given to the transmitter is in the decoded list determined by the receiver, then the transmission is successful, else an error is said to occur. Let us proceed with the formal definitions of these concepts.

Definition 2: A *channel* W is a function from *the input set* $\mathcal{X}$ to the set of all probability measures on *the output space* $(\mathcal{Y},\mathscr{Y})$, i.e.

$$W:\mathcal{X}\to\mathcal{P}(\mathscr{Y}).$$

$\mathcal{Y}$ is called the output set and $\mathscr{Y}$ is called the σ-algebra of the output events. A channel W is a *discrete channel* if both $\mathcal{X}$ and $\mathcal{Y}$ are finite sets.

We denote the set of all channels with the input set $\mathcal{X}$ and the output space $(\mathcal{Y},\mathscr{Y})$ by $\mathcal{P}(\mathscr{Y}|\mathcal{X})$. For the purposes of the channel coding problem, Definition 2 suffices. However, while analyzing other information transmission problems — such as the joint source channel coding problem— one needs to introduce a σ-algebra $\mathscr{X}$ on $\mathcal{X}$ and work with the transition probabilities, described in Definition 1. Note that every transition probability is a channel, i.e. $\mathcal{P}(\mathscr{Y}|\mathscr{X})\subset\mathcal{P}(\mathscr{Y}|\mathcal{X})$ for all σ-algebras $\mathscr{X}$. The converse statement holds only for $\mathscr{X}=2^{\mathcal{X}}$.

Definition 2 describes the channel as introduced in the first paragraph of this subsection accurately and it subsumes a diverse collection of channels as special cases. However, it is not an all-encompassing definition because it might not be possible to model the effect of the channel input on the channel output solely by the probabilistic rule of the channel output. The compound channels and the arbitrarily varying channels fall outside of the framework of Definition 2. Those models, however, are beyond the scope of this article.

Definition 3: An (M,L) *channel code* on $W:\mathcal{X}\to\mathcal{P}(\mathscr{Y})$ is an ordered pair (Ψ,Θ) composed of an *encoding function* Ψ and a *decoding function* Θ:

- An *encoding function* is a function from the message set $\mathcal{M}\triangleq\{1,2,\ldots,M\}$ to the input set $\mathcal{X}$.
- A *decoding function* is a measurable function from the output space $(\mathcal{Y},\mathscr{Y})$ to $\widehat{\mathcal{M}}\triangleq\{\mathcal{L}:\mathcal{L}\subset\mathcal{M}\text{ and }|\mathcal{L}|\leq L\}$ with its power set $2^{\widehat{\mathcal{M}}}$ as the σ-algebra.

In an (M,L) channel code, M is called the message set size and L is called the list size.

The channel codes are customarily defined with the tacit assumption that their list size is one and the channel codes with list sizes larger than one are customarily called list codes. We will neither assume the list size of the codes to be one, nor use the term list code; instead we will be explicit about the list sizes of the codes throughout the manuscript.

Definition 4: Given an (M,L) code (Ψ,Θ) on $W:\mathcal{X}\to\mathcal{P}(\mathscr{Y})$, for each $m\in\mathcal{M}$ *the conditional error probability* $P_{\mathbf{e}}^{m}$ is

$$P_{\mathbf{e}}^{m}\triangleq\mathbf{E}_{W(\Psi(m))}\left[\mathbb{1}_{\{m\notin\Theta(\mathsf{Y})\}}\right].$$

The average error probability $P_{\mathbf{e}}^{av}$ is

$$P_{\mathbf{e}}^{av} \triangleq \frac{1}{M} \sum\nolimits_{m \in \mathcal{M}} P_{\mathbf{e}}^{m}.$$

For a channel W, the triplet $(M, L, P_{\mathbf{e}})$ is achievable if there exists an (M, L) channel code with the average error probability less than or equal to $P_{\mathbf{e}}$. Broadly speaking, the point-to-point channel coding problem aims to characterize the achievable $(M, L, P_{\mathbf{e}})$ triplets. The abstract formulation given above is general enough to subsume a diverse collection of point-to-point channel coding problems as special cases. However, it has scant structure to establish achievability and infeasibility results that are provably close to one another. The product structure, discussed in the following, is commonly assumed in order to establish such bounds.

Definition 5: For any $n \in \mathbb{Z}_{+}$ and $W_t : \mathcal{X}_t \to \mathcal{P}(\mathcal{Y}_t)$ for t in $\{1, \ldots, n\}$, the *length n product channel* $W_{[1,n]} : \mathcal{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ is defined via the following relation:

$$W_{[1,n]}(x_1^n) = \bigotimes\nolimits_{t=1}^{n} W_t(x_t) \qquad \forall x_1^n \in \mathcal{X}_1^n.$$

A product channel is *stationary* iff all W_t's are identical.

Definition 6: A channel $U : \mathcal{Z} \to \mathcal{P}(\mathcal{Y}_1^n)$ is a *memoryless channel*, if there exits a product channel $W_{[1,n]} : \mathcal{X}_1^n \to \mathcal{P}(\mathcal{Y}_1^n)$ satisfying $U(z) = W(z)$ for all $z \in \mathcal{Z}$ and $\mathcal{Z} \subset \mathcal{X}_1^n$.

The preceding definition is consistent with the definition of the memorylessness used by Cover and Thomas [12, p. 184]: "The channel is said to be memoryless if the probability distribution of the output depends only on the input at that time and is conditionally independent of previous channel inputs or outputs." The same property is asserted by Gallager [28, (4.2.1)] and Csiszár and Körner [16, p. 84] while describing the memorylessness. However, the authors of these classic texts and the information theory community at large use the term "the discrete memoryless channel (DMC)" to describe channels that satisfy much more than mere discreteness and memorylessness. In particular, customarily the term "the DMC" stands for the DSPC described in the following paragraph.

For any discrete channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, n 'independent' uses of it —denoted by $W_{[1,n]}$— is not only a memoryless channel, but also a stationary product channel according to Definitions 5 and 6. Thus we call these channels discrete stationary product channels (DSPCs). If the discrete channels at each time instance are not necessarily the same, i.e. if W_t can be different for different values of t, then we call $W_{[1,n]}$ a discrete product channel (DPC). Furthermore, any $U : \mathcal{Z} \to \mathcal{P}(\mathcal{Y}_1^n)$ satisfying $U(z) = W_{[1,n]}(z)$ for all z in $\mathcal{Z}$ is called a discrete memoryless channel (DMC). A commonly considered family of DMCs is the one defined via cost constraints.

Definitions 5 and 6 can be applied to the Poisson channels. For a duration T Poisson channel the input set $\mathcal{F}^T$ is the set of all integrable functions of the form $f : (0, T] \to \Re_{\geq 0}$. The output set is the set of all possible sample paths for the arrival process, i.e. the set of all nondecreasing, right-continuous, integer valued functions on $(0, T]$. The σ-algebra of observable events is the Borel σ-algebra for the Skorokhod metric on the output set and $\Lambda^T(f)$ is the Poisson point process with deterministic intensity function f for all $f \in \mathcal{F}^T$. For any duration $T \in \Re_{+}$, intensity levels $0 \leq a \leq \varrho \leq b \leq \infty$, and integrable intensity function g satisfying $g(t) \geq a$ for all $t \in (0, T]$, the Poisson channels $\Lambda^{T,a,b,\varrho}$, $\Lambda^{T,a,b,\leq\varrho}$, $\Lambda^{T,a,b,\geq\varrho}$, $\Lambda^{T,a,b}$, and $\Lambda^{T,a,g(\cdot)}$ —which are also described in [40, Sec. V-C]— are obtained by curtailing the input set $\mathcal{F}^T$ of the Poisson channel Λ^T as follows:

$$\mathcal{F}^{T,a,b,\varrho} \triangleq \{f \in \mathcal{F}^T : a \leq f \leq b \text{ and } \textstyle\int_0^T f \mathrm{d}t = T\varrho\}, \tag{1a}$$
$$\mathcal{F}^{T,a,b,\leq\varrho} \triangleq \cup_{\gamma \in [a,\varrho]} \mathcal{F}^{T,a,b,\gamma}, \tag{1b}$$
$$\mathcal{F}^{T,a,b,\geq\varrho} \triangleq \cup_{\gamma \in [\varrho,b]} \mathcal{F}^{T,a,b,\gamma}, \tag{1c}$$
$$\mathcal{F}^{T,a,b} \triangleq \cup_{\gamma \in [a,b]} \mathcal{F}^{T,a,b,\gamma}, \tag{1d}$$
$$\mathcal{F}^{T,a,g(\cdot)} \triangleq \{f \in \mathcal{F}^T : a \leq f \leq g\}. \tag{1e}$$

For any $T \in \Re_{+}$ and $n \in \mathbb{Z}_{+}$, the Poisson channel $\Lambda^{T,a,b}$ is a length n stationary product channel (SPC), in particular $\Lambda^{T,a,b} = W_{[1,n]}$ for $W_t = \Lambda^{T/n,a,b}$. Whereas, the Poisson channel $\Lambda^{T,a,g(\cdot)}$ is a length n product channel, which is stationary if $g(\cdot)$ is periodic with period T/n. The Poisson channels $\Lambda^{T,a,b,\varrho}$, $\Lambda^{T,a,b,\leq\varrho}$, and $\Lambda^{T,a,b,\geq\varrho}$ are not product channels, but they are memoryless channels.

In a product channel both the input set and the output space are products. In product channels with feedback, the output space is still a product; but the input set is enlarged by allowing the channel input at any time instance to depend on the previous channel outputs. Thus the channel input at time t is a member of $\mathcal{X}_t^{\mathcal{Y}_1^{t-1}}$ rather than a member of $\mathcal{X}_t$. For channels with uncountable input or output sets, there are additional measurability requirements and this makes the description of the product channels with feedback more nuanced. Thus we will describe the discrete case first.

In a length n discrete product channel with feedback each element $\overrightarrow{x_1^n}$ of the input set $\overrightarrow{\mathcal{X}_1^n}$ is of the form

$$\overrightarrow{x_1^n} = (x_1, \Psi_2, \ldots, \Psi_n)$$

where $x_1 \in \mathcal{X}_1$ and $\Psi_t \in \mathcal{X}_t^{\mathcal{Y}_1^{t-1}}$. We use the symbol Ψ_t, rather than x_t, in order reflect in our notation the fact that Ψ_t is a function from $\mathcal{Y}_1^{t-1}$ to $\mathcal{X}_t$, similar to the encoding functions we have discussed in Definition 3.

Definition 7: For any $n \in \mathbb{Z}_{+}$ 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:

$$W_{\overrightarrow{[1,n]}}(y_1^n | \overrightarrow{x_1^n}) = W_1(y_1|x_1) \prod\nolimits_{t=2}^{n} W_t(y_t | \Psi_t(y_1^{t-1}))$$

for all $\overrightarrow{x_1^n} \in \overrightarrow{\mathcal{X}_1^n}$ and $y_1^n \in \mathcal{Y}_1^n$ where $\overrightarrow{\mathcal{X}_1^n} = \mathcal{X}_1 \times (\times_{t=2}^{n} \mathcal{X}_t^{\mathcal{Y}_1^{t-1}})$.

For describing the product channels with feedback without assuming the discreteness, we use the concept of transition probability described in Definition 1. Let $\Psi : \mathcal{Z} \to \mathcal{X}$ be a $(\mathscr{Z}, \mathscr{X})$-measurable function, W be a transition probability from $(\mathcal{X}, \mathscr{X})$ to $(\mathcal{Y}, \mathscr{Y})$, and $W \circ \Psi : \mathscr{Y} \times \mathcal{Z} \to [0, 1]$ be

$$W \circ \Psi(\mathcal{E}|z) \triangleq W(\mathcal{E}|\Psi(z)) \qquad \forall \mathcal{E} \in \mathscr{Y}, z \in \mathcal{Z}.$$

Then $W \circ \Psi$ is a transition probability from $(\mathcal{Z}, \mathscr{Z})$ to $(\mathcal{Y}, \mathscr{Y})$ (i.e. $W \circ \Psi \in \mathcal{P}(\mathscr{Y}|\mathscr{Z})$) as a result of the definitions of the measurability and the transition probabilities. On the other hand, there exists a unique probability measure $p \circledast U$ for any $U \in \mathcal{P}(\mathcal{Y}|\mathcal{Z})$ and $p \in \mathcal{P}(\mathcal{Z})$ by [9, Th. 10.7.2]. Using these two observation we can generalize Definition 7 as follows.

Definition 8: For any channel $W_1 \in \mathcal{P}(\mathcal{Y}_1|\mathcal{X}_1)$ and transition probabilities $W_t \in \mathcal{P}(\mathcal{Y}_t|\mathcal{X}_t)$ for $t \in \{2, \ldots, n\}$, the *length n product channel with feedback* $W_{\overrightarrow{[1,n]}}$ is defined via the following relation:

$$W_{\overrightarrow{[1,n]}}(\overrightarrow{x_1^n}) = W_1(x_1) \circledast (W_2 \circ \Psi_2) \cdots \circledast (W_n \circ \Psi_n)$$

for all $\overrightarrow{x_1^n} \in \overrightarrow{\mathcal{X}_1^n}$ where $\overrightarrow{\mathcal{X}_1^n} = \mathcal{X}_1 \times (\times_{t=1}^{n} \mathcal{X}_t^{\mathcal{Y}_1^{t-1}})$. A product channel with feedback is *stationary* iff all W_t's are identical.

C. Overview and Main Contributions

In §II, we review Rényi's information measures and the sphere packing exponent.

In §III, we derive preliminary results about Augustin's averaging scheme and tilting.

In §IV, we establish an asymptotic SPB with a prefactor that is polynomial in the block length for product channels, without assuming the input sets to be finite or the channels to be stationary. This asymptotic SPB, given in Theorem 2, is derived by using the non-asymptotic SBP for product channels given in Lemma 20, which can be further simplified to (54) for SPCs.

If $\sup_{t \in (0,T]} g(t)$ is $O(\ln T)$ then the Poisson channel $\varLambda^{T,a,g(\cdot)}$ satisfies the hypothesis of Theorem 2. Without major changes, neither Wyner's approach in [62], nor the approach of Burnashev and Kutoyants in [10] can establish the SPB for these channels. To the best of our knowledge, the SPB has not been proved for any non-stationary Poisson channel before —except for [6, Th. 4.7] and [7, Th. 31.4], which imply the SPB for these channels in the way that Theorem 2 does, albeit with inferior prefactors. Augustin's SPBs in [6] and [7] for product channels are compared with our results in §IV-D.

In §V, we establish an asymptotic SPB for DSPCs with feedback, i.e. Theorem 3, by first deriving a non-asymptotic —but parametric— one in Lemma 26. The stationarity hypothesis can be weakened significantly; Theorem 4 establishes the SPB for (possibly non-stationary, non-periodic, and non-symmetric) DPCs with feedback satisfying Assumption 4. Theorem 4 is the first such result to best of our knowledge. Readers who are only interested Theorems 3 and 4 may bypass §IV.

Proofs of Theorems 3 and 4 rely on the averaging and subblock ideas of Augustin [7], Taylor's expansion idea of Sheverdyaev [55], and the auxiliary channel method of Haroutunian [32]. Nevertheless, they are substantially different from the proofs suggested by Augustin [7] and Sheverdyaev [55]. We compare our results with the previous results and discuss possible extensions in §V-D. Lemmas 24 and 25, presenting preliminary results, are new, to the best of our knowledge. Lemma 25 is used to derive SPB for DSPCs with feedback from Haroutunian's bound in §V-E.

In §VI, we briefly discuss the novel observation underlying Augustin's method and generalizations our results to the memoryless channels.

II. General Preliminaries

Our main aim in this section is to introduce the concepts that we use in the rest of the article. We define Rényi's information measures —i.e. the Rényi divergence, information, mean, capacity, radius, and center— and review their properties that are relevant for our purposes in §II-A-§II-C. All of the propositions in this part of the article, except Lemma 12 of §II-C, are either from [24] or from [40]. We define and analyze sphere packing exponent in §II-D.

A. The Rényi Divergence

Definition 9: For any $\alpha \in \Re_+$ and $w, q \in \mathcal{M}^+(\mathcal{Y})$ *the order α Rényi divergence between w and q* is

$$D_\alpha(w\| q) \triangleq \begin{cases} \frac{1}{\alpha-1} \ln \mathbf{E}_\nu \left[\left(\frac{dw}{d\nu}\right)^\alpha \left(\frac{dq}{d\nu}\right)^{1-\alpha} \right] & \alpha \neq 1 \\ \mathbf{E}_\nu \left[\frac{dw}{d\nu} \left(\ln \frac{dw}{d\nu} - \ln \frac{dq}{d\nu} \right) \right] & \alpha = 1 \end{cases}$$

where ν is any probability measure satisfying $w \prec \nu$ and $q \prec \nu$.

The Rényi divergence is usually defined for the probability measures; the inclusion of the finite measures allows us to invoke Lemma 2 given in the following. The propositions derived for the usual definition will suffice for our purposes most of the time. Thus we borrow them from the recent article of van Erven and Harremoës [24]. The equivalence of Definition 9 and the one used by van Erven and Harremoës [24] for probability measures follows from [24, Th. 5].

Lemma 1 [24, Ths. 3 and 7]: For all $w, q \in \mathcal{P}(\mathcal{Y})$, $D_\alpha(w\| q)$ is a nondecreasing and lower semicontinuous function of α on $\Re_+$ that is continuous on $(0, (1 \vee \chi_{w,q})]$ where $\chi_{w,q} \triangleq \sup\{\alpha : D_\alpha(w\| q) < \infty\}$.

Lemma 2: Let w, q, v be non-zero finite measures on $(\mathcal{Y}, \mathcal{Y})$ and α be any order in $\Re_+$.

- *If $v \leq q$, then $D_\alpha(w\| q) \leq D_\alpha(w\| v)$.*
- *If $q = \gamma v$ for a $\gamma \in \Re_+$, then $D_\alpha(w\| q) = D_\alpha(w\| v) - \ln \gamma$.*

Lemma 2 is an immediate consequence of Definition 9.

Let w and q be two probability measures on the measurable space $(\mathcal{Y}, \mathcal{Y})$ and $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{Y}$. Then the identities $w_{|\mathcal{G}}(\mathcal{E}) = w(\mathcal{E})$ for all $\mathcal{E} \in \mathcal{G}$ and $q_{|\mathcal{G}}(\mathcal{E}) = q(\mathcal{E})$ for all $\mathcal{E} \in \mathcal{G}$ uniquely define probability measures $w_{|\mathcal{G}}$ and $q_{|\mathcal{G}}$ on $(\mathcal{Y}, \mathcal{G})$. In the following, we denote $D_\alpha\left(w_{|\mathcal{G}} \| q_{|\mathcal{G}}\right)$ by $D_\alpha^{\mathcal{G}}(w\| q)$.

Lemma 3 [24, Th. 9]: For any order α in $\Re_+$, probability measures w and q on $(\mathcal{Y}, \mathcal{Y})$, and sub-$\sigma$-algebra $\mathcal{G} \subset \mathcal{Y}$

$$D_\alpha(w\| q) \geq D_\alpha^{\mathcal{G}}(w\| q).$$

Lemma 4 [24, Ths. 3 and 31]: For any order α in $\Re_+$ and probability measures w and q on $(\mathcal{Y}, \mathcal{Y})$

$$D_\alpha(w\| q) \geq \frac{1 \wedge \alpha}{2} \|w - q\|^2. \tag{2}$$

For orders in $(0, 1]$, the bound given in (2) is called the Pinsker's inequality. For orders in $(0, 1)$, it is possible to bound $D_\alpha(w\| q)$ from above in terms of $\|w - q\|$ as well, see [57, eq. (24), p. 365]. We will only need the following identity for $\alpha = 1/2$ case, see [57, eq. (21), p. 364],

$$D_{1/2}\left(\mu \| q\right) \leq 2\ln\frac{2}{2-\|\mu-q\|}. \tag{3}$$

Lemma 5 [24, Th. 12]: For any order α in $\Re_+$, the order α Rényi divergence is convex in its second argument for probability measures, i.e.

$$D_\alpha\left(w\| q_\beta\right) \leq \beta D_\alpha\left(w\| q_1\right) + (1-\beta) D_\alpha\left(w\| q_0\right)$$

for all probability measure w, q_0, q_1 in $\mathcal{P}(\mathcal{Y})$ and $\beta \in (0,1)$ where $q_\beta = \beta q_1 + (1-\beta) q_0$.

Lemma 6 [24, Th. 15]: For any α in $\Re_+$, $D_\alpha\left(w\| q\right)$ is a lower semicontinuous function of the pair of probability measures (w,q) in the topology of setwise convergence.

B. The Rényi Information and Mean

Definition 10: For any $\alpha \in \Re_+$, $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and $p \in \mathcal{P}(\mathcal{X})$, *the order α Rényi information for the input distribution p* is

$$I_\alpha\left(p; W\right) \triangleq \begin{cases} \frac{\alpha}{\alpha-1}\ln \mathbf{E}_\nu\left[\left[\sum_x p(x)\left[\frac{\mathrm{d}W(x)}{\mathrm{d}\nu}\right]^\alpha\right]^{1/\alpha}\right] & \alpha \neq 1 \\ \sum_x p(x)\mathbf{E}_\nu\left[\frac{\mathrm{d}W(x)}{\mathrm{d}\nu}\ln\frac{\mathrm{d}W(x)}{\mathrm{d}q_{1,p}}\right] & \alpha = 1. \end{cases} \tag{4}$$

where ν is any probability measure satisfying $q_{1,p} \prec \nu$ for $q_{1,p} \in \mathcal{P}(\mathcal{Y})$ defined as $q_{1,p} \triangleq \sum_x p(x) W(x)$.

We call $q_{1,p}$ the order one Rényi mean for the input distribution p. For other positive real orders, *the order α Rényi mean for the input distribution p*, is defined via its Radon-Nikodym derivative as follows:

$$\frac{\mathrm{d}q_{\alpha,p}}{\mathrm{d}\nu} \triangleq \frac{1}{\kappa}\left(\sum_x p(x)\left(\frac{\mathrm{d}W(x)}{\mathrm{d}\nu}\right)^\alpha\right)^{1/\alpha} \tag{5}$$

where $q_{1,p} \prec \nu$ and $\kappa = \mathbf{E}_\nu\left[\left(\sum_x p(x)\left(\frac{\mathrm{d}W(x)}{\mathrm{d}\nu}\right)^\alpha\right)^{1/\alpha}\right]$.

Remark 1: Gallager's functions $E_0(\rho, p)$ can be written in terms of the Rényi information as follows:

$$E_0(\rho, p) = \rho I_{\frac{1}{1+\rho}}\left(p; W\right) \quad \forall \rho \in (-1, \infty).$$

One can confirm the following identity by substitution

$$D_\alpha\left(p \circledast W \| p \otimes q\right) = D_\alpha\left(p \circledast W \| p \otimes q_{\alpha,p}\right) + D_\alpha\left(q_{\alpha,p} \| q\right) \tag{6}$$

for all $\alpha \in \Re_+$, $p \in \mathcal{P}(\mathcal{X})$, $q \in \mathcal{M}^+(\mathcal{Y})$ where $p \circledast W$ is the probability measure on $2^{\mathrm{supp}(p)} \otimes \mathcal{Y}$ whose marginal distribution on $\mathrm{supp}(p)$ is p and whose conditional distribution is $W(x)$. Using (6) together with Lemma 4 one obtains the following alternative characterization of the Rényi information.

Lemma 7 [40, Lemma 14]: For any $\alpha \in \Re_+$, $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and $p \in \mathcal{P}(\mathcal{X})$

$$I_\alpha\left(p; W\right) = D_\alpha\left(p \circledast W \| p \otimes q_{\alpha,p}\right) \tag{7}$$

$$= \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha\left(p \circledast W \| p \otimes q\right). \tag{8}$$

Remark 2: We defined the Rényi information, mean, capacity, radius, and center in [40] for subsets of $\mathcal{P}(\mathcal{Y})$, rather than functions from some $\mathcal{X}$ to $\mathcal{P}(\mathcal{Y})$. For functions that are one-to-one, these two approaches are describing same quantities with different notation. Thus the propositions we are borrowing from [40] are merely restated in an alternative notation. The functions we consider, however, are not necessarily one-to-one. Nevertheless, one can show easily that each proposition we are borrowing from [40] for subset of $\mathcal{P}(\mathcal{Y})$ implies the corresponding proposition for functions to $\mathcal{P}(\mathcal{Y})$.

