On Exact Asymptotics of the Error Probability in Channel Coding: Symmetric Channels

Yücel Altuğ and Aaron B. Wagner , *Fellow, IEEE*

Abstract—**The exact order of the optimal sub-exponentially decaying factor in the classical bounds on the error probability of fixed-length codes over a Gallager-symmetric discrete memoryless channel with and without ideal feedback is determined for rates above the critical rate. Regardless of the availability of feedback, it is shown that the order of the optimal sub-exponential factor exhibits a dichotomy. Moreover, the proof technique is used to establish the third-order term in the normal approximation for symmetric channels, where a similar dichotomy is shown to exist.**

Index Terms—**Channel coding, error exponents, exact asymptotics, reliable communication, symmetric channels.**

I. Introduction

IN CHANNEL coding, error exponents describe the rate of decay of the error probability with the rate held fixed below the capacity (e.g., [1]– [10] and references therein). As such, they provide an exponentially fast convergence result in the channel coding theorem, and thereby indicate approximately how large of a blocklength one needs to achieve a target error probability for a given rate. The caveat with classical error exponent results, however, is that they are typically expressed as bounds on the reliability function, which is defined as (e.g., [6, Eq. (5.8.8)])

$$\mathrm{E}(R) := \limsup_{N \to \infty} -\frac{1}{N} \ln \mathrm{P_e}(N, R), \tag{1}$$

where $\mathrm{P_e}(N, R)$ is the minimum error probability of all codes with blocklength N and rate R. Thus, they ignore the subexponential factors in $\mathrm{P_e}(N, R)$, which potentially could be quite significant for small to moderate N. This is especially true for rates near capacity, since typically both the exponent and its first derivative vanish as the rate approaches capacity. Therefore, one would like to have more refined bounds on

$\mathrm{P_e}(N, R)$ that capture the sub-exponential factors, which we will also refer to as the *pre-factor(s)*.

Classical bounds on the pre-factor were quite loose. In particular, until recently the best known upper and lower bounds on the optimal pre-factor that are valid for any DMC were $O(1)$ and $\Omega(N^{-|\mathcal{X}||\mathcal{Y}|})$, respectively, due to Fano [4] and Haroutunian [8], respectively. Here, $|\mathcal{X}|$ and $|\mathcal{Y}|$ denote the cardinality of the input and output alphabet of the channel, respectively. The authors have improved upon these results to obtain relatively tight bounds on the order of the pre-factor, which we summarize next. Specifically, [13] proves that the error probability of any (N, R) constant-composition code, i.e., a code in which all codewords possess the same empirical distribution, is lower bounded by

$$\frac{K_1}{N^{\frac{1}{2}(1+|\mathrm{E}'_{\mathrm{SP}}(R)|)}} \mathrm{e}^{-N\mathrm{E}_{\mathrm{SP}}(R)}, \tag{2}$$

where $\mathrm{E}'_{\mathrm{SP}}(R)$ is the slope of the sphere-packing exponent at R and $K_1 \in \mathbb{R}^+$ is a constant that depends on the channel and R. In [14], it is shown that if the channel satisfies a certain condition, then the optimal error probability is upper bounded by

$$\frac{K_2}{N^{\frac{1}{2}(1+\bar{\rho}_R)}} \mathrm{e}^{-N\mathrm{E}_{\mathrm{r}}(R)}, \tag{3}$$

where $\bar{\rho}_R$ is related to the slope of the random coding exponent and is typically equal to $|\mathrm{E}'_{\mathrm{r}}(R)|$, and $K_2 \in \mathbb{R}^+$ is a constant that depends on the channel and R. For the remaining small class of channels, the following upper bound holds

$$\frac{K_3}{\sqrt{N}} \mathrm{e}^{-N\mathrm{E}_{\mathrm{r}}(R)}, \tag{4}$$

where $K_3 \in \mathbb{R}^+$ is a constant that depends on the channel and R. Note that the order of the aforementioned upper and lower bounds asymptotically coincide as the rate approaches capacity.

Related to the above bounds, one of the classical results of Elias is worth mentioning. In [2], he considered binary symmetric and erasure channels and proved that the order of the optimal pre-factor for the binary symmetric (resp. erasure) channel is $\Theta(N^{-\frac{1}{2}(1+|\mathrm{E}'(R)|)})$ (resp. $\Theta(N^{-\frac{1}{2}})$) for rates above the critical rate, where $\mathrm{E}'(R)$ is the slope of the reliability function.

In this paper, we show that for the class of symmetric channels (see Definition 1 to follow) we can improve the bounds in [13] and [14] to give an exact characterization of the order of the dominant sub-exponential factor above the critical rate. Specifically, we prove a dichotomy of symmetric

channels in terms of the order of their optimal pre-factors. For the typical symmetric channels, which we call *nonsingular channels*, the optimal order is $\Theta(N^{-\frac{1}{2}(1+|\mathrm{E}'(R)|)})$, whereas for the remaining symmetric channels, namely *singular channels*, $\Theta(N^{-\frac{1}{2}})$ is the optimal order. These results imply that every symmetric channel has a pre-factor order that matches either that of the BEC or that of the BSC. Thus, Elias had already found all of the different orders that can occur for symmetric channels.

For both singular and nonsingular channels, the upper bound on the pre-factor follows from [14] (which has been strengthened in several ways [15]–[19]). Our contribution is improving the lower bound on the order of the pre-factor, i.e., obtaining a better pre-factor in the sphere-packing bound. There are multiple ways of proving the sphere-packing bound, some more amenable to obtaining pre-factor bounds than the others. For a comparison of these techniques, see [13, Section III.A]. Among these methods, the one that relates the error probability of a given code to the error probability of a related binary hypothesis test with the aid of an auxiliary output distribution is well suited for pre-factor analysis. This method can be traced back to at least the classical results of Blahut [26] and is the starting point of the derivation of (2). However, the auxiliary output distribution used in [13] does not admit a simple explicit form. Indeed, it is defined by using the saddle-point of a certain optimization problem, which is intimately related to the sphere-packing exponent. This complication is due to the asymmetry of the channel. Once we restrict our attention to symmetric channels, it is possible to show a simple characterization of this distribution (see (58) and Proposition 3 to follow), which is in the form of a *tilted distribution*. Since this distribution is independent of the code, we can dispense with the constant-composition assumption[1] in [13].

For the singular case, we introduce a new method of proving the sphere-packing bound. The idea is the following: consider any singular symmetric channel W and any (N, R) code over W. Let $\mathcal{E}$ denote the event that the code makes an error. Define the *information density*

$$\imath(x; y) := \ln \frac{W(y|x)}{\sum_{z \in \mathcal{X}} \frac{W(y|z)}{|\mathcal{X}|}}. \tag{5}$$

By using Wolfowitz's strong converse (e.g., [28]), one can argue that

$$\Pr\left[\mathcal{E} \,\middle|\, \sum_{n=1}^{N} \imath(X_n; Y_n) \leq R\right] \approx 1, \tag{6}$$

where the probability is induced by the uniform distribution over the messages and the channel, and $\mathbf{X}^N$ (resp. $\mathbf{Y}^N$) denotes the input (resp. output) of the channel. Hence,

$$\Pr[\mathcal{E}] \geq \Pr\left[\sum_{n=1}^{N} \imath(X_n; Y_n) \leq R\right] \tag{7}$$

[1]The possibility of proving the sphere-packing bound without the constant-composition restriction for symmetric channels was first observed in [27], where the proof methodology of Shannon *et al.* [7] was followed.

$$\cdot \Pr\left[\mathcal{E} \,\middle|\, \sum_{n=1}^{N} \imath(X_n; Y_n) \leq R\right]$$

$$\approx \Pr\left[\sum_{n=1}^{N} \imath(X_n; Y_n) \leq R\right]. \tag{8}$$

Due to the symmetry of W, the random variables in (8) can be shown to be independent and identically distributed (i.i.d.), and hence one can apply classical exact asymptotics results (e.g., [29]) to deduce an exponentially decaying lower bound with a pre-factor order of $1/\sqrt{N}$. However, this procedure results in a useful lower bound only if the exponent matches the reliability function, i.e., one needs

$$\lim_{N \to \infty} -\frac{1}{N} \ln \Pr\left[\sum_{n=1}^{N} \imath(X_n; Y_n) \leq R\right] = \mathrm{E}_{\mathrm{SP}}(R). \tag{9}$$

Although (9) is not true in general, it can be shown to be so for singular and symmetric channels, thus we can deduce an exponentially vanishing lower bound with the sphere-packing exponent and $\Theta(1/\sqrt{N})$ as the dominant sub-exponential factor.

Furthermore, we show that for both singular and nonsingular symmetric channels the pre-factor order is not affected by the presence of ideal feedback. It is well known that for symmetric channels, feedback does not improve the reliability function above the critical rate (e.g., [30]). The results herein strengthen this statement to assert that both the exponent and the dominant sub-exponential factor are unaffected by feedback. For asymmetric channels, see Nakiboğlu [20], [21] and others [22]–[25] for the effect of feedback in the error exponent and normal approximation regimes, respectively.

Moreover, we also apply the aforementioned proof technique to characterize the third-order term in the normal approximation for singular channels. Specifically, for singular and symmetric channels, we prove a converse result, which is valid in the presence of feedback, which implies a dichotomy of the third-order term in the normal approximation for symmetric channels once coupled with [31] and [32, Sec. 3.4.5]. A remarkable aspect of this dichotomy is that its defining property is again singularity of the channel.

We conclude this section by noting that the type of symmetry notion is crucial regarding the dichotomy of the optimal pre-factor of the symmetric channels. Specifically, if one considers *strongly symmetric channels*, i.e., if every row (resp. column) of the channel is a permutation of every other row (resp. column), which is a proper subset of symmetric channels we consider in this paper, then one can show that (e.g., [5]) $\Theta(N^{-\frac{1}{2}(1+|\mathrm{E}'(R)|)})$ is the order of the optimal pre-factor for rates above the critical rate. Evidently, there is no dichotomy for this class of channels, since it is not rich enough to include singular channels (see Remark 1(iii) to follow). Finally, it is possible to extract the constants from our proofs to obtain finite blocklength bounds on the error probability. However, the resulting expressions are rather complicated, so we shall state the results in asymptotic form to elucidate the dichotomy.

II. NOTATION, DEFINITIONS AND STATEMENT OF THE RESULTS

A. Notation

Boldface letters denote vectors, and regular letters with subscripts denote individual components of vectors. Furthermore, capital letters represent random variables, and lowercase letters denote individual realizations of the corresponding random variable. For a finite set $\mathcal{A}$, $\mathcal{P}(\mathcal{A})$ (resp. $U_{\mathcal{A}}$) denotes the set of all probability measures (resp. the uniform probability measure) on $\mathcal{A}$. Similarly, for two finite sets $\mathcal{A}$ and $\mathcal{B}$, $\mathcal{P}(\mathcal{B}|\mathcal{A})$ denotes the set of all stochastic matrices from $\mathcal{A}$ to $\mathcal{B}$. Given any $P \in \mathcal{P}(\mathcal{A})$, $\mathrm{supp}(P) := \{a \in \mathcal{A} : P(a) > 0\}$. $\mathbb{1}\{\cdot\}$ denotes the standard indicator function. Given probability measures λ_1 and λ_2, $\lambda_1 \ll \lambda_2$ means that λ_1 is absolutely continuous with respect to λ_2 (that is, λ_2 dominates λ_1) and $\lambda_1 \equiv \lambda_2$ means that $\lambda_1 \ll \lambda_2$ and $\lambda_2 \ll \lambda_1$. $\Phi(\cdot)$ (resp. $\phi(\cdot)$) denotes the cumulative distribution function (resp. probability density function) of the standard Gaussian random variable. $\mathbb{Z}^+, \mathbb{R}, \mathbb{R}^+$ and $\mathbb{R}_+$ denote the set of positive integers, reals, positive reals and non-negative reals, respectively. We follow the notation of the book of Csiszár and Körner [10] for standard information theoretic quantities.

B. Definitions

An (N, R) *code*, say (f, φ), consists of an encoder, i.e., $f \colon \mathcal{M} \to \mathcal{X}^N$, where $\mathcal{M} := \{1, \ldots, \lceil e^{NR} \rceil\}$ is the set of messages to be transmitted, and a decoder, i.e., $\varphi \colon \mathcal{Y}^N \to \mathcal{M}$. Let $\{\mathcal{A}_m\}_{m=1}^{|\mathcal{M}|}$ denote the decoding regions and $\bar{\mathrm{P}}_e(f, \varphi)$ denote the average error probability of (f, φ). Evidently,

$$\bar{\mathrm{P}}_e(f, \varphi) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m^c} P_{\mathbf{Y}^N|\mathbf{X}^N}(\mathbf{y}^N|f(m)). \quad (10)$$

$\bar{\mathrm{P}}_e(N, R)$ denotes the minimum average error probability attainable by any (N, R) code. Similarly, $\mathrm{P}_e(N, R)$ denotes the minimum maximal error probability attainable by any (N, R) code.

For any $\epsilon \in (0, 1)$,

$$M^*(N, \epsilon) := \max\{\lceil e^{NR} \rceil \in \mathbb{R}_+ : \bar{\mathrm{P}}_e(N, R) \le \epsilon\}, \quad (11)$$

$$M_c^*(N, \epsilon) := \max\{\lceil e^{NR} \rceil \in \mathbb{R}_+ : \bar{\mathrm{P}}_{e,c}(N, R) \le \epsilon\}, \quad (12)$$

where $\bar{\mathrm{P}}_{e,c}(N, R)$ denotes the minimum average error probability attainable by any (N, R) constant-composition code.

