Refined Strong Converse for the Constant Composition Codes

Hao-Chung Cheng
Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Cambridge CB3 0WA, U.K.
Email: haochung.ch@gmail.com

Barış Nakiboğlu
Department of Electrical and Electronics Engineering
Middle East Technical University (METU)
06800 Ankara, Turkey
Email:bnakib@metu.edu.tr

***Abstract*—A strong converse bound for constant composition codes of the form $P_{\mathbf{e}}^{(n)} \geq 1 - An^{-0.5(1-E'_{sc}(R,W,p))}e^{-nE_{sc}(R,W,p)}$ is established using the Berry–Esseen theorem through the concepts of Augustin information and Augustin mean, where A is a constant determined by the channel W, the composition p, and the rate R, i.e., A does not depend on the block length n.**

I. Introduction

On discrete stationary product channels, the error probability of codes operating at rates above capacity is not only bounded away from zero but also converging to one. This property, first observed by Wolfowitz [1], is called the strong converse property. For arbitrary stationary product channels, a necessary and sufficient condition for the strong converse property was determined by Augustin [2, §10], [3, §13]. The strong converse property does not hold in general; nevertheless there does exist a universal asymptotic constant that bounds the error probability of codes operating at rates above the capacity of the stationary product channels, according to Beck and Csiszar [4]. In [5], Verdú and Han provided a necessary and sufficient condition for the strong converse property for channels that are not necessarily stationary or memoryless.

Using the concept of Rényi capacity, which was employed earlier by Gallager [6] for analyzing the error probability of codes operating at rates below the channel capacity, Arimoto established in [7] the following lower bound to the error probability of codes on discrete stationary product channels (DPSCs) operating at a rate R above the channel capacity:

$$P_{\mathbf{e}}^{(n)} \geq 1 - e^{-nE_{sc}(R)} \tag{1}$$

where $E_{sc}(\cdot)$ is the strong converse exponent of the channel. Although Arimoto's initial proof in [7] is for DSPCs, Arimoto's lower bound can be proved as a one shot bound for more general channel models using Jensen's inequality or Hölders inequality as noted by Augustin [3] and Sheverdyaev [8], see also [9]–[11]. Arimoto's lower bound is used to establish the strong converse on channels for which alternative derivations of the strong converse is much more tedious, e.g. quantum channels discussed in [11]–[21] and Poisson channels mentioned in [9, Appendices B-B and B-C].

Aritmoto's lower bound to the error probability has been derived for certain constrained codes on memoryless channels as well, see Dueck and Körner [22] for the constant composition codes on DSPCs, Oohama [23] for the Gaussian channel, and Cheng *et al.* [21] and Mosonyi and Ogawa [24] for the constant composition codes on classical-quantum channels.

For codes on DSPCs Omura [25] has shown

$$\lim_{n\to\infty} -\tfrac{1}{n}\ln(1 - P_{\mathbf{e}}^{(n)}) \leq E_{sc}(R) \tag{2}$$

for all rates above the channel capacity and below a certain threshold. Dueck and Körner [22] established (2) for all rates above the channel capacity. Thus Arimoto's bound is tight, in terms of the exponential decay rate of the probability of correct decoding with block length, for all rates above the channel capacity. An analogous result is derived for constant composition codes on DSPCs in [22], for the Gaussian channel in [23], for classical-quantum channels in [18], for classical data compression with quantum side information in [20], and for constant composition codes on classical-quantum channels in [24].

Although Arimoto's bound, given in (1), is tight in terms of the exponential decay rate of the correct decoding probability with block length, the prefactor multiplying the exponentially decaying term can be improved. In particular, for the constant composition codes operating at rates larger than the mutual information of the composition. Theorems 1 and 2, in the following, establish a strong converse bound of the form

$$P_{\mathbf{e}}^{(n)} \geq 1 - An^{-0.5(1-E'_{sc}(R))}e^{-nE_{sc}(R)} \tag{3}$$

where $E_{sc}(\cdot)$ is the strong converse exponent and $E'_{sc}(\cdot)$ is its derivative with respect to the rate. Since $1 \geq E'_{sc}(R) \geq 0$ for all rates R and $E'_{sc}(R) < 1$ for small enough rates R, the bound (3) improves (1) strictly. In accordance with the corresponding improvements of the sphere packing bound for rates below the channel capacity given in [26]–[30], we call the bounds of the form given in (3) refined strong converses.

Proof of Theorem 1 is analogous to the proof of refined sphere packing bound presented in [29], [30]: it relies on a tight characterization of the trade-off between type-I and type-II error probabilities in the hypothesis testing problem with (possibly non-stationary) independent samples through the concepts of Augustin information and mean. However, in [29], [30] for the regime of interest the optimal tilting

The work was done partially while the authors were visiting the Institute for Mathematical Sciences, National University of Singapore in 2017. The visit was supported by the Institute. This work is also supported by the Ministry of Science and Technology, Taiwan (R.O.C.), under Grant 108-2917-I-564-042 and 109WXA0310019, the Science Academy, Turkey, under The Science Academy's Young Scientist Awards Program (BAGEP), and by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Grant 119E053.

parameter is between zero and one; whereas we are now interested in the regime where the optimal tilting parameter is larger than one. Similarly, in [29], [30] Agustin information measures for orders between zero and one are used together with the sphere packing exponent, whereas we employ Agustin information measures for orders larger than one together with the strong converse exponent in our analysis.

We conclude this section with an overview of the paper. In §II, we describe our model and notation. In §III, we employ the concept of tilted probability measure and the Berry–Esseen theorem to obtain a lower bound on the type-II error probability in hypothesis testing problem with independent —but not necessarily identically distributed—samples for the regime where the optimal tilting parameter is larger than one. In §IV, we review Augustin's information measures and the strong converse exponent. In §V, we establish a refined strong converse for the constant composition codes on stationary memoryless channels. We conclude our presentation with a brief discussion of the results and future work in §VI.

II. Model and Notation

We denote the set of all probability mass functions that are positive only for finitely many elements of $\mathcal{X}$ by $\mathcal{P}(\mathcal{X})$ and the set of all probability measures on a measurable space $(\mathcal{Y},\mathscr{Y})$ by $\mathcal{P}(\mathscr{Y})$. The $\mathcal{L}^1$ norm of a measure μ is denoted by $\|\mu\|$. The expected value and variance of a measurable function f under the probability measure μ are denoted by $\mathbf{E}_{\mu}[f]$ and $\mathbf{V}_{\mu}[f]$. The Cartesian product of sets $\mathcal{X}_1,\ldots,\mathcal{X}_n$ is denoted by $\mathcal{X}_1^n$; the product of σ-algebras $\mathscr{Y}_1,\ldots,\mathscr{Y}_n$ is denoted by $\mathscr{Y}_1^n$. The symbol $\otimes$ is used to denote both products of σ-algebras and products of measures.