C. The Rényi Capacity, Radius, and Center

Definition 11: For any α in $\Re_+$ and $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, *the order α Rényi capacity of W* is

$$C_{\alpha,W} \triangleq \sup_{p \in \mathcal{P}(\mathcal{X})} I_\alpha\left(p; W\right).$$

Remark 3: $E_0(\rho, W) = \rho C_{\frac{1}{1+\rho},W}$ for all ρ in $(-1,\infty)$ as a result of the corresponding expression for $E_0(\rho, p)$.

Lemma 8 [40, Lemma 15-(a,c,e,f)]: Let W be a channel of the form $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$. Then

(a) $C_{\alpha,W}$ is a nondecreasing lower semicontinuous function of α on $\Re_+$.

(b) $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ is a nonincreasing continuous function of α on $(0,1)$ and $C_{\alpha,W}$ is a continuous function of α on $(0,1]$.

(c) If $C_{\eta,W} < \infty$ for an $\eta \in (0,1)$, then $C_{\alpha,W}$ is finite for all $\alpha \in (0,1)$.

(d) If $C_{\eta,W} < \infty$ for an $\eta \in \Re_+$, then $C_{\alpha,W}$ is a nondecreasing continuous function of α on $(0,\eta]$.

Note that, since $C_{\alpha,W}$ is continuous and nondecreasing in α on $(0,1)$ by Lemma 8-(a,b), $C_{\alpha,W}$ has a limit as α converges to zero from the right. We denote this limit by $C_{0^+,W}$:

$$C_{0^+,W} \triangleq \lim_{\alpha \downarrow 0} C_{\alpha,W}. \tag{9}$$

We do not denote this limit by $C_{0,W}$ because $C_{0,W}$ is, customarily, defined as the supremum of $I_0\left(p; W\right)$. For the case when the input set is finite, we know that $C_{0,W} = C_{0^+,W}$, see [40, Lemma 16-(g)]. Unfortunately, we do not have a general result establishing this equality for arbitrary channels.

Note, on the other hand that, Lemma 8-(a,b) implies

$$\frac{\alpha \wedge (1-\alpha)}{1-\alpha}C_{1/2,W} \leq C_{\alpha,W} \leq \frac{\alpha \vee (1-\alpha)}{1-\alpha}C_{1/2,W} \tag{10}$$

for all α in $(0,1)$.

For all positive real orders α, the alternative characterization of the order α Rényi information given in Lemma 7 implies the following alternative expression for the order α Rényi capacity

$$C_{\alpha,W} = \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha\left(p \circledast W \| p \otimes q\right).$$

In the preceding expression, the order of the supremum and infimum can be changed without changing the value of the expression.

Theorem 1 [40, Ths. 1 and 3]: For any $\alpha \in \Re$ and $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$

$$C_{\alpha,W} = \sup_{p \in \mathcal{P}(\mathcal{X})} \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha\left(p \circledast W \| p \otimes q\right) \tag{11}$$

$$= \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{p \in \mathcal{P}(\mathcal{X})} D_\alpha\left(p \circledast W \| p \otimes q\right) \tag{12}$$

$$= \inf_{q \in \mathcal{P}(\mathcal{Y})} \sup_{x \in \mathcal{X}} D_\alpha\left(W(x) \| q\right). \tag{13}$$

If $C_{\alpha,W}<\infty$, then there exists a unique $q_{\alpha,W}\in\mathcal{P}(\mathcal{Y})$, called the order α Rényi center, such that

$$C_{\alpha,W} = \sup_{p\in\mathcal{P}(\mathfrak{X})} D_{\alpha}\left(p\circledast W\| p\otimes q_{\alpha,W}\right) \tag{14}$$
$$= \sup_{x\in\mathfrak{X}} D_{\alpha}\left(W(x)\| q_{\alpha,W}\right). \tag{15}$$

Furthermore, for every countably separated σ-algebra $\mathcal{X}$ of subsets of $\mathfrak{X}$ satisfying $W\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, the suprema over $\mathcal{P}(\mathfrak{X})$ in (11), (12), *and* (14) *can be replaced by suprema over $\mathcal{P}(\mathcal{X})$.*

The right hand side of (13) can be interpreted as a radius; it is, in fact, the definition of the order α Rényi radius:

Definition 12: For any $\alpha\in\Re_{+}$, $W:\mathfrak{X}\to\mathcal{P}(\mathcal{Y})$, and $q\in\mathcal{P}(\mathcal{Y})$ *the order α Rényi radius of W relative to q*, i.e. $S_{\alpha,W}(q)$, and *the order α Rényi radius of W*, i.e. $S_{\alpha,W}$, are

$$S_{\alpha,W}(q) \triangleq \sup_{x\in\mathfrak{X}} D_{\alpha}\left(W(x)\| q\right),$$
$$S_{\alpha,W} \triangleq \inf_{q\in\mathcal{P}(\mathcal{Y})}\sup_{x\in\mathfrak{X}} D_{\alpha}\left(W(x)\| q\right).$$

Hence, (15) allows us to interpret $q_{\alpha,W}$ as a center. This is why $q_{\alpha,W}$ is called the order α Rényi center.

The following bound on $S_{\alpha,W}(q)$ is called the van Erven-Harremoës bound.

Lemma 9 [40, Lemma 19]: If $C_{\alpha,W}<\infty$ for an $\alpha\in\Re_{+}$ and $W:\mathfrak{X}\to\mathcal{P}(\mathcal{Y})$, then

$$C_{\alpha,W}+D_{\alpha}\left(q_{\alpha,W}\middle\| q\right)\leq S_{\alpha,W}(q) \qquad \forall q\in\mathcal{P}(\mathcal{Y}).$$

The van Erven-Harremoës bound can be used to establish the continuity of the Rényi center as a function of the order for the total variation topology on $\mathcal{P}(\mathcal{Y})$.

Lemma 10 [40, Lemma 20]: If $C_{\eta,W}<\infty$ for an $\eta\in\Re_{+}$ and $W:\mathfrak{X}\to\mathcal{P}(\mathcal{Y})$, then

$$D_{\alpha}\left(q_{\alpha,W}\middle\| q_{\phi,W}\right)\leq C_{\phi,W}-C_{\alpha,W}$$

for any α and ϕ satisfying $0<\alpha<\phi\leq\eta$. Furthermore, $q_{\alpha,W}$ is a continuous function of α on $(0,\eta]$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$.

A well known fact about the DSPCs is that their Rényi capacities are additive, see [27, Th. 5], [28, eq. (5.6.59)]. In fact, the additivity of the Rényi capacities holds for arbitrary product channels. Furthermore, whenever it exists, the Rényi center of a product channel is equal to the product of the Rényi centers of its component channels. Lemma 11 states these observations formally.

Lemma 11 [40, Lemma 22]: Any length n product channel $W_{[1,n]}:\mathfrak{X}_{1}^{n}\to\mathcal{P}(\mathcal{Y}_{1}^{n})$ satisfies

$$C_{\alpha,W_{[1,n]}}=\sum\nolimits_{t=1}^{n}C_{\alpha,W_{t}} \qquad \forall\alpha\in\Re_{+}. \tag{16}$$

Furthermore, if $C_{\alpha,W_{[1,n]}}$ is finite for an order $\alpha\in\Re_{+}$, then $q_{\alpha,W_{[1,n]}}=\bigotimes_{t=1}^{n}q_{\alpha,W_{t}}$.

Note that the input set of a product channel is a subset of the input set of the corresponding product channel with feedback. An immediate consequence of this observation is that $C_{\alpha,W_{\overrightarrow{[1,n]}}}\geq C_{\alpha,W_{[1,n]}}$. More interestingly, the reverse inequality $C_{\alpha,W_{\overrightarrow{[1,n]}}}\leq C_{\alpha,W_{[1,n]}}$ is also true.

Lemma 12: Any length n product channel with feedback $W_{\overrightarrow{[1,n]}}:\overrightarrow{\mathcal{X}_{1}^{n}}\to\mathcal{P}(\mathcal{Y}_{1}^{n})$ with countably separated $\sigma-$algebras $\mathcal{X}_{2},\ldots,\mathcal{X}_{n}$ satisfies.

$$C_{\alpha,W_{\overrightarrow{[1,n]}}}=\sum\nolimits_{t=1}^{n}C_{\alpha,W_{t}} \quad \forall\alpha\in\Re_{+}. \tag{17}$$

Furthermore, if $C_{\alpha,W_{\overrightarrow{[1,n]}}}$ is finite for an order $\alpha\in\Re_{+}$, then $q_{\alpha,W_{\overrightarrow{[1,n]}}}=\bigotimes_{t=1}^{n}q_{\alpha,W_{t}}$.

For the case when the component channels are discrete, Lemma 12 has been common knowledge among the researcher working on the error exponents with feedback for some time now. Augustin [7, pp. 304–306] mentions the following equivalent claim without a proof for the case when input sets of W_{t} are finite:

"$e^{\frac{\alpha-1}{\alpha}C_{\alpha,W_{\overrightarrow{[1,n]}}}}q_{\alpha,W_{\overrightarrow{[1,n]}}} = e^{\frac{\alpha-1}{\alpha}C_{\alpha,W_{[1,n]}}}q_{\alpha,W_{[1,n]}}$ whenever $q_{\alpha,W_{[1,n]}}$ is defined."

Proof of Lemma 12: We prove the lemma for $\alpha\in\Re_{+}\setminus\{1\}$ in the following. This implies $C_{\alpha,W_{\overrightarrow{[1,n]}}}=\sum_{t=1}^{n}C_{\alpha,W_{t}}$ for all $\alpha\in\Re_{+}$, because the Rényi capacity is a nondecreasing lower semicontinuous function of the order by Lemma 8-(a). The claim about the Rényi centers follows from the corresponding claim in Lemma 11 and the uniqueness of the Rényi centers, established in Theorem 1.

Recall that $W_{[1,n]}(x_{1}^{n})=W_{\overrightarrow{[1,n]}}(x_{1}^{n})$ for all x_{1}^{n} in $\mathfrak{X}_{1}^{n}$, which is a subset of $\overrightarrow{\mathcal{X}_{1}^{n}}$. Then Lemma 11 implies that $\sum_{t=1}^{n}C_{\alpha,W_{t}}\leq C_{\alpha,W_{\overrightarrow{[1,n]}}}$. Hence, (17) holds if $C_{\alpha,W_{t}}$ is infinite for a $t\in\{1,\ldots,n\}$. Thus, we assume that $C_{\alpha,W_{t}}$ is finite for all t for the rest of the proof. Then for each t, W_{t} has a unique Rényi center $q_{\alpha,W_{t}}$ by Theorem 1.

On the other hand, Definition 8 implies that for every $\overrightarrow{x_{1}^{n}}$ in $\overrightarrow{\mathcal{X}_{1}^{n}}$ there exists an $x_{1}\in\mathfrak{X}_{1}$ and $\Psi_{t}\in\mathcal{X}_{t}^{\mathcal{Y}_{1}^{t-1}}$ for all t in $\{2,\ldots,n\}$ such that

$$W_{\overrightarrow{[1,n]}}(\overrightarrow{x_{1}^{n}})=W_{1}(x_{1})\circledast(W_{2}\circ\Psi_{2})\cdots\circledast(W_{n}\circ\Psi_{n})$$

Note that $W_{t}\circ\Psi_{t}\in\mathcal{P}(\mathcal{Y}_{t}|\mathcal{Y}_{1}^{t-1})$ by definition. Furthermore, Theorem 1 implies for all $y_{1}^{t-1}\in\mathfrak{Y}_{1}^{t-1}$ that

$$D_{\alpha}\left(W_{1}(x_{1})\| q_{\alpha,W_{1}}\right)\leq C_{\alpha,W_{1}}$$
$$D_{\alpha}\left(W_{t}\circ\Psi_{t}(\cdot|y_{1}^{t-1})\middle\| q_{\alpha,W_{t}}\right)\leq C_{\alpha,W_{t}}.$$

Then using [9, Th. 10.7.2] we get

$$\begin{aligned}
&D_{\alpha}\left(W_{\overrightarrow{[1,n]}}(\overrightarrow{x_{1}^{n}})\middle\| q_{n}\right)\\
&\quad\leq D_{\alpha}\left(W_{1}(x_{1})\circledast\cdots\circledast(W_{n-1}\circ\Psi_{n-1})\| q_{n-1}\right)+C_{\alpha,W_{n}}\\
&\quad\leq D_{\alpha}\left(W_{1}(x_{1})\circledast\cdots\circledast(W_{\iota}\circ\Psi_{\iota})\| q_{\iota}\right)+\sum\nolimits_{t=\iota+1}^{n}C_{\alpha,W_{t}}\\
&\quad\leq\sum\nolimits_{t=1}^{n}C_{\alpha,W_{t}}
\end{aligned}$$

where $q_{\iota}=\bigotimes_{t=1}^{\iota}q_{\alpha,W_{t}}$ for $\iota\in\{1,\ldots,n\}$. Hence,

$$\sup_{\overrightarrow{x_{1}^{n}}\in\overrightarrow{\mathcal{X}_{1}^{n}}}D_{\alpha}\left(W_{\overrightarrow{[1,n]}}(\overrightarrow{x_{1}^{n}})\middle\| q_{n}\right)\leq\sum\nolimits_{t=1}^{n}C_{\alpha,W_{t}}.$$

Then $C_{\alpha,W_{\overrightarrow{[1,n]}}}\leq\sum_{t=1}^{n}C_{\alpha,W_{t}}$ by (13). Thus (17) holds and $q_{\alpha,W_{\overrightarrow{[1,n]}}}=\bigotimes_{t=1}^{n}q_{\alpha,W_{t}}$ by Theorem 1. □

In the following we discuss the operational significance of the Rényi capacity in the point to point communication only in terms of the SPB. The implications of Gallager's bound [27] and Arimoto's bound [5] for the operational significance of the Rényi capacity in the point to point communication are discussed in [37, Appendix B].

D. The Sphere Packing Exponent

Definition 13: For any channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and rate $R \in \Re_{\geq 0}$ *the sphere packing exponent* is

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

Lemma 13: For any channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $E_{sp}(R, W)$ is convex and nonincreasing in R on $\Re_{\geq 0}$, finite on $(C_{0^+,W}, \infty)$, and continuous on $[C_{0^+,W}, \infty)$ for $C_{0^+,W}$ is defined in (9). In particular,

$$E_{sp}(R, W) = \begin{cases} \infty & R < C_{0^+,W} \\ \sup\limits_{\alpha\in(0,1)} \frac{1-\alpha}{\alpha}\left(C_{\alpha,W} - R\right) & R = C_{0^+,W} \\ \sup\limits_{\alpha\in[\phi,1)} \frac{1-\alpha}{\alpha}\left(C_{\alpha,W} - R\right) & R = C_{\phi,W} \text{ for a } \phi \in (0,1) \\ 0 & R \geq C_{1,W} \end{cases} \tag{18}$$

Proof of Lemma 13: $E_{sp}(R, W)$ is convex/nonincreasing in R, because $\frac{1-\alpha}{\alpha}(C_{\alpha,W} - R)$ is convex/nonincreasing in R for any $\alpha \in (0, 1)$ and the pointwise supremum of a family of convex/nonincreasing functions is convex/nonincreasing.

Recall that $C_{\alpha,W}$ is a nondecreasing in α by Lemma 8-(a).

- If $C_{0^+,W} = \infty$, then $C_{1/2,W} = \infty$ and $E_{sp}(R, W) = \infty$ for all $R \in \Re_{\geq 0}$. On the other hand $R < C_{0^+,W}$ for all $R \in \Re_{\geq 0}$. Hence (18) holds.
- If $C_{0^+,W} < \infty$ and $C_{0^+,W} = C_{1,W}$, then $E_{sp}(R, W) = \infty$ for all $R < C_{1,W}$ and $E_{sp}(R, W) = 0$ for all $R \geq C_{1,W}$. Thus (18) holds.
- If $C_{0^+,W} < \infty$ and $C_{0^+,W} \neq C_{1,W}$, then $E_{sp}(R, W) = \infty$ for all $R < C_{0^+,W}$. For $R \geq C_{0^+,W}$, the non-negativity of $\frac{1-\alpha}{\alpha}(C_{\alpha,W} - R)$ implies the restrictions given in (18) for different intervals.

$E_{sp}(R, W)$ is continuous on $(C_{0^+,W}, \infty)$ by [21, Th. 6.3.3] because it is finite on $(C_{0^+,W}, \infty)$ by (18) and convex on $\Re_{\geq 0}$. On the other hand $E_{sp}(R, W)$ is lower semicontinuous because it is the pointwise supremum of continuous functions. Thus it is continuous from the right because it is nonincreasing. Thus $E_{sp}(R, W)$ is continuous on $[C_{0^+,W}, \infty)$, as well. □

III. Preliminaries for Augustin's Method

The propositions proved in this section are used in §IV and §V to derive SPBs. In §III-A, we define the average order α Rényi center $q^{\epsilon}_{\alpha,W}$ as the average of the Rényi centers on a specific length ϵ interval around α. Using the convexity and the monotonicity properties of the Rényi divergence, we bound the order α Rényi radius of W relative to $q^{\epsilon}_{\alpha,W}$, i.e. $S_{\alpha,W}(q^{\epsilon}_{\alpha,W})$, from above and call the bound the average Rényi capacity $\widetilde{C}^{\epsilon}_{\alpha,W}$. Then we show that both $\widetilde{C}^{\epsilon}_{\alpha,W}$ and the associated sphere packing exponent $\widetilde{E}^{\epsilon}_{sp}(R, W)$ differ from the corresponding quantities $C_{\alpha,W}$ and $E_{sp}(R, W)$ at most by a factor proportional to ϵ. In §III-B, we consider the tilted probability measure between a probability measure w and a family of probability measures q_α that is continuous in α for the total variation topology on $\mathcal{P}(\mathcal{Y})$. We show that both the tilted probability measure $w^{q_\alpha}_{\alpha}$ and the Rényi divergences $D_\alpha(w\| q_\alpha)$, $D_1\left(w^{q_\alpha}_{\alpha} \| w\right)$, and $D_1\left(w^{q_\alpha}_{\alpha} \| q_\alpha\right)$ are continuous in α on $(0, 1)$.

A. The Augustin's Averaging

If the order ϕ Rényi capacity of a channel is finite for a $\phi \in \Re_+$, then the Rényi centers of the channel form a transition probability from $((0, 1), \mathcal{B}((0, 1)))$ to $(\mathcal{Y}, \mathcal{Y})$. We define the average Rényi center using this transition probability. In order to see why such a transition probability structure exists, first note that $C_{\alpha,W}$ is finite for all $\alpha \in (0, 1)$ by Lemma 8-(a,c) because $C_{\phi,W}$ is finite for a $\phi \in \Re_+$. This implies the existence of a unique order α Rényi center $q_{\alpha,W}$ for each $\alpha \in (0, 1)$ by Theorem 1. Furthermore, $q_{\alpha,W}$ is continuous in α on $(0, 1)$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$, by Lemma 10. As a result, $q_{\cdot,W}(\mathcal{E}) : (0, 1) \to [0, 1]$ is a continuous and hence a $(\mathcal{B}((0, 1)), \mathcal{B}([0, 1]))$-measurable function for any $\mathcal{E} \in \mathcal{Y}$.

Remark 4: The continuity for the topology of setwise convergence is sufficient for ensuring the continuity of $q_{\alpha,W}(\mathcal{E})$ in α for all $\mathcal{E} \in \mathcal{Y}$ and hence for ensuring the existence of the transition probability structure.

Definition 14: For any $\alpha, \epsilon \in (0, 1)$ and $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ satisfying $C_{1/2,W} < \infty$, *the average Rényi center* $q^{\epsilon}_{\alpha,W}$ is the $\mathcal{Y}$ marginal of the probability measure $u_{\alpha,\epsilon} \circledast q_{\cdot,W}$ where $u_{\alpha,\epsilon}$ is the uniform probability distribution on $(\alpha - \epsilon\alpha, \alpha + \epsilon(1-\alpha))$:

$$q^{\epsilon}_{\alpha,W} \triangleq \frac{1}{\epsilon}\int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} q_{\eta,W}\,\mathrm{d}\eta.$$

The order α Rényi radius relative to the order α Rényi center, i.e. $S_{\alpha,W}(q_{\alpha,W})$, is $C_{\alpha,W}$ by Theorem 1. What can we say about $S_{\alpha,W}(q^{\epsilon}_{\alpha,W})$? For channels with certain symmetries such as the ones in [40, Examples 5–8], $q_{\alpha,W}$ is the same probability measure for all α for which it exists. For such channels $S_{\alpha,W}(q^{\epsilon}_{\alpha,W}) = C_{\alpha,W}$ for all $\alpha, \epsilon \in (0, 1)$ because $q^{\epsilon}_{\alpha,W} = q_{\alpha,W}$ for all $\alpha, \epsilon \in (0, 1)$. For certain other channels, such as $\mathcal{W}$ of [40, Example 1], $q_{\alpha,W}$ is same for all α on an interval and $S_{\alpha,W}(q^{\epsilon}_{\alpha,W}) = C_{\alpha,W}$ at least for some α for small enough ϵ. However, we cannot assert the equality of $q_{\alpha,W}$ and $q^{\epsilon}_{\alpha,W}$ in general and $S_{\alpha,W}(q^{\epsilon}_{\alpha,W}) > C_{\alpha,W}$ whenever $q^{\epsilon}_{\alpha,W} \neq q_{\alpha,W}$. In particular,

$$S_{\alpha,W}(q^{\epsilon}_{\alpha,W}) \geq C_{\alpha,W} + D_\alpha\left(q_{\alpha,W} \| q^{\epsilon}_{\alpha,W}\right)$$

by Lemma 9. Lemma 14 bounds $S_{\alpha,W}(q^{\epsilon}_{\alpha,W})$ from above in terms of an integral of the Rényi capacity, which converges to $C_{\alpha,W}$ as ϵ converges to zero for any $\alpha \in (0, 1)$.

Lemma 14: For any $\alpha, \epsilon \in (0, 1)$ and $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ satisfying $C_{1/2,W} < \infty$,

$$\sup_{x\in\mathcal{X}} D_\alpha\left(W(x) \| q^{\epsilon}_{\alpha,W}\right) \leq \widetilde{C}^{\epsilon}_{\alpha,W} \tag{19}$$

$$\leq \frac{C_{1/2,W}}{(1-\alpha)(1-\epsilon)} \tag{20}$$

where the average Rényi capacity $\widetilde{C}^{\epsilon}_{\alpha,W}$ is

$$\widetilde{C}^{\epsilon}_{\alpha,W} \triangleq \frac{1}{\epsilon}\int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \left[1 \vee \left(\frac{\alpha}{1-\alpha}\frac{1-\eta}{\eta}\right)\right] C_{\eta,W}\,\mathrm{d}\eta. \tag{21}$$

Before presenting the proof of Lemma 14, we point out certain properties of $\widetilde{C}^{\epsilon}_{\alpha,W}$. As a result of the continuity of $C_{\alpha,W}$ in α on $(0, 1)$, i.e. Lemma 8-(b), we have

$$\lim_{\epsilon\downarrow 0} \widetilde{C}^{\epsilon}_{\alpha,W} = C_{\alpha,W} \quad \forall \alpha \in (0, 1).$$

In fact, we can bound $\widetilde{C}^{\epsilon}_{\alpha,W}$ from above using the monotonicity of $C_{\alpha,W}$ and $\frac{1-\alpha}{\alpha}C_{\alpha,W}$, i.e. Lemma 8-(a,b), as follows

$$\begin{aligned}\widetilde{C}^{\epsilon}_{\alpha,W} &\leq \frac{C_{\alpha,W}}{\epsilon}\int_{\alpha-\epsilon\alpha}^{\alpha}\frac{1-\alpha(1-\epsilon)}{(1-\alpha)(1-\epsilon)}\mathrm{d}\eta \\ &\quad+\frac{C_{\alpha,W}}{\epsilon}\int_{\alpha}^{\alpha+\epsilon(1-\alpha)}\frac{\alpha+(1-\alpha)\epsilon}{\alpha(1-\epsilon)}\mathrm{d}\eta \\ &\leq C_{\alpha,W}+\frac{\epsilon}{1-\epsilon}\frac{C_{\alpha,W}}{\alpha(1-\alpha)}. \end{aligned} \tag{22}$$

On the other hand $\widetilde{C}^{\epsilon}_{\alpha,W}\geq C_{\alpha,W}$ because $S_{\alpha,W}(q^{\epsilon}_{\alpha,W})\geq C_{\alpha,W}$ by Theorem 1. Thus $C_{\alpha,W}$ can be approximated by $\widetilde{C}^{\epsilon}_{\alpha,W}$ at any α in $(0,1)$. The expression in (22), however, suggests that it might not be possible to do this approximation uniformly on $(0,1)$. In order to show this formally, we bound $\widetilde{C}^{\epsilon}_{\alpha,W}$ from below using the monotonicity of $C_{\alpha,W}$ and $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ as follows,

$$\begin{aligned}\widetilde{C}^{\epsilon}_{\alpha,W} &\geq \frac{1}{\epsilon}\int_{\alpha-\epsilon\alpha}^{\alpha-\frac{\epsilon}{2}\alpha}\frac{\alpha}{1-\alpha}\frac{1-\eta}{\eta}C_{\eta,W}\mathrm{d}\eta \\ &\geq \frac{\alpha}{2-\epsilon}(1+\frac{\alpha\epsilon}{2(1-\alpha)})C_{\alpha-\epsilon\alpha,W}.\end{aligned}$$

Note that this lower bound is true even when $C_{1,W}$ is finite. Thus, $C_{\alpha,W}$ can be approximated by $\widetilde{C}^{\epsilon}_{\alpha,W}$ uniformly only on compact subsets of $(0,1)$, but not on $(0,1)$ in itself.