An (N, R) *code with ideal feedback*, say (f, φ), consists of an encoder, i.e., $\{f_n \colon \mathcal{M} \times \mathcal{Y}^{n-1} \to \mathcal{X}\}_{n=1}^N$, where $\mathcal{M} := \{1, \ldots, \lceil e^{NR} \rceil\}$ is the set of messages to be transmitted, and a decoder, i.e., $\varphi \colon \mathcal{Y}^N \to \mathcal{M}$. Let $\{\mathcal{A}_m\}_{m=1}^{|\mathcal{M}|}$ denote the decoding regions and $\bar{\mathrm{P}}_e(f, \varphi)$ denote the average error probability of (f, φ). Define

$$P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m) := \prod_{n=1}^N W(y_n|f_n(m, \mathbf{y}^{n-1})), \quad (13)$$

where $f_n(m, \mathbf{y}^{n-1})$ denotes the output of the encoder at time n if message m is transmitted, and $\mathbf{y}^{n-1}$ denotes the previous channel outputs, with the usual convention $\mathbf{y}^0 := \emptyset$. Again,

$$\bar{\mathrm{P}}_{e,\mathrm{fb}}(f, \varphi) = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m^c} P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m). \quad (14)$$

$\bar{\mathrm{P}}_{e,\mathrm{fb}}(N, R)$ denotes the minimum average error probability attainable by any (N, R) code with ideal feedback.

Given any channel $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $R \in \mathbb{R}_+$, we recall the following classical quantities (e.g., [10, Sec. 2.5])

$$\mathrm{E}_{\mathrm{SP}}(R, Q) := \min_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X}) \colon I(Q;V) \le R} D(V\|W|Q), \quad (15)$$

$$\mathrm{E}_{\mathrm{SP}}(R) := \max_{Q \in \mathcal{P}(\mathcal{X})} \mathrm{E}_{\mathrm{SP}}(R, Q), \quad (16)$$

$$\tilde{\mathrm{E}}_{\mathrm{SP}}(R, Q) := \sup_{\rho \ge 0} \{\mathrm{E}_o(\rho, Q) - \rho R\}, \quad (17)$$

$$\tilde{\mathrm{E}}_{\mathrm{SP}}(R) := \max_{Q \in \mathcal{P}(\mathcal{X})} \tilde{\mathrm{E}}_{\mathrm{SP}}(R, Q), \quad (18)$$

$$\mathrm{E}_{\mathrm{r}}(R, Q) := \max_{0 \le \rho \le 1} \{\mathrm{E}_o(\rho, Q) - \rho R\}, \quad (19)$$

$$\mathrm{E}_{\mathrm{r}}(R) := \max_{Q \in \mathcal{P}(\mathcal{X})} \mathrm{E}_{\mathrm{r}}(R, Q), \quad (20)$$

where

$$\mathrm{E}_o(\rho, Q) := -\ln \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} Q(x) W(y|x)^{1/(1+\rho)} \right)^{1+\rho}. \quad (21)$$

It is well known that given any $R \in \mathbb{R}_+$, $\mathrm{E}_{\mathrm{SP}}(R, Q) \ge \tilde{\mathrm{E}}_{\mathrm{SP}}(R, Q)$ for all $Q \in \mathcal{P}(\mathcal{X})$ and $\mathrm{E}_{\mathrm{SP}}(R) = \tilde{\mathrm{E}}_{\mathrm{SP}}(R)$ (e.g., [10, Ex. 2.5.23]). R_∞ denotes the maximum rate such that for all rates below it, $\mathrm{E}_{\mathrm{SP}}(R) = \infty$ (e.g., [9, pg. 158]). Also, R_{cr} denotes the *critical rate* of the channel, i.e., the value such that $\mathrm{E}_{\mathrm{r}}(R) = \mathrm{E}_{\mathrm{SP}}(R)$ if and only if $R \ge R_{\mathrm{cr}}$ (e.g., [9, pg. 160]). Evidently, $\mathrm{E}_{\mathrm{r}}(R) = \mathrm{E}_{\mathrm{SP}}(R) = \tilde{\mathrm{E}}_{\mathrm{SP}}(R)$ for all $R \ge R_{\mathrm{cr}}$.

Given $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, $C(W)$ denotes the capacity of the channel. For any $P \in \mathcal{P}(\mathcal{X})$, define

$$q_P(y) := \sum_{x \in \mathcal{X}} P(x) W(y|x). \quad (22)$$

For notational convenience, let q denote $q_{U_{\mathcal{X}}}$. Given any $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, $P \in \mathcal{P}(\mathcal{X})$ and $\epsilon \in (0, 1)$, define (e.g., [32, Sec. 3.4])

$$V(P, W) := \sum_{x,y} P(x) W(y|x) \left[\ln \frac{W(y|x)}{q_P(y)} \right.$$
$$\left. - \sum_b W(b|x) \ln \frac{W(b|x)}{q_P(b)} \right]^2, \quad (23)$$

and

$$V_\epsilon(W) := \begin{cases} \min_{Q \colon I(Q;W)=C(W)} V(Q, W), & \epsilon \in (0, 1/2), \\ \max_{Q \colon I(Q;W)=C(W)} V(Q, W), & \epsilon \in [1/2, 1). \end{cases} \quad (24)$$

We call $V_\epsilon(W)$ the ϵ-*dispersion* of the channel W. The *dispersion* refers to $V_\epsilon(W)$ for $\epsilon < 1/2$.

The following definition is the type of symmetry we use in this work.

Definition 1 (Gallager [9, p. 94]): A discrete channel is *symmetric* if the channel outputs can be partitioned into subsets such that within each subset, the matrix of transition probabilities satisfies the following: each row (resp. column) is a permutation of each other row (resp. column).

Weaker notions of symmetry are available, such as *weakly input-symmetric* [32] and *quasi-symmetric* [12]. The arguments in this paper do not extend to these weaker notions,

however. In particular, we rely on the optimal input distribution being uniform, which precludes weakly input-symmetric channels, and we require that a tilted version of the channel remain symmetric in some sense (see (78) to follow), which precludes quasi-symmetric channels. We delineate symmetric channels with respect to the order of their optimal pre-factors by using the following notion.

Definition 2 (Singularity): A symmetric channel $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ is *singular* if for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $z \in \mathcal{X}$ such that $W(y|x)W(y|z) > 0$, we have $W(y|x) = W(y|z)$. That is, all inputs that lead to a given output do so with the same probability. Otherwise, the channel is called *nonsingular*.

For general channels, the definition of singularity is more involved [14, Definition 1]. That definition reduces to Definition 2 for symmetric channels, however. More precisely, if a symmetric channel is singular according to Definition 2, then it is singular at all rates according to [14, Definition 1], and, if it is nonsingular according to Definition 2, then it is nonsingular at all rates according to [14, Definition 1].

An equivalent definition of singularity can be given in terms of the following quantity, which is defined in [32, Sec. 3.4],

$$
V^r(P, W) := \sum_{x,y} P(x)W(y|x) \left[\ln \frac{W(y|x)}{q_P(y)} \right.
$$
$$
\left. - \sum_z \frac{P(z)W(y|z)}{q_P(y)} \ln \frac{W(y|z)}{q_P(y)} \right]^2. \quad (25)
$$

Specifically, for a symmetric channel W and $P \in \mathcal{P}(\mathcal{X})$ with $P(x) > 0$ for all $x \in \mathcal{X}$, $V^r(P, W) = 0$ if and only if W is singular. To see this, note that if P has full support then $V^r(P, W) = 0$ if and only if

$$
\ln W(y|x) = \sum_z \frac{P(z)W(y|z)}{q_P(y)} \ln W(y|z), \quad (26)
$$

for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ such that $W(y|x) > 0$. In light of Definition 2, the latter condition is equivalent to saying that W is singular.

In [32, Lemma 52], it is claimed that $V^r(P, W) = 0$ if and only if

$$
\forall (x, y, y'): \ W(y|x) = W(y'|x) \text{ or } P(x)W(y|x) = 0. \quad (27)
$$

By choosing $P = U_{\mathcal{X}}$ and W to be a BEC with parameter $\delta \in (0, 1)$, one can verify that $V^r(P, W) = 0$ by elementary calculation. Evidently, this (P, W) pair does not satisfy (27). For more on singularity, see [14, Remark 1].

C. Statement of the Results

Theorem 1: Let W be a symmetric and nonsingular channel with $R_{cr} < C(W)$.

(i) For any $R_{cr} < R < C(W)$ and any N,

$$
P_e(N, R) \leq \frac{K_1}{N^{\frac{1}{2}(1+|E_r'(R)|)}} \exp\{-NE_r(R)\}, \quad (28)
$$

where K_1 is a positive constant that depends on W and R.

(ii) For any $R_\infty < R < C(W)$ and any N,

$$
\bar{P}_{e,fb}(N, R) \geq \frac{\tilde{K}_1}{N^{\frac{1}{2}(1+|E_{SP}'(R)|)}} \exp\{-NE_{SP}(R)\}, \quad (29)
$$

where $\tilde{K}_1$ is a positive constant that depends on W and R.

Proof: Theorem 1 is proven in Section III-A.

Theorem 2: Let W be a symmetric and singular channel with $R_{cr} < C(W)$.

(i) For any $R_{cr} < R < C(W)$ and any N,

$$
P_e(N, R) \leq \frac{K_2}{\sqrt{N}} \exp\{-NE_r(R)\}, \quad (30)
$$

where K_2 is a positive constant that depends on W and R.

(ii) For any $R_\infty < R < C(W)$ and any N,

$$
\bar{P}_{e,fb}(N, R) \geq \frac{\tilde{K}_2}{\sqrt{N}} \exp\{-NE_{SP}(R)\}, \quad (31)
$$

where $\tilde{K}_2$ is a positive constant that depends on W and R.

Proof: Theorem 2 is proven in Section III-B.

Remark 1: (i) For any $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, the following three statements are equivalent (e.g., [9, pg. 160]): $R_{cr} < C$, $R_\infty < C$, and the dispersion of W is positive.

(ii) Recall that at rates above the critical rate, $E_{SP}(R) = E_r(R)$ by definition. Thus the exponents in (28)–(31) are all the same in this regime.

(iii) As mentioned in Section I, if every row (resp. column) of the channel is a permutation of every other row (resp. column), then we call it a *strongly symmetric channel*. When particularized to this class of channels without feedback, Theorem 1 reduces to a result of Dobrushin [5] by noting the fact that any strongly symmetric channel with $R_{cr} < C$ is necessarily nonsingular (e.g., [14, Footnote 3]).

(iv) For rates above the critical rate, the ratios of the upper and lower bounds in Theorems 1 and 2 are bounded away from 0 and ∞ as $N \to \infty$. Indeed, we can explicitly deduce the constants in both theorems from their proofs, although they are not optimized since our goal in this work is to prove an order-optimal pre-factor. Nevertheless, it would be interesting to refine the bounds so that their ratio converges to 1. A first step in this direction is the work of Scarlett *et al.* [18], in which the rate dependence of the pre-factor's constant is investigated for the random coding (i.e., upper) bound. See Font-Segura *et al.* [33] for an analogous, though nonrigorous, study of the sphere-packing bound.

(v) The achievability parts of Theorems 1 and 2, i.e., Theorem 1(i) and Theorem 2(i), follow almost immediately from their asymmetric counterparts [14]. The contribution of the paper lies in the impossibility parts, i.e., Theorem 1(ii) and Theorem 2(ii), and the normal approximation results to follow.

The technique used to prove part (ii) of Theorem 2 can also be used to prove the next two results, the first of which

fills a gap in the literature on the normal approximation (see Theorem 5 to follow).

Theorem 3: Given $\epsilon \in (0,1)$ and a singular, symmetric W with $V_\epsilon(W) > 0$, for any N,

$$\ln M_{\text{fb}}^*(N, \epsilon) \leq N \cdot C(W) + \sqrt{N \cdot V_\epsilon(W)}\Phi^{-1}(\epsilon) + K(\epsilon, W), \tag{32}$$

where $K(\epsilon, W) \in \mathbb{R}^+$ is a constant that depends on ϵ and W.

Proof: Given in Section III-C.

Theorem 4: Given a singular and asymmetric W,

(i) If $\epsilon \in (0, 1/2)$, then for all N,

$$\ln M_{\text{c}}^*(N, \epsilon) \leq N \cdot C(W)$$
$$+ \sqrt{N \cdot V_\epsilon(W)}\Phi^{-1}(\epsilon) + \tilde{K}(\epsilon, W), \quad (33)$$

where $\tilde{K}(\epsilon, W) \in \mathbb{R}^+$ is a constant that depends on ϵ and W.

(ii) If $\epsilon \in (1/2, 1)$ and $V_\epsilon(W) > 0$, then for all N,

$$\ln M_{\text{c}}^*(N, \epsilon) \leq N \cdot C(W)$$
$$+ \sqrt{N \cdot V_\epsilon(W)}\Phi^{-1}(\epsilon) + \hat{K}(\epsilon, W), \quad (34)$$

where $\hat{K}(\epsilon, W) \in \mathbb{R}^+$ is a constant that depends on ϵ and W.

Proof: Given in Section III-D.