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

$$W:\mathcal{X}\to\mathcal{P}(\mathscr{Y}). \tag{4}$$

If $\mathcal{X}$ and $\mathcal{Y}$ are both finite sets, then W is a *discrete channel*. The product of $W_t:\mathcal{X}_t\to\mathcal{P}(\mathscr{Y}_t)$ for $t\in\{1,\ldots,n\}$ is a channel of the form $W_{[1,n]}:\mathcal{X}_1^n\to\mathcal{P}(\mathscr{Y}_1^n)$ satisfying

$$W_{[1,n]}(x_1^n)=\bigotimes\nolimits_{t=1}^{n} W_t(x_t) \qquad \forall x_1^n\in\mathcal{X}_1^n. \tag{5}$$

Any channel obtained by curtailing the input set of a length n product channel is called a length n *memoryless channel*. A product channel $W_{[1,n]}$ is *stationary* iff $W_t=W$ for all t's for some W. On a stationary channel, we denote the composition (i.e. the empirical distribution, the type) of each x_1^n by $\Upsilon(x_1^n)$; thus $\Upsilon(x_1^n)\in\mathcal{P}(\mathcal{X})$.

An (M,L) *channel code* on $W_{[1,n]}$ is composed of an *encoding function* Ψ from the message set $\mathcal{M}\triangleq\{1,2,\ldots,M\}$ to the input set $\mathcal{X}_1^n$ and a $\mathscr{Y}_1^n$-measurable *decoding function* Θ from the output set $\mathcal{Y}_1^n$ to $\widehat{\mathcal{M}}\triangleq\{\mathcal{L}:\mathcal{L}\subset\mathcal{M}\text{ and }|\mathcal{L}|\leq L\}$. For any channel code (Ψ,Θ) on $W_{[1,n]}$, *the conditional error probability* $P_{\mathbf{e}}^m$ for $m\in\mathcal{M}$ and *the average error probability* $P_{\mathbf{e}}$ are defined as

$$P_{\mathbf{e}}^m\triangleq\mathbf{E}_{W_{[1,n]}(\Psi(m))}\left[\mathbb{1}_{\{m\notin\Theta(\mathsf{Y}_1^n)\}}\right],$$
$$P_{\mathbf{e}}\triangleq\tfrac{1}{M}\sum\nolimits_{m\in\mathcal{M}}P_{\mathbf{e}}^m.$$

A channel code is a constant composition code iff all of its codewords have the same composition, i.e. $\exists p\in\mathcal{P}(\mathcal{X})$ such that $\Upsilon(\Psi(m))=p$, $\forall m\in\mathcal{M}$.

III. Hypothesis Testing Problem, Tilted Probability Measure, and Berry Esseen Theorem

Our main aim in this section is to characterize the trade-off between type-I and type-II error probabilities using the Berry-Essen theorem and the tilted probability measure. This trade-off can be studied in various regimes; in order to specify the regime of interest, let us first recall the definition of Rényi divergence and define the tilted probability measure.

Definition 1. For any $\alpha\in\mathbb{R}_+$ and $w,q\in\mathcal{P}(\mathscr{Y})$, *the order* α *Rényi divergence between* w *and* q is

$$D_{\alpha}(w\|\,q)\triangleq\begin{cases}\frac{1}{\alpha-1}\ln\int(\frac{\mathrm{d}w}{\mathrm{d}\nu})^{\alpha}(\frac{\mathrm{d}q}{\mathrm{d}\nu})^{1-\alpha}\nu(\mathrm{d}y) & \alpha\neq1\\ \int\frac{\mathrm{d}w}{\mathrm{d}\nu}\left[\ln\frac{\mathrm{d}w}{\mathrm{d}\nu}-\ln\frac{\mathrm{d}q}{\mathrm{d}\nu}\right]\nu(\mathrm{d}y) & \alpha=1\end{cases}$$

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

Definition 2. For any $w,q\in\mathcal{P}(\mathscr{Y})$, let w_{ac} be the component of w that is absolutely continuous in q. If $\|w_{ac}\|\neq0$, then *the order* 1 *tilted probability measure* w_1^q is

$$w_1^q\triangleq\tfrac{w_{ac}}{\|w_{ac}\|}.$$

Furthermore, for any $\alpha\in\mathbb{R}_+$ satisfying $D_{\alpha}(w_1^q\|\,q)<\infty$, *the order* α *tilted probability measure* w_{α}^q is defined in terms of its Radon-Nikodym derivative with respect to q as follows

$$\tfrac{\mathrm{d}}{\mathrm{d}q}w_{\alpha}^q\triangleq e^{(1-\alpha)D_{\alpha}\left(w_1^q\|\,q\right)}\left(\tfrac{\mathrm{d}w_1^q}{\mathrm{d}q}\right)^{\alpha}.$$

The definition of tilted probability measure used in [9], [30], [31], employs w in the place of w_1^q. Whenever $w_1^q=w$, i.e. whenever $w\prec q$, it is equivalent to Definition 2. For orders in $(0,1)$ these two definitions are equivalent even if $w_1^q\neq w$. For orders larger than or equal to one, they differ only when $w_1^q\neq w$ and $D_{\alpha}(w_1^q\|\,q)<\infty$. In this case, $D_{\alpha}(w\|\,q)=\infty$ for all α's in $[1,\infty)$ and w_{α}^q is not defined according to the definition used [9], [30], [31] but w_{α}^q is defined according to Definition 2.

In order to see why Definition 2 can be more relevant than the one used in [9], [30], [31], let us consider two probability measures w and q for which $w_1^q\neq w$ and $D_{\beta}(w_1^q\|\,q)$ is finite for some $\beta>1$. Then both $D_1(w_{\alpha}^q\|\,q)$ and $D_1(w_{\alpha}^q\|\,w_1^q)$ are analytic, and hence continuous, functions of the order α on $(0,\beta)$ by [31, Lemma 11]. On the other hand as, a result of Pinsker's inequality [32, Thm. 31] we have

$$\|w_{\alpha}^q-w_1^q\|\leq\sqrt{2D_1(w_{\alpha}^q\|\,w_1^q)}.$$

Thus w_{α}^q converges in total variation to w_1^q, rather than w, as α converges to one by the continuity of $D_1(w_{\alpha}^q\|\,w_1^q)$ in α. Furthermore, the continuity of $D_1(w_{\alpha}^q\|\,q)$ in α implies that

$$\lim\nolimits_{\alpha\uparrow1}D_1(w_{\alpha}^q\|\,q)=D_1(w_1^q\|\,q).$$

This convergence provides further justification to Definition 2 because $D_1(w\|\,q)=\infty$. Recall that both definitions of the tilted probability measure lead to the same w_{α}^q for $\alpha\in(0,1)$.