The additivity of Rényi capacity for product channels, i.e. Lemma 11, implies the additivity of average Rényi capacity for product channels:

$$\widetilde{C}^{\epsilon}_{\alpha,W_{[1,n]}}=\sum\nolimits_{t=1}^{n}\widetilde{C}^{\epsilon}_{\alpha,W_t}. \tag{23}$$

Lemma 11 also states that $q_{\alpha,W_{[1,n]}}=\bigotimes_{t=1}^{n}q_{\alpha,W_t}$. The average Rényi center $q^{\epsilon}_{\alpha,W_{[1,n]}}$, however, does not satisfy such a product structure, in general.

Proof of Lemma 14: The convexity of $D_{\alpha}(w\|q)$ in q, i.e. Lemma 5, and the Jensen's inequality imply

$$D_{\alpha}\left(W(x)\|\, q^{\epsilon}_{\alpha,W}\right)\leq\int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)}\frac{D_{\alpha}\left(W(x)\|\, q_{\eta,W}\right)}{\epsilon}\mathrm{d}\eta. \tag{24}$$

Note that $D_{\alpha}\left(W(x)\|\, q_{\eta,W}\right)=\frac{\alpha}{1-\alpha}D_{1-\alpha}\left(q_{\eta,W}\|\, W(x)\right)$ for any $\alpha\in(0,1)$ by definition and the Rényi divergence is nondecreasing in its order by Lemma 1. Thus

$$\begin{aligned}D_{\alpha}\left(W(x)\|\, q_{\eta,W}\right) &\leq \mathbb{1}_{\{\eta\geq\alpha\}}D_{\eta}\left(W(x)\|\, q_{\eta,W}\right) \\ &\quad+\mathbb{1}_{\{\eta<\alpha\}}\frac{\alpha}{1-\alpha}\frac{1-\eta}{\eta}D_{\eta}\left(W(x)\|\, q_{\eta,W}\right) \\ &=\left(1\vee\frac{\alpha}{1-\alpha}\frac{1-\eta}{\eta}\right)D_{\eta}\left(W(x)\|\, q_{\eta,W}\right).\end{aligned} \tag{25}$$

Recall that $D_{\eta}\left(W(x)\|\, q_{\eta,W}\right)\leq C_{\eta,W}$ for all $x\in\mathcal{X}$ by Theorem 1. Then (19) follows from using (21), (24), and (25).

In order to obtain (20) from (19) recall that $C_{\alpha,W}$ is nondecreasing in α by Lemma 8-(a) and $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ is nonincreasing in α on $(0,1)$ by Lemma 8-(b). Thus we have,

$$C_{\eta,W}\leq\frac{\eta}{1-\eta}C_{1/2,W}\mathbb{1}_{\{\eta>1/2\}}+C_{1/2,W}\mathbb{1}_{\{\eta\leq 1/2\}}\quad\forall\eta\in(0,1).$$

Then using the definition of $\widetilde{C}^{\epsilon}_{\alpha,W}$ given in (21) we get

$$\begin{aligned}\widetilde{C}^{\epsilon}_{\alpha,W} &\leq \frac{C_{1/2,W}}{\epsilon}\int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)}\left(1\vee\frac{\alpha}{1-\alpha}\frac{1-\eta}{\eta}\right)\left(1\vee\frac{\eta}{1-\eta}\right)\mathrm{d}\eta \\ &\leq \frac{C_{1/2,W}}{\epsilon}\int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)}\left(\frac{1}{1-\eta}\vee\frac{\alpha}{1-\alpha}\frac{1}{\eta}\right)\mathrm{d}\eta \\ &\leq \frac{C_{1/2,W}}{\epsilon}\int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)}\frac{1}{(1-\alpha)(1-\epsilon)}\mathrm{d}\eta.\end{aligned}$$

□

Definition 15: For any $\epsilon\in(0,1)$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ with finite $C_{1/2,W}$, and $R\in\Re_{\geq 0}$, *the average sphere packing exponent* $\widetilde{E}^{\epsilon}_{sp}(R,W)$ is

$$\widetilde{E}^{\epsilon}_{sp}(R,W)\triangleq\sup\nolimits_{\alpha\in(0,1)}\frac{1-\alpha}{\alpha}\left(\widetilde{C}^{\epsilon}_{\alpha,W}-R\right). \tag{26}$$

$\widetilde{E}^{\epsilon}_{sp}(R,W)$ is nonincreasing and convex in R on $\Re_{+}$ because it is the pointwise supremum of nonincreasing and convex functions of R. One can show that $\widetilde{C}^{\epsilon}_{\alpha,W}$ is nondecreasing and continuous in α on $(0,1)$ for any $\epsilon\in(0,1)$ using the continuity and monotonicity of $C_{\alpha,W}$ in α on $(0,1)$. Since we do not need this observation in our analysis, we leave its proof to the interested reader. Using the monotonicity of $\widetilde{C}^{\epsilon}_{\alpha,W}$ one can also show that $\widetilde{E}^{\epsilon}_{sp}(R,W)$ is finite and continuous in R on $(\lim_{\alpha\downarrow 0}\widetilde{C}^{\epsilon}_{\alpha,W},\infty)$.

Lemma 15: For any $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ with finite $C_{1/2,W}$, $\phi\in(0,1)$, $R\in[C_{\phi,W},\infty)$, and $\epsilon\in(0,\phi)$,

$$\begin{aligned}0\leq\widetilde{E}^{\epsilon}_{sp}(R,W)-E_{sp}(R,W) &\leq \frac{\epsilon}{1-\epsilon}\frac{R\vee E_{sp}(R,W)}{\phi} &(27)\\ &\leq \frac{\epsilon}{1-\epsilon}\frac{R}{\phi^{2}}. &(28)\end{aligned}$$

Proof of Lemma 15: $C_{\alpha,W}\leq S_{\alpha,W}(q^{\epsilon}_{\alpha,W})\leq\widetilde{C}^{\epsilon}_{\alpha,W}$ by Lemma 14 and Theorem 1. Then as a result the definitions of $E_{sp}(R,W)$ and $\widetilde{E}^{\epsilon}_{sp}(R,W)$ we have

$$E_{sp}(R,W)\leq\widetilde{E}^{\epsilon}_{sp}(R,W)\qquad\forall R\in\Re_{\geq 0}. \tag{29}$$

Let us proceed with bounding $\widetilde{E}^{\epsilon}_{sp}(R,W)$ from above for R's greater than or equal to $C_{\phi,W}$. First note that

$$\begin{aligned}&\int_{\alpha-\epsilon\alpha}^{\alpha}\left(\frac{1-\eta}{\eta}C_{\eta,W}-\frac{1-\alpha}{\alpha}R\right)\mathrm{d}\eta \\ &\quad=\int_{\alpha-\epsilon\alpha}^{\alpha}\frac{1-\eta}{\eta}(C_{\eta,W}-R)\mathrm{d}\eta+\int_{\alpha-\epsilon\alpha}^{\alpha}\frac{\alpha-\eta}{\eta\alpha}R\mathrm{d}\eta \\ &\quad\leq\int_{\alpha-\epsilon\alpha}^{\alpha}\frac{1-\eta}{\eta}(C_{\eta,W}-R)\mathrm{d}\eta+\frac{\epsilon^{2}}{1-\epsilon}R.\end{aligned}$$

Then as a result of the definition of $\widetilde{C}^{\epsilon}_{\alpha,W}$ we have

$$\begin{aligned}\frac{1-\alpha}{\alpha}\left(\widetilde{C}^{\epsilon}_{\alpha,W}-R\right) &\leq \frac{1}{\epsilon}\int_{\alpha-\epsilon\alpha}^{\alpha}\frac{1-\eta}{\eta}(C_{\eta,W}-R)\mathrm{d}\eta+\frac{\epsilon}{1-\epsilon}R \\ &\quad+\frac{1}{\epsilon}\int_{\alpha}^{\alpha+\epsilon(1-\alpha)}\frac{1-\alpha}{\alpha}(C_{\eta,W}-R)\mathrm{d}\eta.\end{aligned} \tag{30}$$

We bound $\widetilde{E}^{\epsilon}_{sp}(R,W)$ by bounding the expression in (30) separately on two intervals for α.

Note that $\frac{1-\eta}{\eta}(C_{\eta,W}-R) \le E_{sp}(R,W)$ for all $\eta \in (0,1)$ and $(1-\alpha)E_{sp}(R,W)+\alpha R \le R \vee E_{sp}(R,W)$ for all $\alpha \in (0,1)$. Thus for $\alpha \in [\phi, 1)$, (30) implies

$$
\begin{aligned}
\frac{1-\alpha}{\alpha}&\left(\widetilde{C}^{\epsilon}_{\alpha,W}-R\right)\\
&\le \frac{1}{\epsilon}\int_{\alpha-\epsilon\alpha}^{\alpha} E_{sp}(R,W)\,\mathrm{d}\eta+\frac{\epsilon}{1-\epsilon}R\\
&\quad+\frac{1}{\epsilon}\frac{1-\alpha}{\alpha}\int_{\alpha}^{\alpha+\epsilon(1-\alpha)}\frac{\eta}{1-\eta}E_{sp}(R,W)\,\mathrm{d}\eta\\
&\le E_{sp}(R,W)+\frac{\epsilon}{1-\epsilon}\frac{1-\alpha}{\alpha}E_{sp}(R,W)+\frac{\epsilon}{1-\epsilon}R\\
&\le E_{sp}(R,W)+\frac{\epsilon}{1-\epsilon}\frac{R\vee E_{sp}(R,W)}{\phi}.
\end{aligned} \tag{31}
$$

On the other hand, $R \ge C_{\phi,w}$ by the hypothesis and the Rényi capacity is nondecreasing in its order by Lemma 8−(a). Thus for $\alpha\in(0,\phi]$, (30) implies

$$
\begin{aligned}
&\frac{1-\alpha}{\alpha}(\tilde{C}^{\epsilon}_{\alpha,W}-R)\\
&\le \frac{\epsilon}{1-\epsilon}R+\frac{1}{\epsilon}\int_{\phi}^{\alpha+\epsilon(1-\alpha)}\frac{(1-\alpha)\eta}{\alpha(1-\eta)}E_{sp}(R,W)\mathrm{d}\eta\mathbb{1}_{\{\alpha\in[\frac{\phi-\epsilon}{1-\epsilon},\phi]\}}\\
&\le \frac{\epsilon}{1-\epsilon}R+\frac{1}{\epsilon}\frac{\alpha(1-\epsilon)+\epsilon}{\alpha(1-\epsilon)}\int_{\phi}^{\alpha(1-\epsilon)+\epsilon}E_{sp}(R,W)\mathrm{d}\eta\mathbb{1}_{\{\alpha\in[\frac{\phi-\epsilon}{1-\epsilon},\phi]\}}\\
&= \frac{\epsilon}{1-\epsilon}R\\
&\quad+\left[\frac{\alpha(1-\epsilon)+2\epsilon-\phi}{\epsilon}-\frac{\phi-\epsilon}{\alpha(1-\epsilon)}\right]E_{sp}(R,W)\mathbb{1}_{\{\alpha\in[\frac{\phi-\epsilon}{1-\epsilon},\phi]\}}\\
&\le \frac{\epsilon}{1-\epsilon}\frac{\phi R+(1-\phi)E_{\mathrm{sp}}(R,W)}{\phi}+(1-\phi)E_{sp}(R,W).
\end{aligned} \tag{32}
$$

Note that (27) follows from (29), (31), and (32). In order to obtain (28) from (27), recall that $C_{\alpha,W}$ is nondecreasing in α by Lemma 8-(a) and $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ is nonincreasing in α on $(0,1)$ by Lemma 8-(b). Then $E_{sp}(R,W) \le \frac{1-\phi}{\phi}R$ and hence $R \vee E_{sp}(R,W) \le R/\phi$ for all $R \ge C_{\phi,W}$ by Lemma 13. □

B. Tilting With a Family of Measures

Definition 16: For any $\alpha \in \Re_{+}$ and w, q in $\mathcal{P}(\mathcal{Y})$ satisfying $D_{\alpha}(w\|q) < \infty$, *the order α tilted probability measure w_{α}^{q}* is

$$
\frac{\mathrm{d}w_{\alpha}^{q}}{\mathrm{d}\nu} \triangleq e^{(1-\alpha)D_{\alpha}(w\|q)}\left(\frac{\mathrm{d}w}{\mathrm{d}\nu}\right)^{\alpha}\left(\frac{\mathrm{d}q}{\mathrm{d}\nu}\right)^{1-\alpha} \tag{33}
$$

where $\nu \in \mathcal{P}(\mathcal{Y})$ satisfies $w \prec \nu$ and $q \prec \nu$.

In many applications, the tilted probability measure between two fixed probability measures is of interest for orders in $(0,1)$. In our analysis we need to allow one of those probability measures to change with the order, as well. Lemma 16, in the following, considers the tilted probability measure $w_{\alpha}^{q_{\alpha}}$ as a function of the order α for the case when q_{α} is a continuous function of α from $(0,1)$ to $\mathcal{P}(\mathcal{Y})$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$.

Lemma 16: Let q_{α} be a continuous function of α from $(0,1)$ to $\mathcal{P}(\mathcal{Y})$ for the total variation topology and $w \in \mathcal{P}(\mathcal{Y})$ 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 $\mathcal{P}(\mathcal{Y})$ for the total variation topology,

(b) $D_{\alpha}(w\|q_{\alpha})$, $D_{1}\left(w_{\alpha}^{q_{\alpha}}\|w\right)$, and $D_{1}\left(w_{\alpha}^{q_{\alpha}}\|q_{\alpha}\right)$ are continuous functions of α from $(0,1)$ to $\Re_{\ge 0}$.

Remark 5: The continuity of q_{α} in the order α for the total variation topology on $\mathcal{P}(\mathcal{Y})$ does not imply the continuity of the corresponding Radon-Nikodym derivatives $\frac{\mathrm{d}q_{\alpha}}{\mathrm{d}\nu}$ for some reference measure ν, see [40, Remark 2]. Thus Lemma 16 is not a corollary of standard results on the continuity of integrals, such as [9, Corollary 2.8.7].

Lemma 16 does not assume q_{α} to be any particular family of probability measures, such as Rényi centers; q_{α} is unspecified except for its continuity in α and finiteness of $D_{\alpha}(w\|q_{\alpha})$ for all α in $(0,1)$. However, for any channel W with finite $C_{1/2,W}$, the Rényi center $q_{\alpha,W}$ satisfies the hypothesis of Lemma 16 for $w = W(x)$ for all $x \in \mathcal{X}$ because $C_{\alpha,W} < \infty$ for all $\alpha \in (0,1)$ by Lemma 8-(c), $D_{\alpha}\left(W(x)\|q_{\alpha,W}\right) \le C_{\alpha,W}$ by Theorem 1, and $q_{\alpha,W}$ is continuous in α for the total variation topology on $\mathcal{P}(\mathcal{Y})$ by Lemma 10.

Remark 6: We believe $q_{\alpha,W}$ satisfies the monotonicity described in [40, Conjecture 1]. If that is the case, we do not need Lemma 16, we can establish the continuity of $w_{\alpha}^{q_{\alpha,W}}$, $D_{\alpha}\left(w\|q_{\alpha,W}\right)$, $D_{1}\left(w_{\alpha}^{q_{\alpha,W}}\|w\right)$, and $D_{1}\left(w_{\alpha}^{q_{\alpha,W}}\|q_{\alpha,W}\right)$ for $w = W(x)$ for all $x \in \mathcal{X}$ using standard results on the continuity of integrals, such as [9, Cor. 2.8.7].

Proof of Lemma 16: Any function g on $(0,1)$ is continuous iff for every convergent sequence $\alpha_n \to \alpha$ in $(0,1)$, the sequence $g(\alpha_n)$ converges to $g(\alpha)$ by [36, Th. 21.3] because $(0,1)$ is metrizable. Let $\{\alpha_n\}_{n\in\mathbb{Z}_+}$ be a convergent sequence such that $\lim_{n\to\infty}\alpha_n = \alpha$ and ν be

$$
\nu = \frac{w}{4}+\frac{q_{\alpha}}{4}+\frac{1}{2}\sum\nolimits_{n\in\mathbb{Z}_+}\frac{q_{\alpha_n}}{2^n}.
$$

Instead of working with measures as members of $\mathcal{M}(\mathcal{Y})$ for the total variation topology, we work with the corresponding Radon-Nikodym derivatives with respect to ν as members of $\mathcal{L}^1(\nu)$. We can do so because all of the measures we are considering are absolutely continuous with respect to ν and for any sequence $\{\xi_n\}_{n\in\mathbb{Z}_+} \subset \mathcal{L}^1(\nu)$, $\xi_n \xrightarrow{\mathcal{L}^1(\nu)} \xi$ iff the corresponding sequence of measures $\{\xi_n\nu\}_{n\in\mathbb{Z}_+}$ converges to $\xi\nu$ in $\mathcal{M}(\mathcal{Y})$ for the total variation topology.

For any finite signed measure μ such that $\mu\prec\nu$, we denote its Radon-Nikodym derivative with respect to ν by ξ_{μ}:

$$
\xi_{\mu} = \frac{\mathrm{d}\mu}{\mathrm{d}\nu} \qquad \forall \mu\in\mathcal{M}(\mathcal{Y}) \text{ such that } \mu\prec\nu.
$$

We make an exception and denote the Radon-Nikodym derivative of $w_{\alpha}^{q_{\alpha}}$ by ξ_{α} rather than $\xi_{w_{\alpha}^{q_{\alpha}}}$.

(a) For any $\alpha \in (0,1)$ let $\xi_{s_{\alpha}}$ be $\xi_{s_{\alpha}} \triangleq \xi_{w}^{\alpha}\xi_{q_{\alpha}}^{1-\alpha}$. Using the triangle equality we get:

$$
\begin{aligned}
&\left|\xi_{s_{\alpha}}-\xi_{s_{\alpha_n}}\right|\\
&\quad=\left|\xi_{s_{\alpha}}-\xi_{w}^{\alpha_n}\xi_{q_{\alpha_n}}^{1-\alpha_n}\right|\\
&\quad\le\left|\xi_{s_{\alpha}}-\xi_{w}^{\alpha_n}\xi_{q_{\alpha}}^{1-\alpha_n}\right|+\xi_{w}^{\alpha_n}\left|\xi_{q_{\alpha}}^{1-\alpha_n}-\xi_{q_{\alpha_n}}^{1-\alpha_n}\right|.
\end{aligned} \tag{34}
$$

- $\{\xi_{w}^{\alpha_n}\xi_{q_{\alpha}}^{1-\alpha_n}\}_{n\in\mathbb{Z}_+}$ is uniformly integrable because $\mathbf{E}_{\nu}\left[\mathbb{1}_{\{\mathcal{E}\}}\xi_{w}^{\alpha_n}\xi_{q_{\alpha}}^{1-\alpha_n}\right] \le w(\mathcal{E})^{\alpha_n}q_{\alpha}(\mathcal{E})^{1-\alpha_n}$ by the

Hölder's inequality and $w(\mathcal{E})^{\alpha_n} q_\alpha(\mathcal{E})^{1-\alpha_n}$ is bounded above by $w(\mathcal{E}) + q_\alpha(\mathcal{E})$.

- $\xi_w^{\alpha_n} \xi_{q_\alpha}^{1-\alpha_n} \xrightarrow{\nu} \xi_{s_\alpha}$ because almost everywhere convergence implies convergence in measure for finite measures by [9, Th. 2.2.3] and $\xi_w^{\alpha_n} \xi_{q_\alpha}^{1-\alpha_n} \xrightarrow{\nu-a.e.} \xi_{s_\alpha}$ by definition.

Then

$$\xi_w^{\alpha_n} \xi_{q_\alpha}^{1-\alpha_n} \xrightarrow{\mathcal{L}^1(\nu)} \xi_{s_\alpha} \tag{35}$$

by the Lebesgue-Vitali convergence theorem [9, 4.5.4]. Using the derivative test one can confirm that $(z+\tau)^\beta - z^\beta$ is a nonincreasing function of z for any $z \geq 0$, $\tau \geq 0$, and $\beta \in (0,1)$. Thus $(z+\tau)^\beta - z^\beta \leq \tau^\beta$ for any $z \geq 0$, $\tau \geq 0$, and $\beta \in (0,1)$. Then using the Hölder's inequality we get,

$$\begin{aligned}\mathbf{E}_\nu\left[\xi_w^{\alpha_n}\left|\xi_{q_\alpha}^{1-\alpha_n} - \xi_{q_{\alpha_n}}^{1-\alpha_n}\right|\right] &\leq \mathbf{E}_\nu\left[\xi_w^{\alpha_n}\left|\xi_{q_\alpha} - \xi_{q_{\alpha_n}}\right|^{1-\alpha_n}\right] \\ &\leq \mathbf{E}_\nu\left[\xi_w\right]^{\alpha_n} \mathbf{E}_\nu\left[\left|\xi_{q_\alpha} - \xi_{q_{\alpha_n}}\right|\right]^{1-\alpha_n}\end{aligned}$$

Then using $\xi_{q_{\alpha_n}} \xrightarrow{\mathcal{L}^1(\nu)} \xi_{q_\alpha}$, $\alpha_n \to \alpha$, and $\alpha \in (0,1)$ we get

$$\lim_{n\to\infty} \mathbf{E}_\nu\left[\xi_w^{\alpha_n}\left|\xi_{q_\alpha}^{1-\alpha_n} - \xi_{q_{\alpha_n}}^{1-\alpha_n}\right|\right] = 0. \tag{36}$$

(34), (35), and (36) imply $\xi_{s_{\alpha_n}} \xrightarrow{\mathcal{L}^1(\nu)} \xi_{s_\alpha}$. Thus s_α is continuous in α on $(0,1)$ for the total variation topology on $\mathcal{M}^+(\mathcal{Y})$. Then $\|s_\alpha\| = \mathbf{E}_\nu\left[\xi_{s_\alpha}\right]$ is continuous in α, as well. Furthermore, $\|s_\alpha\|$ is positive because $\|s_\alpha\| = e^{(\alpha-1)D_\alpha(w\|q_\alpha)}$ and $D_\alpha(w\| q_\alpha)$ is finite by the hypothesis of the lemma. On the other hand, $\xi_\alpha = \xi_{s_\alpha}/\|s_\alpha\|$ and the triangle inequality implies

$$\left|\xi_\alpha - \xi_{\alpha_n}\right| \leq \frac{1}{\|s_\alpha\|}\left|\xi_{s_\alpha} - \xi_{s_{\alpha_n}}\right| + \left|\frac{\|s_\alpha\| - \|s_{\alpha_n}\|}{\|s_\alpha\|}\right|.$$

Since $\|s_\alpha\|$ is positive for all $\alpha \in (0,1)$, the continuity of $\|s_\alpha\|$ in α and $\xi_{s_{\alpha_n}} \xrightarrow{\mathcal{L}^1(\nu)} \xi_{s_\alpha}$ imply $\xi_{\alpha_n} \xrightarrow{\mathcal{L}^1(\nu)} \xi_\alpha$. Thus $w_\alpha^{q_\alpha}$ is continuous in α on $(0,1)$ for the total variation topology on $\mathcal{P}(\mathcal{Y})$.

(b) $\|s_\alpha\|$ is positive for all $\alpha \in (0,1)$ by the hypothesis because $D_\alpha(w\| q_\alpha) = \frac{\ln\|s_\alpha\|}{\alpha-1}$. Furthermore, $D_\alpha(w\| q_\alpha)$ is continuous in α because product and composition of continuous functions are continuous.

$D_1\left(w_\alpha^{q_\alpha}\middle\| w\right)$ and $D_1\left(w_\alpha^{q_\alpha}\middle\| q_\alpha\right)$ are both lower semicontinuous in α because the Rényi divergence is jointly lower semicontinuous in its arguments for the topology of setwise convergence by Lemma 6 and $w_\alpha^{q_\alpha}$ and q_α are continuous in the topology of setwise convergence.