Note that the set of asymmetric and singular channels is not empty. For an example, let $\mathcal{X} := \{0,1,2\}$, $\mathcal{Y} := \{0,1,2,3\}$ and consider

$$W(y|x) := \begin{cases} 2/3, & (x,y) = (0,0), \\ 1/6, & (x,y) \in \{(0,1),(0,3),(1,3),(2,1)\}, \\ 5/6, & (x,y) \in \{(1,2),(2,2)\}, \\ 0, & \text{else.} \end{cases} \tag{35}$$

Theorem 3 completes the proof of the following assertion:

Theorem 5: Given a symmetric W and $\epsilon \in (0,1)$,

(a) If W is nonsingular and $V_\epsilon(W) > 0$, then

$$\ln M^*(N, \epsilon) = N \cdot C(W) + \sqrt{N \cdot V_\epsilon(W)}\Phi^{-1}(\epsilon)$$
$$+ \ln \sqrt{N} + \Theta(1). \quad (36)$$

(b) If W is singular and $V_\epsilon(W) > 0$, then

$$\ln M^*(N, \epsilon) = N \cdot C(W) + \sqrt{N \cdot V_\epsilon(W)}\Phi^{-1}(\epsilon) + \Theta(1). \tag{37}$$

(c) If $V_\epsilon(W) = 0$, then

$$\ln M^*(N, \epsilon) = N \cdot C(W) + \Theta(1). \tag{38}$$

Specifically, achievability of item (a) follows from [32, Corollary 54]. The converse of item (a) follows from [32, Theorem 55]. Achievability of item (b) follows from [32, Theorem 47], coupled with Lemma 10(ii) to follow. The converse for item (b) is proven in Theorem 3. Item (c) is proven in [32, Corollary 57].

For bounds on the constant in (36), see Moulin [34].

We assume that ϵ-dispersion is positive in Theorem 4(ii) in order to exclude exotic channels; this allows us to focus on the role of singularity. See [32, p. 68] and [31, Section III] for a discussion of exotic channels.

III. PROOFS

We begin with a result that will be used in the proofs of Theorems 1, 2, and 3, and can be interpreted as the essential reason behind ideal feedback's inability to improve even the sub-exponential factor (resp. third-order term) in the fixed-rate (resp. fixed-error probability) regime for symmetric channels. In particular, (40) to follow demonstrates that even if the channel inputs can be chosen by using all the previous channel outputs, the induced distribution of the information density would be the same as picking any input repeatedly. Thus, for the symmetric channels we consider in this paper, one can intuitively posit that no input symbol is different from any other.

To state the result, fix some symmetric channel W and let $\{\mathcal{Y}_l\}_{l=1}^L$ be a partition of the columns of W mentioned in Definition 1, whose choice is immaterial in what follows. Consider any sequence of mappings $\{\zeta_n : \mathcal{M} \times \mathcal{Y}^{n-1} \to \mathcal{X}\}_{n=1}^N$ and $Q_Y \in \mathcal{P}(\mathcal{Y})$ that satisfies the following:

(i) For any $x \in \mathcal{X}$, $Q_Y \gg W(\cdot|x)$.

(ii) For any $l \in \{1, \ldots, L\}$, $Q_Y(y)$ has the same value for all $y \in \mathcal{Y}_l$.

Finally, for any $x \in \mathcal{X}$ and $\lambda \in \mathbb{R}$, define

$$M_x(\lambda) := \sum_{y \in \text{supp}(W(\cdot|x))} W(y|x)^{1-\lambda}Q_Y(y)^\lambda. \tag{39}$$

Proposition 1: Assume the two conditions before (39) are satisfied. Then

(i) For any $\lambda \in \mathbb{R}$, $M_x(\lambda)$ is finite and constant over $x \in \mathcal{X}$.

(ii) Let $\mathbf{Y}$ be generated via $\{\zeta_n(m, \cdot)\}$ for some fixed m and the channel $W(\cdot|\cdot)$. Then the distribution of

$$\sum_{n=1}^N \ln \frac{W(Y_n|\zeta_n(m, \mathbf{Y}^{n-1}))}{Q_Y(Y_n)} \tag{40}$$

is invariant to m and $\{\zeta_n(m, \cdot)\}$. In particular, the choice $\zeta_n(m, \mathbf{y}^{n-1}) = x_o$ for all m, n, and $\mathbf{y}^{n-1}$ and some fixed x_o induces the same distribution as does every other $\{\zeta_n(m, \cdot)\}$.

Proof: (i) $M_x(\lambda) \in \mathbb{R}$ directly follows from the assumption that $W(\cdot|x) \ll Q_Y$ for any $x \in \mathcal{X}$, along with the fact that $|\mathcal{Y}| < \infty$. Further, the assumption that $Q_Y(y)$ has the same value for all $y \in \mathcal{Y}_l$, coupled with the fact that every row is a permutation of every other row for any sub-channel defined by the aforementioned partition, i.e., $\{\mathcal{Y}_l\}_{l=1}^L$, allows us to conclude the proof of the second assertion.

(ii) For notational convenience, define

$$R_{\mathbf{Y}^N|M}(\mathbf{y}^N|m) := \prod_{n=1}^N W(y_n|\zeta_n(m, \mathbf{y}^{n-1})). \tag{41}$$

For any $\lambda \in \mathbb{R}$, define

$$M_m(\lambda) := \mathrm{E}_{R_{\mathbf{Y}^N|M}(\cdot|m)} \left[\exp \left\{ \lambda \ln \frac{Q_{\mathbf{Y}^N}(\mathbf{Y}^N)}{R_{\mathbf{Y}^N|M}(\mathbf{Y}^N|m)} \right\} \right], \tag{42}$$

where $Q_{\mathbf{Y}^N}(\mathbf{y}^N) := \prod_{n=1}^N Q_Y(y_n)$. We have

$$M_m(\lambda) = \sum_{y_1 \in \mathcal{Y}} \cdots \sum_{y_N \in \mathcal{Y}} \prod_{n=1}^N W(y_n|\zeta_n(m, \mathbf{y}^{n-1})) \tag{43}$$

$$\cdot \exp \left\{ \lambda \ln \frac{Q_Y(y_n)}{W(y_n|\zeta_n(m, \mathbf{y}^{n-1}))} \right\} \tag{44}$$

$$= [M_{x_o}(\lambda)]^N, \tag{45}$$

where (45) follows from the first assertion of this lemma. Since

$$\mathrm{E}_{W(\cdot|\mathbf{x}_o^N)} \left[\exp \left\{ \lambda \ln \frac{Q_{\mathbf{Y}^N}(\mathbf{Y}^N)}{W(\mathbf{Y}^N|\mathbf{x}_o^N)} \right\} \right] = [M_{x_o}(\lambda)]^N, \tag{46}$$

(45) and the uniqueness theorem for the moment generating function (e.g., [35, Ex. 26.7]) imply the claim. ∎

We continue with a result that unifies the crucial step of the converse proofs of Theorems 1 and 2. To this end, fix any (N, R) feedback code, (f, φ), with $R_\infty < R < C(W)$.

Proposition 2: Fix some $Q_Y \in \mathcal{P}(\mathcal{Y})$, $r \in (0, R]$, a sequence of sets $\{\mathcal{S}(m, r)\}_{m \in \mathcal{M}}$ with $\mathcal{S}(m, r) \subset \mathcal{Y}^N$ and assume the following:

(i) $Q_Y \gg W(\cdot|x)$ for all $x \in \mathcal{X}$.
(ii) For all $m \in \mathcal{M}$,

$$P_{\mathbf{Y}^N|M} \{\mathcal{S}(m, r)|m\} \geq \frac{k_1}{\sqrt{N}} e^{-N\xi(r, R)} > 0, \tag{47}$$

for some $k_1 \in \mathbb{R}^+$ and $\xi(r, R) \in \mathbb{R}^+$.
(iii) For any $m \in \mathcal{M}$ and $\mathbf{y}^N \in \mathcal{S}(m, r)$ with $P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m) > 0$,

$$\frac{1}{N} \ln \left(\frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|M} \{\mathcal{S}(m, r)|m\}} \cdot \frac{Q_{\mathbf{Y}^N} \{\mathcal{S}(m, r)\}}{Q_{\mathbf{Y}^N}(\mathbf{y}^N)} \right)$$

$$\leq R - \frac{k_2}{N}, \tag{48}$$

for some $k_2 \in \mathbb{R}^+$.
(iv) For some $k_3 \in \mathbb{R}^+$ such that $\frac{e^{-k_2}}{k_3} < 1$,

$$\sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m} \frac{Q_{\mathbf{Y}^N}(\mathbf{y}^N)}{Q_{\mathbf{Y}^N} \{\mathcal{S}(m, r)\}} \mathbb{1}(\mathbf{y}^N \in \mathcal{S}(m, r)) \leq \frac{1}{k_3}. \tag{49}$$

Then we have

$$\bar{\mathrm{P}}_{\mathrm{e,fb}}(f, \varphi) \geq \left(1 - \frac{e^{-k_2}}{k_3} \right) \frac{k_1}{\sqrt{N}} e^{-N\xi(r, R)}. \tag{50}$$

Proof: Owing to the assumptions $Q_Y \gg W(\cdot|x)$ for all $x \in \mathcal{X}$, and (47), we note that $Q_{\mathbf{Y}^N} \{\mathcal{S}(m, r)\} > 0$, which,

in turn, ensures that the expressions in (48) and (49) are well-defined. We then proceed as follows:

$$\bar{\mathrm{P}}_{\mathrm{e,fb}}(f, \varphi)$$

$$= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m^c} P_{\mathbf{Y}^N|M} \{\mathcal{S}(m, r)|m\} \tag{51}$$

$$\cdot \frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|M} \{\mathcal{S}(m, r)|m\}}$$

$$\geq \frac{k_1}{\sqrt{N}} e^{-N\xi(r, R)} \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m^c} \frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|M} \{\mathcal{S}(m, r)|m\}} \tag{52}$$

$$\geq \frac{k_1}{\sqrt{N}} e^{-N\xi(r, R)} \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m^c} \frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|M} \{\mathcal{S}(m, r)|m\}} \tag{53}$$

$$\cdot \mathbb{1}(\mathbf{y}^N \in \mathcal{S}(m, r)),$$

where (51) follows from the definition of $\bar{\mathrm{P}}_{\mathrm{e,fb}}(f, \varphi)$, i.e., (14), along with (47) and (52) follows from (47). Continuing,

$$\bar{\mathrm{P}}_{\mathrm{e,fb}}(f, \varphi) \frac{\sqrt{N}}{k_1} e^{N\xi(r, R)}$$

$$\geq 1 - \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m} \frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|M} \{\mathcal{S}(m, r)|m\}} \tag{54}$$

$$\cdot \mathbb{1}(\mathbf{y}^N \in \mathcal{S}(m, r))$$

$$\geq 1 - \frac{e^{NR - k_2}}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m} \frac{Q_{\mathbf{Y}^N}(\mathbf{y}^N)}{Q_{\mathbf{Y}^N} \{\mathcal{S}(m, r)\}} \tag{55}$$

$$\cdot \mathbb{1}(\mathbf{y}^N \in \mathcal{S}(m, r))$$

$$\geq 1 - \frac{e^{-k_2}}{k_3}, \tag{56}$$

where (55) follows from (48), and (56) follows from (49), along with the fact that $|\mathcal{M}| \geq e^{NR}$. ∎

Further, we state two results that are used in the proofs of both Theorems 1 and 2. To this end, for any symmetric channel $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ with $R_{\mathrm{cr}} < C(W)$ and any $R_\infty < R < C(W)$, define

$$\mathbb{R}^+ \ni \rho_R := - \left. \frac{\partial \mathrm{E}_{\mathrm{SP}}(r, U_\mathcal{X})}{\partial r} \right|_{r=R}, \tag{57}$$

$$\forall y \in \mathcal{Y}, q_R(y) := \frac{\left(\sum_{x \in \mathcal{X}} U_\mathcal{X}(x) W(y|x)^{\frac{1}{1+\rho_R}} \right)^{1+\rho_R}}{\sum_{b \in \mathcal{Y}} \left(\sum_{a \in \mathcal{X}} U_\mathcal{X}(a) W(b|a)^{\frac{1}{1+\rho_R}} \right)^{1+\rho_R}}, \tag{58}$$

where (57) is well-defined thanks to [13, Proposition 3], and its positivity can be verified by using the fact that $\mathrm{E}_{\mathrm{SP}}(R) > 0$.

Proposition 3: Fix a symmetric channel $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ with $R_{\mathrm{cr}} < C(W)$. Consider any $R_\infty < R < C(W)$.

(i) We have

$$\mathrm{E}_{\mathrm{SP}}(R) = \mathrm{E}_{\mathrm{SP}}(R, U_\mathcal{X}) = \tilde{\mathrm{E}}_{\mathrm{SP}}(R, U_\mathcal{X}) = \tilde{\mathrm{E}}_{\mathrm{SP}}(R). \tag{59}$$

(ii) For any $\rho \in \mathbb{R}_+$,

$$
\sum_{y \in \mathcal{Y}} W(y|x)^{\frac{1}{1+\rho}} \left(\sum_{z \in \mathcal{X}} U_{\mathcal{X}}(z) W(y|z)^{\frac{1}{1+\rho}} \right)^{\rho}
$$
$$
= \sum_{y \in \mathcal{Y}} \left(\sum_{z \in \mathcal{X}} U_{\mathcal{X}}(z) W(y|z)^{\frac{1}{1+\rho}} \right)^{1+\rho}, \quad (60)
$$

for all $x \in \mathcal{X}$.

(iii) ρ_R attains the supremum in the definition of $\tilde{\mathrm{E}}_{\mathrm{SP}}(R, U_{\mathcal{X}})$, i.e., (17).