Lemma 1. *For any* $\alpha\in(1,\infty)$, $n\in\mathbb{Z}_{+}$, $w_t, q_t\in\mathcal{P}(\mathcal{Y}_t)$, *let* $w_{t,ac}$ *be the component of* w_t *that is absolutely continuous in* q_t *and let* a_2, a_3, *and* Δ *be*

$$a_2 \triangleq \tfrac{1}{n}\sum\nolimits_{t=1}^{n}\mathbf{E}_{w_\alpha^q}\left[\left(\ln\tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t}-\mathbf{E}_{w_\alpha^q}\left[\ln\tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t}\right]\right)^2\right],$$
$$a_3 \triangleq \tfrac{1}{n}\sum\nolimits_{t=1}^{n}\mathbf{E}_{w_\alpha^q}\left[\left|\ln\tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t}-\mathbf{E}_{w_\alpha^q}\left[\ln\tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t}\right]\right|^3\right],$$
$$\Delta \triangleq \tfrac{1}{e\sqrt{a_2}}\left(\tfrac{1}{\sqrt{2\pi}}+2\tfrac{0.56a_3}{a_2}\right),$$

where $w=\otimes_{t=1}^{n}w_t$ *and* $q=\otimes_{t=1}^{n}q_t$. *Then for any* $\mathcal{E}\in\mathcal{Y}_1^n$ *and* $\beta\in\mathbb{R}_+$ *satisfying* $q(\mathcal{E})\le\beta e^{-D_1(w_\alpha^q\|q)}$, *we have*

$$w(\mathcal{Y}_1^n\setminus\mathcal{E})\ge\left[\prod_{t=1}^{n}\|w_{t,ac}\|\right]-\tfrac{2e^{\alpha}\Delta^{\frac{1}{\alpha}}\beta^{\frac{\alpha-1}{\alpha}}}{(\alpha-1)^{1/\alpha}}n^{-\frac{1}{2\alpha}}e^{-D_1(w_\alpha^q\|w)}, \quad (6)$$
$$=\left[1-\tfrac{2e^{\alpha}\Delta^{\frac{1}{\alpha}}\beta^{\frac{\alpha-1}{\alpha}}}{(\alpha-1)^{1/\alpha}}\tfrac{e^{-D_1(w_\alpha^q\|w_1^q)}}{n^{1/2\alpha}}\right]\prod_{t=1}^{n}\|w_{t,ac}\|. \quad (7)$$

Lemma 1 is often applied for the case when w_t is absolutely continuous in q_t for all t, i.e. in the case when $w_t\prec q_t$ for all t. In that case $w_{t,ac}=w_t$ for all t and thus $\prod_{t=1}^{n}\|w_{t,ac}\|=1$.

The lower bound asserted in Lemma 1 is tight in the sense that for any $\alpha\in(1,\infty)$ and $\beta\in\left[\tfrac{9\Delta e^{\alpha\delta}}{\sqrt{n}}e^{-\alpha\sqrt{a_2 n}},\tfrac{9\Delta}{\sqrt{n}}e^{\alpha\sqrt{a_2 n}}\right]$, there exists an $\mathcal{E}\in\mathcal{Y}_1^n$ such that

$$q(\mathcal{E})\le\beta e^{-D_1(w_\alpha^q\|q)},$$
$$w(\mathcal{Y}_1^n\setminus\mathcal{E})\le\left[\prod_{t=1}^{n}\|w_{t,ac}\|\right]-\tfrac{e^{(1-\alpha)\delta}}{\sqrt{2\pi a_2}}\left(\tfrac{\beta}{9\Delta}\right)^{\frac{\alpha-1}{\alpha}}\tfrac{e^{-D_1(w_\alpha^q\|w)}}{n^{1/2\alpha}},$$

where $\delta=e\sqrt{2\pi e a_2}\Delta$, see [33, Appendix A] for a proof.

One can calculate exact asymptotic value of the constant in the trade-off between error probabilities in the hypothesis testing problem in this regime, under stronger hypotheses. For the stationary case —i.e. the case when $w_t=w_1$, $q_t=q_1$ for all t— assuming $w\prec q$, first Csiszár and[1] Longo [34] and more recently Vazquez-Vilar *et al.* [35] have discussed this problem.

Proof of Lemma 1. Let the random variables ξ_t and ξ be[2]

$$\xi_t\triangleq\ln\tfrac{\mathrm{d}w_{t,ac}}{\mathrm{d}q_t},$$
$$\xi\triangleq\sum\nolimits_{t=1}^{n}\xi_t.$$

Then $\xi=\ln\frac{\mathrm{d}w_{ac}}{\mathrm{d}q}$, and hence $\xi=\ln\frac{\mathrm{d}w}{\mathrm{d}q}$, holds q-a.s., and the Radon-Nikodym derivatives $\frac{\mathrm{d}w_\alpha^q}{\mathrm{d}q}$ and $\frac{\mathrm{d}w_\alpha^q}{\mathrm{d}w}$ can be expressed in terms of ξ as follows

$$\ln\tfrac{\mathrm{d}w_\alpha^q}{\mathrm{d}q}=D_1(w_\alpha^q\|q)+\alpha\left(\xi-\mathbf{E}_{w_\alpha^q}[\xi]\right), \quad (8)$$
$$\ln\tfrac{\mathrm{d}w_\alpha^q}{\mathrm{d}w}=D_1(w_\alpha^q\|w)+(\alpha-1)\left(\xi-\mathbf{E}_{w_\alpha^q}[\xi]\right). \quad (9)$$

For each integer κ, let the set $\mathcal{B}_\kappa$ be

$$\mathcal{B}_\kappa\triangleq\left\{y_1^n:\tau+\kappa\le\xi-\mathbf{E}_{w_\alpha^q}[\xi]<\tau+\kappa+1\right\}. \quad (10)$$

[1]The approach of [34] is sound, but its calculations seem to have some mistakes.

[2]ξ_t and ξ are implicitly assumed to be zero outside the support of q.