$D_1\left(w_\alpha^{q_\alpha}\middle\| w\right)$ is upper semicontinuous in α because $D_\alpha(w\| q_\alpha)$ is continuous in α, $D_1\left(w_\alpha^{q_\alpha}\middle\| q_\alpha\right)$ is lower semicontinuous in α, and $D_1\left(w_\alpha^{q_\alpha}\middle\| w\right)$ satisfies

$$D_1\left(w_\alpha^{q_\alpha}\middle\| w\right) = \frac{1-\alpha}{\alpha} D_\alpha(w\| q_\alpha) - \frac{1-\alpha}{\alpha} D_1\left(w_\alpha^{q_\alpha}\middle\| q_\alpha\right).$$

Then $D_1\left(w_\alpha^{q_\alpha}\middle\| w\right)$ is continuous in α because it is both lower semicontinuous and upper semicontinuous in α.

Expressing $D_1\left(w_\alpha^{q_\alpha}\middle\| q_\alpha\right)$ in terms of $D_1\left(w_\alpha^{q_\alpha}\middle\| w\right)$ and following a similar reasoning, we deduce that $D_1\left(w_\alpha^{q_\alpha}\middle\| q_\alpha\right)$ is continuous in α, as well. □

IV. The SPB for Product Channels

Assumption 1: $\{W_t\}_{t\in\mathbb{Z}_+}$ is a sequence of channels such that the maximum $C_{1/2,W_t}$ among the first n W_t's is $O(\ln n)$: there exists $n_0 \in \mathbb{Z}_+$ and $K \in \Re_+$ such that

$$\max_{t\in[1,n]} C_{1/2,W_t} \leq K \ln n \qquad \forall n \geq n_0.$$

Theorem 2: Let $\{W_t\}_{t\in\mathbb{Z}_+}$ be a sequence of channels satisfying Assumption 1, ε be a positive real number, and α_0, α_1 be orders satisfying $0 < \alpha_0 < \alpha_1 < 1$. Then for any sequence of codes on the product channels $\{W_{[1,n]}\}_{n\in\mathbb{Z}_+}$ satisfying

$$C_{\alpha_1,W_{[1,n]}} \geq \ln\frac{M_n}{L_n} \geq C_{\alpha_0,W_{[1,n]}} + \varepsilon(\ln n)^2 \qquad \forall n \geq n_0 \tag{37}$$

there exists a $\tau \in \Re_+$ and an $n_1 \geq n_0$ such that

$$P_{\mathbf{e}}^{av(n)} \geq n^{-\tau} e^{-E_{sp}\left(\ln\frac{M_n}{L_n}, W_{[1,n]}\right)} \qquad \forall n \geq n_1. \tag{38}$$

The main aim of this section is to prove the asymptotic SPB given in Theorem 2; we do so following the program put forward by Augustin [6]. In §IV-A, we bound the moments of certain zero mean random variables related to the tilted probability measures. In §IV-B, we bound the small deviation probability for sums of independent random variables using the Berry-Esseen theorem. In §IV-C, we first derive non-asymptotic, but parametric, SPBs for codes on arbitrary product channels and on certain product channels with feedback; then we prove Theorem 2 using the bound for codes on arbitrary product channels. In §IV-D, we compare our SPBs with the SBPs derived by Augustin [6], [7] for the product channels.

We make a brief digression to discuss the implications of Theorem 2, before starting its proof. Theorem 2 and the list decoding variant of Gallager's bound, i.e. [37, Lemma 29], determine the optimal $P_{\mathbf{e}}^{av(n)}$ up to a polynomial factor for all sequences of product channels satisfying Assumption 1. In order to see why, note that if there exists an $\alpha_0 \in [\frac{1}{1+L_n}, 1]$ satisfying $\ln\frac{M_n}{L_n} \geq C_{\alpha_0,W_{[1,n]}}$, then

$$P_{\mathbf{e}}^{av(n)} \leq e^{\frac{1-\alpha_0}{\alpha_0}} e^{-E_{sp}\left(\ln\frac{M_n}{L_n}, W_{[1,n]}\right)} \tag{39}$$

by [37, Lemma 29] because $E_{sp}(R,W) \leq \frac{1-\alpha}{\alpha} + E_{sp}(R+1,W)$ for any $\alpha \in (0,1)$ satisfying $R \geq C_{\alpha,W}$ by Lemma 13.

If the sequence of channels satisfying Assumption 1 is composed of channels with bounded order $1/2$ Rényi capacity, i.e. if $\sup_{t\in\mathbb{Z}_+} C_{1/2,W_t} \leq K$ for some $K \in \Re_+$, then we can bound τ in Theorem 2 from above, as well. But, our bounds are too crude to recover the right polynomial prefactor.

Assumption 1 holds for all stationary product channels and many non-stationary product channels. As an example consider the Poisson channel $\Lambda^{T,a,g(\cdot)}$ whose input set is described in (1e). The Rényi capacity of $\Lambda^{T,a,g(\cdot)}$ is determined in [40, eq. (92)] to be:

$$C_{\alpha,\Lambda^{T,a,g(\cdot)}}=\int_0^T\left[\left(\alpha\frac{g-a}{g^\alpha-a^\alpha}\right)^{\frac{1}{1-\alpha}}-\frac{\alpha}{\alpha-1}\frac{ag^\alpha-ga^\alpha}{g^\alpha-a^\alpha}\right]\mathrm{d}t$$

Then $C_{\frac{1}{2},\Lambda^{T,a,g(\cdot)}}=\frac{1}{4}\int_0^T(\sqrt{g(t)}-\sqrt{a})^2\mathrm{d}t$ and the Poisson channels $W_{[1,n]}=\Lambda^{n,a,g(\cdot)}$ satisfy Assumption 1 provided that $\sup_{t\in(0,T]}g(t)$ is $O(\ln T)$. Thus Theorem 2 implies the SPB for the Poisson channel $\Lambda^{T,a,g(\cdot)}$ asymptotically, provided that $\sup_{t\in(0,T]}g(t)$ is $O(\ln T)$.

A. Moment Bounds for Tilting

The tilted probability measures arise naturally in the trade off between the exponents of the false alarm and the missed detection probabilities in the binary hypothesis testing problem with independent samples. We use them in the same vein with the help of the following bound.

Lemma 17: Let w and q be two probability measures on the measurable space $(\mathcal{Y},\mathcal{Y})$ such that $D_{1/2}(w\|q)<\infty$. Then

$$\mathbf{E}_{w_\alpha^q}\left[|\xi_\alpha|^\kappa\right]^{1/\kappa}\leq 3^{\frac{1}{\kappa}}\frac{((1-\alpha)D_\alpha(w\|q))\vee\kappa}{\alpha(1-\alpha)} \tag{40}$$

for all $\kappa\in\Re_+$ and $\alpha\in(0,1)$ where w_α^q is the tilted probability measure given in (33) and ξ_α is defined using the Radon-Nikodym derivative of w_{ac}, i.e. the component[3] of w that is absolutely continuous in q, as follows

$$\xi_\alpha\triangleq\ln\frac{\mathrm{d}w_{ac}}{\mathrm{d}q}-\mathbf{E}_{w_\alpha^q}\left[\ln\frac{\mathrm{d}w_{ac}}{\mathrm{d}q}\right].$$

Proof: Note that for any $\gamma>0$,

$$\mathbf{E}_{w_\alpha^q}\left[|\xi_\alpha|^\kappa\right]=\mathbf{E}_{w_\alpha^q}\left[\mathbb{1}_{\{\xi_\alpha>\gamma\}}|\xi_\alpha|^\kappa\right]+\mathbf{E}_{w_\alpha^q}\left[\mathbb{1}_{\{|\xi_\alpha|\leq\gamma\}}|\xi_\alpha|^\kappa\right]+\mathbf{E}_{w_\alpha^q}\left[\mathbb{1}_{\{\xi_\alpha<-\gamma\}}|\xi_\alpha|^\kappa\right]. \tag{41}$$

In the following, we bound the expectations in the preceding expression from above for an arbitrary γ and show that these bounds are not larger than $(\gamma_0)^\kappa$ for $\gamma=\gamma_0$, see (44) and (46), where

$$\gamma_0\triangleq\frac{((1-\alpha)D_\alpha(w\|q))\vee\kappa}{\alpha(1-\alpha)}.$$

This will imply $\mathbf{E}_{w_\alpha^q}\left[|\xi_\alpha|^\kappa\right]\leq 3(\gamma_0)^\kappa$ and thus (40).

Using the identity $\frac{\mathrm{d}w_\alpha^q}{\mathrm{d}w}=e^{(\alpha-1)\xi_\alpha+D_1(w_\alpha^q\|w)}$, we can bound the first expectation in (41) for all $\gamma\geq0$ and $\kappa\geq0$ as follows

$$\begin{aligned}\mathbf{E}_{w_\alpha^q}\left[\mathbb{1}_{\{\xi_\alpha>\gamma\}}|\xi_\alpha|^\kappa\right]&=\mathbf{E}_w\left[\mathbb{1}_{\{\xi_\alpha>\gamma\}}|\xi_\alpha|^\kappa e^{(\alpha-1)\xi_\alpha+D_1(w_\alpha^q\|w)}\right]\\&\leq e^{D_1(w_\alpha^q\|w)}\sup\nolimits_{z>\gamma}e^{-(1-\alpha)z}z^\kappa.\end{aligned} \tag{42}$$

On the other hand, for any $\beta>0$, $\kappa\geq0$, and $\gamma\geq0$ we have

$$\sup\nolimits_{z>\gamma}e^{-\beta z}z^\kappa=\begin{cases}(\frac{\kappa}{e\beta})^\kappa & \gamma\leq\frac{\kappa}{\beta}\\ e^{-\beta\gamma}\gamma^\kappa & \gamma>\frac{\kappa}{\beta}\end{cases}. \tag{43}$$

Using (42) and (43) for $\beta=(1-\alpha)$ and $\gamma=\gamma_0$ and invoking $(1-\alpha)D_\alpha(w\|q)=\alpha D_1(w_\alpha^q\|w)+(1-\alpha)D_1(w_\alpha^q\|q)$ we get

$$\begin{aligned}\mathbf{E}_{w_\alpha^q}\left[\mathbb{1}_{\{\xi_\alpha>\gamma\}}|\xi_\alpha|^\kappa\right]&\leq e^{D_1(w_\alpha^q\|w)-\frac{1-\alpha}{\alpha}D_\alpha(w\|q)}(\gamma_0)^\kappa\\&=e^{-\frac{1-\alpha}{\alpha}D_1(w_\alpha^q\|q)}(\gamma_0)^\kappa.\end{aligned} \tag{44}$$

Using the identity $\frac{\mathrm{d}w_\alpha^q}{\mathrm{d}q}=e^{\alpha\xi_\alpha+D_1(w_\alpha^q\|q)}$, we can bound the third expectation in (41) for all $\gamma\geq0$ and $\kappa\geq0$ as follows

$$\begin{aligned}\mathbf{E}_{w_\alpha^q}\left[\mathbb{1}_{\{\xi_\alpha<-\gamma\}}|\xi_\alpha|^\kappa\right]&=\mathbf{E}_q\left[\mathbb{1}_{\{\xi_\alpha<-\gamma\}}|\xi_\alpha|^\kappa e^{\alpha\xi_\alpha+D_1(w_\alpha^q\|q)}\right]\\&\leq e^{D_1(w_\alpha^q\|q)}\sup\nolimits_{z>\gamma}e^{-\alpha z}z^\kappa.\end{aligned} \tag{45}$$

Using (43) and (45) for $\beta=\alpha$ and $\gamma=\gamma_0$ and invoking $(1-\alpha)D_\alpha(w\|q)=\alpha D_1(w_\alpha^q\|w)+(1-\alpha)D_1(w_\alpha^q\|q)$ we get

$$\begin{aligned}\mathbf{E}_{w_\alpha^q}\left[\mathbb{1}_{\{\xi_\alpha<-\gamma\}}|\xi_\alpha|^\kappa\right]&\leq e^{D_1(w_\alpha^q\|q)-D_\alpha(w\|q)}(\gamma_0)^\kappa\\&=e^{-\frac{\alpha}{1-\alpha}D_1(w_\alpha^q\|w)}(\gamma_0)^\kappa.\end{aligned} \tag{46}$$

□

B. A Corollary of the Berry-Esseen Theorem

In this subsection we derive a lower bound to the probability of having a small deviation from the mean for sums of independent random variables using the Berry-Esseen theorem. Let us start with recalling the Berry-Esseen theorem.

Lemma 18 (Berry-Esseen Theorem [8], [25], [56]): Let $\{\xi_t\}_{t\in\mathbb{Z}_+}$ be independent random variables satisfying

$$\mathbf{E}[\xi_t]=0\quad\forall t \text{ and } g_2\in\Re_+$$

where $g_\kappa=\left(\sum_{t=1}^n\mathbf{E}\left[|\xi_t|^\kappa\right]\right)^{1/\kappa}$. Then there exists an absolute constant $\omega\leq0.5600$ such that

$$\left|\mathbf{P}\left[\sum\nolimits_{t=1}^n\xi_t<\tau g_2\right]-\Phi(\tau)\right|\leq\omega\left(\frac{g_3}{g_2}\right)^3$$

where $\Phi(\tau)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\tau e^{-\frac{z^2}{2}}\mathrm{d}z$.

The following lemma is obtained by applying the Berry-Esseen theorem for appropriately chosen values of τ; thus it is merely a corollary of the Berry-Esseen theorem.

Lemma 19: Let $\{\xi_t\}_{t\in\mathbb{Z}_+}$ be independent zero mean random variables. Then

$$\mathbf{P}\left[\left|\sum\nolimits_{t=1}^n\xi_t\right|<3g_\kappa\right]\geq\frac{1}{2\sqrt{n}}\qquad\forall n\in\mathbb{Z}_+,\kappa\geq3. \tag{47}$$

Augustin [7, Th. 18.2] derived a similar bound; the proof of Lemma 19 is similar to the proof of that bound.

Proof of Lemma 19: If $g_\kappa/g_2\geq\sqrt{2}/3$, then using the Markov inequality we get

$$\begin{aligned}\mathbf{P}\left[\left|\sum\nolimits_{t=1}^n\xi_t\right|\leq3g_\kappa\right]&\geq1-\mathbf{P}\left[\left|\sum\nolimits_{t=1}^n\xi_t\right|>g_2\sqrt{2}\right]\\&\geq\frac{1}{2}.\end{aligned}$$

Thus (47) holds. Hence, we assume that $g_\kappa/g_2<\sqrt{2}/3$ for the rest of the proof. By the Berry-Esseen theorem we have

$$\begin{aligned}&\mathbf{P}\left[\left|\sum\nolimits_{t=1}^n\xi_t\right|\leq3g_\kappa\right]\\&\geq\left(\Phi\left(\frac{3g_\kappa}{g_2}\right)-\omega\left(\frac{g_3}{g_2}\right)^3\right)-\left(\Phi\left(\frac{-3g_\kappa}{g_2}\right)+\omega\left(\frac{g_3}{g_2}\right)^3\right)\\&=2\left[\int_0^{\frac{3g_\kappa}{g_2}}\frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\mathrm{d}z-\omega\left(\frac{g_3}{g_2}\right)^3\right].\end{aligned}$$

[3]Any w has a unique decomposition $w=w_{ac}+w_s$ such that $w_{ac}\prec q$ and $w_s\perp q$ by the Lebesgue decomposition theorem, [21, 5.5.3].

On the other hand $\int_0^\tau e^{-\frac{z^2}{2}}\mathrm{d}z \geq \tau e^{-\tau^2/6}$ by the Jensen's inequality because the exponential function is convex. Thus

$$\mathbf{P}\left[\left|\sum\nolimits_{t=1}^{n}\xi_t\right| \leq 3g_\kappa\right] \geq 2\left[\frac{3g_\kappa/g_2}{\sqrt{2\pi}}e^{-\frac{(3g_\kappa/g_2)^2}{6}} - \omega\left(\frac{g_3}{g_2}\right)^3\right]. \tag{48}$$

Since $\kappa \geq 3$, the Hölder's inequality implies

$$\mathbf{E}\left[\sum\nolimits_{t=1}^{n}|\xi_t|^3\right] \leq \mathbf{E}\left[\sum\nolimits_{t=1}^{n}|\xi_t|^\kappa\right]^{\frac{1}{\kappa-2}}\mathbf{E}\left[\sum\nolimits_{t=1}^{n}|\xi_t|^2\right]^{\frac{\kappa-3}{\kappa-2}}.$$

Then $(\frac{g_3}{g_2})^3 \leq (\frac{g_\kappa}{g_2})^{\frac{\kappa}{\kappa-2}}$. Thus using $\frac{g_\kappa}{g_2} < \frac{\sqrt{2}}{3}$ and (48) we get

$$\begin{aligned}\mathbf{P}\left[\left|\sum\nolimits_{t=1}^{n}\xi_t\right| \leq 3g_\kappa\right] &\geq 2\left[\frac{e^{-1/3}}{\sqrt{2\pi}}3 - 0.56\right]\frac{g_\kappa}{g_2}\\ &\geq 0.595\frac{g_\kappa}{g_2}.\end{aligned} \tag{49}$$

In order to bound g_κ/g_2 we use the Jensen's inequality and the concavity of the functions z^α in z for $\alpha \in (0,1]$.

$$\mathbf{E}\left[\sum\nolimits_{t=1}^{n}\frac{1}{n}|\xi_t|^2\right]^{1/2} \leq \mathbf{E}\left[\sum\nolimits_{t=1}^{n}\frac{1}{n}|\xi_t|^\kappa\right]^{1/\kappa} \qquad \forall\kappa\geq 2.$$

Then $g_\kappa/g_2 \geq n^{\frac{1}{\kappa}-\frac{1}{2}} \geq 1/\sqrt{n}$ and (47) follows from (49). $\square$

C. Non-Asymptotic SPBs for Product Channels

The ultimate aim of this subsection is to establish the asymptotic SPB given in Theorem 2. First we establish a non-asymptotic SPB for product channels, i.e. Lemma 20, using Lemmas 17 and 19, the intermediate value theorem, and pigeon hole arguments. We prove Theorem 2 using Lemma 20 at the end of this subsection. We make a brief digression before that proof and point out three variants of Lemma 20 that are proved without invoking the averaging scheme described in §III-A. Lemma 21 is for product channels satisfying

$$e^{\frac{\alpha-1}{\alpha}C_{\alpha,W_{[1,n]}}}q_{\alpha,W_{[1,n]}} \leq e^{\frac{z-1}{z}C_{z,W_{[1,n]}}}q_{z,W_{[1,n]}} \tag{50}$$

for all $\alpha \leq z$ in $(0,1)$. Lemma 22 is for product channels whose Rényi centers do not change with the order. Lemma 23 establishes the SPB given in Lemma 22 for product channels with feedback, under a stronger hypothesis.

Lemma 20: Let $W_{[1,n]}$ be a product channel for an $n \in \mathbb{Z}_+$, $\phi \in (0,1)$, $\epsilon \in (0, \frac{n}{n+1})$, $\kappa \geq 3$, and γ be

$$\gamma \triangleq \frac{3\sqrt[\kappa]{3}}{1-\epsilon}\left(\sum\nolimits_{t=1}^{n}\left(C_{1/2,W_t}\vee\kappa\right)^\kappa\right)^{1/\kappa}. \tag{51}$$

If M and L are integers such that $\frac{M}{L} > 16\sqrt{n}e^{\widetilde{C}^{\epsilon}_{\phi,W_{[1,n]}}+\frac{\gamma}{1-\phi}}$, then any (M,L) channel code on $W_{[1,n]}$ satisfies

$$P_{\mathbf{e}}^{av} \geq \left(\frac{\epsilon e^{-2\gamma}}{16e^2n^{3/2}}\right)^{1/\phi}e^{-\widetilde{E}^{\epsilon}_{sp}\left(\ln\frac{M}{L},W_{[1,n]}\right)} \tag{52}$$

for $\widetilde{E}^{\epsilon}_{sp}(R,W)$ defined in (26).

We have presented Lemma 20 via (52) in order to emphasize its similarity to the Gallager's bound, i.e. [37, Lemma 29]. However, the expression on the right hand side of (52) converges to zero as ϕ converges to zero because $\frac{\epsilon e^{-2\gamma}}{16e^2n^{3/2}} < 1$. By changing the analysis slightly it is possible to obtain the following alternative bound which does not have that problem:

$$P_{\mathbf{e}}^{av} \geq \frac{\epsilon e^{-2\gamma}}{16n^{3/2}}e^{-\widetilde{E}^{\epsilon}_{sp}(R,W_{[1,n]})} \tag{53}$$

where $R = \ln\frac{M}{L} - 2\gamma - \ln\frac{16e^2n^{3/2}}{\epsilon}$. The bound given in (53) is preferable especially for codes with low rates on channels satisfying $\lim_{R\downarrow 0}E_{sp}\left(R,W_{[1,n]}\right) < \infty$.

We can make $\widetilde{C}^{\epsilon}_{\phi,W_{[1,n]}}$ and $\widetilde{E}^{\epsilon}_{sp}\left(\ln\frac{M}{L},W_{[1,n]}\right)$ as close as we please to $C_{\phi,W_{[1,n]}}$ and $E_{sp}\left(\ln\frac{M}{L},W_{[1,n]}\right)$ by choosing ϵ small enough. But as we decrease ϵ, the first term of the lower bound in (52) also decreases. The appropriate choice of ϵ balances these two effects. The choice of κ influences the constraint on $\ln\frac{M}{L}$ and the lower bound in (52) only through the value of γ. The appropriate choice of κ minimizes the value of γ. The constraint on $\ln\frac{M}{L}$ becomes easier to satisfy as ϕ decreases; however, the smaller values of ϕ also lead to smaller, i.e. worse, lower bounds in (52). Thus for a given $\ln\frac{M}{L}$ we desire to have the greatest possible value for ϕ to have the best bound in (52). As an example consider the stationary case when $W_t = W$ for all t and set $\kappa = \ln n$ and $\epsilon = \frac{1}{n}$. Invoking (22) to bound $\widetilde{C}^{\epsilon}_{\phi,W_{[1,n]}}$ and Lemma 15 to bound $\widetilde{E}^{\epsilon}_{sp}\left(R,W_{[1,n]}\right)$, we get the following: For any $n \geq 10$, if

$$\frac{1}{n}\ln\frac{M}{L} \geq \frac{\ln 16\sqrt{n}}{n} + C_{\phi,W} + \frac{C_{\phi,W}+13.2\phi(C_{1/2,W}\vee\ln n)}{(n-1)\phi(1-\phi)}$$

for a $\phi \in (1/n, 1)$ then any (M,L) channel code on $W_{[1,n]}$ satisfies

$$P_{\mathbf{e}} \geq \left(\frac{(L/M)^{\frac{1}{(n-1)\phi}}}{16e^2(n\vee e^{C_{1/2,W}})^{29}}\right)^{1/\phi}e^{-nE_{sp}\left(\frac{1}{n}\ln\frac{M}{L},W\right)}. \tag{54}$$

Proof of Lemma 20 and (53)*:* Let (Ψ,Θ) be an (M,L) channel code on $W_{[1,n]}$. In order to avoid lengthy expressions we denote $W_{[1,n]}(\Psi(m))$ by w^m and its marginal in $\mathcal{P}(\mathcal{Y}_t)$ by w^m_t. Let us describe $v^m_{\alpha,t} \in \mathcal{P}(\mathcal{Y}_t)$ through its Radon-Nikodym derivative:

$$\frac{\mathrm{d}v^m_{\alpha,t}}{\mathrm{d}\nu} \triangleq e^{(1-\alpha)D_\alpha\left(w^m_t\|q^{\epsilon}_{\alpha,W_t}\right)}\left(\frac{\mathrm{d}w^m_t}{\mathrm{d}\nu}\right)^\alpha\left(\frac{\mathrm{d}q^{\epsilon}_{\alpha,W_t}}{\mathrm{d}\nu}\right)^{1-\alpha}$$

where $q^{\epsilon}_{\alpha,W_t}$ is the average Rényi center of W_t and ν is any probability measure satisfying both $w^m_t \prec \nu$ and $q^{\epsilon}_{\alpha,W_t} \prec \nu$.