(iv) We have

$$
\mathrm{E}_{\mathrm{SP}}(R, U_{\mathcal{X}}) = \sup_{\rho \in \mathbb{R}_+} \min_{q \in \mathcal{P}(\mathcal{Y})} \left\{ -\rho R \right.
$$
$$
\left. - (1+\rho) \sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) \ln \sum_{y \in \mathcal{Y}} W(y|x)^{\frac{1}{1+\rho}} q(y)^{\frac{\rho}{1+\rho}} \right\},
$$
$$(61)$$

and (ρ_R, q_R) is the unique saddle-point of (61).

Proof: The proof is given in Appendix A.

Next, we state a concentration result, which is proven in [36, Lemma 5] and reproduced here for completeness. Although there are various bounds of this sort, the classical versions in probability theory literature are stated in asymptotic form.

To state the result, let $\{Z_n\}_{n=1}^N$ be independent, real-valued random variables with law ν_n, and assume

$$
\sum_{n=1}^N \mathrm{Var}_{\nu_n}[Z_n] > 0. \quad (62)
$$

Define $\Lambda_n(\lambda) := \ln \mathrm{E}_{\nu_n}\left[e^{\lambda Z_n} \right]$ and assume the existence of a $c \in \mathbb{R}$ with a corresponding $\eta > 0$ satisfying:

(i) There exists a neighborhood of η such that $\frac{1}{N} \sum_{n=1}^N \Lambda_n(\lambda) < \infty$, for all λ in this neighborhood.

(ii) $\frac{1}{N} \sum_{n=1}^N \Lambda_n'(\eta) = c$.

For any $b \in \mathbb{R}$, $\Lambda_N^*(b)$ denotes the Fenchel-Legendre transform of $\frac{1}{N} \sum_{n=1}^N \Lambda_n(\cdot)$ at b, i.e.,

$$
\Lambda_N^*(b) := \sup_{\lambda \in \mathbb{R}} \left\{ \lambda b - \frac{1}{N} \sum_{n=1}^N \Lambda_n(\lambda) \right\}. \quad (63)
$$

Define

$$
\frac{\mathrm{d}\tilde{\nu}_n}{\mathrm{d}\nu_n}(z) := e^{\eta z - \Lambda_n(\eta)}, \quad (64)
$$
$$
T_n := Z_n - \mathrm{E}_{\tilde{\nu}_n}[Z_n], \quad (65)
$$
$$
m_{2,N} := \sum_{n=1}^N \mathrm{Var}_{\tilde{\nu}_n}[T_n], \quad (66)
$$
$$
m_{3,N} := \sum_{n=1}^N \mathrm{E}_{\tilde{\nu}_n}[|T_n|^3], \quad (67)
$$
$$
t_N := \eta 2\sqrt{2\pi} \frac{m_{3,N}}{m_{2,N}}. \quad (68)
$$

Lemma 1: Assume conditions (i) and (ii) following (62) are satisfied. Then for any $N \in \mathbb{Z}^+$ and $a > 1$,

$$
\Pr\left[\frac{1}{N} \sum_{n=1}^N Z_n \geq c \right] \frac{\eta \sqrt{2\pi m_{2,N}} e^{at_n}}{(1 + at_N)(1 - 1/a)} \geq
$$
$$
\left\{ 1 - \frac{[1+(1+at_N)^2]}{(1+at_N)\eta(1-1/a)2\sqrt{em_{2,N}}} \right\} \exp\left\{ -N\Lambda_N^*(c) \right\}. \quad (69)
$$

Proof: For completeness, we provide an outline of the proof in Appendix B.

We conclude this section with a simple result for sums of independent random variables, which is used in the proofs of both Theorem 3 and Theorem 4. Its derivation is inspired by the proof of [11, Lemma 47]; it is tighter than that result by at least a factor of 2.

Lemma 2: Let $\{Z_n\}_{n=1}^N$ be independent with

$$
m_{2,N} := \sum_{n=1}^N \mathrm{Var}[Z_n] > 0, \quad (70)
$$
$$
m_{3,N} := \sum_{n=1}^N \mathrm{E}\left[|Z_n - \mathrm{E}[Z_n]|^3 \right] < \infty. \quad (71)
$$

Then for any $r \in \mathbb{R}$,

$$
\mathrm{E}\left[\mathbb{1}\left\{ \sum_{n=1}^N Z_n \leq r \right\} \exp\left\{ -\left[r - \sum_{n=1}^N Z_n \right] \right\} \right]
$$
$$
\leq \frac{1}{\sqrt{2\pi m_{2,N}}} + \frac{2m_{3,N}}{m_{2,N}^{3/2}}. \quad (72)
$$

Further, if the random variables are also identically distributed, then

$$
\mathrm{E}\left[\mathbb{1}\left\{ \sum_{n=1}^N Z_n \leq r \right\} \exp\left\{ -\left[r - \sum_{n=1}^N Z_n \right] \right\} \right]
$$
$$
\leq \frac{1}{\sqrt{2\pi m_{2,N}}} + \frac{m_{3,N}}{m_{2,N}^{3/2}}. \quad (73)
$$

Proof: The proof is given in Appendix C.

A. Proof of Theorem 1

The upper bound, (28), follows from an application of [14, Theorem 2(ii)] with the pair $(U_{\mathcal{X}}, W)$, which is nonsingular under [14, Definition 1] by Definition 2.

The first step of the proof of (29) is the application of Proposition 2. To this end, let (f_N, φ_N) denote an arbitrary (N, R) code with ideal feedback, and ρ_R (resp. q_R) be as defined in (57) (resp. (58)). Evidently, $q_R(y) > 0$ for all $y \in \mathcal{Y}$, since without loss of generality we can assume that W has no all-zero columns, which, in turn, ensures that $q_R \gg W(\cdot|x)$ for any $x \in \mathcal{X}$. Note that condition (i) in Proposition 2 is satisfied with this choice. To define the set to be used in the desired particularization, for any $R_\infty < r \leq R$, define

$$
\mathrm{e}_{\mathrm{SP}}(r, R) := \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X}) : \, \mathrm{D}(V\|q_R|U_{\mathcal{X}}) \leq r} \mathrm{D}(V\|W|U_{\mathcal{X}}). \quad (74)
$$

For any $\mathbf{x}^N \in \mathcal{X}^N$, $m \in \mathcal{M}$ and $r \in \mathbb{R}_+$, let

$$\mathcal{S}(m,r) :=$$
$$\left\{ \mathbf{y}^N \in \mathcal{Y}^N : \frac{1}{N} \sum_{n=1}^N \ln \frac{W(y_n|f_n(m, \mathbf{y}^{n-1}))}{q_R(y_n)} \right.$$
$$\left. \leq r - e_{\mathrm{SP}}(r, R) \right\}. \quad (75)$$

We also use the notation $\mathcal{S}(m,r)$ to refer to the event

$$\left\{ \mathbf{Y}^N \in \mathcal{S}(m,r) \right\}. \quad (76)$$

This convention will be used with other similar quantities that are introduced later.

To verify conditions (ii)–(iv) in Proposition 2, we need the following definitions: Fix some $x_o \in \mathcal{X}$ whose choice is immaterial in what follows, for any $\lambda \in \mathbb{R}$, we define

$$\Lambda(\lambda) := \ln \mathrm{E}_{W(\cdot|x_o)} \left[\exp \left\{ \lambda \ln \frac{q_R(Y)}{W(Y|x_o)} \right\} \right]. \quad (77)$$

Let $\{\mathcal{Y}_l\}_{l=1}^L$ be a partition of the columns of W mentioned in Definition 1, whose choice is immaterial in what follows. Since each column is a permutation of any other column for any sub-channel defined by this partition,

$$\left(\sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) W(y|x)^{\frac{1}{1+\rho_R}} \right)^{1+\rho_R} \quad (78)$$

has the same value for any $y \in \mathcal{Y}_l$. This observation, coupled with the definition of q_R, i.e., (58), allows us to invoke Proposition 1 in what follows.

Specifically, particularization of Proposition 1(i) with $Q_Y \leftarrow q_R$ implies that $\Lambda(\cdot)$ is finite over the entire real line, which, in turn, ensures that $\Lambda(\cdot)$ is a smooth function on $\mathbb{R}$ [42, Ex. 2.2.24].

Next, for any $x \in \mathcal{X}$, let

$$W_R(y|x) := \frac{q_R(y)}{q_R\{\mathrm{supp}(W(\cdot|x))\}} \mathbb{1}\left\{ y \in \mathrm{supp}(W(\cdot|x)) \right\}, \quad (79)$$

where $q_R\{\mathrm{supp}(W(\cdot|x))\} := \sum_{y \in \mathrm{supp}(W(\cdot|x))} W(y|x)$ and we will use similar convention in what follows. Further, define $\bar{R} := \frac{1}{2}(R + \mathrm{D}(W_R\|q_R|U_{\mathcal{X}}))$ and, for any $b \in \mathbb{R}$, let $\Lambda^*(b)$ denote the Fenchel-Legendre transform of $\Lambda(\cdot)$ at b, i.e.,

$$\Lambda^*(b) = \sup_{\lambda \in \mathbb{R}} \left\{ \lambda b - \Lambda(\lambda) \right\}. \quad (80)$$

The next result collects useful properties of the aforementioned quantities.

Lemma 3: (i) $R > \mathrm{D}(W_R\|q_R|U_{\mathcal{X}})$.
(ii) $e_{\mathrm{SP}}(R, R) = E_{\mathrm{SP}}(R)$.
(iii) $\Lambda''(\lambda) > 0$, for any $\lambda \in [0, 1)$.
(iv) $s_{(\cdot)} : (\mathrm{D}(W_R\|q_R|U_{\mathcal{X}}), R] \rightarrow \mathbb{R}$ s.t. $s_r := -\left. \frac{\partial e_{\mathrm{SP}}(a, R)}{\partial a} \right|_{a=r}$ is a well-defined, continuous, positive and strictly decreasing function.
(v) Fix some $r \in (\mathrm{D}(W_R\|q_R|U_{\mathcal{X}}), R]$. We have

$$\Lambda^*(e_{\mathrm{SP}}(r, R) - r) = e_{\mathrm{SP}}(r, R). \quad (81)$$

Moreover, $\eta_r := \frac{s_r}{1+s_r} \in (0, 1)$ is the unique real number that satisfies

$$\Lambda'(\eta_r) = e_{\mathrm{SP}}(r, R) - r. \quad (82)$$

(vi) $s_R = \rho_R$.
Proof: The proof is given in Appendix D.
Lemma 4: For some $k_1 \in \mathbb{R}^+$ that is a function of $R, \bar{R}$, and W, $k_o \in \mathbb{R}^+$ that is a function of $R, \bar{R}$ and W, such that $k_o + \ln k_1 > 0$, and for any $m \in \mathcal{M}$, we have

$$P_{\mathbf{Y}^N|M} \left\{ \mathcal{S}(m, R_N)|m \right\} \geq \frac{k_1}{\sqrt{N}} e^{-N e_{\mathrm{SP}}(R_N, R)} > 0, \quad (83)$$

for all sufficiently large N, where $R_N := R - \frac{k_o + \ln \sqrt{N}}{N}$.
Proof: We begin by noticing that $W_R(\cdot|x) \equiv W(\cdot|x)$ for all $x \in \mathcal{X}$. For any $x \in \mathcal{X}$ and $\lambda \in [0, 1)$, define

$$\tilde{W}_\lambda(y|x) := \frac{W(y|x)^{1-\lambda} q_R(y)^\lambda}{\sum_{b \in \mathcal{Y}} W(b|x)^{1-\lambda} q_R(b)^\lambda}. \quad (84)$$

Via routine calculations, we deduce that

$$\Lambda'(\lambda) = \mathrm{E}_{\tilde{W}_\lambda(\cdot|x_o)} \left[\ln \frac{q_R(Y)}{W(Y|x_o)} \right], \quad (85)$$

$$\Lambda''(\lambda) = \mathrm{Var}_{\tilde{W}_\lambda(\cdot|x_o)} \left[\ln \frac{q_R(Y)}{W(Y|x_o)} \right]. \quad (86)$$

Similarly, for any $\lambda \in [0, 1)$, define

$$m_3(\lambda) := \mathrm{E}_{\tilde{W}_\lambda(\cdot|x_o)} \left[\left| \ln \frac{q_R(Y)}{W(Y|x_o)} - \Lambda'(\lambda) \right|^3 \right]. \quad (87)$$

From (84)–(87), one can verify that $\Lambda'(\cdot), \Lambda''(\cdot)$ and $m_3(\cdot)$ are continuous over $[0, 1)$.