Then for any $\mathcal{E}\in\mathcal{Y}_1^n$ and $\kappa\in\mathbb{Z}$, we can bound $w_\alpha^q(\mathcal{E}\cap\mathcal{B}_\kappa)$ from above in terms of $q(\mathcal{E}\cap\mathcal{B}_\kappa)$ using (8) and from below in terms of $w(\mathcal{E}\cap\mathcal{B}_\kappa)$ using (9), as follows

$$w_\alpha^q(\mathcal{E}\cap\mathcal{B}_\kappa)\le q(\mathcal{E}\cap\mathcal{B}_\kappa)e^{D_1(w_\alpha^q\|q)+\alpha\tau+\alpha(\kappa+1)}, \quad (11)$$
$$w_\alpha^q(\mathcal{E}\cap\mathcal{B}_\kappa)\ge w(\mathcal{E}\cap\mathcal{B}_\kappa)e^{D_1(w_\alpha^q\|w)+(\alpha-1)\tau+(\alpha-1)\kappa}. \quad (12)$$

In order bound $w(\mathcal{Y}_1^n\setminus\mathcal{E})$ we use $w(\mathcal{Y}_1^n\setminus\mathcal{E})\ge w(\mathcal{B}_\mathbb{Z}\setminus\mathcal{E})$ and $w(\mathcal{B}_\mathbb{Z}\setminus\mathcal{E})=w(\mathcal{B}_\mathbb{Z})-w(\mathcal{B}_\mathbb{Z}\cap\mathcal{E})$ where $\mathcal{B}_\mathbb{Z}\triangleq\cup_{\kappa\in\mathbb{Z}}\mathcal{B}_\kappa$. First note that for any family of reference measures $\{\nu_t\}$ satisfying $w_t\prec\nu_t$ and $q_t\prec\nu_t$ for all t we have

$$w(\mathcal{B}_\mathbb{Z})=\left(\bigotimes\nolimits_{\tau=1}^{n}w_\tau\right)\left(\left\{y_1^n:\tfrac{\mathrm{d}w_t}{\mathrm{d}\nu_t}>0\text{ and }\tfrac{\mathrm{d}q_t}{\mathrm{d}\nu_t}>0\quad\forall t\right\}\right)$$
$$=\prod\nolimits_{t=1}^{n}w_t\left(\left\{y_t:\tfrac{\mathrm{d}w_t}{\mathrm{d}\nu_t}>0\text{ and }\tfrac{\mathrm{d}q_t}{\mathrm{d}\nu_t}>0\right\}\right)$$
$$=\prod\nolimits_{t=1}^{n}\|w_{t,ac}\|. \quad (13)$$

Thus for $\mathcal{B}_{\mathbb{Z}_{\le0}}\triangleq\cup_{\kappa\in\mathbb{Z}_{\le0}}\mathcal{B}_\kappa$ and $\mathcal{B}_{\mathbb{Z}_+}\triangleq\cup_{\kappa\in\mathbb{Z}_+}\mathcal{B}_\kappa$, we have

$$w(\mathcal{Y}_1^n\setminus\mathcal{E})\ge\prod\nolimits_{t=1}^{n}\|w_{t,ac}\|-w(\mathcal{E}\cap\mathcal{B}_{\mathbb{Z}_{\le0}})-w(\mathcal{E}\cap\mathcal{B}_{\mathbb{Z}_+}). \quad (14)$$

In order to bound $w(\mathcal{E}\cap\mathcal{B}_{\mathbb{Z}_{\le0}})$ we use (11), (12), the identity $q(\mathcal{E}\cap\mathcal{B}_\kappa)\le q(\mathcal{E})$, the hypothesis $q(\mathcal{E})\le\beta e^{-D_1(w_\alpha^q\|q)}$, and the formula for the sum of a geometric series:

$$w(\mathcal{E}\cap\mathcal{B}_{\mathbb{Z}_{\le0}})=\sum\nolimits_{\kappa\in\mathbb{Z}_{\le0}}w(\mathcal{E}\cap\mathcal{B}_\kappa)$$
$$\le\sum\nolimits_{\kappa\in\mathbb{Z}_{\le0}}\beta e^{\tau+\kappa+\alpha-D_1(w_\alpha^q\|w)}$$
$$\le\beta e^{\tau+\alpha-D_1(w_\alpha^q\|w)}\tfrac{1}{1-e^{-1}}. \quad (15)$$

On the other hand ξ_t's are jointly independent under w_α^q. Thus the Berry–Esseen theorem [36]–[38] implies

$$w_\alpha^q(\mathcal{B}_\kappa)\le\Phi\left(\tfrac{\tau+\kappa+1}{\sqrt{a_2 n}}\right)-\Phi\left(\tfrac{\tau+\kappa}{\sqrt{a_2 n}}\right)+2\tfrac{0.56}{\sqrt{n}}\tfrac{a_3}{a_2\sqrt{a_2}}$$
$$\le\tfrac{1}{\sqrt{a_2 n}}\left(\tfrac{1}{\sqrt{2\pi}}+2\tfrac{0.56a_3}{a_2}\right).$$
$$\le e\Delta n^{-1/2}.$$

Thus we can bound $w(\mathcal{E}\cap\mathcal{B}_{\mathbb{Z}_+})$ using (12), the fact that $w_\alpha^q(\mathcal{E}\cap\mathcal{B}_\kappa)\le w_\alpha^q(\mathcal{B}_\kappa)$, and the formula for the sum of a geometric series, as well:

$$w(\mathcal{E}\cap\mathcal{B}_{\mathbb{Z}_+})=\sum\nolimits_{\kappa\in\mathbb{Z}_+}w(\mathcal{E}\cap\mathcal{B}_\kappa)$$
$$\le\sum\nolimits_{\kappa\in\mathbb{Z}_+}e\Delta n^{-1/2}e^{-D_1(w_\alpha^q\|w)+(1-\alpha)\tau+(1-\alpha)\kappa}$$
$$\le e\Delta n^{-1/2}e^{-D_1(w_\alpha^q\|w)+(1-\alpha)\tau}\tfrac{e^{1-\alpha}}{1-e^{1-\alpha}}. \quad (16)$$

For $\tau=\frac{2\ln\Delta-2\ln\beta-\ln n}{2\alpha}+\frac{1}{\alpha}\ln\frac{e-1}{e^{\alpha-1}-1}-1$, (6) follows from (14), (15), (16), and the identity $\frac{(e-1)^{1-\alpha}}{e^{\alpha-1}-1}\le\frac{1}{\alpha-1}$. ∎

IV. Augustin Information, Augustin Mean and The Strong Converse Exponent

Our primary goal in this section is to define the Augustin information and mean and the strong converse exponent and review those properties of them that will be useful in our analysis. Let us start by defining the conditional Rényi divergence:

Definition 3. For any $\alpha \in \mathbb{R}_+$, $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $q \in \mathcal{P}(\mathcal{Y})$, and $p \in \mathcal{P}(\mathcal{X})$ *the order α conditional Rényi divergence for the input distribution p* is

$$D_\alpha(W\| q| p) \triangleq \sum\nolimits_{x\in\mathcal{X}} p(x) D_\alpha(W(x)\| q) . \quad (17)$$

Definition 4. For any $\alpha \in \mathbb{R}_+$, $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and $p \in \mathcal{P}(\mathcal{X})$ *the order α Augustin information for the input distribution p* is

$$I_\alpha(p;W) \triangleq \inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})} D_\alpha(W\| q| p) . \quad (18)$$

The infimum in (18) is achieved by a unique probability measure $q_{\alpha,p}$, called *the order α Augustin mean for the input distribution p*, by [31, Lemma 13-(b,c,d)]. Furthermore,

$$D_{1\vee\alpha}(q_{\alpha,p}\| q) \geq D_\alpha(W\| q| p) - I_\alpha(p;W) \geq D_{1\wedge\alpha}(q_{\alpha,p}\| q) \quad (19)$$

for all $q \in \mathcal{P}(\mathcal{Y})$ by [31, Lemma 13-(b,c,d)], as well. $I_\alpha(p;W)$ is continuously differentiable in α on $\mathbb{R}_+$ and

$$\tfrac{\partial}{\partial\alpha} I_\alpha(p;W) = \begin{cases} \frac{1}{(\alpha-1)^2} D_1\left(W_\alpha^{q_{\alpha,p}}\middle\| W \middle| p\right) & \alpha \neq 1 \\ \sum_x \frac{p(x)}{2} \mathbf{V}_{W(x)}\left[\ln \frac{\mathrm{d}W(x)}{\mathrm{d}q_{1,p}}\right] & \alpha = 1 \end{cases} \quad (20)$$

by [31, Lemma 17-(e)], where $W_\alpha^{q_{\alpha,p}}(x)$ is the order α tilted probability measure between $W(x)$ and $q_{\alpha,p}$.

$W_\alpha^{q_{\alpha,p}}$ is called *the order α tilted channel* for the channel W and the output distribution $q_{\alpha,p}$. The tilted channel is also used to express $I_\alpha(p;W)$ in terms of the Kullback–Leibler divergences in [31, Lemma 13-(e)]:

$$I_\alpha(p;W) = \tfrac{\alpha}{1-\alpha} D_1(W_\alpha^{q_{\alpha,p}}\| W| p) + I_1(p;W_\alpha^{q_{\alpha,p}}) . \quad (21)$$

Since $\sum_x p(x) W_\alpha^{q_{\alpha,p}}(x) = q_{\alpha,p}$ by[3] [31, Lemma 13-(b,c,d)], we also have the following identity for all $\alpha \in \mathbb{R}_+$

$$I_1(p;W_\alpha^{q_{\alpha,p}}) = D_1(W_\alpha^{q_{\alpha,p}}\| q_{\alpha,p}| p) . \quad (22)$$

A more comprehensive discussion of Augustin's information measures can be found in [31].

Definition 5. For any $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, $p \in \mathcal{P}(\mathcal{X})$, and $R \in \mathbb{R}_+$, the strong converse exponent (SCE) is

$$E_{sc}(R, W, p) \triangleq \sup\nolimits_{\alpha\in(1,\infty)} \tfrac{1-\alpha}{\alpha} (I_\alpha(p;W) - R) . \quad (23)$$

We can apply the derivative test to determine $E_{sc}(R, W, p)$, because $I_\alpha(p;W)$ is continuously differentiable in the order α by [31, Lemma 17-(e)]. Equations (20) and (21) imply

$$\tfrac{\partial}{\partial\alpha} \tfrac{1-\alpha}{\alpha} (I_\alpha(p;W) - R) = \tfrac{1}{\alpha^2} (R - I_1(p;W_\alpha^{q_{\alpha,p}})) . \quad (24)$$

On the other hand, either $I_1\left(p;W_\alpha^{q_{\alpha,p}}\right)$ is increasing and continuous in α on $\mathbb{R}_+$, or $I_1\left(p;W_\alpha^{q_{\alpha,p}}\right) = I_1(p;W)$ for all positive α by [31, Lemma 17-(f)]. Furthermore, $I_1\left(p;W_1^{q_{1,p}}\right)$ is equal to $I_1(p;W)$. Thus for any rate R in $(I_1(p;W), \lim_{\alpha\uparrow\infty} I_1\left(p;W_\alpha^{q_{\alpha,p}}\right))$, there exists an order α^* in $(1,\infty)$ satisfying

$$R = I_1\left(p;W_{\alpha^*}^{q_{\alpha^*,p}}\right) \quad (25)$$

[3] In fact the Augustin mean is the only probability measure satisfying such a fixed point property by [31, Lemma 13], as well.

by the intermediate value theorem [39, 4.23]. The α^* satisfying (25) is unique because $I_1\left(p;W_\alpha^{q_{\alpha,p}}\right)$ is increasing in α. The monotonicity of $I_1\left(p;W_\alpha^{q_{\alpha,p}}\right)$ in α and (24) also implies $E_{sc}(R, W, p) = \frac{1-\alpha^*}{\alpha^*} (I_{\alpha^*}(p;W) - R)$. Thus as a result of (21), the unique α^* satisfying (25) also satisfies

$$E_{sc}(R, W, p) = D_1\left(W_{\alpha^*}^{q_{\alpha^*,p}}\middle\| W \middle| p\right) . \quad (26)$$

Since $D_1\left(W_\alpha^{q_{\alpha,p}}\middle\| q_{\alpha,p} \middle| p\right)$ is continuous and increasing in α, its inverse is increasing and continuous, as well. Thus the definition of SCE given in (23) and the definition of derivative as a limit imply that for any R in $\left(I_1(p;W), \lim_{\alpha\uparrow\infty} I_1\left(p;W_\alpha^{q_{\alpha,p}}\right)\right)$ the unique α^* satisfying (25) also satisfies

$$\tfrac{\partial}{\partial R} E_{sc}(R, W, p) = \tfrac{\alpha^*-1}{\alpha^*} . \quad (27)$$

If $R \geq \lim_{\alpha\uparrow\infty} I_1\left(p;W_\alpha^{q_{\alpha,p}}\right)$, then the derivative given in (24) is positive for all $\alpha \in (1,\infty)$ and thus

$$\begin{aligned} E_{sc}(R, W, p) &= \lim\nolimits_{\alpha\uparrow\infty} \tfrac{1-\alpha}{\alpha} (I_\alpha(p;W) - R) \\ &= R - I_\infty(p;W) \end{aligned} \quad (28)$$

for all $R \geq \lim_{\alpha\uparrow\infty} I_1\left(p;W_\alpha^{q_{\alpha,p}}\right)$.