Let $\xi^m_{\alpha,t}$ be a random variable defined as

$$\xi^m_{\alpha,t} \triangleq \ln\frac{\mathrm{d}(w^m_t)_{ac}}{\mathrm{d}q^{\epsilon}_{\alpha,W_t}} - \mathbf{E}_{v^m_{\alpha,t}}\left[\ln\frac{\mathrm{d}(w^m_t)_{ac}}{\mathrm{d}q^{\epsilon}_{\alpha,W_t}}\right]$$

where $(w^m_t)_{ac}$ is the component of w^m_t that is absolutely continuous in $q^{\epsilon}_{\alpha,W_t}$. Note that $\xi^m_{\alpha,t}$ can also be written as follows:

$$\begin{aligned}\xi^m_{\alpha,t} &= \frac{1}{\alpha-1}\left(\ln\frac{\mathrm{d}v^m_{\alpha,t}}{\mathrm{d}w^m_t} - D_1\left(v^m_{\alpha,t}\middle\|\,w^m_t\right)\right) &(55)\\ &= \frac{1}{\alpha}\left(\ln\frac{\mathrm{d}v^m_{\alpha,t}}{\mathrm{d}q^{\epsilon}_{\alpha,W_t}} - D_1\left(v^m_{\alpha,t}\middle\|\,q^{\epsilon}_{\alpha,W_t}\right)\right). &(56)\end{aligned}$$

Let q_α and v^m_α be the probability measures defied as

$$\begin{aligned}q_\alpha &\triangleq \bigotimes\nolimits_{t=1}^{n}q^{\epsilon}_{\alpha,W_t},\\ v^m_\alpha &\triangleq \bigotimes\nolimits_{t=1}^{n}v^m_{\alpha,t}.\end{aligned}$$

Let ξ_α^m be a random variable in $(\mathcal{Y}_1^n, \mathcal{Y}_1^n, v_\alpha^m)$

$$\xi_\alpha^m \triangleq \sum\nolimits_{t=1}^{n} \xi_{\alpha,t}^m.$$

As a result of (55), (56), and the product structure of q_α, v_α^m, and w^m we have

$$\xi_\alpha^m = \frac{1}{\alpha-1}\left[\ln \frac{\mathrm{d}v_\alpha^m}{\mathrm{d}w^m} - D_1\left(v_\alpha^m \middle\| w^m\right)\right], \tag{57}$$

$$= \frac{1}{\alpha}\left[\ln \frac{\mathrm{d}v_\alpha^m}{\mathrm{d}q_\alpha} - D_1\left(v_\alpha^m \middle\| q_\alpha\right)\right]. \tag{58}$$

Let $\mathcal{E}_m \in \mathcal{Y}_1^n$ be $\mathcal{E}_m \triangleq \{y \in \mathcal{Y}_1^n : m \in \Theta(y)\}$. Then for any $\alpha \in (0,1)$ and real numbers τ_1 and τ_2 we have

$$P_{\mathbf{e}}^m \geq e^{-D_1\left(v_\alpha^m \| w^m\right)-\tau_1} \mathbf{E}_{v_\alpha^m}\left[\mathbb{1}_{\{\mathcal{Y}_1^n \setminus \mathcal{E}_m\}} \mathbb{1}_{\{\xi_\alpha^m \geq \frac{\tau_1}{\alpha-1}\}}\right],$$

$$q_\alpha(\mathcal{E}_m) \geq e^{-D_1\left(v_\alpha^m \| q_\alpha\right)-\tau_2} \mathbf{E}_{v_\alpha^m}\left[\mathbb{1}_{\{\mathcal{E}_m\}} \mathbb{1}_{\{\xi_\alpha^m \leq \frac{\tau_2}{\alpha}\}}\right].$$

Then for $\tau_1 = \frac{\gamma}{\alpha}$ and $\tau_2 = \frac{\gamma}{1-\alpha}$ using (57) and (58) we get

$$P_{\mathbf{e}}^m e^{D_1\left(v_\alpha^m \| w^m\right)+\frac{\gamma}{\alpha}} + q_\alpha(\mathcal{E}_m) e^{D_1\left(v_\alpha^m \| q_\alpha\right)+\frac{\gamma}{1-\alpha}} \geq \mathbf{E}_{v_\alpha^m}\left[\mathbb{1}_{\{\frac{-\gamma}{(1-\alpha)\alpha} \leq \xi_\alpha^m \leq \frac{\gamma}{(1-\alpha)\alpha}\}}\right]. \tag{59}$$

The random variables $\xi_{\alpha,t}^m$ are zero mean in the probability space $(\mathcal{Y}_1^n, \mathcal{Y}_1^n, v_\alpha^m)$ by construction. Furthermore, they are jointly independent because of the product structure of the probability measures w^m, q_α, and v_α^m. Thus we can apply Lemma 19 to bound the right hand side of (59) from below, if we can show that γ defined in (51) is large enough. To that end, first note that $\mathbf{E}_{v_\alpha^m}\left[\left|\xi_{\alpha,t}^m\right|^\kappa\right] = \mathbf{E}_{v_{\alpha,t}^m}\left[\left|\xi_{\alpha,t}^m\right|^\kappa\right]$ by construction. Then Lemma 17 implies

$$\mathbf{E}_{v_\alpha^m}\left[\left|\xi_{\alpha,t}^m\right|^\kappa\right] \leq 3\left[\frac{\left((1-\alpha)D_\alpha\left(w_t^m \middle\| q_{\alpha,W_t}^\epsilon\right)\right) \vee \kappa}{\alpha(1-\alpha)}\right]^\kappa. \tag{60}$$

We can bound $D_\alpha\left(w_t^m \middle\| q_{\alpha,W_t}^\epsilon\right)$ using (20) of Lemma 14

$$(1-\alpha)D_\alpha\left(w_t^m \middle\| q_{\alpha,W_t}^\epsilon\right) \leq \frac{C_{1/2,W_t}}{1-\epsilon}. \tag{61}$$

Using the definition of γ given in (51) and together with (60) and (61) we get

$$3\left(\sum\nolimits_{t=1}^{n} \mathbf{E}_{v_\alpha^m}\left[\left|\xi_{\alpha,t}^m\right|^\kappa\right]\right)^{1/\kappa} \leq \frac{\gamma}{\alpha(1-\alpha)}.$$

Then (59) and Lemma 19 implies

$$P_{\mathbf{e}}^m e^{D_1\left(v_\alpha^m \| w^m\right)+\frac{\gamma}{\alpha}} + q_\alpha(\mathcal{E}_m) e^{D_1\left(v_\alpha^m \| q_\alpha\right)+\frac{\gamma}{1-\alpha}} \geq \frac{1}{2\sqrt{n}}. \tag{62}$$

On the other hand, the product structure of the probability measures w^m and q_α implies

$$D_\alpha\left(w^m \middle\| q_\alpha\right) = \sum\nolimits_{t=1}^{n} D_\alpha\left(w_t^m \middle\| q_{\alpha,W_t}^\epsilon\right).$$

Bounding each term in the sum using (19) of Lemma 14 and then invoking (23) we get

$$D_\alpha\left(w^m \middle\| q_\alpha\right) \leq \widetilde{C}_{\alpha,W_{[1,n]}}^\epsilon. \tag{63}$$

In the following, we show that the message set has a size $\approx \frac{M\epsilon}{n}$ subset in which all the messages has a conditional error probability greater than $\approx (\frac{e^{-2\gamma}}{\sqrt{n}})^{\frac{1}{\phi}} (\frac{\epsilon}{n})^{\frac{1-\phi}{\phi}} e^{-\widetilde{E}_{sp}^\epsilon(R, W_{[1,n]})}$. The existence of such a subset will imply (52). We prove the existence of such a subset using (62), (63), the intermediate value theorem, and pigeon hole arguments. Let us consider the subset of the message set, $\mathcal{M}_1$ defined as follows:

$$\mathcal{M}_1 \triangleq \left\{m : \inf_{\alpha \in (\phi,1)} \left[(q_\alpha(\mathcal{E}_m) + \frac{L}{M}) e^{D_1\left(v_\alpha^m \| q_\alpha\right)+\frac{\gamma}{1-\alpha}}\right] \geq \frac{1}{4\sqrt{n}}\right\}.$$

First, we bound the size of $\mathcal{M}_1$ from above. We can bound $D_1\left(v_\alpha^m \middle\| q_\alpha\right)$ using the definitions of $v_{\alpha,t}^m$, v_α^m, q_α, the non-negativity of the Rényi divergence for probability measures, which is implied by Lemma 4, and (63), as follows

$$D_1\left(v_\alpha^m \middle\| q_\alpha\right) = D_\alpha\left(w^m \middle\| q_\alpha\right) - \frac{\alpha}{1-\alpha} D_1\left(v_\alpha^m \middle\| w^m\right) \tag{64}$$

$$\leq \widetilde{C}_{\alpha,W_{[1,n]}}^\epsilon \tag{65}$$

for all $m \in \mathcal{M}$, $\alpha \in (0,1)$. Then summing the inequality in the condition for membership of $\mathcal{M}_1$ over the members of $\mathcal{M}_1$ we get

$$2Le^{\widetilde{C}_{\alpha,W_{[1,n]}}^\epsilon + \frac{\gamma}{1-\alpha}} \geq |\mathcal{M}_1| \frac{1}{4\sqrt{n}} \qquad \forall \alpha \in (\phi, 1).$$

Then $\frac{|\mathcal{M}_1|}{L} \leq 8\sqrt{n} e^{\widetilde{C}_{\phi,W_{[1,n]}}^\epsilon + \frac{\gamma}{1-\phi}}$. Consequently $|\mathcal{M}_1| < \frac{M}{2}$ because $\frac{M}{L} > 16\sqrt{n} e^{\widetilde{C}_{\phi,W_{[1,n]}}^\epsilon + \frac{\gamma}{1-\phi}}$ by the hypothesis.

On the other hand, as a result of the definition of $\mathcal{M}_1$, for each $m \in \mathcal{M} \setminus \mathcal{M}_1$ there is an $\alpha \in (\phi, 1)$ satisfying

$$(q_\alpha(\mathcal{E}_m) + \frac{L}{M}) e^{D_1\left(v_\alpha^m \| q_\alpha\right)+\frac{\gamma}{1-\alpha}} < \frac{1}{4\sqrt{n}}.$$

Furthermore, q_α is continuous in α for the total variation topology on $\mathcal{P}(\mathcal{Y}_1^n)$ by construction.[4] Then $D_1\left(v_\alpha^m \middle\| q_\alpha\right)$ is continuous in α by Lemma 16 and $q_\alpha(\mathcal{E}_m)$ is continuous in α, as well. Thus $(q_\alpha(\mathcal{E}_m) + \frac{L}{M}) e^{D_1\left(v_\alpha^m \| q_\alpha\right)+\frac{\gamma}{1-\alpha}}$ is continuous in α. Since $\lim_{\alpha \uparrow 1} (q_\alpha(\mathcal{E}_m) + \frac{L}{M}) e^{D_1\left(v_\alpha^m \| q_\alpha\right)+\frac{\gamma}{1-\alpha}} = \infty$, using the intermediate value theorem [51, 4.23] we can conclude that for each $m \in \mathcal{M} \setminus \mathcal{M}_1$ there exists an $\alpha_m \in (\phi, 1)$ such that

$$(q_{\alpha_m}(\mathcal{E}_m) + \frac{L}{M}) e^{D_1\left(v_{\alpha_m}^m \| q_{\alpha_m}\right)+\frac{\gamma}{1-\alpha_m}} = \frac{1}{4\sqrt{n}}. \tag{66}$$

Then for any positive integer K, there exists a length $\frac{1}{K}$ interval with $\lceil \frac{M}{2K} \rceil$ or more α_m's. Let $[\eta, \eta + \frac{1}{K}]$ be the aforementioned interval, $\tilde{\epsilon}$ and $\tilde{\alpha}$ be real numbers in $(0,1)$

$$\tilde{\epsilon} \triangleq \frac{1}{K} + \epsilon(1 - \frac{1}{K}), \qquad \tilde{\alpha} \triangleq \frac{1-\epsilon}{1-\tilde{\epsilon}} \eta.$$

Then $q_{\alpha,W_t}^\epsilon \leq \frac{\tilde{\epsilon}}{\epsilon} q_{\tilde{\alpha},W_t}^{\tilde{\epsilon}}$ for all $\alpha \in [\eta, \eta + \frac{1}{K}]$, by the definition of the average Rényi center. Thus

$$q_\alpha \leq (\frac{\tilde{\epsilon}}{\epsilon})^n \tilde{q} \qquad \forall \alpha \in [\eta, \eta + \frac{1}{K}]$$

where $\tilde{q} \in \mathcal{P}(\mathcal{Y}_1^n)$ is defined as follows

$$\tilde{q} \triangleq \bigotimes\nolimits_{t=1}^{n} q_{\tilde{\alpha},W_t}^{\tilde{\epsilon}}.$$

[4]In particular $\left\|q_\alpha - q_\eta\right\| \leq \sqrt{8n \ln \frac{\epsilon}{\epsilon - |\alpha-\eta|}}$ for all α and η in $(0,1)$ such that $|\eta - \alpha| < \epsilon$ by (3), the product structure of q_α, and the definition of the average Rényi center, which implies $\left\|q_{\alpha,W_t}^\epsilon - q_{\eta,W_t}^\epsilon\right\| \leq \frac{2|\alpha-\eta|}{\epsilon}$ for all W_t.

On the other hand, at least half of the messages with α_m's in $[\eta, \eta+\frac{1}{K}]$ satisfy $\tilde{q}(\mathcal{E}_m) \leq 2\frac{L}{\lceil M/2\ K\rceil}$. Then at least $\lceil \frac{1}{2}\lceil \frac{M}{2\ K}\rceil\rceil$ messages with α_m's in $[\eta, \eta+\frac{1}{K}]$ satisfy

$$q_{\alpha_m}(\mathcal{E}_m) \leq \frac{4L}{M}K\left(1+\frac{1}{K}\frac{1-\epsilon}{\epsilon}\right)^n$$

Note that $\frac{n(1-\epsilon)}{\epsilon} > 1$ because $\epsilon < \frac{n}{n+1}$. Then we can set K to $\lfloor \frac{n(1-\epsilon)}{\epsilon}\rfloor$ and use the identity $(1+z)^{1/z} < e$ to get

$$q_{\alpha_m}(\mathcal{E}_m) \leq \frac{4L}{M}\frac{n(1-\epsilon)}{\epsilon}e^2. \tag{67}$$

Then using (66), we can bound $D_1\left(v^m_{\alpha_m}\middle\| q_{\alpha_m}\right)$ for all m satisfying (67) as follows

$$e^{D_1\left(v^m_{\alpha_m}\| q_{\alpha_m}\right)+\frac{\gamma}{1-\alpha_m}} \geq \frac{1}{4\sqrt{n}}\frac{\epsilon}{4e^2 n}\frac{M}{L}. \tag{68}$$

On the other hand we can bound $P^m_{\mathbf{e}}$ using (62) and (66)

$$P^m_{\mathbf{e}} e^{D_1\left(v^m_{\alpha_m}\| \Psi(m)\right)+\frac{\gamma}{\alpha_m}} \geq \frac{1}{4\sqrt{n}}. \tag{69}$$

Using (63), (64), (68), and (69) we get

$$P^m_{\mathbf{e}} e^{\frac{1-\alpha_m}{\alpha_m}\widetilde{C}^{\epsilon}_{\alpha_m,W_{[1,n]}}+2\frac{\gamma}{\alpha_m}} \geq \left(\frac{1}{4\sqrt{n}}\right)^{\frac{1}{\alpha_m}}\left(\frac{\epsilon}{4e^2 n}\frac{M}{L}\right)^{\frac{1-\alpha_m}{\alpha_m}}. \tag{70}$$

Hence, for all m satisfying (67) as a result of the definition of $\widetilde{E}^{\epsilon}_{sp}(R,W)$ given in (26) we have

$$P^m_{\mathbf{e}} \geq e^{-2\frac{\gamma}{\phi}}\left(\frac{1}{4\sqrt{n}}\right)^{\frac{1}{\phi}}\left(\frac{\epsilon}{4e^2 n}\right)^{\frac{1-\phi}{\phi}} e^{-\widetilde{E}^{\epsilon}_{sp}(R,W_{[1,n]})}$$

where $R = \ln\frac{M}{L}$. Since there are at least $\lceil \frac{M}{4\ K}\rceil$ such messages we get the inequality given in (52).

In order to obtain (53), we change the analysis after (70). For all m satisfying (67), as a result of (70) and the definition of $\widetilde{E}^{\epsilon}_{sp}(R,W)$ given in (26) we have

$$P^m_{\mathbf{e}} \geq \left(\frac{1}{4\sqrt{n}}\right)e^{-2\gamma}\, e^{-\widetilde{E}^{\epsilon}_{sp}(R,W_{[1,n]})}$$

where $R = \ln\frac{M}{L} - 2\gamma - \ln\frac{16\ e^2 n^{3/2}}{\epsilon}$.

Since there are at least $\lceil \frac{M}{4\ K}\rceil$ such messages we get the bound given in (53). ☐

The Rényi centers and capacities of certain channels satisfy (50) for all $\alpha \leq z$ in $(0,1)$, e.g. the Poisson channel $\Lambda^{T,a,b,\varrho}$ whose input set is described in (1a). The Rényi capacity and center of $\Lambda^{T,a,b,\varrho}$ are determined in [40, Example 9]:

$$C_{\alpha,\Lambda^{T,a,b,\varrho}} = \begin{cases} \frac{\alpha}{\alpha-1}\left[\left(\frac{\varrho-a}{b-a}b^{\alpha}+\frac{b-\varrho}{b-a}a^{\alpha}\right)^{1/\alpha}-\varrho\right]T & \alpha\neq 1 \\ \left[\frac{\varrho-a}{b-a}b\ln\frac{b}{\varrho}+\frac{b-\varrho}{b-a}a\ln\frac{a}{\varrho}\right]T & \alpha = 1. \end{cases}$$

The order α Rényi center of $\Lambda^{T,a,b,\varrho}$ is the stationary Poisson processes with intensity $\left(\frac{\varrho-a}{b-a}b^{\alpha}+\frac{b-\varrho}{b-a}a^{\alpha}\right)^{1/\alpha}$. For channels satisfying (50) for all $\alpha \leq z$ in $(0,1)$, the averaging scheme is not needed to establish the SPB. In addition, the resulting bound is sharper than the one given in Lemma 20.

Lemma 21: Let $W_{[1,n]}$ be a product channel for an $n \in \mathbb{Z}_+$ satisfying $C_{1,W_{[1,n]}} \geq \frac{\phi^2}{2}$ for a $\phi \in (0,1)$ and (50) for all α, z satisfying $\phi \leq \alpha \leq z < 1$, κ satisfy $\kappa \geq 3$, and γ be

$$\gamma \triangleq 3\sqrt[\kappa]{3}\left(\sum\nolimits_{t=1}^{n}\left(C_{1/2,W_t}\vee\kappa\right)^{\kappa}\right)^{1/\kappa}. \tag{71}$$

If M, L are integers satisfying $\frac{M}{L} > 16\sqrt{n}e^{C_{\phi,W_{[1,n]}}+\frac{\gamma}{1-\phi}}$, then any (M,L) channel code on $W_{[1,n]}$ satisfies

$$P^{av}_{\mathbf{e}} \geq \frac{\phi^2 e^{-2\gamma}}{32n^{1/2}C_{1,W_{[1,n]}}}e^{-E_{sp}(R,W)} \tag{72}$$

where $R = \ln\frac{M}{L} - \ln\frac{95n^{1/2}C_{1,W_{[1,n]}}}{\phi^2 e^{-2\gamma}}$.

Proof of Lemma 21: We use q_{α,W_t}'s rather than $q^{\epsilon}_{\alpha,W_t}$'s to define q_{α}; thus q_{α} is equal to $q_{\alpha,W_{[1,n]}}$. We repeat the analysis of Lemma 20 up to (66): There are at least $\lceil\frac{M}{2}\rceil$ messages satisfying the following identity for some $\alpha_m \in (\phi,1)$

$$\left(q_{\alpha_m}(\mathcal{E}_m)+\frac{L}{M}\right)e^{D_1\left(v^m_{\alpha_m}\| q_{\alpha_m}\right)+\frac{\gamma}{1-\alpha_m}} = \frac{1}{4\sqrt{n}}.$$

Let K be $\lfloor\frac{2C_{1,W_{[1,n]}}}{\phi^2}\rfloor$. Note that there exists a length $\frac{1}{K}$ interval with $\lceil\frac{M}{2\ K}\rceil$ or more α_m's. Let $[\eta-\frac{1}{K},\eta]$ be the aforementioned interval; then for all $\alpha \in [\eta-\frac{1}{K},\eta]$ we have $\frac{1-\alpha}{\alpha}C_{\alpha,W_{[1,n]}} - \frac{1-\eta}{\eta}C_{\eta,W_{[1,n]}} \leq 1$ by the monotonicity of $C_{\alpha,W}$ in α, i.e. Lemma 8-(a). Then as a result of the hypothesis of the lemma we have $q_{\alpha} \leq e q_{\eta}$ for all α in $[\eta-\frac{1}{K},\eta]$. On the other hand at least half of the messages with α_m's in $[\eta-\frac{1}{K},\eta]$, satisfy $q_{\eta}(\mathcal{E}_m) \leq 2\frac{L}{\lceil M/2\ K\rceil}$. Then at least $\lceil\frac{1}{2}\lceil\frac{M}{2\ K}\rceil\rceil$ messages with α_m's in $[\eta-\frac{1}{K},\eta]$ satisfy

$$q_{\alpha_m}(\mathcal{E}_m) \leq \frac{4eLK}{M}. \tag{73}$$

Using (73) instead of (67) and repeating the rest of the analysis we get (72) using $8(4e+1) \leq 95$. ☐

For certain channels the Rényi center does not change with the order on the interval that it exits, e.g. [40, Example 8], the binary symmetric channels. The hypothesis of Lemma 21, described in (50), is satisfied for these channels as a result of the monotonicity of $\frac{1-\alpha}{\alpha}C_{\alpha,W}$, i.e. Lemma 8-(b). But it is possible to derive the following sharper bound for these channels.

Lemma 22: Let $W_{[1,n]}$ be a product channel for an $n \in \mathbb{Z}_+$ satisfying

$$q_{\alpha,W_{[1,n]}} = q_{\phi,W_{[1,n]}} \qquad \forall\alpha\in(\phi,1) \tag{74}$$

for a $\phi \in (0,1)$ and $\kappa \geq 3$. If M, L are integers satisfying $\frac{M}{L} > 16\sqrt{n}e^{C_{\phi,W_{[1,n]}}+\frac{\gamma}{1-\phi}}$ for γ described in (71), then any (M,L) channel code on $W_{[1,n]}$ satisfies

$$P^{av}_{\mathbf{e}} \geq \frac{e^{-2\gamma}}{16n^{1/2}}e^{-E_{sp}(R,W_{[1,n]})} \tag{75}$$

where $R = \ln\frac{M}{L} - \ln\frac{20n^{1/2}}{e^{-2\gamma}}$.

Proof of Lemma 22: $q_{\alpha,W_t} = q_{\phi,W_t}$ for all $\alpha \in [\phi,1)$ by the hypothesis of the lemma and Lemma 11. We use q_{ϕ,W_t}'s rather than $q^{\epsilon}_{\alpha,W_t}$'s to define the probability measure q_{α}. Since q_{α} is the same probability measure for all $\alpha \in [\phi,1)$, we denote it by q. We repeat the analysis of Lemma 20 up to (66): There are at least $\lceil \frac{M}{2} \rceil$ messages satisfying

$$(q(\mathcal{E}_m)+\frac{L}{M})e^{D_1(v_{\alpha_m}^m\|q)+\frac{\gamma}{1-\alpha_m}}=\frac{1}{4\sqrt{n}}$$

for some $\alpha_m \in (\phi,1)$. Among these $\lceil \frac{M}{2} \rceil$ messages, there exists at least $\lceil \frac{M}{4} \rceil$ messages satisfying

$$q(\mathcal{E}_m) \le \frac{4L}{M}. \tag{76}$$

Using (76) instead of (67) and repeating the rest of the analysis we get (75). □

Lemma 19 is a key ingredient of the proof of Lemma 20. The independence hypothesis of Lemma 19 is implied by the product structure of each $W_{[1,n]}(\Psi(m))$ and q_α. However, the product structure is not necessary for the independence, provided that the channel has certain symmetries.