Due to Lemma 3(i), $\bar{R} \in (\mathrm{D}(W_R\|q_R|U_{\mathcal{X}}), R)$. Moreover, as a direct consequence of Lemma 3(iv) and (v),

$$0 < \eta_R < \eta_r < \eta_{\bar{R}} < 1, \quad (88)$$

for any $r \in (\bar{R}, R)$. Fix an arbitrary $a > 1$ and define

$$t_{\max} := a 2 \sqrt{2\pi} \eta_{\bar{R}} \max_{\lambda \in [0, \eta_{\bar{R}}]} \frac{m_3(\lambda)}{\Lambda''(\lambda)}, \quad (89)$$

$$m_{2,\min} := \min_{\lambda \in [0, \eta_{\bar{R}}]} \Lambda''(\lambda), \quad (90)$$

$$m_{2,\max} := \max_{\lambda \in [0, \eta_{\bar{R}}]} \Lambda''(\lambda). \quad (91)$$

Evidently all of the aforementioned quantities are well-defined, positive and finite. Finally, define

$$\frac{e^{-t_{\max}} \left(1 - \frac{1}{a} \right)}{\eta_{\bar{R}} 2 \sqrt{2\pi m_{2,\max}}} =: k_1 \in \mathbb{R}^+. \quad (92)$$

Fix some $k_o \in \mathbb{R}^+$ such that $k_o + \ln k_1 > 0$. For any $N \in \mathbb{Z}^+$, define $R_N := R - \frac{1}{N}(k_o + \ln \sqrt{N})$. Consider a sufficiently large $N \in \mathbb{Z}^+$, such that

$$R_N \geq \bar{R}, \quad (93)$$

$$\frac{1 + (1 + t_{\max})^2}{\eta_R (1 - 1/a) 2 \sqrt{eN m_{2,\min}}} \leq 1/2. \quad (94)$$

Particularization of Proposition 1(ii) with $Q_Y \leftarrow q_R$ implies that for any $m \in \mathcal{M}$, we have

$$P_{\mathbf{Y}^N|M}\{\mathcal{S}(m, R_N)|m\} = W\{\mathcal{S}(\mathbf{x}_o^N, R_N)|\mathbf{x}_o^N\} \qquad (95)$$

$$\geq \frac{k_1}{\sqrt{N}} \exp\{-N\mathrm{e}_{\mathrm{SP}}(R_N, R)\} \quad (96)$$

$$> 0, \qquad (97)$$

where (96) follows from Lemma 1, whose application is ensured by Lemma 3(iii) and (v), coupled with (92), (93) and (94), and (97) follows by noticing

$$\mathrm{e}_{\mathrm{SP}}(R_N, R) \leq \mathrm{e}_{\mathrm{SP}}(\bar{R}, R) < \infty. \qquad (98)$$

∎

Lemma 5: For any $m \in \mathcal{M}$ and $\mathbf{y}^N \in \mathcal{Y}^N$ with $P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m) > 0$

$$\frac{1}{N} \ln\left(\frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|M}\{\mathcal{S}(m, R_N)|m\}}\frac{Q_{\mathbf{Y}^N}\{\mathcal{S}(m, R_N)\}}{Q_{\mathbf{Y}^N}(\mathbf{y}^N)}\right)$$
$$\leq R - \frac{k_2}{N}, \qquad (99)$$

where $k_2 := k_o + \ln k_1$ and R_N, k_o, and k_1 are as defined in Lemma 4.

Proof: For the sake of notational convenience, we define

$$P_{\mathbf{Y}^N|M,\mathcal{S}(m,R_N)}(\mathbf{y}^N|m) := \qquad (100)$$
$$\frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|M}\{\mathcal{S}(m, R_N)|m\}}\mathbb{1}\{\mathbf{y}^N \in \mathcal{S}(m, R_N)\},$$

$$P_{\mathbf{Y}^N|\mathcal{S}(m,R_N)}(\mathbf{y}^N) := \qquad (101)$$
$$\frac{q_R(\mathbf{y}^N)}{q_R\{\mathcal{S}(m, R_N)\}}\mathbb{1}\{\mathbf{y}^N \in \mathcal{S}(m, R_N)\},$$

and note that since $q_R \gg W(\cdot|x)$, (97) ensures that both (100) and (101) are well-defined probability measures.

Next, fix any $m \in \mathcal{M}$ and $\mathbf{y}^N \in \mathcal{S}(m, R_N)$ with $P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m) > 0$. We have

$$\frac{1}{N} \ln \frac{P_{\mathbf{Y}^N|M,\mathcal{S}(m,R_N)}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|\mathcal{S}(m,R_N)}(\mathbf{y}^N)} \qquad (102)$$

$$= \frac{1}{N} \ln \frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{q_R(\mathbf{y}^N)} + \frac{1}{N} \ln \frac{q_R\{\mathcal{S}(m, R_N)\}}{P_{\mathbf{Y}^N|M}\{\mathcal{S}(m, R_N)|m\}} \qquad (103)$$

$$\leq \frac{1}{N} \ln \frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{q_R(\mathbf{y}^N)} + \mathrm{e}_{\mathrm{SP}}(R_N, R) + \frac{\ln\sqrt{N}}{N} - \frac{\ln k_1}{N} \qquad (104)$$

$$\leq R - \frac{k_2}{N}, \qquad (105)$$

where (103) follows from the definitions of $P_{\mathbf{Y}^N|M,\mathcal{S}(m,R_N)}$ and $P_{\mathbf{Y}^N|\mathcal{S}(m,R_N)}$, i.e., (100) and (101), (104) follows from (96) and (105) follows from the definition of $\mathcal{S}(m, R_N)$, i.e., (75), along with the fact that $k_2 - k_o = \ln k_1$. ∎

Lemma 6: Consider any $k_3 \in (0, 1)$ that satisfies $\mathrm{e}^{-k_2} < k_3$, where k_2 is as given in Lemma 5. For any $\{\psi_n : \mathcal{M} \times \mathcal{Y}^{n-1} \to$

$\mathcal{X}\}_{n=1}^N$ and $r \in (\mathrm{D}(W_R\|q_R|U_{\mathcal{X}}), R]$,

$$q_R\left\{\frac{1}{N}\sum_{n=1}^N \ln \frac{W(Y_n|\psi_n(m, \mathbf{Y}^{n-1}))}{q_R(Y_n)} \leq r - \mathrm{e}_{\mathrm{SP}}(r, R)\right\}$$
$$\geq k_3, \quad (106)$$

for all sufficiently large $N \in \mathbb{Z}^+$, independent of $m \in \mathcal{M}$.

Proof: Let $x_o \in \mathcal{X}$ be as given prior to (77). First, note that

$$q_R\left\{\frac{1}{N}\sum_{n=1}^N \ln \frac{W(Y_n|\psi_n(m, \mathbf{Y}^{n-1}))}{q_R(Y_n)} \leq r - \mathrm{e}_{\mathrm{SP}}(r, R)\right\}$$
$$= q_R\left\{\frac{1}{N}\sum_{n=1}^N \ln \frac{W(Y_n|x_o)}{q_R(Y_n)} \leq r - \mathrm{e}_{\mathrm{SP}}(r, R)\right\}, \quad (107)$$

which follows from the fact that, by the symmetry of the channel, for any $x \in \mathcal{X}$, $\ln\frac{W(Y|x)}{q_R(Y)}$ and $\ln\frac{W(Y|x_o)}{q_R(Y)}$ have the same distribution when Y has distribution q_R.

We conclude the proof of Lemma 6 as follows: first, assume that there exists a pair $(x, y) \in \mathcal{X} \times \mathcal{Y}$ with $W(y|x) = 0$. The symmetry of the channel ensures that there exists $y_o \in \mathcal{Y}$ such that $W(y_o|x_o) = 0$. Note that

$$\left\{\mathbf{y}^N \in \mathcal{Y}^N : \frac{1}{N}\sum_{n=1}^N \ln \frac{W(y_n|x_o)}{q_R(y_n)} > r - \mathrm{e}_{\mathrm{SP}}(r, R)\right\}$$
$$\subseteq \{\mathcal{Y}\backslash\{y_o\}\}^N, \qquad (108)$$
$$q_R\{\mathcal{Y}\backslash\{y_o\}\} < 1, \qquad (109)$$

which are direct consequences of the fact that $\mathrm{supp}(q_R) = \mathcal{Y}$. From (108) and (109), we conclude that

$$q_R\left\{\frac{1}{N}\sum_{n=1}^N \ln \frac{W(Y_n|x_o)}{q_R(Y_n)} \leq r - \mathrm{e}_{\mathrm{SP}}(r, R)\right\} \geq k_3, \quad (110)$$

for all sufficiently large $N \in \mathbb{Z}^+$.

Next, assume that for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, $W(y|x) > 0$. For any $\lambda \in \mathbb{R}$,

$$\Lambda_1(\lambda) := \ln \mathrm{E}_{q_R}\left[\exp\left\{\lambda \ln \frac{W(Y|x_o)}{q_R(Y)}\right\}\right] = \Lambda(1-\lambda), \quad (111)$$

as a direct consequence of the positivity of W. Equation (111), along with Lemma 3(v), implies that there exists $\eta_r \in (0, 1)$ with

$$[\Lambda'(\eta_r) = \mathrm{e}_{\mathrm{SP}}(r, R) - r] \iff [\Lambda_1'(1 - \eta_r) = r - \mathrm{e}_{\mathrm{SP}}(r, R)]. \qquad (112)$$

Further, Lemma 3(iii) ensures that for any $\lambda \in (0, 1)$,

$$[\Lambda''(\lambda) > 0] \iff [\Lambda_1''(1 - \lambda) > 0] \iff [\Lambda_1''(\lambda) > 0]. \quad (113)$$

From (112) and (113), we infer that

$$\mu_{x_o} := \mathrm{E}_{q_R}\left[\ln \frac{W(Y|x_o)}{q_R(Y)}\right] \qquad (114)$$

$$= \Lambda_1'(0) \qquad (115)$$

$$< \Lambda_1'(1 - \eta_r) \qquad (116)$$

$$= r - \mathrm{e}_{\mathrm{SP}}(r, R), \qquad (117)$$

$$\sigma_{x_o}^2 := \mathrm{Var}_{q_R}\left[\ln \frac{W(Y|x_o)}{q_R(Y)}\right] \qquad (118)$$

$$= \Lambda_1''(0) \in \mathbb{R}, \qquad (119)$$

where the boundedness of $\Lambda_1''(0)$ is an immediate consequence of the positivity of W and the fact that the input and output alphabets are finite. Hence, Chebyshev's inequality, coupled with (117) and (119), implies that

$$q_R \left\{ \frac{1}{N} \sum_{n=1}^{N} \ln \frac{W(Y_n|x_o)}{q_R(Y_n)} \leq r - e_{\mathrm{SP}}(r, R) \right\}$$
$$\geq 1 - \frac{\sigma_{x_o}^2}{N[\Lambda_1'(1 - \eta_r) - \mu_{x_o}]^2} \geq k_3, \quad (120)$$

for all sufficiently large $N \in \mathbb{Z}^+$. Equations (107), (110) and (120) imply (106). ∎

Lemmas 4, 5, and 6 ensure the conditions (ii), (iii), and (iv) of Proposition 2 are satisfied. Hence, we invoke Proposition 2 with the corresponding values to conclude that

$$\bar{P}_{e,\mathrm{fb}}(f_N, \varphi_N) \geq \left(1 - \frac{e^{-k_2}}{k_3} \right) \frac{k_1}{\sqrt{N}} \exp\{-N e_{\mathrm{SP}}(R_N, R)\}, \quad (121)$$

for all sufficiently large N.

Lemma 7: Let $\varepsilon_N := \frac{k_o + \ln \sqrt{N}}{N}$.

$$e_{\mathrm{SP}}(R_N, R) \leq E_{\mathrm{SP}}(R) + \varepsilon_N |E_{\mathrm{SP}}'(R)|$$
$$+ \varepsilon_N^2 \frac{(1 + s_{\bar{R}})^2}{2 m_{2,\min}} (1 + |E_{\mathrm{SP}}'(R)|). \quad (122)$$

Proof: The proof is given in Appendix E.
Let $N \in \mathbb{Z}^+$ be sufficiently large such that

$$\exp\left\{ -N \varepsilon_N^2 \frac{(1 + s_{\bar{R}})^2}{2 m_{2,\min}} (1 + |E_{\mathrm{SP}}'(R)|) \right\} \geq \frac{1}{2}. \quad (123)$$

Then Lemma 7 and (121) imply that

$$\bar{P}_{e,\mathrm{fb}}(f_N, \varphi_N) \geq$$
$$\frac{k_1}{2} \left(1 - \frac{e^{-k_2}}{k_3} \right) \exp\{-k_o |E_{\mathrm{SP}}'(R)|\} \frac{\exp\{-N E_{\mathrm{SP}}(R)\}}{N^{\frac{1}{2}(1 + |E_{\mathrm{SP}}'(R)|)}}. \quad (124)$$

Since the code is arbitrary, (124) implies (29). ∎

B. Proof of Theorem 2

The achievability proof is similar to its counterpart in Theorem 1. In particular, we begin by invoking [14, Corollary 1(i)] with the pair $(U_{\mathcal{X}}, W)$. However, in that result the singularity of the pairs in $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}|\mathcal{X})$, which differs from the singularity of symmetric channels in Definition 2, is the crucial assumption. As we note next, however, the fact that W is a singular symmetric channel implies that the pair $(U_{\mathcal{X}}, W)$ is singular. Specifically, note that since $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ is a singular symmetric channel, for all $(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{X}$ such that

$$U_{\mathcal{X}}(x) U_{\mathcal{X}}(z) W(y|x) W(y|z) > 0, \quad (125)$$

we have that $W(y|x) = W(y|z)$, which, in light of [14, Definition 1], ensures that the pair $(U_{\mathcal{X}}, W)$ is singular. Owing to the symmetry of the channel, $E_r(\cdot, U_{\mathcal{X}}) = E_r(\cdot)$ on $(R_{\mathrm{cr}}, C(W))$ (e.g., [9, p. 145]). Since $(U_{\mathcal{X}}, W)$ pair is singular, (30) is a direct consequence of [14, Corollary 1(i)].