On the other hand, if $R \leq I_1\left(p;W_1^{q_{1,p}}\right)$, then the derivative given in (24) is negative for all $\alpha \in (1,\infty)$ and thus

$$\begin{aligned} E_{sc}(R, W, p) &= \lim\nolimits_{\alpha\downarrow 1} \tfrac{1-\alpha}{\alpha} (I_\alpha(p;W) - R) \\ &= 0. \end{aligned} \quad (29)$$

for all $R \leq I_1(p;W)$.

Equations (25), (26), (27), (28), and (29) characterize the strong converse exponent $E_{sc}(R, W, p)$ defined in (23) as a non-decreasing continuously differentiable convex function that is strictly convex on $(I_1(p;W), \lim_{\alpha\uparrow\infty} I_1\left(p;W_\alpha^{q_{\alpha,p}}\right))$ and increasing on $(I_1(p;W), \infty)$.

Remark 1. *The definition of $E_{sc}(R, W, p)$ given in* (23) *is equivalent to the one used by Dueck and Körner [22]. In order to see why, recall that the Augustin information satisfies the following variational characterization by [31, Lemma 13-(e)]*

$$\tfrac{1-\alpha}{\alpha} I_\alpha(p;W) = \inf_{V\in\mathcal{P}(\mathcal{Y}|\mathcal{X})} D_1(V\| W| p) + \tfrac{1-\alpha}{\alpha} I_1(p;V) .$$

Thus $E_{sc}(R, W, p)$ can be written as follows for $s = \frac{\alpha-1}{\alpha}$:

$$\begin{aligned} &E_{sc}(R, W, p) \\ &\quad= \sup_{s\in(0,1)} \inf_{V\in\mathcal{P}(\mathcal{Y}|\mathcal{X})} D_1(V\| W| p) + s(R - I_1(p;V)) \\ &\quad= \inf_{V\in\mathcal{P}(\mathcal{Y}|\mathcal{X})} \sup_{s\in(0,1)} D_1(V\| W| p) + s(R - I_1(p;V)) \\ &\quad= \inf_{V\in\mathcal{P}(\mathcal{Y}|\mathcal{X})} D_1(V\| W| p) + |R - I_1(p;V)|^+ . \end{aligned}$$

We can change the order of the infimum and supremum using Sion's minimax theorem [40], [41] because we can replace $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ by the set of elements of $\mathcal{P}(\mathcal{Y}|\mathtt{supp}(p))$ satisfying $D_1(V\| W| p) \leq R$ and the latter set is compact in the topology of setwise convergence by the necessary and sufficient condition for the uniform integrability given by de la Vallee Poussin [42, Thm. 4.5.9], see [31, (d-iii) on p.36] for a similar argument.

V. The Refined Strong Converse

Theorem 1. *For any* $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $M,L,n\in\mathbb{Z}_{+}$, $p\in\mathcal{P}(\mathcal{X})$ *satisfying* $I_1(p;W)<\frac{1}{n}\ln\frac{M}{L}<\lim_{\alpha\uparrow\infty}I_1\big(p;W_{\alpha}^{q_{\alpha,p}}\big)$ *and* $np(x)\in\mathbb{Z}_{\geq 0}$ *for all* $x\in\mathcal{X}$, *the order* $\alpha^{*}\triangleq\frac{1}{1-E_{sc}'\left(\frac{1}{n}\ln\frac{M}{L},W,p\right)}$ *satisfies*

$$I_1\big(p;W_{\alpha^*}^{q_{\alpha^*,p}}\big)=\tfrac{1}{n}\ln\tfrac{M}{L}. \tag{30}$$

Furthermore, any (M,L) *channel code of length* n *whose codewords all have the same composition* p *satisfies*

$$P_{\mathbf{e}}^{(n)}\geq 1-2e^{\alpha^*}\left(\tfrac{\Delta}{\alpha^*-1}\right)^{\frac{1}{\alpha^*}}n^{-1/2\alpha^*}e^{-nE_{sc}\left(\frac{1}{n}\ln\frac{M}{L},W,p\right)} \tag{31}$$

where

$$a_2=\mathbf{E}_{p\circledast W_{\alpha^*}^{q_{\alpha^*,p}}}\left[\left|\ln\tfrac{\mathrm{d}W}{\mathrm{d}q_{\alpha^*,p}}-\mathbf{E}_{W_{\alpha^*}^{q_{\alpha^*,p}}}\left[\ln\tfrac{\mathrm{d}W}{\mathrm{d}q_{\alpha^*,p}}\right]\right|^{2}\right], \tag{32}$$

$$a_3=\mathbf{E}_{p\circledast W_{\alpha^*}^{q_{\alpha^*,p}}}\left[\left|\ln\tfrac{\mathrm{d}W}{\mathrm{d}q_{\alpha^*,p}}-\mathbf{E}_{W_{\alpha^*}^{q_{\alpha^*,p}}}\left[\ln\tfrac{\mathrm{d}W}{\mathrm{d}q_{\alpha^*,p}}\right]\right|^{3}\right], \tag{33}$$

$$\Delta\triangleq\tfrac{1}{e\sqrt{a_2}}\left(\tfrac{1}{\sqrt{2\pi}}+2\tfrac{0.56a_3}{a_2}\right). \tag{34}$$

Theorem 2. *For any* $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $M,L,n\in\mathbb{Z}_{+}$, $p\in\mathcal{P}(\mathcal{X})$ *satisfying* $\frac{1}{n}\ln\frac{M}{L}\geq\lim_{\alpha\uparrow\infty}I_1\big(p;W_{\alpha}^{q_{\alpha,p}}\big)$ *and* $np(x)\in\mathbb{Z}_{\geq 0}$ *for all* $x\in\mathcal{X}$, *any* (M,L) *channel code of length* n *whose codewords all have the same composition* p *satisfies*

$$P_{\mathbf{e}}^{(n)}\geq 1-e^{-nE_{sc}\left(\frac{1}{n}\ln\frac{M}{L},W,p\right)}. \tag{35}$$

Theorems 1 and 2 collectively imply for all $n\in\mathbb{Z}_{+}$ a strong converse of the form (3) for $E_{sc}(R)=E_{sc}\left(\frac{1}{n}\ln\frac{M}{L},W,p\right)$, for a constant A determined by the rate R, the channel W, and the composition p. Following the convention used for the corresponding improvement of the sphere packing bound in [26]–[30], we call these bounds refined strong converses.