Lemma 23: Let $W_{\overrightarrow{[1,n]}}$ be a product channel with feedback for an $n \in \mathbb{Z}_+$ satisfying[5] for each $t \in \{1,\ldots,n\}$

$$q_{\alpha,W_t} = q_t \qquad \forall \alpha \in (0,1), \tag{77}$$

$$q_t(\frac{\mathrm{d}(W_t(x))_{ac}}{\mathrm{d}q_t} \le z) = g_t(z) \qquad \forall x \in \mathcal{X}_t,\ z \in \Re_{\ge 0} \tag{78}$$

for a $q_t \in \mathcal{P}(\mathcal{Y}_t)$ and a cumulative distribution function g_t, $\phi \in (0,1)$, $\kappa \ge 3$, and γ be the constant defined in (71)*. If M, L are integers satisfying $\frac{M}{L} > 16\sqrt{n}e^{C_{\phi,W_{[1,n]}}+\frac{\gamma}{1-\phi}}$, then any (M,L) channel code on $W_{\overrightarrow{[1,n]}}$ satisfies* (75).

Any product channel whose component channels are modular shift channels described in [40, Example 5], satisfy the constraints given in (77) and (78). Products of more general shift invariant channels described in [40, Example 8], do satisfy the constraint given in (77) but they may or may not satisfy the constraint given in (78) depending on $\mathcal{F}$.

Proof of Lemma 23: As we have done for Lemma 22 we use q_t's, rather than $q_{\alpha,W_t}^{\epsilon}$'s, to define q. Although $W_{\overrightarrow{[1,n]}}(\Psi(m))$ is not necessarily a product measure, $\xi_{\alpha,t}^m$'s are jointly independent random variables in the probability space $(\mathcal{Y}_1^n, \mathcal{Y}_1^n, v_\alpha^m)$ for any $\alpha \in (0,1)$ and $m \in \mathcal{M}$, as a result of the hypothesis of the lemma given in (78). The rest of the proof is identical to the proof of Lemma 22. □

Proof of Theorem 2: We prove Theorem 2 using Lemmas 15 and 20. Note that we are free to choose different values for ϵ and κ for different values of n, provided that the hypotheses of Lemmas 15 and 20 are satisfied.

As a result of Assumption 1 there exists a $K \in [1,\infty)$ and an $n_0 \in \mathbb{Z}_+$ such that $\max_{t\in[1,n]} C_{1/2,W_t} \le K \ln n$ for all $n \ge n_0$. Let κ_n be $K \ln n$ and ϵ_n be $1/n$. Then

$$\gamma_n \le 4eK \ln n \tag{79}$$

for all n large enough. Furthermore, (22) and (79) imply

$$16\sqrt{n}e^{\widetilde{C}_{\alpha_0,W_{[1,n]}}^{\epsilon}+\frac{\gamma_n}{1-\alpha_0}} \le 16e^{C_{\alpha_0,W_{[1,n]}}+(\frac{1}{2}+\frac{4eK}{1-\alpha_0}+\frac{K}{\alpha_0(1-\alpha_0)})\ln n}$$

for all n large enough. Thus as a result of the hypothesis of the theorem, hypotheses of Lemma 20 is satisfied for all n large enough. Thus using (79) we can conclude that

$$P_{\mathbf{e}}^{av} \ge \left(\frac{n^{-1-8eK}}{16e^2n^{3/2}}\right)^{\frac{1}{\alpha_0}} e^{-\widetilde{E}_{sp}^{1/n}\left(\ln\frac{M_n}{L_n},W_{[1,n]}\right)} \tag{80}$$

for all n large enough.

On the other hand Lemma 15, the hypothesis given in (37), and the monotonicity of $C_{\alpha,W}$ in α imply that

$$\widetilde{E}_{sp}^{1/n}\left(\ln\frac{M_n}{L_n},W_{[1,n]}\right) \le E_{sp}\left(\ln\frac{M_n}{L_n},W_{[1,n]}\right)+\frac{C_{\alpha_1,W_{[1,n]}}}{(n-1)\alpha_0^2}$$

for all n large enough. Then using the monotonicity of $C_{\alpha,W}$ and $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ in α we can conclude that

$$\widetilde{E}_{sp}^{1/n}\left(\ln\frac{M_n}{L_n},W_{[1,n]}\right) \le E_{sp}\left(\ln\frac{M_n}{L_n},W_{[1,n]}\right) +\frac{(\frac{\alpha_1}{1-\alpha_1}\vee 1)nK\ln n}{(n-1)\alpha_0^2}$$

for all n large enough. Then (38) follows from (80). □

D. Augustin's SPBs for Product Channels

In the following, we compare our results with the SPBs derived by Augustin [6], [7] for the product channels. Augustin works with the maximum error probability, $P_{\mathbf{e}}^{\max} \triangleq \max_{m\in\mathcal{M}} P_{\mathbf{e}}^m$, rather than the average error probability. This, however, is inconsequential for our purposes because any SPB for $P_{\mathbf{e}}^{av}$ holds for $P_{\mathbf{e}}^{\max}$ as is and any SPB for $P_{\mathbf{e}}^{\max}$ can be converted into a SPB for $P_{\mathbf{e}}^{av}$ through a standard application of Markov inequality for channel codes, with definite and essentially inconsequential correction terms.

The main advantage of Theorem 2 over the SPBs in [6] and [7], is its polynomial prefactor. Augustin did establish a SPB with a polynomial prefactor, but only under considerably stronger hypotheses, [6, Th. 4.8]. In addition all of the asymptotic SPBs in [6] and [7] assume the uniform continuity condition described in Assumption 2, given in the following. Theorem 2, on the other hand, does not have such a hypothesis.

Assumption 2: $\{W_t\}_{t\in\mathbb{Z}_+}$ and $C_{0^+,W}$ defined in (9) satisfy

$$\lim_{\alpha\downarrow 0}\sup_{n\in\mathbb{Z}_+}\frac{1}{n}\left[C_{\alpha,W_{[1,n]}}-C_{0^+,W_{[1,n]}}\right]=0.$$

Remark 7: This assumption is given as in [6, eq. (7)] and [7, Condition 31.3a]. In [7], the condition is stated without $1/n$ factor; we believe that it is a typo.

After this general overview, let us continue with a discussion of the individual results.

Reference [6, Th. 4.7b] bounds $P_{\mathbf{e}}^{\max(n)}$ from below by $e^{-e^K\sqrt{32n}-E_{sp}\left(\ln\frac{M_n}{L_n},W_{[1,n]}\right)}$ for large enough n for any sequence of channels satisfying $\sup_{t\in\mathbb{Z}_+} C_{1,W_t} < K$ for some $K \in \Re_+$ and Assumption 2. Thus [6, Th. 4.7b] proves a claim weaker than Theorem 2 under a hypothesis stronger than Assumption 1.

Reference [6, Th. 4.8] is a SPB with a polynomial prefactor for product channels satisfying Assumptions 2 and 3.

Assumption 3: $\exists K \in \Re_+, \{\nu_t\}_{t\in\mathbb{Z}_+}$ satisfying

$$\frac{1}{K} \le \frac{\mathrm{d}W_t(x)}{\mathrm{d}\nu_t} \le K \qquad W_t(x)-\text{a.e.} \qquad \forall x \in \mathcal{X}_t.$$

[5] $(W_t(x))_{ac}$ stands for the component that is absolutely continuous in q_t.

Assumption 3 implies Assumption 1, but the converse is not true, e.g. if $W_t = \Lambda^{T,a,b}$ then Assumption 1 holds but Assumption 3 does not hold. Thus [6, Th. 4.8] is weaker than Theorem 2 because it establishes the same claim under a stronger hypothesis.

References [6, Th. 4.7a] and [7, Th. 31.4] bound $P_{\mathbf{e}}^{\max(n)}$ from below by $e^{-O(\sqrt{n})-E_{sp}\left(\ln\frac{M_n}{L_n},W_{[1,n]}\right)}$ for large enough n. These SPBs are not comparable with Theorem 2 because their hypotheses are not comparable with the hypotheses of Theorem 2. However, these SPBs can be proved without relying on Assumption 2, using a variant of Lemma 20, which is obtained by applying Chebyshev's inequality instead of Lemma 19.

Reference [7, Lemma 31.2] is quite similar to Lemma 20; the main difference is the infimum taken over (0, 1). In order to remove this infimum and obtain a bound similar to Lemma 20, one needs to assume an equicontinuity condition similar to the one in Assumption 2.

V. The SPB for Product Channels with Feedback

Theorem 3: Let $\{W_t\}_{t\in\mathbb{Z}_+}$ be a sequence of discrete channels satisfying $W_t = W$ for all $t \in \mathbb{Z}_+$ and α_0, α_1 be orders satisfying $0 < \alpha_0 < \alpha_1 < 1$. Then for any sequence of codes on the discrete stationary product channels with feedback $\{W_{\overrightarrow{[1,n]}}\}_{n\in\mathbb{Z}_+}$ satisfying

$$C_{\alpha_1,W} \geq \frac{1}{n}\ln\frac{M_n}{L_n} \geq C_{\alpha_0,W} + \frac{\ln n}{n^{1/4}} \quad \forall n\geq n_0 \tag{81}$$

there exists an $n_1 \geq n_0$ such that

$$P_{\mathbf{e}}^{av(n)} \geq e^{-n\left[E_{sp}\left(\frac{1}{n}\ln\frac{M_n}{L_n},W\right)+\frac{1}{\alpha_0}\frac{\ln n}{n^{1/4}}\right]} \quad \forall n\geq n_1. \tag{82}$$

We prove Theorem 3 using ideas from Sheverdyaev [55], Haroutunian [31], [32], and Augustin [6], [7]. In §V-A, we establish a Taylor's expansion for $D_\alpha(w\| q)$ around $\alpha = 1$ assuming $D_\lambda(w\| q)$ is finite for a $\lambda > 1$. In §V-B, we recall the auxiliary channel method and prove that for any channel W satisfying $\lim_{\alpha\uparrow 1}\frac{1-\alpha}{\alpha}C_{\alpha,W} = 0$ and rate $R \in (C_{0^+,W}, C_{1,W})$ there exists a channel V and a constant $\beta > 1$ satisfying both $C_{\beta,V} \lesssim R$ and $\sup_{x\in\mathcal{X}} D_1(V(x)\| W(x)) \lesssim E_{sp}(R, W)$. In §V-C, we first introduce the concept of subblocks and derive a non-asymptotic SPB using it; then we prove Theorem 3 using this SPB. In §V-D, we provide an asymptotic SPB for (possibly non-stationary) DPCs, i.e. Theorem 4, and compare our results with the earlier ones. In §V-E, we show that Haroutunian's bound, the results of §V-B, and the concept of subblocks imply an asymptotic SPB for DSPCs with feedback, as well.

A. A Taylor's Expansion for the Rényi Divergence

Sheverdyaev employed the Taylor's expansion —albeit with approximation error terms that are not explicit— for his attempt to prove the SPB for the codes on the DSPCs with feedback in [55]. Recently, Fong and Tan [26, Proposition 11] bounded $D_\beta(w\| q)$ for $\beta \in [1, \frac{5}{4}]$ using Taylor's expansion for the case when $\mathcal{Y}$ is a finite set and $\mathcal{Y}$ is $2^{\mathcal{Y}}$. The bound by Fong and Tan, however, is not appropriate for our purposes because its approximation error term is proportional to $|\mathcal{Y}|$. Assuming $\frac{dw}{dq}$ to be bounded Sason and Verdú [52, Th. 35-(b), (469)] derived a similar bound.[6] In Lemma 24 we bound $D_\beta(w\| q)$ for $\beta \in (1, \lambda)$ using Taylor's expansion assuming only $D_\lambda(w\| q)$ to be finite.

Lemma 24: Let w and q be two probability measures on the measurable space $(\mathcal{Y}, \mathcal{Y})$ satisfying $D_\lambda(w\| q) \leq \gamma$ for a $\gamma \in \Re_+$ and a $\lambda \in (1,\infty)$. Then for any $\beta \in (1, \lambda)$

$$\begin{aligned}0 &\leq D_\beta(w\| q) - D_1(w\| q)\\ &\leq \frac{2(\beta-1)}{e^2}\left[1 + e^{(\beta-1)\gamma}\left(\frac{\gamma\, e^{\tau}}{2\tau}\right)^2\right]\end{aligned} \tag{83}$$

where $\tau = \frac{(\lambda-\beta)\gamma}{2} \wedge 1$.

The Rényi divergences with orders greater than one are not customarily used for establishing the SPB; Sheverdyaev's proof [55] is an exception in this respect.[7] In §V-B, we use Lemma 24 to construct an auxiliary channel with certain properties desirable for our purposes, see Lemma 25.

Proof of Lemma 24: $D_\beta(w\| q) - D_1(w\| q)$ is nonnegative because the Rényi divergence is a nondecreasing function of the order by Lemma 1. In order to bound $D_\beta(w\| q) - D_1(w\| q)$ from above we use Taylor's theorem on the function $g(\alpha)$ defined as follows:

$$g(\alpha) \triangleq \mathbf{E}_q\left[\left(\frac{dw}{dq}\right)^\alpha\right].$$

$g(\alpha)$ is continuous in α on $(0, \lambda)$ by [9, Corollary 2.8.7] because $\mathbf{E}_q\left[\left(\frac{dw}{dq}\right)^\lambda\right] = e^{(\lambda-1)D_\lambda(w\|q)} < \infty$ by the hypothesis and $\left(\frac{dw}{dq}\right)^\alpha \leq 1 + \left(\frac{dw}{dq}\right)^\lambda$. In order to apply Taylor's theorem to $g(\alpha)$, we show that $g(\alpha)$ is twice differentiable and bound its second derivative. To that end, first note that we can bound $x^\alpha|\ln x|^\kappa$ for any $\alpha \in (0, \lambda)$ and $\kappa \in \{1, 2\}$, using the derivative test as follows

$$x^\alpha|\ln x|^\kappa \leq \left(\frac{\kappa}{e\alpha}\right)^\kappa \mathbb{1}_{\{x\in[0,1]\}} + \left(\frac{\kappa}{e(\lambda-\alpha)}\right)^\kappa x^\lambda \mathbb{1}_{\{x\in(1,\infty)\}}.$$

Hence, for all $\alpha \in (0, \lambda)$ and $\kappa \in \{1, 2\}$ we have

$$\begin{aligned}\left|\frac{d^\kappa}{d\alpha^\kappa}\left(\frac{dw}{dq}\right)^\alpha\right| &= \left(\frac{dw}{dq}\right)^\alpha\left|\ln\frac{dw}{dq}\right|^\kappa\\ &\leq \left(\frac{\kappa}{e\alpha}\right)^\kappa + \left(\frac{\kappa}{e(\lambda-\alpha)}\right)^\kappa\left(\frac{dw}{dq}\right)^\lambda.\end{aligned} \tag{84}$$

The expression on the right hand side has finite expectation under q for any $\alpha \in (0, \lambda)$ and $\kappa \in \{1, 2\}$. Thus $g(\alpha)$ is twice differentiable in α on $(0, \lambda)$ by [9, Corollary 2.8.7]. Furthermore, for α in $(0, \lambda)$ and $\kappa \in \{1, 2\}$ we have

$$\frac{d^\kappa}{d\alpha^\kappa}g(\alpha) = \mathbf{E}_q\left[\left(\frac{dw}{dq}\right)^\alpha\left(\ln\frac{dw}{dq}\right)^\kappa\right]. \tag{85}$$

[6] Reference [52, Th. 35-(b)] is obtained by expressing $D_\alpha(w\| q)$, which is not an f-divergence, as a monotonically increasing function of the order α Hellinger divergence between w and q, which is an f-divergence. Guntuboyina *et al.* [29] presented a general method for establishing sharp bounds among f-divergences, without assuming either $\frac{dw}{dq}$ or $\frac{dq}{dw}$ to be bounded. Yet such conditions can easily be included in the framework proposed in [29].

[7] To be precise Sheverdyaev does not explicitly use Rényi divergences in [55], but his analysis can be easily expressed via Rényi divergences.

Since $g(\alpha)$ is twice differentiable applying Taylor's theorem [21, Appendix B4] around $\alpha = 1$ we get

$$g(\beta) \leq 1 + (\beta - 1) \left. \frac{\mathrm{d}}{\mathrm{d}\alpha} g(\alpha) \right|_{\alpha=1} + \frac{(\beta-1)^2}{2!} \sup_{\alpha \in (1,\beta)} \frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} g(\alpha). \tag{86}$$

On the other hand using (84) and (85) we get

$$\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} g(\alpha) \leq \left(\frac{2}{e}\right)^2 + \left(\frac{2}{e(\lambda-\beta)}\right)^2 g(\lambda) \quad \forall \alpha \in (1, \beta). \tag{87}$$

Then using the identity $\ln z \leq z - 1$ together with (85), (86), and (87) we get the following inequality $\beta \in (1, \lambda)$

$$\ln g(\beta) \leq (\beta - 1) D_1 (w \| q) + (\beta - 1)^2 \frac{2}{e^2} \left(1 + \frac{g(\lambda)}{(\lambda-\beta)^2}\right).$$

On the other hand $g(\alpha) = e^{(\alpha-1)D_\alpha(w\|q)}$ by definition and $D_\lambda (w\| q) \leq \gamma$ by the hypothesis. Thus

$$D_\beta (w\| q) - D_1 (w\| q) \leq (\beta - 1) \frac{2}{e^2} \left(1 + \frac{e^{(\lambda-1)\gamma}}{(\lambda-\beta)^2}\right). \tag{88}$$

Note that $D_\alpha (w\| q) \leq \gamma$ for any $\alpha \in (\beta, \lambda)$ because $D_\lambda (w\| q) \leq \gamma$ and the Rényi divergence is a nondecreasing in its order by Lemma 1. Thus using the analysis leading to (88), we get the following inequality $\forall \alpha \in (\beta, \lambda]$

$$D_\beta (w\| q) - D_1 (w\| q) \leq (\beta - 1) \frac{2}{e^2} \left(1 + \frac{e^{(\alpha-1)\gamma}}{(\alpha-\beta)^2}\right). \tag{89}$$

Using the derivative test we can confirm that the least upper bound among the upper bounds given in (89) is the one at $\alpha = \lambda \wedge (\frac{2}{\gamma} + \beta)$ and the resulting upper bound is the one given in (83). As a side note, let us point out that the least upper bound is strictly less than the upper bound at $\alpha = \lambda$ iff $\gamma(\lambda - \beta) > 2$. $\square$

B. The Auxiliary Channel Method

Haroutunian's seminal paper [31], establishing the SPB for the stationary product channels with finite input sets, used the performance of a code on an auxiliary channel as an anchor to bound its performance on the actual channel. To the best of our knowledge, this is the first explicit use of the auxiliary channel method. In a nutshell, auxiliary channel method can be described as follows:

(i) Choose an auxiliary channel $V : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ based on the actual channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and the code (Ψ, Θ).
(ii) Bound the performance of (Ψ, Θ) on V.
(iii) Bound the performance of (Ψ, Θ) on W using the bound derived in part (ii) and the features of V.

Many infeasibility results that are derived without using the auxiliary channel method, can be interpreted as an implicit application of the auxiliary channel method, as well. As an example, let us consider a version of Arimoto's bound, due to Augustin [7, Th. 27.2-(ii)], given in the following: If M and L are positive integers satisfying $\ln \frac{M}{L} > C_{1,W}$ for a channel W, then the average error probability $P_{\mathbf{e}}^{av}$ of any (M, L) channel code on W satisfies

$$P_{\mathbf{e}}^{av} \geq 1 - e^{\frac{\alpha-1}{\alpha}(C_{\alpha,W} - \ln \frac{M}{L})} \qquad \forall \alpha > 1. \tag{90}$$

Augustin obtained (90) by a convexity argument in [7]; but it can be derived using the auxiliary channel method as well: Let $V : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ be such that $V(x) = q_{\alpha,W}$ for all $x \in \mathcal{X}$ and $P_{\mathbf{e}}^{V}$ be the average error probability of (Ψ, Θ) on V, then $P_{\mathbf{e}}^{V} \geq 1 - \frac{L}{M}$ for any (M, L) channel code (Ψ, Θ). Furthermore, $C_{\alpha,W} \geq D_\alpha (p \circledast W \| p \circledast V)$ by Theorem 1 and $D_\alpha (p \circledast W \| p \circledast V) \geq \frac{\ln[(P_{\mathbf{e}}^{av})^\alpha (P_{\mathbf{e}}^{V})^{1-\alpha} + (1-P_{\mathbf{e}}^{av})^\alpha (1-P_{\mathbf{e}}^{V})^{1-\alpha}]}{\alpha-1}$ by Lemma 3. Thus $C_{\alpha,W} \geq \frac{\ln[(1-P_e^{av})^\alpha (1-P_e^{V})^{1-\alpha}]}{\alpha-1}$ and (90) follows.

Haroutunian [32] applied the auxiliary channel method to bound the error probability of codes on DSPCs with feedback from below. The exponential decay rate of Haroutunian's bound with block length, however, is greater than the sphere packing exponent for most channels. In §V-C, we use the auxiliary channel method—via subblocks — to establish a SPB. To do that, we employ the auxiliary channel described in Lemma 25-(b,c), given in the following. In §V-E, we demonstrate that one can establish a SPB for codes on DSPCs with feedback by applying Lemma 25-(b,c) to subblocks and invoking Haroutunian's bound [32], as well.

Lemma 25 describes its auxiliary channels using the order α Rényi center $q_{\alpha,W}$ described in Theorem 1, the average Rényi center $q_{\alpha,W}^{\epsilon}$ described in Definition 14, and the tilted channel defined in the following.

Definition 17: For any $\alpha \in \Re_+$, $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and $q \in \mathcal{P}(\mathcal{Y})$ such that $\sup_{x\in\mathcal{X}} D_\alpha (W(x)\| q) < \infty$, *the order α tilted channel $W_\alpha^q : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$* is defined as

$$\frac{\mathrm{d}W_\alpha^q(x)}{\mathrm{d}\nu} \triangleq e^{(1-\alpha)D_\alpha(W(x)\|q)} (\frac{\mathrm{d}W(x)}{\mathrm{d}\nu})^\alpha (\frac{\mathrm{d}q}{\mathrm{d}\nu})^{1-\alpha} \tag{91}$$

for all $x \in \mathcal{X}$ where $\nu \in \mathcal{P}(\mathcal{Y})$ satisfies $W(x) \prec \nu$ and $q \prec \nu$.

Lemma 25: For any channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ satisfying both $C_{0^+,W} \neq C_{1,W}$ and $\lim_{\alpha\uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha,W} = 0$ and rate R in $(C_{0^+,W}, C_{1,W})$ there exist a $\phi \in (0, 1)$ such that

$$R = C_{\phi,W}$$

and an $\eta \in (\phi, 1)$ such that

$$E_{sp}(R, W) = \frac{1-\eta}{\eta} C_{\eta,W}.$$

Furthermore W, R, ϕ, η satisfy the following assertions.

(a) There exists an $f : \mathcal{X} \to [\phi, \eta]$ satisfying both (92) *and* (93) *for all $x \in \mathcal{X}$.*

$$D_1 \left(W_{f(x)}^{q_{f(x),W}}(x) \middle\| q_{f(x),W} \right) \leq R \tag{92}$$

$$D_1 \left(W_{f(x)}^{q_{f(x),W}}(x) \middle\| W(x) \right) \leq E_{sp}(R, W) \tag{93}$$

(b) For all $\epsilon \in (0, \phi/2)$ there exists an $f_\epsilon : \mathcal{X} \to [\phi, \eta]$ satisfying both (94) *and* (95) *for all $x \in \mathcal{X}$.*

$$D_1 \left(W_{f_\epsilon(x)}^{q_{f_\epsilon(x),W}^{\epsilon}}(x) \middle\| q_{f_\epsilon(x),W}^{\epsilon} \right) \leq R + \frac{2\epsilon C_{1/2,W}}{\phi(1-\phi)^2} \tag{94}$$

$$D_1 \left(W_{f_\epsilon(x)}^{q_{f_\epsilon(x),W}^{\epsilon}}(x) \middle\| W(x) \right) \leq E_{sp}(R, W) + \frac{2\epsilon C_{1/2,W}}{\phi^2(1-\eta)} \tag{95}$$

(c) If $\epsilon \in (0, \phi/2)$ then for all $\beta \in (1, \frac{1+\eta}{2\eta})$ we have

$$C_{\beta, W_{f_\epsilon}^{q_{f_\epsilon,W}^\epsilon}} \leq R + \frac{2\epsilon C_{1/2,W}}{\phi(1-\phi)^2} + \ln\frac{1}{\epsilon} + (\beta-1)e^{(\beta-1)\frac{2C_{1/2,W}}{1-\eta}}\left[\frac{4\vee 2C_{1/2,W}}{1-\eta}\right]^2. \tag{96}$$

Remark 8: There is a slight abuse of notation in the symbol $W_{f_\epsilon}^{q_{f_\epsilon,W}^\epsilon}$ in Lemma 25-(c). It stands for a $V : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ satisfying $V(x) = W_{f_\epsilon(x)}^{q_{f_\epsilon(x),W}^\epsilon}(x)$ for all $x \in \mathcal{X}$

Proof of Lemma 25: $C_{\alpha,W}$ is continuous in α on $(0, 1]$ by Lemma 8-(b). Then for any $R \in (C_{0^+,W}, C_{1,W})$ there exists a $\phi \in (0, 1)$ such that $R = C_{\phi,W}$ by the intermediate value theorem [51, 4.23]. Furthermore, $E_{sp}(R, W) \leq \frac{1-\phi}{\phi}C_{\phi,W}$ by the expression for $E_{sp}(R, W)$ given in Lemma 13 and the monotonicity of $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ in α established in Lemma 8-(b). Then the existence of η follows from the intermediate value theorem, [51, 4.23], and the hypothesis of the lemma because $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ is continuous in α by Lemma 8-(b).