In order to prove the converse, we apply Proposition 2. To this end, let (f_N, φ_N) denote an arbitrary (N, R) code with ideal feedback, and recall that $q(y) := \sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) W(y|x)$. Due to the singularity of W, given any $y \in \mathcal{Y}$, $W(y|\cdot)$ is either zero or a positive constant that only depends on y, say ξ_y. Hence,

$$q(y) = \xi_y \alpha_y \quad \text{with} \quad \alpha_y := \sum_{x: W(y|x) > 0} U_{\mathcal{X}}(x). \quad (126)$$

Since, without loss of generality, we can assume that W has no all-zero columns, $q(y) > 0$ for all $y \in \mathcal{Y}$ and hence $q \gg W(\cdot|x)$ for any $x \in \mathcal{X}$. Therefore, condition (i) of Proposition 2 is satisfied. For any $r \in \mathbb{R}_+$, define

$$\mathcal{S}(r) := \left\{ \mathbf{y}^N \in \mathcal{Y}^N : \frac{1}{N} \sum_{n=1}^{N} \ln \frac{1}{\alpha_{y_n}} \leq r \right\} \quad (127)$$
$$= \left\{ \mathbf{y}^N \in \mathcal{Y}^N : \frac{1}{N} \sum_{n=1}^{N} \ln \frac{W(y_i|x_i)}{q(y_n)} \leq r \quad \text{for some} \right.$$
$$\left. \mathbf{x}^N \text{ such that } W(\mathbf{y}^N|\mathbf{x}^N) > 0 \right\}. \quad (128)$$

Let $\bar{R} := \frac{R + R_\infty}{2}$. Define $R_N := R - \frac{1}{N}$. Consider a sufficiently large N, such that $R_N \geq \bar{R}$.

Lemma 8: Let $\mathbf{x}_o^N$ denote the sequence consisting of $x_o \in \mathcal{X}$ repeated N times for some x_o, whose choice is immaterial in what follows. Consider any $\{\psi_n\}_{n=1}^{N}$ with $\psi_1 \in \mathcal{X}$ and $\psi_n : \mathcal{Y}^{n-1} \to \mathcal{X}$ for all $n \in \{2, \ldots, N\}$.

(i) For any $r \in \mathbb{R}^+$,

$$\sum_{\mathbf{y}^N \in \mathcal{S}(r)} W(y_1|\psi_1) \prod_{n=2}^{N} W(y_n|\psi_n(\mathbf{y}^{n-1}))$$
$$= W\{\mathcal{S}(r)|\mathbf{x}_o^N\}. \quad (129)$$

(ii) For some $\tilde{K} \in \mathbb{R}^+$ that depends on $R, \bar{R}$ and W,

$$W\{\mathcal{S}(R_N)|\mathbf{x}_o^N\} \geq \frac{\tilde{K}}{\sqrt{N}} \exp\{-N E_{\mathrm{SP}}(R)\} > 0, \quad (130)$$

for all sufficiently large N.

Proof: The proof is given in Appendix F.
For any $m \in \mathcal{M}$, by particularizing Lemma 8 with $\{\psi_n\}_{n=1}^{N} \leftarrow \{f_n(m, \cdot)\}_{n=1}^{N}$, we conclude that condition (ii) of Proposition 2 is satisfied.

Lemma 9: For any $m \in \mathcal{M}$ and $\mathbf{y}^N \in \mathcal{Y}^N$ with $P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m) > 0$

$$\frac{1}{N} \ln \left(\frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|M}\{\mathcal{S}(R_N)|m\}} \frac{q\{\mathcal{S}(R_N)\}}{q(\mathbf{y}^N)} \right) \leq R - \frac{1}{N}. \quad (131)$$

Proof: We begin by noticing the fact that $q \gg W(\cdot|x)$ for any $x \in \mathcal{X}$ implies that the left side of (131) is finite for any $m \in \mathcal{M}$ and $\mathbf{y}^N \in \mathcal{Y}^N$ with $P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m) > 0$.

Fix any $m \in \mathcal{M}$ and $\mathbf{y}^N \in \mathcal{S}(R_N)$ with $P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m) > 0$. First, we claim that

$$q\{\mathcal{S}(R_N)\} = P_{\mathbf{Y}^N|M}\{\mathcal{S}(R_N)|m\}. \quad (132)$$

To see this, note that

$$q\{\mathcal{S}(R_N)\} \tag{133}$$

$$= \sum_{\mathbf{x}^N \in \mathcal{X}^N} U_{\mathcal{X}^N}(\mathbf{x}^N) \sum_{\mathbf{y}^N \in \mathcal{Y}^N} W(\mathbf{y}^N|\mathbf{x}^N) \mathbb{1}\left\{\mathbf{y}^N \in \mathcal{S}(R_N)\right\} \tag{134}$$

$$= \sum_{\mathbf{x}^N \in \mathcal{X}^N} U_{\mathcal{X}^N}(\mathbf{x}^N) W\left\{\mathcal{S}(R_N)|\mathbf{x}^N\right\} \tag{135}$$

$$= \sum_{\mathbf{x}^N \in \mathcal{X}^N} U_{\mathcal{X}^N}(\mathbf{x}^N) W\left\{\mathcal{S}(R_N)|\mathbf{x}_{\mathrm{o}}^N\right\} \tag{136}$$

$$= P_{\mathbf{Y}^N|M}\left\{\mathcal{S}(R_N)|m\right\}, \tag{137}$$

where (136) and (137) follow from Lemma 8(i). Hence,

$$\frac{1}{N}\ln\left(\frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|M}\left\{\mathcal{S}(R_N)|m\right\}}\frac{q\{\mathcal{S}(R_N)\}}{q(\mathbf{y}^N)}\right)$$

$$= \frac{1}{N}\ln\frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{q(\mathbf{y}^N)} \tag{138}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\ln\frac{1}{\alpha_{y_n}} \tag{139}$$

$$\leq R - \frac{1}{N}, \tag{140}$$

where (138) follows from (132), (139) follows from the fact that whenever $W(y|x) > 0$, $\frac{W(y|x)}{q(y)} = \frac{1}{\alpha_y}$, which is a direct consequence of the singularity of the channel, and (140) follows from the definition of $\mathcal{S}(R_N)$, i.e., (127). $\blacksquare$

Clearly, Lemma 9 ensures condition (iii) of Proposition 2 is satisfied. Finally, we note that the following is a probability measure

$$\frac{q(\mathbf{y}^N)}{q\{\mathcal{S}(R_N)\}}\mathbb{1}\left\{\mathbf{y}^N \in \mathcal{S}(R_N)\right\}, \tag{141}$$

which, in turn, ensures that condition (iv) of Proposition 2 is satisfied with $k_3 = 1$, owing to the fact that the decoding regions are disjoint. Thus, invoking Proposition 2 with these particularizations, we get

$$\bar{\mathrm{P}}_{\mathrm{e,fb}}(f_N, \varphi_N) \geq \tilde{K}\left(1 - \mathrm{e}^{-1}\right)\frac{1}{\sqrt{N}}\exp\left\{-N\mathrm{E}_{\mathrm{SP}}(R)\right\}. \tag{142}$$

Since the code is arbitrary, (142) implies (31). $\blacksquare$

C. Proof of Theorem 3

Let $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ be a symmetric and singular channel with $V_\epsilon(W) > 0$. Without loss of generality, assume W has no all-zero columns. Consider any $\epsilon \in (0, 1)$. Similar to Section III-B, define, for all $x \in \mathcal{X}$,

$$M_x(\lambda) := \mathrm{E}_{W(\cdot|x)}\left[\mathrm{e}^{\lambda \ln \frac{W(Y|x)}{q(Y)}}\right], \tag{143}$$

$$m_3(x) := \mathrm{E}_{W(\cdot|x)}\left[\left|\ln\frac{W(Y|x)}{q(Y)} - C(W)\right|^3\right], \tag{144}$$

for any $\lambda \in \mathbb{R}$ (recall that $q(\cdot)$ is the output distribution induced by the uniform input distribution). In the proof to follow, we essentially use the same idea given in Section III-B, and in particular the set $\mathcal{S}(R)$, which is defined in (127).

Lemma 10: Let $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ be a symmetric and singular channel. Let α_y be as defined in (126). Fix an arbitrary $x_{\mathrm{o}} \in \mathcal{X}$.

(i) For any $x \in \mathcal{X}$, $M_x(\lambda) = M_{x_{\mathrm{o}}}(\lambda)$ for all $\lambda \in \mathbb{R}$.

(ii) For all $x \in \mathcal{X}$,

$$\mathrm{E}_{W(\cdot|x)}\left[\ln\frac{W(Y|x)}{q(Y)}\right] = \mathrm{E}_{W(\cdot|x_{\mathrm{o}})}\left[\ln\frac{W(Y|x_{\mathrm{o}})}{q(Y)}\right] \tag{145}$$

$$= C(W), \tag{146}$$

$$\mathrm{Var}_{W(\cdot|x)}\left[\ln\frac{W(Y|x)}{q(Y)}\right] = \mathrm{Var}_{W(\cdot|x_{\mathrm{o}})}\left[\ln\frac{W(Y|x_{\mathrm{o}})}{q(Y)}\right] \tag{147}$$

$$=: V(W) \tag{148}$$

$$= V_\epsilon(W), \tag{149}$$

$$m_3(x) = m_3(x_{\mathrm{o}}). \tag{150}$$

(iii) For any $m \in \mathcal{M}$,

$$P_{\mathbf{Y}^N|M}\left\{\mathcal{S}(R)|m\right\} = W\left\{\mathcal{S}(R)|\mathbf{x}_{\mathrm{o}}^N\right\}. \tag{151}$$

(iv) We have

$$\mathrm{E}_q[-\ln\alpha_Y] = C(W), \tag{152}$$

$$\mathrm{Var}_q[-\ln\alpha_Y] = V(W), \tag{153}$$

$$\mathrm{E}_q[|-\ln\alpha_Y - C(W)|^3] = m_3(x_{\mathrm{o}}). \tag{154}$$

Proof: Since $U_{\mathcal{X}}$ is a capacity-achieving input distribution of W (e.g., [9, Theorem 4.5.2]) and the unique capacity-achieving output distribution has full support (e.g., [9, Corollary 1 and 2 to Theorem 4.5.1]), we conclude that $\alpha_y > 0$, for all $y \in \mathcal{Y}$.

(i) The assertion follows by invoking Proposition 1(i) with $\lambda \leftarrow -\lambda$.

(ii) The first assertion of this lemma, along with the uniqueness theorem for the moment generating function (e.g., [35, Ex. 26.7]), directly implies (145), (146), (147), and (150). (149) is evident in light of (147) and the fact that q is the unique capacity-achieving output distribution of W.

(iii) The assertion is a direct consequence of Lemma 8(i) by particularizing it to $\{\psi_n(\cdot)\}_{n=1}^N \leftarrow \{f_n(m, \cdot)\}_{n=1}^N$ and $r \leftarrow R$.

(iv) The claim directly follows from the second assertion of this lemma on account of the definition of q and the fact that $q(y) = \xi_y\alpha_y$, as noted in (126). $\blacksquare$

Returning to the proof of Theorem 3, we first define

$$k(W) := \frac{m_3(x_{\mathrm{o}})}{V(W)^{3/2}}, \tag{155}$$

$$K(\epsilon, W) := \frac{k(W)\sqrt{V(W)}}{\phi(\Phi^{-1}(\epsilon))} \tag{156}$$

$$+ \frac{2}{\phi(\Phi^{-1}(\epsilon))}\left(\frac{1}{\sqrt{2\pi}} + \frac{m_3(x_{\mathrm{o}})}{V(W)}\right).$$

Evidently, $K(\epsilon, W) \in \mathbb{R}^+$. Choose some $N_o(\epsilon, W) \in \mathbb{Z}^+$ such that for all $N \geq N_o(\epsilon, W)$,

$$1 - \frac{K(\epsilon, W)}{2\phi(\Phi^{-1}(\epsilon))\sqrt{N \cdot V(W)}} > 1/2. \tag{157}$$

Consider any $N \geq N_o(\epsilon, W)$ and define

$$R := C(W) + \sqrt{\frac{V(W)}{N}}\Phi^{-1}(\epsilon) + \frac{K(\epsilon, W)}{N}. \tag{158}$$

Let (f, φ) be an arbitrary (N, R) code with feedback. We claim that

$$\bar{P}_{e,fb}(f, \varphi) \geq W\{\mathcal{S}(R)|\mathbf{x}_o^N\}$$
$$- \sum_{\mathbf{y}^N \in \mathcal{S}(R)} q(\mathbf{y}^N) \exp\left\{-N\left[R - \frac{1}{N}\sum_{k=1}^N \ln\frac{1}{\alpha_{y_k}}\right]\right\}, \tag{159}$$

where $\bar{P}_{e,fb}(f, \varphi)$ denotes the average error probability of the code (f, φ). To see (159), assume $W\{\mathcal{S}(R)|\mathbf{x}_o^N\} > 0$, because otherwise (159) is trivially true. Also, recall that $\mathcal{A}_m \in \mathcal{Y}^n$ denotes the decoding region corresponding to the message $m \in \mathcal{M}$. Define the following probability distributions

$$P_{\mathbf{Y}^N|M,\mathcal{S}(R)}(\mathbf{y}^N|m, \mathcal{S}(R)) := \frac{P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)}{P_{\mathbf{Y}^N|M}\{\mathcal{S}(R)|m\}} \tag{160}$$
$$\cdot \mathbb{1}\{\mathbf{y}^N \in \mathcal{S}(R)\}$$
$$P_{D|\mathbf{Y}^N}(m|\mathbf{y}^N) := \mathbb{1}\{\mathbf{y}^N \in \mathcal{A}_m\}, \tag{161}$$

and note that

$$\bar{P}_{e,fb}(f, \varphi) \tag{162}$$
$$= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m^c} P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m) \tag{163}$$
$$\geq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m^c \cap \mathcal{S}(R)} P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m) \tag{164}$$
$$= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} P_{\mathbf{Y}^N|M}\{\mathcal{S}(R)|m\} \tag{165}$$
$$\cdot \sum_{\mathbf{y}^N \in \mathcal{A}_m^c} P_{\mathbf{Y}^N|M,\mathcal{S}(R)}(\mathbf{y}^N|m, \mathcal{S}(R)).$$

Thus we have

$$- \bar{P}_{e,fb}(f, \varphi) + W\{\mathcal{S}(R)|\mathbf{x}_o^N\}$$
$$\leq \frac{W\{\mathcal{S}(R)|\mathbf{x}_o^N\}}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m} P_{\mathbf{Y}^N|M,\mathcal{S}(R)}(\mathbf{y}^N|m, \mathcal{S}(R))$$
$$\tag{166}$$
$$= e^{-NR} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{S}(R)} P_{D|\mathbf{Y}^N}(m|\mathbf{y}^N)P_{\mathbf{Y}^N|M}(\mathbf{y}^N|m)$$
$$\tag{167}$$
$$\leq \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{S}(R)} \frac{P_{D|\mathbf{Y}^N}(m|\mathbf{y}^N)q(\mathbf{y}^N)}{e^{NR}} \exp\left\{\sum_{k=1}^N \ln\frac{1}{\alpha_{y_k}}\right\}$$
$$\tag{168}$$

$$\leq \sum_{\mathbf{y}^N \in \mathcal{S}(R)} q(\mathbf{y}^N) \exp\left\{-N\left[R - \frac{1}{N}\sum_{k=1}^N \ln\frac{1}{\alpha_{y_k}}\right]\right\},$$
$$\tag{169}$$

where in (166) and (167) we use Lemma 10(iii), and (168) follows from the fact that q dominates $W(\cdot|x)$ for any $x \in \mathcal{X}$, along with the singularity of the channel. This establishes (159).