Proof of Theorem 1. The existence of a unique order α^* satisfying (30) was proved and its value was determined in §IV, see (25), (26), and (27).

Let the probability measures w_m, q, and v_m in $\mathcal{P}(\mathcal{Y}_1^n)$ be

$$w_m\triangleq\bigotimes\nolimits_{t=1}^{n}W(\Psi_t(m)),$$
$$q\triangleq\bigotimes\nolimits_{t=1}^{n}q_{\alpha^*,p},$$
$$v_m\triangleq\bigotimes\nolimits_{t=1}^{n}W_{\alpha^*}^{q_{\alpha^*,p}}(\Psi_t(m)).$$

Then v_m is equal to the order α^* tilted probability measure between w_m and q. Furthermore,[4]

$$D_1(v_m\|\,q)=nD_1\big(W_{\alpha^*}^{q_{\alpha^*,p}}\big\|\,q_{\alpha^*,p}\big|\,p\big)\qquad m\in\mathcal{M},$$
$$D_1(v_m\|\,w_m)=nD_1\big(W_{\alpha^*}^{q_{\alpha^*,p}}\big\|\,W\big|\,p\big)\qquad m\in\mathcal{M}.$$

Note that $D_1\big(W_{\alpha^*}^{q_{\alpha^*,p}}\big\|\,q_{\alpha^*,p}\big|\,p\big)=\frac{1}{n}\ln\frac{M}{L}$ by (22) and (30) and $D_1\big(W_{\alpha^*}^{q_{\alpha^*,p}}\big\|\,W\big|\,p\big)=E_{sc}\left(\frac{1}{n}\ln\frac{M}{L},W,p\right)$ by (25), (26), and (30). Thus applying Lemma 1, for $\mathcal{E}=\{y_1^n:m\in\Theta(y_1^n)\}$ and $\beta=q(m\in\Theta)\frac{M}{L}$ we get

$$P_{\mathbf{e}}^{m}\geq 1-\tfrac{2e^{\alpha^*}\Delta^{1/\alpha^*}}{(\alpha^*-1)^{1/\alpha^*}}\left(\tfrac{q(m\in\Theta)M}{L}\right)^{\frac{\alpha^*-1}{\alpha^*}}\tfrac{e^{-nE_{sc}\left(\frac{1}{n}\ln\frac{M}{L},W,p\right)}}{n^{1/2\alpha^*}}. \tag{36}$$

On the other hand $\sum_{m\in\mathcal{M}}q(m\in\Theta)\leq L$, as a result of the definition of the list decoding. Thus using the concavity of the function $z^{\frac{\alpha^*-1}{\alpha^*}}$ in z together with the Jensen's inequality get

$$\sum\nolimits_{m\in\mathcal{M}}\tfrac{1}{M}\left(\tfrac{q(m\in\Theta)M}{L}\right)^{\frac{\alpha^*-1}{\alpha^*}}\leq\left(\sum\nolimits_{m\in\mathcal{M}}\tfrac{1}{M}\tfrac{q(m\in\Theta)M}{L}\right)^{\frac{\alpha^*-1}{\alpha^*}}=1.$$

Then (31) follows from (36) and the definition error probability as the average of the conditional error probabilities. □

Proof of Theorem 2. Let the probability measures w_m, q_α be

$$w_m\triangleq\bigotimes\nolimits_{t=1}^{n}W(\Psi_t(m)),\quad\text{and}\quad q_\alpha\triangleq\bigotimes\nolimits_{t=1}^{n}q_{\alpha,p}.$$

Then $D_\alpha(w_m\|\,q_\alpha)=nI_\alpha(p;W)$ for all m because all $\Psi(m)$'s have the composition p. On the other hand the data processing inequality of the Rényi divergence, [32, Thm 9], imply

$$D_\alpha(w_m\|\,q_\alpha)\geq\tfrac{\ln\left[(P_{\mathbf{e}}^{m})^{\alpha}(q_\alpha(m\in\Theta))^{1-\alpha}+(1-P_{\mathbf{e}}^{m})^{\alpha}(q_\alpha(m\in\Theta))^{1-\alpha}\right]}{\alpha-1}$$
$$\geq\tfrac{\ln\left[(1-P_{\mathbf{e}}^{m})^{\alpha}(q_\alpha(m\in\Theta))^{1-\alpha}\right]}{\alpha-1}.$$

Thus $P_{\mathbf{e}}^{m}\geq 1-\left(q_\alpha(m\in\Theta)\right)^{\frac{\alpha-1}{\alpha}}e^{\frac{\alpha-1}{\alpha}nI_\alpha(p;W)}$. On the other hand the concavity of the function $z^{\frac{\alpha-1}{\alpha}}$ in z for $\alpha>1$, the Jensen's inequality, and $\sum_{m\in\mathcal{M}}q_\alpha(m\in\Theta)\leq L$, imply $\sum_{m\in\mathcal{M}}\frac{1}{M}\left(q_\alpha(m\in\Theta)\right)^{\frac{\alpha-1}{\alpha}}\leq\left(L/M\right)^{\frac{\alpha-1}{\alpha}}$. Hence

$$P_{\mathbf{e}}^{(n)}\geq 1-e^{-\frac{1-\alpha}{\alpha}n\left(I_\alpha(p;W)-\frac{1}{n}\ln\frac{M}{L}\right)}\qquad\forall\alpha\in(1,\infty).$$

Then (35) follows from (23). □

VI. Discussion

Although we have confined our analysis to the constant composition codes for brevity, using the Augustin capacity and center —instead of Augustin information and mean— one can obtain analogous results for additive white Gaussian noise channels with quadratic cost functions and Rényi symmetric channels defined in [30]. For Rényi symmetric channels the refined strong converse (3), can be established with smaller, i.e., better, constant A using the saddle point approximation. Such a result has been reported in [35, (36)], assuming a common support for all output distributions of the channel and a non-lattice structure for the random variables involved. Establishing refined strong converses without any symmetry hypothesis is the main technical challenge in this line of work.

We believe the refined strong converses of the form (3) are the best possible bounds for derivations of the strong converse relying on the asymptotic behavior of sums of independent random variables. Nevertheless for the singular symmetric channels considered in [43], it should be possible to improve (3) as $P_{\mathbf{e}}^{(n)}\geq 1-An^{-0.5}e^{-nE_{sc}(R)}$.

[4]It is worth mentioning that both $D_1(v_m\|\,q)$ and $D_1(v_m\|\,w_m)$ can be expressed in this form for all messages because all $\Psi(m)$'s have the same composition p.