(a) $q_{\alpha,W}$ is continuous in α by Lemma 10. Thus we can replace $q_{\alpha,W}^\epsilon$ with $q_{\alpha,W}$ in the proof of part (b) to prove this part.

(b) We prove the existence of the function f_ϵ by showing that (94) and (95) are satisfied for some $\alpha \in [\phi, \eta]$ for each $x \in \mathcal{X}$. We denote $W_\alpha^{q_{\alpha,W}^\epsilon}(x)$ —which is $W_\alpha^q(x)$ defined in (91) for $q = q_{\alpha,W}^\epsilon$— by v_α in the proof of this part. Note that v_α satisfies

$$D_1\left(v_\alpha \| q_{\alpha,W}^\epsilon\right) + \frac{\alpha}{1-\alpha}D_1\left(v_\alpha \| W(x)\right) = D_\alpha\left(W(x)\| q_{\alpha,W}^\epsilon\right)$$

for all $\alpha \in (0, 1)$. Then (19) of Lemma 14 implies that

$$D_1\left(v_\alpha \| q_{\alpha,W}^\epsilon\right) + \frac{\alpha}{1-\alpha}D_1\left(v_\alpha \| W(x)\right) \leq \widetilde{C}_{\alpha,W}^\epsilon \tag{97}$$

for all $\alpha \in (0, 1)$. Then using the non-negativity of the Rényi divergence, which is implied by Lemma 4, we get

$$D_1\left(v_\phi \| q_{\phi,W}^\epsilon\right) \leq \widetilde{C}_{\phi,W}^\epsilon, \tag{98}$$

$$D_1\left(v_\eta \| W(x)\right) \leq \frac{1-\eta}{\eta}\widetilde{C}_{\eta,W}^\epsilon. \tag{99}$$

As a result of (98), $D_1\left(v_\phi \| q_{\phi,W}^\epsilon\right)$ and $D_1\left(v_\eta \| q_{\eta,W}^\epsilon\right)$ satisfy one of the following three cases:

(i) If $D_1\left(v_\phi \| q_{\phi,W}^\epsilon\right) = \widetilde{C}_{\phi,W}^\epsilon$, then $D_1\left(v_\phi \| W(x)\right) = 0$ by (97). Then (94) and (95) hold for $\alpha = \phi$ as a result of (10) and (22).

(ii) If $D_1\left(v_\eta \| q_{\eta,W}^\epsilon\right) \leq \widetilde{C}_{\phi,W}^\epsilon$, then (94) and (95) hold for $\alpha = \eta$ as a result of (10), (22), and (99).

(iii) If $D_1\left(v_\phi \| q_{\phi,W}^\epsilon\right) < \widetilde{C}_{\phi,W}^\epsilon$ and $D_1\left(v_\eta \| q_{\eta,W}^\epsilon\right) > \widetilde{C}_{\phi,W}^\epsilon$, then $D_1\left(v_\alpha \| q_{\alpha,W}^\epsilon\right) = \widetilde{C}_{\phi,W}^\epsilon$ for some $\alpha \in (\phi, \eta)$ by the intermediate value theorem [51, 4.23] provided that $D_1\left(v_\alpha \| q_{\alpha,W}^\epsilon\right)$ is continuous in α. The continuity of $D_1\left(v_\alpha \| q_{\alpha,W}^\epsilon\right)$, on the other hand, follows from $\left\| q_{\alpha,W}^\epsilon - q_{\alpha',W}^\epsilon\right\| \leq \frac{1-\epsilon}{\epsilon}\left|\alpha - \alpha'\right|$, which holds for all $\alpha, \alpha' \in (0, 1)$, and Lemma 16-(b). The α satisfying $D_1\left(v_\alpha \| q_{\alpha,W}^\epsilon\right) = \widetilde{C}_{\phi,W}^\epsilon$ satisfies (94) as a result of (10) and (22). Furthermore, $D_1\left(v_\alpha \| W(x)\right) \leq \widetilde{E}_{sp}^\epsilon\left(\widetilde{C}_{\phi,W}^\epsilon, W\right)$ for the same α by (97) and the definition of the average sphere packing exponent given in (26). On the other hand, $\widetilde{E}_{sp}^\epsilon(R, W)$ is a nonincreasing in R because it is the pointwise supremum of such functions. Then $\widetilde{E}_{sp}^\epsilon\left(\widetilde{C}_{\phi,W}^\epsilon, W\right) \leq E_{sp}(R, W) + \frac{2\epsilon}{\phi^2}R$ by Lemma 15. Thus (95) holds for α satisfying $D_1\left(v_\alpha \| q_{\alpha,W}^\epsilon\right) = \widetilde{C}_{\phi,W}^\epsilon$ by (10).

(c) We introduce two shorthands for notational brevity:

$$V(x) = W_{f_\epsilon(x)}^{q_{f_\epsilon(x),W}^\epsilon}(x),$$
$$Q(x) = q_{f_\epsilon(x),W}^\epsilon.$$

The Rényi divergence is a nondecreasing function of the order by Lemma 1 and $f_\epsilon(x) \in [\phi, \eta]$ for all $x \in \mathcal{X}$ by part (b); then

$$D_{\frac{1}{\eta}}\left(V(x)\| Q(x)\right) \leq D_{\frac{1}{f_\epsilon(x)}}\left(V(x)\| Q(x)\right). \tag{100}$$

The definitions of the Rényi divergence, V, and Q imply

$$\begin{aligned} D_{\frac{1}{f_\epsilon(x)}}&\left(V(x)\| Q(x)\right) \\ &= D_{f_\epsilon(x)}\left(W(x)\| Q(x)\right) + \frac{1}{\frac{1}{f_\epsilon(x)}-1}\ln \mathbf{E}_\nu\left[\frac{\mathrm{d}W(x)}{\mathrm{d}\nu}\mathbb{1}_{\{\frac{\mathrm{d}Q(x)}{\mathrm{d}\nu}>0\}}\right] \\ &\leq D_{f_\epsilon(x)}\left(W(x)\| Q(x)\right). \end{aligned} \tag{101}$$

Using (20) of Lemma 14, together with $f_\epsilon(x) \leq \eta$ and $\epsilon \leq \phi/2 \leq 1/2$, we get

$$D_{f_\epsilon(x)}\left(W(x)\| Q(x)\right) \leq \frac{2C_{1/2,W}}{1-\eta}. \tag{102}$$

First bounding $D_{\frac{1}{\eta}}\left(V(x)\| Q(x)\right)$ using (100), (101), (102) and then applying Lemma 24 we get,

$$D_\beta\left(V(x)\| Q(x)\right) - D_1\left(V(x)\| Q(x)\right) \leq \frac{2(\beta-1)}{e^2}\left(1 + e^{(\beta-1)\frac{2C_{1/2,W}}{1-\eta}}\left(\frac{2C_{1/2,W}}{1-\eta}\frac{e^{\tau_\beta}}{2\tau_\beta}\right)^2\right)$$

for all $\beta \in (1, \frac{1}{\eta})$, $x \in \mathcal{X}$ where $\tau_\beta = (\frac{1}{\eta} - \beta)\frac{C_{1/2,W}}{1-\eta} \wedge 1$. Since e^z/z is a decreasing function of z on $(0, 1)$,

$$D_\beta\left(V(x)\| Q(x)\right) - D_1\left(V(x)\| Q(x)\right) \leq \frac{2(\beta-1)}{e^2}\left(1 + e^{(\beta-1)\frac{2C_{1/2,W}}{1-\eta}}\left(\frac{C_{1/2,W}}{1-\eta}\frac{e^{\tau}}{\tau}\right)^2\right)$$

for all $\beta \in (1, \frac{1+\eta}{2\eta})$, $x \in \mathcal{X}$ where $\tau = \frac{C_{1/2,W}}{2\eta} \wedge 1$. Thus

$$\begin{aligned} D_\beta\left(V(x)\| Q(x)\right) &- D_1\left(V(x)\| Q(x)\right) \\ &\leq 2(\beta-1)\left(1 + e^{(\beta-1)\frac{2C_{1/2,W}}{1-\eta}}\left(\frac{2\vee C_{1/2,W}}{1-\eta}\right)^2\right) \\ &\leq (\beta-1)e^{(\beta-1)\frac{2C_{1/2,W}}{1-\eta}}\left[\frac{4\vee 2C_{1/2,W}}{1-\eta}\right]^2 \end{aligned} \tag{103}$$

On the other hand $Q(x) \leq \frac{1}{\epsilon} q_W$ for all $x \in \mathcal{X}$ by the definition of $Q(x)$ where $q_W \triangleq \int_0^1 q_{z,W} \mathrm{d}z$. Then as a result of Lemma 2 we have

$$D_\beta(V(x) \| q_W) \leq D_\beta(V(x) \| Q(x)) + \ln \frac{1}{\epsilon}. \quad (104)$$

Since $S_{\beta,V} \leq S_{\beta,V}(q_W)$ by definition, (94), (103) and (104) imply

$$S_{\beta,V} \leq R + \frac{2\epsilon R}{\phi(1-\phi)^2} + \ln \frac{1}{\epsilon} + (\beta-1) e^{(\beta-1)\frac{2C_{1/2,W}}{1-\eta}} \left[\frac{4 \vee 2C_{1/2,W}}{1-\eta}\right]^2.$$

Then (96) follows from $C_{\beta,V} = S_{\beta,V}$ established in Theorem 1.

□

C. A Non-Asymptotic SPB for Product Channels With Feedback

The ultimate aim of this subsection is to prove Theorem 3. To that end we first derive the following parametric bound on the error probability of codes on DSPCs with feedback.

Lemma 26: Let n be a positive integer, $W_{\overrightarrow{[1,n]}}$ be a DSPC with feedback satisfying $W_t = W$ for all t for a W for which $C_{0^+,W} \neq C_{1,W}$, $\alpha_0 < \alpha_1 < z$ be orders in $(0,1)$ satisfying[8] $\frac{1-z}{z} C_{z,W} = E_{sp}(C_{\alpha_1,W}, W)$, and M, L, κ be positive integers satisfying $\lfloor \frac{n}{\kappa} \rfloor C_{1/2,W} \geq 2$ and

$$C_{\alpha_1,W} \geq \frac{1}{n} \ln \frac{M}{L} \geq C_{\alpha_0,W} + \frac{C_{1/2,W}}{1-z} \left[\frac{2\epsilon}{\alpha_0(1-z)} + \frac{14}{\sqrt[3]{\kappa}}\right] + \frac{\kappa}{n} \ln \frac{1}{\epsilon} \quad (105)$$

for an $\epsilon \in (0, \frac{\alpha_0}{2})$. Then any (M, L) channel code on $W_{\overrightarrow{[1,n]}}$ satisfies

$$P_{\mathbf{e}}^{av} \geq \frac{1}{4} e^{-n\left[E_{sp}\left(\frac{1}{n}\ln\frac{M}{L}, W\right) + \frac{C_{1/2,W}}{\alpha_0(1-z)}\left[\frac{6\epsilon}{\alpha_0(1-z)} + \frac{15}{\sqrt[3]{\kappa}}\right] - \frac{\kappa \ln \epsilon}{n\alpha_0}\right]}. \quad (106)$$

Lemma 26 is proved using the auxiliary channel method:
(i) Apply Lemma 25 on subblocks to choose V
(ii) Use (90) to bound the error probability on V, i.e. $P_{\mathbf{e}}^V$.
(iii) Use Lemma 3 to bound $P_{\mathbf{e}}^{av}$ in terms of $P_{\mathbf{e}}^V$.

We have described all ingredients of the proof strategy given above, except the concept of subblocks. Before the proof of Lemma 26, let us revisit the DPC with feedback and introduce the concept of subblocks.

Any DPC with feedback can be reinterpreted as a shorter DPC with feedback with larger component channels, which we call subblocks. Consider for example a length n DPC with feedback $W_{\overrightarrow{[1,n]}}$. Recall that the input set of $W_{\overrightarrow{[1,n]}}$ can be written in terms of the input and output sets of the component channels as

$$\bigtimes_{t=1}^{n} \mathcal{X}_t^{\mathcal{Y}_1^{t-1}}$$

where $\mathcal{A}^{\mathcal{B}}$ is the set of all functions from the set $\mathcal{B}$ to the set $\mathcal{A}$, $\mathcal{A}^{\emptyset} = \mathcal{A}$, $\mathcal{Y}_\imath^\jmath = \bigtimes_{t=\imath}^{\jmath} \mathcal{Y}_t$ for all integers $\imath \leq \jmath$, and $\mathcal{Y}_\imath^\jmath = \emptyset$ for all integers $\imath > \jmath$. Furthermore, the output set of $W_{\overrightarrow{[1,n]}}$ is $\mathcal{Y}_1^n$ and the transition probabilities of $W_{\overrightarrow{[1,n]}}$ can be written as

$$W_{\overrightarrow{[1,n]}}(y_1^n | \Psi_1^n) = \prod_{t=1}^{n} W_t(y_t | \Psi_t(y_1^{t-1})).$$

where $\Psi_t \in \mathcal{X}_t^{\mathcal{Y}_1^{t-1}}$.

The preceding description can be modified to define a subblock $W_{\overrightarrow{[\tau,t]}}$ for any $t > \tau$, analogously. Furthermore, these subblocks can be used to construct alternative descriptions of the DPC with feedback. Let $t_0, \ldots, t_\kappa$ a sequence of integers satisfying $t_0 = 0$, $t_\kappa = n$, and $t_\jmath < t_\imath$ for all $\jmath < \imath$ and $U_\imath : \mathcal{A}_\imath \to \mathcal{P}(\mathcal{B}_\imath)$ be $W_{\overrightarrow{[1+t_{\imath-1},t_\imath]}}$ for each $\imath \in \{1, \ldots, \kappa\}$:

$$\mathcal{A}_\imath = \bigtimes_{\jmath=1+t_{\imath-1}}^{t_\imath} \mathcal{X}_\jmath^{\mathcal{Y}_{1+t_{\imath-1}}^{\jmath-1}}$$
$$\mathcal{B}_\imath = \mathcal{Y}_{1+t_{\imath-1}}^{t_\imath}$$
$$U_\imath(b_\imath | a_\imath) = \prod_{\jmath=1+t_{\imath-1}}^{t_\imath} W_\jmath(y_\jmath | \Psi_\jmath(y_{1+t_{\imath-1}}^{\jmath-1}))$$

where $\Psi_\jmath \in \mathcal{X}_\jmath^{\mathcal{Y}_{1+t_{\imath-1}}^{\jmath-1}}$, $a_\imath = \Psi_{1+t_{\imath-1}}^{t_\imath}$, and $b_\imath = y_{1+t_{\imath-1}}^{t_\imath}$.

Then the length n DPC with feedback $W_{\overrightarrow{[1,n]}}$ and the length κ DPC with feedback $U_{\overrightarrow{[1,\kappa]}}$ are representing the same channel:

$$\bigtimes_{t=1}^{n} \mathcal{X}_t^{\mathcal{Y}_1^{t-1}} = \bigtimes_{\imath=1}^{\kappa} \mathcal{A}_\imath^{\mathcal{B}_1^{\imath-1}}$$
$$\mathcal{Y}_1^n = \mathcal{B}_1^\kappa$$
$$W_{\overrightarrow{[1,n]}}(y_1^n | \Psi_1^n) = U_{\overrightarrow{[1,\kappa]}}(b_1^\kappa | \widetilde{\Psi}_1^\kappa)$$

where $\widetilde{\Psi}_\imath(b_1^{\imath-1}) = (\Psi_{1+t_{\imath-1}}(b_1^{\imath-1}), \ldots, \Psi_{t_\imath}(\cdot, b_1^{\imath-1}))$ and $b_\imath = y_{1+t_{\imath-1}}^{t_\imath}$. This observation plays a crucial role in the proof of Lemma 26 and hence in establishing the SPB for codes on the DPCs with feedback.

Proof of Lemma 26: We divide the interval $[1, n]$ into κ subintervals of, approximately, equal length: we set t_0 to zero and define $\ell_\imath$ and $t_\imath$ for $\imath \in \{1, \ldots, \kappa\}$ as follows

$$\ell_\imath \triangleq \lceil n/\kappa \rceil \mathbb{1}_{\{\imath \leq n - \lfloor n/\kappa \rfloor \kappa\}} + \lfloor n/\kappa \rfloor \mathbb{1}_{\{\imath > n - \lfloor n/\kappa \rfloor \kappa\}},$$
$$t_\imath \triangleq t_{\imath-1} + \ell_\imath.$$

The length n DSPC with feedback $W_{\overrightarrow{[1,n]}}$ can be interpreted as a length κ DPC[9] with feedback $U_{\overrightarrow{[1,\kappa]}}$ for $U_\imath : \mathcal{A}_\imath \to \mathcal{B}_\imath$ defined as follows

$$U_\imath \triangleq W_{\overrightarrow{[1+t_{\imath-1},t_\imath]}} \quad \forall \imath \in \{1, \ldots, \kappa\}.$$

As a result any (M, L) channel code (Ψ, Θ) on the channel $W_{\overrightarrow{[1,n]}} : (\bigtimes_{t=1}^{n} \mathcal{X}_t^{\mathcal{Y}_1^{t-1}}) \to \mathcal{P}(\mathcal{Y}_1^n)$ is also an (M, L) channel code on $U_{\overrightarrow{[1,\kappa]}} : (\bigtimes_{\imath=1}^{\kappa} \mathcal{A}_\imath^{\mathcal{B}_1^{\imath-1}}) \to \mathcal{P}(\mathcal{B}_1^\kappa)$ with exactly the same error probability. In the rest of the proof we work with the latter interpretation.

Since $W_t = W$ for all t, Lemma 12 and the definition of the sphere packing exponent imply

$$C_{\alpha,U_\imath} = \ell_\imath C_{\alpha,W} \quad (107)$$
$$E_{sp}(C_{\alpha,U_\imath}, U_\imath) = \ell_\imath E_{sp}(C_{\alpha,W}, W) \quad (108)$$

for all $\imath \in \{1, \ldots, \kappa\}$ and $\alpha \in (0, 1)$.

[8] The existence of such a z is established in Lemma 25.

[9] $U_{\overrightarrow{[1,n]}}$ is stationary iff $\ell_\imath$ is same for all $\imath$, i.e. iff n/κ is an integer.

We define ϕ and η by applying Lemma 25 to W for R defined as

$$R \triangleq \frac{1}{n}\ln\frac{M}{L} - \frac{C_{1/2,W}}{1-z}\left[\frac{2\epsilon}{\alpha_0(1-z)} + \frac{14}{\sqrt[3]{\kappa}}\right] - \frac{\kappa}{n}\ln\frac{1}{\epsilon}. \tag{109}$$

Then $\phi \in [\alpha_0, \alpha_1]$ by (105) and the monotonicity of $C_{\alpha,W}$ in α, i.e. Lemma 8-(a). Hence, the definition of z, the monotonicity of $\frac{1-\alpha}{\alpha}C_{\alpha,W}$ in α, i.e. Lemma 8-(b), and the monotonicity of $E_{sp}(R,W)$ in R, i.e. Lemma 13, imply $\eta \in [\alpha_0, z]$.

For each $\imath \in \{1,\ldots,\kappa\}$, we define $\phi_\imath$ and $\eta_\imath$ by applying Lemma 25 to $U_\imath$ for $\ell_\imath R$. Then $\phi_\imath = \phi$ and $\eta_\imath = \eta$ for all $\imath$ by (107) and (108). We denote $W_{f_\epsilon}^{q^\epsilon_{f_\epsilon},W}$ resulting from applying Lemma 25-(b,c) to $U_\imath$ by $V_\imath : \mathcal{A}_\imath \to \mathcal{P}(\mathcal{B}_\imath)$, i.e.

$$V_\imath(a) = U_{\imath\, f_\epsilon(a)}^{q^\epsilon_{f_\epsilon(a)},U_\imath}(a) \qquad \forall a \in \mathcal{A}_\imath. \tag{110}$$

Then $\phi \in [\alpha_0, \alpha_1]$, $\eta \in [\alpha_0, z]$, and Lemma 25-(b) imply

$$D_1(V_\imath(a)\| U_\imath(a)) \le E_{sp}(\ell_\imath R, U_\imath) + \frac{2\epsilon C_{1/2,U_\imath}}{\alpha_0^2(1-z)} \quad \forall a \in \mathcal{A}_\imath \tag{111}$$

On the other hand, $C_{1/2,U_\imath} \ge 2$ by (107) and the hypothesis $\lfloor\frac{n}{\kappa}\rfloor C_{1/2,W} \ge 2$. Thus Lemma 25-(c) implies

$$C_{\beta,V_\imath} \le \ell_\imath R + \frac{2\epsilon C_{1/2,U_\imath}}{\alpha_0(1-z)^2} + \ln\frac{1}{\epsilon} + (\beta-1)e^{(\beta-1)\frac{2C_{1/2,U_\imath}}{1-z}}\left[\frac{2C_{1/2,U_\imath}}{1-z}\right]^2$$

for all $\beta \in (1, \frac{1+z}{2z})$. Furthermore, $\lfloor\frac{n}{\kappa}\rfloor C_{1/2,W} \ge 2$ and $\kappa \ge 1$ imply $1 + \frac{\kappa^{2/3}(1-z)}{4nC_{1/2,W}} \le \frac{1+z}{2z}$. Hence

$$\begin{aligned} C_{\beta,V_\imath} &\le \ell_\imath R + \frac{2\epsilon C_{1/2,U_\imath}}{\alpha_0(1-z)^2} + \ln\frac{1}{\epsilon} + \frac{e^{1/\sqrt[3]{\kappa}}}{\sqrt[3]{\kappa}}\frac{2C_{1/2,U_\imath}}{1-z} \\ &\le \ell_\imath R + \frac{2\epsilon C_{1/2,U_\imath}}{\alpha_0(1-z)^2} + \ln\frac{1}{\epsilon} + \frac{6}{\sqrt[3]{\kappa}}\frac{C_{1/2,U_\imath}}{1-z} \end{aligned} \tag{112}$$

for $\beta = 1 + \frac{\kappa^{2/3}(1-z)}{4nC_{1/2,W}}$.

We use $V_\imath$'s described in (110) to define the length κ DPC with feedback $V_{\overrightarrow{[1,\kappa]}} : \left(\bigtimes_{\imath=1}^{\kappa}\mathcal{A}_\imath^{\mathcal{B}_1^{\imath-1}}\right) \to \mathcal{P}(\mathcal{B}_1^\kappa)$. Then using Lemma 12, (107), (109), and (112) we get

$$C_{\beta,V_{\overrightarrow{[1,n]}}} \le \ln\frac{M}{L} - n\frac{C_{1/2,W}}{1-z}\frac{8}{\sqrt[3]{\kappa}}$$

for $\beta = 1 + \frac{\kappa^{2/3}(1-z)}{4nC_{1/2,W}}$. We bound the average error probability of (Ψ, Θ) on $V_{\overrightarrow{[1,n]}}$, i.e. $P_{\mathbf{e}}^{V_{\overrightarrow{[1,n]}}}$, using (90) and $\tau \ge \ln(1+\tau)$:

$$\begin{aligned} P_{\mathbf{e}}^{V_{\overrightarrow{[1,n]}}} &\ge 1 - e^{-\sqrt[3]{\kappa}} \\ &\ge \frac{\sqrt[3]{\kappa}}{1+\sqrt[3]{\kappa}}. \end{aligned} \tag{113}$$

On the other hand (107), (108), and (111) imply

$$D_1\left(V_{\overrightarrow{[1,\kappa]}}(x)\middle\| U_{\overrightarrow{[1,\kappa]}}(x)\right) \le nE_{sp}(R,W) + n\frac{2\epsilon C_{1/2,W}}{\alpha_0^2(1-z)} \tag{114}$$

for all $x \in \left(\bigtimes_{\imath=1}^{\kappa}\mathcal{A}_\imath^{\mathcal{B}_1^{\imath-1}}\right)$.