Since $V(W) > 0$, Lemma 10(iv) enables us to apply Lemma 2 to deduce that

$$\sum_{\mathbf{y}^N \in \mathcal{S}(R)} q(\mathbf{y}^N) \exp\left\{-N\left[R - \frac{1}{N}\sum_{i=1}^N \ln\frac{1}{\alpha_{y_i}}\right]\right\}$$
$$\leq \frac{1}{\sqrt{2\pi N \cdot V(W)}} + \frac{k(W)}{\sqrt{N}}. \tag{170}$$

Next, we claim that

$$W(\mathcal{S}(R)|\mathbf{x}_o^N) \geq \epsilon - \frac{k(W)}{2\sqrt{N}} +$$
$$\frac{K(\epsilon, W)\phi(\Phi^{-1}(\epsilon))}{\sqrt{N \cdot V(W)}}\left\{1 - \frac{K(\epsilon, W)}{\phi(\Phi^{-1}(\epsilon))2\sqrt{N \cdot V(W)}}\right\}. \tag{171}$$

To see (171), we note that

$$W(\mathcal{S}(R)|\mathbf{x}_o^N)$$
$$= W\left\{\frac{1}{N}\sum_{i=1}^N \ln\frac{W(Y_i|x_o)}{q(Y_i)} \leq R \,\Big|\, \mathbf{x}_o^N\right\} \tag{172}$$
$$= W\left\{\frac{1}{\sqrt{N \cdot V(W)}}\sum_{i=1}^N \left[\ln\frac{W(Y_i|x_o)}{q(Y_i)} - C(W)\right] \leq \Phi^{-1}(\epsilon)\right.$$
$$\tag{173}$$
$$\left. + \frac{K(\epsilon, W)}{\sqrt{N \cdot V(W)}} \,\Big|\, \mathbf{x}_o^N\right\}$$
$$\geq \Phi\left(\Phi^{-1}(\epsilon) + \frac{K(\epsilon, W)}{\sqrt{N \cdot V(W)}}\right) - \frac{k(W)}{2\sqrt{N}}, \tag{174}$$

where (172) follows since $q(y) = \xi_y\alpha_y$, along with the singularity of the channel, (173) follows from the definition of R, i.e., (158), and (174) follows from the Berry-Esseen Theorem[2], whose applicability is ensured by Lemma 10(ii) and the fact that $V(W) > 0$. For any $a \in \mathbb{R}$, a second-order power series expansion implies that

$$\Phi(\Phi^{-1}(\epsilon) + a) = \epsilon + \phi(\Phi^{-1}(\epsilon))a + \frac{a^2}{2}\phi'(\bar{a}), \tag{175}$$

for some $\Phi^{-1}(\epsilon) \leq \bar{a} \leq \Phi^{-1}(\epsilon) + a$. Using the fact that $x\phi(x) \leq 1$, which can be verified via elementary calculus,

[2]For convenience, we take the universal constant as 1, although it is not the best possible for independent random variables. See [44] for a survey on the constants of this theorem.

(175) further implies that

$$\Phi(\Phi^{-1}(\epsilon) + a) \geq \epsilon + \phi(\Phi^{-1}(\epsilon))a - \frac{a^2}{2} \quad (176)$$

$$= \epsilon + \phi(\Phi^{-1}(\epsilon))a \left\{ 1 - \frac{a}{2\phi(\Phi^{-1}(\epsilon))} \right\}. \quad (177)$$

By particularizing (177) with $a \leftarrow \frac{K(\epsilon,W)}{\sqrt{NV(W)}}$, one can check that (174) implies (171).

By substituting (170) and (171) into (159), along with (157), we note that, eventually

$$\bar{\mathrm{P}}_{\mathrm{e,fb}}(f,\varphi) > \epsilon + \frac{K(\epsilon,W)\phi(\Phi^{-1}(\epsilon))}{2\sqrt{NV(W)}} - \frac{3k(W)}{2\sqrt{N}} - \frac{1}{\sqrt{2\pi NV(W)}}. \quad (178)$$

Since the code is arbitrary, this implies that eventually,

$$\bar{\mathrm{P}}_{\mathrm{e}}(N,R) > \epsilon, \quad (179)$$

which implies that eventually,

$$\ln M_{\mathrm{fb}}^*(N,\epsilon) \leq N \cdot C(W) + \sqrt{N \cdot V(W)}\Phi^{-1}(\epsilon) + K(\epsilon,W), \quad (180)$$

which, in turn, implies the desired result. ∎

D. Proof of Theorem 4

For any $Q \in \mathcal{P}(\mathcal{X})$, define

$$\alpha_y(Q) := \sum_{x:W(y|x)>0} Q(x), \quad (181)$$

and consider any singular $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$. As mentioned before, the singularity ensures that for any $y \in \mathcal{Y}$, $W(y|x)$ is either 0 or a column-specific positive constant ξ_y. For any $y \in \mathcal{Y}$,

$$q_Q(y) = \xi_y \alpha_y(Q). \quad (182)$$

The following set, which is a generalization of (127), is instrumental in our analysis:

$$\mathcal{S}_R(Q) := \left\{ \mathbf{y}^N : \frac{1}{N} \sum_{i=1}^N \ln \frac{1}{\alpha_{y_i}(Q)} \leq R \right\}, \quad (183)$$

for any $R \in \mathbb{R}_+$.

Lemma 11: Consider a singular $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$. Consider any (N,R) code, say (f,φ), with codewords $\{\mathbf{x}^n(m)\}_{m=1}^{|\mathcal{M}|}$. Let $\bar{\mathrm{P}}_{\mathrm{e}}(f,\varphi)$ denote the average error probability of this code. Fix some $Q \in \mathcal{P}(\mathcal{X})$ and $\mathbf{z}^N \in \mathcal{X}^N$ and assume that for all $m \in \mathcal{M}$, $W(\mathcal{S}_R(Q)|\mathbf{x}^n(m)) = W(\mathcal{S}_R(Q)|\mathbf{z}^N)$ and q_Q dominates $W(\cdot|x)$ for all $x \in \mathrm{supp}(P_{\mathbf{x}^N(m)})$, where $P_{\mathbf{x}^N(m)}$ denotes the empirical distribution of $\mathbf{x}^N(m)$. Then

$$\bar{\mathrm{P}}_{\mathrm{e}}(f,\varphi) \geq W(\mathcal{S}_R(Q)|\mathbf{z}^N) - \sum_{\mathbf{y}^N \in \mathcal{S}_R(Q)} q_Q(\mathbf{y}^N) \exp \left\{ -N \left[R - \frac{1}{N} \sum_{i=1}^N \ln \frac{1}{\alpha_{y_i}(Q)} \right] \right\}. \quad (184)$$

Proof: Assume $W(\mathcal{S}_R(Q)|\mathbf{z}^N) > 0$, otherwise (184) is trivial. For any $\mathbf{x}^N \in \mathcal{X}^N$ with $W(\mathcal{S}_R(Q)|\mathbf{x}^N) > 0$, define

$$P_{\mathbf{Y}^N|\mathbf{X}^N,\mathcal{S}_R(Q)}(\mathbf{y}^N|\mathbf{x}^N,\mathcal{S}_R(Q))$$
$$:= \frac{W(\mathbf{y}^N|\mathbf{x}^N)}{W(\mathcal{S}_R(Q)|\mathbf{x}^N)} \mathbb{1}\{\mathbf{y}^N \in \mathcal{S}_R(Q)\}. \quad (185)$$

Evidently, $P_{\mathbf{Y}^N|\mathbf{X}^N,\mathcal{S}_R(Q)}(\cdot|\mathbf{x}^N,\mathcal{S}_R(Q))$ is a well-defined probability measure. As before, $\{\mathcal{A}_m\}_{m=1}^{|\mathcal{M}|}$ denote the decoding regions of the code and

$$\bar{\mathrm{P}}_{\mathrm{e}}(f,\varphi)$$
$$= \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m^c} W(\mathbf{y}^N|\mathbf{x}^N(m)) \quad (186)$$
$$\geq \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m^c} W(\mathcal{S}_R(Q)|\mathbf{x}^N(m)) \quad (187)$$
$$\cdot P_{\mathbf{Y}^N|\mathbf{X}^N,\mathcal{S}_R(Q)}(\mathbf{y}^N|\mathbf{x}^N(m),\mathcal{S}_R(Q))$$
$$\geq W(\mathcal{S}_R(Q)|\mathbf{z}^N) \cdot \Bigg[1 - \quad (188)$$
$$\frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{A}_m} P_{\mathbf{Y}^N|\mathbf{X}^N,\mathcal{S}_R(Q)}(\mathbf{y}^N|\mathbf{x}^N(m),\mathcal{S}_R(Q)) \Bigg],$$

where (187) follows from (185) and (188) follows from the assumption that $W(\mathcal{S}_R(Q)|\mathbf{x}^N(m)) = W(\mathcal{S}_R(Q)|\mathbf{z}^N)$, for all $m \in \mathcal{M}$. As before, define $P_{D|Y}(m|\mathbf{y}^N) := \mathbb{1}\{\mathbf{y}^N \in \mathcal{A}_m\}$, for all $m \in \mathcal{M}$. Since the decoding regions are mutually exclusive and collectively exhaustive on $\mathcal{M}$, $P_{D|Y}(\cdot|\mathbf{y}^N)$ is a well-defined probability measure. Hence, (188) implies that

$$(\bar{\mathrm{P}}_{\mathrm{e}}(f,\varphi) - W(\mathcal{S}_R(Q)|\mathbf{z}^N))\mathrm{e}^{NR}$$
$$\geq -\sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{S}_R(Q)} P_{D|Y}(m|\mathbf{y}^N)W(\mathbf{y}^N|\mathbf{x}^N(m)) \quad (189)$$
$$\geq -\sum_{m \in \mathcal{M}} \sum_{\mathbf{y}^N \in \mathcal{S}_R(Q)} P_{D|Y}(m|\mathbf{y}^N)q_Q(\mathbf{y}^N)\mathrm{e}^{\sum_{i=1}^N \ln \frac{1}{\alpha_{y_i}(Q)}} \quad (190)$$
$$\geq -\sum_{\mathbf{y}^N \in \mathcal{S}_R(Q)} q_Q(\mathbf{y}^N)\mathrm{e}^{\sum_{i=1}^N \ln \frac{1}{\alpha_{y_i}(Q)}}. \quad (191)$$

where (189) uses again the fact that $W(\mathcal{S}_R(Q)|\mathbf{x}^N(m)) = W(\mathcal{S}_R(Q)|\mathbf{z}^N)$ for all m, (190) follows from the fact that $q_Q(y) = \xi_y \alpha_y(Q)$ and the assumption that for all $m \in \mathcal{M}$, q_Q dominates $W(\cdot|x)$ for all $x \in \mathrm{supp}(P_{\mathbf{x}^N(m)})$. ∎

We analyze three different possibilities for the composition of the code P: large $I(P;W)$ with large $V(P,W)$, large $I(P;W)$ with small $V(P,W)$, and small $I(P;W)$. This idea originated in Strassen [37] and is frequently used in the normal approximation regime.

Specifically, given any $\delta, \nu \in \mathbb{R}^+$, we define

$$\mathcal{S}_3(\delta) := \left\{ P \in \mathcal{P}(\mathcal{X}): \min_{P^* \in \mathcal{P}_W^*} ||P - P^*||_2 > \delta \right\}, \quad (192)$$
$$\mathcal{S}_2(\delta,\nu) := \mathcal{S}_3(\delta)^c \cap \{P \in \mathcal{P}(\mathcal{X}): V(P,W) < \nu\}, \quad (193)$$
$$\mathcal{S}_1(\delta,\nu) := \mathcal{S}_3(\delta)^c \cap \{P \in \mathcal{P}(\mathcal{X}): V(P,W) \geq \nu\}, \quad (194)$$

where $\mathcal{P}_W^* := \{P \in \mathcal{P}(\mathcal{X}): I(P;W) = C(W)\}$.