REFERENCES

[1] J. Wolfowitz, "The coding of messages subject to chance errors," *Illinois Journal of Mathematics*, vol. 1, no. 4, pp. 591–606, 12 1957.

[2] U. Augustin, "Gedächtnisfreie Kanäle für diskrete Zeit," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 6, no. 1, pp. 10–61, Mar 1966.

[3] ——, "Noisy channels," Habilitation Thesis, Universität Erlangen-Nürnberg, 1978, (http://bit.ly/2ID8h7m).

[4] J. Beck and I. Csiszár, "Medium converse for memoryless channels with arbitrary alphabets," *Problems of control and information theory*, vol. 7, no. 3, pp. 199–202, 1978.

[5] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, Jul 1994.

[6] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.

[7] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels (corresp.)," *IEEE Transactions on Information Theory*, vol. 19, no. 3, pp. 357–359, May 1973.

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

[9] B. Nakiboğlu, "The Sphere Packing Bound via Augustin's Method," *IEEE Transactions on Information Theory*, vol. 65, no. 2, pp. 816–840, Feb 2019, (arXiv:1611.06924 [cs.IT]).

[10] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, Oct 2010, pp. 1327 –1333.

[11] H. Nagaoka, "Strong converse theorems in quantum information theory," in *Proceedings of the ERATO Conference on Quantum Information Science (EQIS)*, vol. 33, 2001, (also appeared as Chap. 4 of Asymptotic Theory of Quantum Statistical Inference: Selected Papers, ed. by M. Hayashi).

[12] T. Ogawa and H. Nagaoka, "Strong converse to the quantum channel coding theorem," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2486–2489, Nov 1999.

[13] R. König and S. Wehner, "A strong converse for classical channel coding using entangled inputs," *Phys. Rev. Lett.*, vol. 103, no. 7, p. 070504, Aug 2009.

[14] M. Mosonyi and F. Hiai, "On the quantum Rényi relative entropies and related capacity formulas," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2474–2487, April 2011.

[15] N. Sharma and N. A. Warsi, "Fundamental bound on the reliability of quantum information transmission," *Phys. Rev. Lett.*, vol. 110, no. 8, p. 080501, Feb 2013.

[16] M. M. Wilde, A. Winter, and D. Yang, "Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy," *Communications in Mathematical Physics*, vol. 331, no. 2, pp. 593–622, Oct 2014.

[17] M. Gupta and M. Wilde, "Multiplicativity of completely bounded p-norms implies a strong converse for entanglement-assisted capacity," *Communications in Mathematical Physics*, vol. 334, pp. 867–887, 2015.

[18] M. Mosonyi and T. Ogawa, "Strong converse exponent for classical-quantum channel coding," *Communications in Mathematical Physics*, vol. 355, no. 1, pp. 373–426, Oct 2017.

[19] M. Tomamichel, M. M. Wilde, and A. Winter, "Strong converse rates for quantum communication," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 715–727, Jan 2017.

[20] H.-C. Cheng, E. P. Hanson, N. Datta, and M. Hsieh, "Non-asymptotic classical data compression with quantum side informations," *arXiv:1803.07505 [quant-ph]*, 2018.

[21] ——, "Duality between source coding with quantum side information and c-q channel coding," *arXiv:1809.11143 [quant-ph]*, 2018.

[22] G. Dueck and J. Korner, "Reliability function of a discrete memoryless channel at rates above capacity (corresp.)," *IEEE Transactions on Information Theory*, vol. 25, no. 1, pp. 82–85, Jan 1979.

[23] Y. Oohama, "The optimal exponent function for the additive white gaussian noise channel at rates above the capacity," in *2017 IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, June 2017, pp. 1053–1057.

[24] Mosonyi and Ogawa, "Divergence radii and the strong converse exponent of classical-quantum channel coding with constant compositions," *arXiv:1811.10599v6 [cs.IT]*, 2018.

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

[26] Y. Altuğ and A. B. Wagner, "Refinement of the sphere packing bound for symmetric channels," in *49th Annual Allerton Conference on Communication, Control, and Computing*, Sept 2011, pp. 30–37.

[27] ——, "Refinement of the sphere-packing bound: Asymmetric channels," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1592–1614, March 2014.

[28] H.-C. Cheng, M. H. Hsieh, and M. Tomamichel, "Quantum sphere-packing bounds with polynomial prefactors," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 2872–2898, May 2019, (arXiv:1704.05703 [quant-ph]).

[29] B. Nakiboğlu, "A Simple Derivation of the Refined SPB for the Constant Composition Codes," in *2019 IEEE International Symposium on Information Theory (ISIT)*, Paris, France, July 2019, pp. 2659–2663.

[30] ——, "A simple derivation of the refined sphere packing bound under certain symmetry hypotheses," *Turkish Journal Of Mathematics*, vol. 44, no. 3, pp. 919–948, 2020, (arXiv:1904.12780 [cs.IT]).

[31] ——, "The Augustin Capacity and Center," *Problems of Information Transmission*, vol. 55, no. 4, pp. 299–342, October 2019, (arXiv:1803.07937 [cs.IT]).

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

[33] H.-C. Cheng and B. Nakiboğlu, "Refined strong converse for the constant composition codes," *arXiv:2002.11414v2*, 2020.

[34] I. Csiszár and G. Longo, "On the error exponent for source coding and for testing simple statistical hypotheses," *Studia Scientiarum Mathematicarum Hungarica*, vol. 6, pp. 181–191, 1971, (http://real-j.mtak.hu/id/eprint/5457).

[35] G. Vazquez-Vilar, A. G. i. Fabregas, T. Koch, and A. Lancho, "Saddlepoint approximation of the error probability of binary hypothesis testing," in *2018 IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 2306–2310.

[36] A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Transactions of the American Mathematical Society*, vol. 49, no. 1, pp. 122–136, 1941.

[37] C. G. Esseen, "On the Liapunoff limit of error in the theory of probability," *Arkiv för Matematik, Astronomi Och Fysik*, vol. A28, pp. 1–19, 1942.

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

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

[40] M. Sion, "On general minimax theorems," *Pacific Journal of Mathematics*, vol. 8, no. 1, pp. 171–176, 1958.

[41] H. Komiya, "Elementary proof for Sion's minimax theorem," *Kodai Mathematical Journal*, vol. 11, no. 1, pp. 5–7, 1988.

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

[43] Y. Altuğ and A. B. Wagner, "On exact asymptotics of the error probability in channel coding: symmetric channels," *arXiv:1908.11419 [cs.IT]*, 2019.