Let p be the probability distribution generated by the encoder Ψ on the input set of $U_{\overrightarrow{[1,\kappa]}}$, i.e. on $\left(\bigtimes_{\imath=1}^{\kappa}\mathcal{A}_\imath^{\mathcal{B}_1^{\imath-1}}\right)$, for the uniform distribution over the message set. Then Lemma 3 and the identity $\tau\ln\tau + (1-\tau)\ln(1-\tau) \ge \ln{}^1\!/_2$, which holds for all $\tau \in [0,1]$, imply

$$D_1\left(p \circledast V_{\overrightarrow{[1,\kappa]}}\middle\| p \circledast U_{\overrightarrow{[1,\kappa]}}\right) \ge \ln{}^1\!/_2 - P_{\mathbf{e}}^{V_{\overrightarrow{[1,n]}}}\ln P_{\mathbf{e}}^{av}. \tag{115}$$

Note that $D_1\left(p \circledast V_{\overrightarrow{[1,\kappa]}}\middle\| p \circledast U_{\overrightarrow{[1,\kappa]}}\right)$ is bounded from above by the supremum of $D_1\left(V_{\overrightarrow{[1,\kappa]}}(x)\middle\| U_{\overrightarrow{[1,\kappa]}}(x)\right)$ over the common input set of $V_{\overrightarrow{[1,\kappa]}}$ and $U_{\overrightarrow{[1,\kappa]}}$, i.e. $\bigtimes_{\imath=1}^{\kappa}\mathcal{A}_\imath^{\mathcal{B}_1^{\imath-1}}$. Then using (113), (114), and (115) we get

$$\ln P_{\mathbf{e}}^{av} \ge -\frac{1+\sqrt[3]{\kappa}}{\sqrt[3]{\kappa}}\left(nE_{sp}(R,W) + n\frac{2\epsilon C_{1/2,W}}{\alpha_0^2(1-z)} + \ln 2\right).$$

Then using the identity $E_{sp}\left(C_{\phi,W}, W\right) \le \frac{(1-\phi)C_{\phi,W}}{\phi}$, which is implied by Lemma 13, together with (10), (105), and (109) we get

$$\ln P_{\mathbf{e}}^{av} \ge -nE_{sp}(R,W) + n\frac{C_{1/2,W}}{\alpha_0}\left[\frac{4\epsilon}{\alpha_0(1-z)} + \frac{1}{\sqrt[3]{\kappa}}\right] + 2\ln 2.$$

Then (106) is implied by (109) via the following consequence of Lemma 13: If $R = C_{\alpha,W}$ for an $\alpha \in [\alpha_0, \alpha_1]$,

$$E_{sp}(R,W) \le E_{sp}(R+\delta, W) + \frac{(1-\alpha_0)}{\alpha_0}\delta \qquad \forall \delta \ge 0.$$

□

Remark 9: The input sets of the subblocks grow rapidly with their length; in particular

$$\ln|\mathcal{A}_\imath| = \left(\sum\nolimits_{\jmath=0}^{\ell_\imath-1}|\mathcal{Y}|^\jmath\right)\ln|\mathcal{X}|.$$

This rapid growth would have made our bounds useless, at least for establishing a result in the spirit Lemma 26, if the approximation error terms in Lemma 25-(b,c) were in terms of $\ln|\mathcal{X}|$ rather than $C_{1/2,W}$, which grows only linearly with the length of the subblock.

One is initially inclined to use Lemma 26 either for $\kappa = n$ or for $\kappa = 1$, i.e. apply Lemma 25 either to $W_{\overrightarrow{[1,n]}}$ or to the component channel W. Both of these choices, however, lead to poor approximation error terms. Instead we use Lemma 26 for $\kappa \approx n^{3/4}$ to prove Theorem 3. In [7], while proving a statement similar to Theorem 3, Augustin used subblocks in a similar fashion; other ingredients of Augustin's analysis, however, are quite different. Palaiyanur discussed Augustin's proof sketch in more detail in his thesis [48, A.8]. A complete proof following Augustin's sketch can be found in [41].

Proof of Theorem 3: We prove Theorem 3 by applying Lemma 26 for appropriately chosen ϵ_n and κ_n. Note that ϵ and κ can take any value as long as the hypothesis of Lemma 26 is satisfied. For $\epsilon_n = \frac{\alpha_0(1-\eta)}{\sqrt[4]{n}}$ and $\kappa_n = \lfloor n^{3/4}\rfloor$, the hypothesis $\lfloor\frac{\kappa}{n}\rfloor C_{1/2,W} \ge 2$ holds for all n large enough. Furthermore, the other hypothesis of Lemma 26 given in (105) is satisfied for n large enough because of (81). Thus we can apply Lemma 26 with $\epsilon_n = \frac{\alpha_0(1-\eta)}{\sqrt[4]{n}}$ and $\kappa_n = \lfloor n^{3/4}\rfloor$ for all n large enough. Then (106) implies (82) for n large enough. □

D. Extensions and Comparisons

Theorem 3 is stated for stationary sequences of channels, but it holds for periodic sequences of channels too. In other words, Theorem 3 assumed $W_t = W$ for all $t \in \mathbb{Z}_+$; but its assertions hold whenever there exists a $\tau \in \mathbb{Z}_+$ satisfying $W_t = W_{t+\tau}$ for all $t \in \mathbb{Z}_+$. Thus the SPB holds for codes on the periodic discrete product channels, as well.

It is possible establish similar results under weaker stationarity hypotheses. In order to prove the SPB for codes on the DPCs with feedback using the approach employed for proving Theorem 3, we need the Rényi capacity of the subblocks to be approximately equal to one another as functions, i.e.

$$C_{\alpha,U_\imath} \approx \frac{\ell_\imath}{n} C_{\alpha,W_{[1,n]}}$$

uniformly over $\imath$ and α. This condition is a stationarity hypotheses too; but it is considerably weaker than assuming all W_t's to be identical. There is not just one but many precise ways to impose this condition and each one of them leads to a slightly different result. Assumption 4 and Theorem 4 are provided as examples. In order to prove Theorem 4 we need to modify Lemmas 25 and 26, slightly. We present those modifications, their proofs, and the proof of Theorem 4 in [37, Appendix A].

Assumption 4: $\{W_t\}_{t\in\mathbb{Z}_+}$ is a sequence of channels satisfying the following three conditions for some $\varphi : (0,1) \to \Re_+$

i. $\lim_{n\to\infty} \frac{1}{n} C_{\alpha,W_{[1,n]}} = \varphi(\alpha)$ for all $\alpha \in (0,1)$.
ii. $\lim_{\alpha\uparrow 1} \frac{1-\alpha}{\alpha}\varphi(\alpha) = 0$.
iii. There exists $K \in \Re_+$ and $n_0 \in \mathbb{Z}_+$ such that

$$\sup_{\alpha\in(0,1)} \sup_{t\in\mathbb{Z}_+} \left| C_{\alpha,W_{[t,t+n-1]}} - n\varphi(\alpha)\right| \le K \ln n \quad \forall n \ge n_0.$$

Theorem 4: Let $\{W_t\}_{t\in\mathbb{Z}_+}$ be a sequence of discrete channels satisfying Assumption 4 and α_0, α_1 be orders satisfying $0 < \alpha_0 < \alpha_1 < 1$. Then for any sequence of codes on the discrete product channels with feedback $\{W_{\overrightarrow{[1,n]}}\}_{n\in\mathbb{Z}_+}$ satisfying

$$C_{\alpha_1,W_{[1,n]}} \ge \ln \frac{M_n}{L_n} \ge C_{\alpha_0,W_{[1,n]}} + (K+1)n^{3/4} \ln n \quad \forall n \ge n_0 \tag{116}$$

there exists an $n_1 \ge n_0$ such that

$$P_{\mathbf{e}}^{av(n)} \ge e^{-E_{sp}\left(\ln \frac{M_n}{L_n}, W_{[1,n]}\right) - \frac{6K+1}{\alpha_0} n^{3/4} \ln n} \quad \forall n \ge n_1. \tag{117}$$

We have confined the claims of Theorem 4 to discrete channels in order avoid certain measurability issues. We believe, however, it should be possible to resolve those issues and to extend Theorem 4 to any sequence of channels satisfying Assumption 4. Augustin [7, Corollary 41.9] makes the same conjecture for the stationary channels.

Augustin sketches a derivation of the SPB for codes on finite input set SPCs with feedback in [7, Sec. 41]. The approximation error terms in Augustin's asymptotic SPB [7, Th. 41.7] are $O\left(n^{-1/3} \ln n\right)$ rather than $O\left(n^{-1/4} \ln n\right)$. A complete proof of SPB for codes on DSPCs with feedback following Augustin's sketch can be found in [41].

Throughout this section, we have refrained from making any assumptions on the Rényi centers of the component channels or their relation to the output distributions of the component channels. Such assumptions may lead to sharper bounds under milder stationarity hypotheses. For example, Lemma 23 establishes a non-asymptotic bound for certain product channels with feedback that can be used, in place of Lemma 20, to prove the asymptotic SPB given in Theorem 2 under Assumption 1. Thus if the sequence $\{W_t\}_{t\in\mathbb{Z}_+}$ satisfies Assumption 1 and every channel in $\{W_t\}_{t\in\mathbb{Z}_+}$ satisfies the hypothesis of Lemma 23, then the SPB holds with a polynomial prefactor for codes on $W_{\overrightarrow{[1,n]}}$. First Dobrushin [20] and then Haroutunian [32] employed similar observations to establish the SPB for codes on certain DSPCs with feedback. Later, Augustin [7, p. 318] did the same for codes on certain product channels with feedback.

E. Haroutunian's Bound and Subblocks

Haroutunian's article [32] is probably the most celebrated work on the exponential lower bounds to the error probability of channel codes on DSPCs with feedback. In the rest of this section we discuss [32] in light of Lemma 25 and the concept of subblocks.

Haroutunian [32] considers (M, L) channel codes satisfying $R = \frac{1}{n} \ln \frac{M}{L}$ on DSPCs with feedback $W_{\overrightarrow{[1,n]}}$ satisfying $W_t = W$ for a $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ to prove that for any rate $R \ge 0$ and $\varepsilon > 0$ the following bound holds for large enough n

$$P_{\mathbf{e}}^{av(n)} \ge (1-\varepsilon) e^{-n(E_h(R-\varepsilon,W)+\varepsilon)} \tag{118}$$

where $E_h(R, W)$, which is customarily called Haroutunian's exponent, is defined as, [16, p. 180], [32, eq. (15)],

$$E_h(R, W) \triangleq \inf_{V:C_{1,V}\le R} \sup_{x\in\mathcal{X}} D_1\left(V(x) \| W(x)\right). \tag{119}$$

Haroutunian points out not only that $E_h(R, W)$ is greater than or equal to $E_{sp}(R, W)$ for all R, but also that the inequality is strict on $(C_{0,W}, C_{1,W})$ even for most of the binary input binary output channels, see [32, Th. 3.1]. Thus, for certain W's there does not exist any V satisfying both $C_{1,V} \le R$ and $\sup_{x\in\mathcal{X}} D_1\left(V(x) \| W(x)\right) \le E_{sp}(R, W)$ at the same time.

On the other hand, both inequalities are satisfied approximately for $V = W_{f_\epsilon}^{q_{f_\epsilon,W}^\epsilon}$ by Lemma 25-(b,c). In particular,

$$\sup_{x\in\mathcal{X}} D_1\left(V(x) \| W(x)\right) \le E_{sp}(R, W) + \frac{2\epsilon C_{1/2,W}}{\phi^2(1-\eta)} \tag{120}$$

$$C_{1,V} \le R + \frac{2\epsilon C_{1/2,W}}{\phi(1-\phi)^2} + \ln\frac{1}{\epsilon} \tag{121}$$

for any $\epsilon \in (0, \phi/2)$ where ϕ and η are determined uniquely by $C_{\phi,W} = R$ and $\frac{1-\eta}{\eta} = E_{sp}(R, W)$.

If we apply (120) and (121) to $W_{\overrightarrow{[1,n]}}$ for $R_n = n(R-\varepsilon)$ and $\epsilon_n = 1/n$, then the additivity of the Rényi capacity for the product channels with feedback, i.e. Lemma 12, and monotonicity of $E_h(R, W)$ in R, i.e. [32, Th. 3.5], imply

$$\limsup_{n\to\infty} \frac{1}{n} E_h(nR, W_{\overrightarrow{[1,n]}}) \le E_{sp}(R-\varepsilon, W) \quad \forall \varepsilon > 0.$$

Then the continuity of $E_{sp}(R, W)$ in R, i.e Lemma 13, and the identity $E_{sp}(R, W) \le E_h(R, W)$ imply

$$\lim_{n\to\infty} \frac{1}{n} E_h(nR, W_{\overrightarrow{[1,n]}}) = E_{sp}(R, W). \tag{122}$$

Recall that any channel code on $W_{\overrightarrow{[1,n\ell]}}$ is also a channel code on $U_{\overrightarrow{[1,n]}}$ where $U_t = W_{\overrightarrow{[1,\ell]}}$ for all t, see the discussion of the concept of subblocks in the beginning of §V-C. Thus (118) implies that for any $\varepsilon > 0$ for large enough n

$$P_{\mathbf{e}}^{av(\ell n)} \geq (1-\varepsilon) e^{-n(E_h(\ell R-\varepsilon, W_{\overrightarrow{[1,\ell]}})+\varepsilon)}.$$

Then (122) implies

$$\limsup_{n\to\infty} \frac{1}{n} \ln \frac{1}{P_{\mathbf{e}}^{av(n)}} \leq E_{sp}(R, W).$$

Thus Lemma 25, the concept of subblocks, and Haroutunian's bound imply the most important asymptotic conclusion of Theorem 3, i.e. the reliability function of the DSPC with feedback is bounded from above by the sphere packing exponent.

Remark 10: Palaiyanur and Sahai [47] used the method of types to establish the following relation for all discrete channels W,

$$\lim_{n\to\infty} \frac{1}{n} E_h(nR, W_{[1,n]}) = E_{sp}(R, W). \tag{123}$$

This was first reported in Palaiyanur's thesis [48, Lemma 7]. Note that (123) cannot be used in the preceding argument to establish the sphere packing exponent as an upper bound to the reliability function of the DSPCs with feedback because $W_{\overrightarrow{[1,n\ell]}}$ is not equivalent to $B_{\overrightarrow{[1,n]}}$ for $B_t = W_{[1,\ell]}$.

It is worth mentioning that (122) implies (123) by the definition of Haroutunian's exponent.

VI. Discussion

We have established SPBs with approximation error terms that are polynomial in the block length for a class of product channels, which includes all stationary product channels. Our results hold for a large class of non-stationary product channels, which might have infinite channel capacity.

We have presented a new proof of the SPB for the codes on DSPCs with feedback that can be applied to the codes on DPCs with feedback satisfying a milder stationarity hypothesis, see §V-D and [37, Appendix A]. The validity of SPB for codes on DSPCs with feedback implies improvements in the bounds for codes with errors-and-erasures decoding on DSPCs with feedback that were previously derived using Haroutunian's bound in [43, Secs. IV and V] and [44, Secs. 2.4 and 2.5].

In our judgment, the averaging described in §III-A is one way of employing the following more fundamental observation

$$\lim_{\phi\to\alpha} S_{\alpha,W}(q_{\phi,W}) = S_{\alpha,W} \qquad \forall \alpha \in (0,1).$$

The preceding observation and Theorem 1, are at the heart of Augustin's method. However, only the preceding observation can be interpreted as a novelty of Augustin's method because Theorem 1 is employed by Shannon *et al.* [54], albeit in an indirect way and for discrete channels only.[10]

In §IV and §V, we have confined our discussion of the SPB to the product channels. The Rényi capacity and center, as defined in §II-C, served our purposes satisfactorily. For studying the SPB on the memoryless channels, however, the Augustin capacity and center, described below, are better suited. The Rényi information has multiple non-equivalent definitions. The following definition was proposed and analyzed by Augustin [7, Sec. 34] and later popularized by Csiszár [13]:

$$I_\alpha^c(p; W) \triangleq \inf_{q\in\mathcal{P}(\mathcal{Y})} \sum\nolimits_x p(x) D_\alpha(W(x)\| q).$$

We have called this quantity Augustin information in [38] and [39]. The Augustin capacity and center are defined analogously to the Rényi capacity and center.[11] Using these concepts and assuming a bounded cost function Augustin derived a SPB for channel codes on the cost constrained memoryless channels in [7, Ch. VII]. Augustin's framework is general enough to subsume the Poisson channels described in (1) as special cases in the way that the framework of Theorem 2 subsumed the Poisson channels described in (1d) and (1e) as special cases. The Gaussian channels studied by Ebert [22], Richters [50], and Shannon [53], however, are not subsumed by Augustin's framework because the quadratic cost function used for these channels is not bounded. To remedy this situation, we have recently derived a SPB with a polynomial prefactor for codes on the cost constrained (possibly non-stationary) memoryless channels, [38], [42], without assuming the boundedness of the cost function. We have also derived the SPB for codes on the stationary memoryless channels with convex composition constraints on the codewords in [38] and [42]. It seems extending the results to the channels with memory is the pressing issue in this line of work; but that is likely to be more challenging than the case of memoryless channels.

Acknowledgment

The author would like to thank Fatma Nakiboğlu and Mehmet Nakiboğlu for their hospitality; this work simply would not have been possible without it. 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 [7], Vincent Tan for pointing out [52, Th. 35-(b), (469)], Reviewer III for pointing out [19], and Güneş Nakiboğlu and the reviewers for their suggestions on the manuscript.

References

[1] Y. Altuğ and A. B. Wagner, "Refinement of the sphere packing bound for symmetric channels," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2011, pp. 30–37.

[2] Y. Altuğ and A. B. Wagner, "Moderate deviations in channel coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4417–4426, Aug. 2014.

[3] Y. Altuğ and A. B. Wagner, "Refinement of the random coding bound," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6005–6023, Oct. 2014.

[4] Y. Altuğ and A. B. Wagner, "Refinement of the sphere-packing bound: Asymmetric channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1592–1614, Mar. 2014.

[5] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 3, pp. 357–359, May 1973.

[10]The equality of the Rényi capacity to the Rényi radius and the existence of a Rényi center is invoked via [54, eq. (4.22)]. The uniqueness of the Rényi center is implicit in the analysis of $\mathtt{f}_\mathtt{s}$ as a function of $\mathtt{s}$; it is established in the discussion between [54, (A27) and (A28)].

[11]The constrained Rényi capacity $C_{\alpha,W,\mathcal{A}}$ is defined by taking the supremum of the Rényi information over the priors in a subset $\mathcal{A}$ of $\mathcal{P}(\mathcal{X})$, rather than $\mathcal{P}(\mathcal{X})$ itself, see [40, Appendix A]. The unconstrained Rényi and Augustin capacities are equal, see [13, Propostion 1] or [39, Ths. 2 and 3]; however, this is not the case in general for the constrained capacities.

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

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

[8] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Trans. Amer. Math. Soc.*, vol. 49, no. 1, pp. 122–136, Jan. 1941.

[9] V. I. Bogachev, *Measure Theory*. Berlin, Germany: Springer-Verlag, 2007.

[10] M. V. Burnashev and Y. A. Kutoyants, "On the sphere-packing bound, capacity, and similar results for Poisson channels," *Problems Inf. Transmiss.*, vol. 35, no. 2, pp. 95–111, 1999.

[11] 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.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

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

[14] I. Csiszár, "The method of types [information theory]," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

[15] 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. IT-28, no. 1, pp. 92–93, Jan. 1982.

[16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[17] M. Dalai, "Lower bounds on the probability of error for classical and classical-quantum channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8027–8056, Dec. 2013.

[18] M. Dalai, "Some remarks on classical and classical-quantum sphere packing bounds: Rényi vs. Kullback–Leibler," *Entropy*, vol. 19, no. 7, p. 355, 2017.

[19] M. Dalai and A. Winter, "Constant compositions in the sphere packing bound for classical-quantum channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5603–5617, Sep. 2017.

[20] 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.

[21] R. M. Dudley, *Real Analysis and Probability*, vol. 74. New York, NY, USA: Cambridge Univ. Press, 2002.

[22] 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

[23] P. Elias, "List decoding for noisy channels," Res. Lab. Electron., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 335, 1957. [Online]. Available: http://hdl.handle.net/1721.1/4484

[24] T. van Erven and P. Harremos, "Rényi divergence and Kullback-Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014.

[25] C. G. Esseen, "On the Liapunoff limit of error in the theory of probability," *Ark. Mat. Astr. Fys.*, vol. A28, pp. 1–19, 1942.

[26] S. L. Fong and V. Y. F. Tan, "Strong converse theorems for classes of multimessage multicast networks: A Rényi divergence approach," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 4953–4967, Sep. 2016.

[27] 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.

[28] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[29] A. Guntuboyina, S. Saha, and G. Schiebinger, "Sharp inequalities for f-divergences," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 104–121, Jan. 2014.

[30] T. S. Han, *Information-Spectrum Methods in Information Theory* (Stochastic Modelling and Applied Probability). Berlin, Germany: Springer, 2003.

[31] 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.

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

[33] E. A. Haroutunian, M. E. Haroutunian, and A. N. Harutyunyan, *Reliability Criteria in Information Theory and in Statistical Hypothesis Testing* (Foundations and Trends in Communications and Information), vol. 4, nos. 2–3. Delft, The Netherlands: Now, 2007, pp. 97–263.

[34] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[35] P. Moulin, "The log-volume of optimal codes for memoryless channels, asymptotically within a few nats," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, pp. 2278–2313, Apr. 2017.

[36] J. Munkres, *Topology*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[37] B. Nakiboğlu. (2016). "The sphere packing bound via Augustin's method." [Online]. Available: https://arxiv.org/abs/1611.06924

[38] B. Nakiboğlu, "The Augustin center and the sphere packing bound for memoryless channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 1401–1405.

[39] B. Nakiboğlu. (2018). "The Augustin capacity and center." [Online]. Available: https://arxiv.org/abs/1803.07937

[40] B. Nakiboğlu. "The Rényi capacity and center," *IEEE Trans. Inf. Theory*, to be published. [Online]. Available: https://arxiv.org/abs/1608.02424, doi: 10.1109/TIT.2018.2861002.

[41] B. Nakiboğlu. (2018). "The sphere packing bound for DSPCs with feedback a la Augustin." [Online]. Available: https://arxiv.org/abs/1806.11531

[42] B. Nakiboğlu. (2018). "The sphere packing bound for memoryless channels." [Online]. Available: https://arxiv.org/abs/1804.06372

[43] B. Nakiboğlu and L. Zheng, "Errors-and-erasures decoding for block codes with feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 24–49, Jan. 2012.

[44] B. Nakiboğlu, "Exponential bounds on error probability with feedback," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, Feb. 2011. [Online]. Available: http://hdl.handle.net/1721.1/64485

[45] J. K. Omura, "A lower bounding method for channel and source coding probabilities," *Inf. Control*, vol. 27, no. 2, pp. 148–177, Feb. 1975.

[46] 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.

[47] H. Palaiyanur and A. Sahai, "On Haroutunian's exponent for parallel channels and an application to fixed-delay codes without feedback," *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1298–1308, Mar. 2015.

[48] 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

[49] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[50] 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

[51] W. Rudin, *Principles of Mathematical Analysis*. New York, NY, USA: McGraw-Hill, 1976.

[52] I. Sason and S. Verdú, "f-divergence inequalities," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 5973–6006, Nov. 2016.

[53] 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.

[54] 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.

[55] 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.

[56] I. G. Shevtsova, "An improvement of convergence rate estimates in the Lyapunov theorem," *Doklady Math.*, vol. 82, no. 3, pp. 862–864, 2010.

[57] A. N. Shiryaev, *Probability*. New York, NY, USA: Springer-Verlag, 1995.

[58] V. Strassen, "Asymptotic estimates in Shannon's information," in *Proc. Trans. 9th Prague Conf.*, 1962, pp. 689–723. [Online]. Available: http://www.math.cornell.edu/~pmlut/strassen.pdf

[59] M. Tomamichel and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics for most discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7041–7051, Nov. 2013.

[60] A. Valembois and M. P. C. Fossorier, "Sphere-packing bounds revisited for moderate block lengths," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 2998–3014, Dec. 2004.

[61] G. Wiechman and I. Sason, "An improved sphere-packing bound for finite-length codes over symmetric memoryless channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1962–1990, May 2008.

[62] 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.

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, Massachusetts, in 2005 and 2011, respectively. He was a postdoctoral scholar at the University of California Berkeley during the 2012-2013 academic year. He is with the Department of Electrical and Electronics Engineering at METU.