Lemma 12: Fix some $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ with $C(W) > 0$, $\delta \in \mathbb{R}^+$ and $\epsilon \in (0,1)$. Consider a sequence of constant-composition (N, R_N) codes $\{(f_N, \varphi_N)\}_{N \geq 1}$ with the common composition $Q_N \in \mathcal{S}_3(\delta)$ and

$$R_N := C(W) + \sqrt{\frac{V_\epsilon(W)}{N}} \Phi^{-1}(\epsilon). \qquad (195)$$

Then

$$\bar{P}_e(f_N, \varphi_N) > \epsilon, \qquad (196)$$

for some $N_o(W, \epsilon, \delta) \in \mathbb{Z}^+$ and for all $N \geq N_o(W, \epsilon, \delta)$.

Proof: Define

$$\mathbb{R}^+ \ni \gamma(\delta) := C(W) - \sup_{Q \in \mathcal{S}_3(\delta)} I(Q; W). \qquad (197)$$

Since $I(\cdot, W)$ is continuous over $\mathcal{P}(\mathcal{X})$, $\gamma(\delta)$ is a well-defined and positive real number. For any message m, let

$$G_N(m) :=$$
$$\left\{ \mathbf{y}^N : \frac{1}{N} \sum_{i=1}^N \ln \frac{W(y_i|x_i(m))}{q_{Q_N}(y_i)} > I(Q_N; W) + \frac{\gamma(\delta)}{2} \right\}. \qquad (198)$$

Define

$$\sigma_{\max}^2 := \max_{P \in \mathcal{P}(\mathcal{X})} V(P, W) \in \mathbb{R}^+. \qquad (199)$$

Since $V(\cdot, W)$ is continuous over the compact set $\mathcal{P}(\mathcal{X})$ (e.g., [11, Lemma 62]), $\sigma_{\max}^2$ is a well-defined and positive real number.

The following arguments are essentially the ones used in [38, Appendix B], which we outline here for completeness. First,

$$\bar{P}_e(f_N, \varphi_N) =$$
$$1 - \frac{1}{|\mathcal{M}_N|} \sum_{m \in \mathcal{M}_N} \sum_{\mathbf{y}^N \in \mathcal{A}_m \cap G_N(m)} W(\mathbf{y}^N|\mathbf{x}^N(m))$$
$$- \frac{1}{|\mathcal{M}_N|} \sum_{m \in \mathcal{M}_N} \sum_{\mathbf{y}^N \in \mathcal{A}_m \cap G_N^c(m)} W(\mathbf{y}^N|\mathbf{x}^N(m)). \qquad (200)$$

Since q_{Q_N} is a probability measure on $\mathcal{Y}^N$ and the decoding regions are disjoint, one can verify that

$$\frac{1}{|\mathcal{M}_N|} \sum_{m \in \mathcal{M}_N} \sum_{\mathbf{y}^N \in \mathcal{A}_m \cap G_N^c(m)} W(\mathbf{y}^N|\mathbf{x}^N(m))$$
$$\leq \exp \left\{ -N \left[\frac{\gamma(\delta)}{2} + \sqrt{\frac{V_\epsilon(W)}{N}} \Phi^{-1}(\epsilon) \right] \right\}. \qquad (201)$$

Further, since the code is constant composition with common composition Q_N, one can verify that

$$E \left[\sum_{i=1}^N \ln \frac{W(Y_i|x_i(m))}{q_{Q_N}(Y_i)} \right] = N \cdot I(Q_N; W), \qquad (202)$$

$$\text{Var} \left[\sum_{i=1}^N \ln \frac{W(Y_i|x_i(m))}{q_{Q_N}(Y_i)} \right] = N \cdot V(Q_N; W). \qquad (203)$$

Hence, via an application of Chebyshev's inequality, and recalling (199), one can verify that

$$\frac{1}{|\mathcal{M}_N|} \sum_{m \in \mathcal{M}_N} \sum_{\mathbf{y}^N \in \mathcal{A}_m \cap G_N(m)} W(\mathbf{y}^N|\mathbf{x}^N(m))$$
$$\leq \frac{N \cdot V(Q_N; W)}{\frac{(N\gamma(\delta))^2}{4}} \qquad (204)$$
$$\leq \frac{4\sigma_{\max}^2}{N\gamma(\delta)^2}. \qquad (205)$$

By substituting (201) and (205) into (200) we get

$$\bar{P}_e(f_N, \varphi_N) \geq$$
$$1 - \exp \left\{ -N \left[\frac{\gamma(\delta)}{2} + \sqrt{\frac{V_\epsilon(W)}{N}} \Phi^{-1}(\epsilon) \right] \right\} - \frac{4\sigma_{\max}^2}{N\gamma(\delta)^2}. \qquad (206)$$

Since this tends to one as $N \to \infty$, this implies the conclusion. ∎

Lemma 13: Fix some $\epsilon \in (\frac{1}{2}, 1)$, $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ with $V_\epsilon(W) > 0$, and $a \in \mathbb{R}^+$ with $a > \frac{2}{1-\epsilon}$. Consider an (N, R_N) constant-composition code (f, φ) with

$$R_N = C(W) + \sqrt{\frac{V_\epsilon(W)}{N}} \Phi^{-1}(\epsilon) - \frac{1}{N} \ln \left(1 - \epsilon - \frac{2}{a} \right), \qquad (207)$$

and the common composition Q satisfying

$$V(Q, W) < \frac{1}{a} V_\epsilon(W) \left[\Phi^{-1}(\epsilon) \right]^2. \qquad (208)$$

Then

$$\bar{P}_e(f, \varphi) > \epsilon. \qquad (209)$$

Proof: Via arguments similar to the ones given in the proof of Lemma 12, by using the set of $\mathbf{y}^N$ such that

$$\frac{1}{N} \sum_{i=1}^N \ln \frac{W(y_i|x_i(m))}{q_Q(y_i)} > C(W) + \sqrt{\frac{V_\epsilon(W)}{N}} \Phi^{-1}(\epsilon) \qquad (210)$$

in lieu of $G_N(m)$, i.e., (198), in the said arguments, one can verify that

$$\bar{P}_e(f, \varphi)$$
$$\geq 1 - \left(1 - \epsilon - \frac{2}{a} \right) \qquad (211)$$
$$- \frac{N \cdot V(Q, W)}{\left[N[C(W) - I(Q; W)] + \sqrt{N \cdot V_\epsilon(W)} \Phi^{-1}(\epsilon) \right]^2}$$
$$\geq \epsilon + \frac{1}{a} \qquad (212)$$
$$> \epsilon. \qquad (213)$$
∎

For any $Q \in \mathcal{P}(\mathcal{X})$, define

$$U(Q, W) := \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} Q(x) W(y|x) \tag{214}$$

$$\cdot \left[\ln \frac{W(y|x)}{q_Q(y)} - I(Q; W) \right]^2,$$

$$m_3(Q, W) := \sum_{x \in \mathcal{X}} Q(x) \tag{215}$$

$$\cdot \mathrm{E}_{W(\cdot|x)} \left[\left| \ln \frac{W(Y|x)}{q_Q(Y)} - \mathrm{E}_{W(\cdot|x)} \ln \frac{W(Y|x)}{q_Q(Y)} \right|^3 \right].$$

Choose $\delta > 0$ such that[3]

$$\mathrm{supp}(q_Q) = \mathcal{Y}, \text{ for all } Q \in \mathcal{P}(\mathcal{X}) \backslash \mathcal{S}_3(\delta). \tag{216}$$

Such a choice is possible due to the evident continuity of $\alpha_y(\cdot)$ for any $y \in \mathcal{Y}$ and the fact that the unique capacity-achieving output distribution has full support, as noted before. The following has been shown by Polyanskiy *et al.* [11, Lemma 46]

$$\tilde{m}_3(Q, W)$$

$$:= \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} Q(x) W(y|x) \left| \ln \frac{W(Y|X)}{q_Q(Y)} - I(Q; W) \right|^3 \tag{217}$$

$$\leq \left(\frac{3}{\mathrm{e}} \left(|\mathcal{X}|^{1/3} + |\mathcal{Y}|^{1/3} \right) + \ln \min\{|\mathcal{X}|, |\mathcal{Y}|\} \right)^3 \tag{218}$$

$$=: \kappa(W) \in \mathbb{R}^+. \tag{219}$$

Fix some $\nu \in \mathbb{R}^+$ and $\epsilon \in (0, 1)$. Assume $\mathcal{S}_1(\delta, \nu) \neq \emptyset$ and define

$$K(W, \epsilon, \delta, \nu) := \frac{2}{\phi(\Phi^{-1}(\epsilon))} \left[\max_{P \in \mathcal{S}_1(\delta, \nu)} \frac{m_3(P, W)}{V(P, W)} \right.$$

$$\left. + \left(\frac{1}{\sqrt{2\pi}} + \frac{\kappa(W)}{\nu} \right) \right] \in \mathbb{R}^+. \tag{220}$$

Since $m_3(\cdot, W)$ and $V(\cdot, W)$ are continuous over $\mathcal{P}(\mathcal{X})$ (e.g., [11, Lemma 62]), $K(W, \epsilon, \delta, \nu)$ is a well-defined and positive real number.

Lemma 14: Fix an asymmetric and singular $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, $\epsilon \in (0, 1)$ and $\nu \in \mathbb{R}^+$. Choose $\delta \in \mathbb{R}^+$ such that (216) holds. For some $\tilde{N}_\mathrm{o}(W, \epsilon, \delta, \nu) \in \mathbb{Z}^+$ and any $N \geq \tilde{N}_\mathrm{o}(W, \epsilon, \delta, \nu)$, consider an (N, R_N) constant-composition code (f, φ) with common composition $Q \in \mathcal{S}_1(\delta, \nu)$ and

$$R_N = I(Q; W) + \sqrt{\frac{V(Q, W)}{N}} \Phi^{-1}(\epsilon) + \frac{1}{N} K(W, \epsilon, \delta, \nu). \tag{221}$$

Then $\bar{\mathrm{P}}_\mathrm{e}(f, \varphi) > \epsilon$.

Proof: Assume $\mathcal{S}_1(\delta, \nu) \neq \emptyset$, because otherwise the claim is void. The proof is similar to the proof of Theorem 3. Let

---

$\tilde{N}_\mathrm{o}(W, \epsilon, \delta, \nu) \in \mathbb{Z}^+$ be such that for all $N \geq \tilde{N}_\mathrm{o}(W, \epsilon, \delta, \nu)$,

$$\sqrt{N} > \frac{2K(W, \epsilon, \delta, \nu)}{\phi(\Phi^{-1}(\epsilon))\sqrt{\nu}}. \tag{222}$$

In light of (220), the existence of such a choice is evident.

Consider any (N, R_N) constant-composition code, say (f, φ), with the common composition Q. Assume Q and R_N are as in the statement of the lemma. Consider any $\mathbf{x}^N \in \mathcal{X}^N$ and define

$$M_{\mathbf{x}^N}(\lambda) := \mathrm{E}_{W(\cdot|\mathbf{x}^N)} \left[\mathrm{e}^{\lambda \ln \frac{W(\mathbf{Y}^N|\mathbf{x}^N)}{q_{P_{\mathbf{x}^N}}(\mathbf{Y}^N)}} \right], \forall \lambda \in \mathbb{R}. \tag{223}$$

We claim that for any $\mathbf{x}^N, \mathbf{z}^N \in \mathcal{X}^N$ with $P_{\mathbf{x}^N} = P_{\mathbf{z}^N}$, we have

$$M_{\mathbf{x}^N}(\lambda) = M_{\mathbf{z}^N}(\lambda), \forall \lambda \in \mathbb{R}. \tag{224}$$

To see this, we simply note that

$$M_{\mathbf{x}^N}(\lambda)$$

$$= \sum_{\mathbf{y}^N : W(\mathbf{y}^N|\mathbf{x}^N) > 0} \frac{\mathrm{e}^{N \sum_y P_{\mathbf{y}^N}(y) \ln \xi_y}}{\mathrm{e}^{\lambda N \sum_y P_{\mathbf{y}^N}(y) \ln \alpha_y(P_{\mathbf{x}^N})}} \tag{225}$$

$$= \sum_{P \in \mathcal{P}_N(\mathcal{Y})} \frac{\mathrm{e}^{N \sum_y P(y) \ln \xi_y}}{\mathrm{e}^{\lambda N \sum_y P(y) \ln \alpha_y(P_{\mathbf{x}^N})}} \cdot |T(P, \mathbf{x}^N)| \tag{226}$$

$$= \sum_{P \in \mathcal{P}_N(\mathcal{Y})} \frac{\mathrm{e}^{N \sum_y P(y) \ln \xi_y}}{\mathrm{e}^{\lambda N \sum_y P(y) \ln \alpha_y(P_{\mathbf{z}^N})}} \cdot |T(P, \mathbf{z}^N)| \tag{227}$$

$$= M_{\mathbf{z}^N}(\lambda), \tag{228}$$

where

$$T(P, \mathbf{x}^N) = \{\mathbf{y}^N : P_{\mathbf{y}^N} = P \text{ and } W(\mathbf{y}^N|\mathbf{x}^N) > 0\}, \tag{229}$$

and (227) follows from the fact that $P_{\mathbf{x}^N} = P_{\mathbf{z}^N}$. Equation (224), along with the uniqueness theorem for the moment generating function (e.g., [35, Ex. 26.7]), and the fact that q_Q is of full support, enables us to invoke Lemma 11 to deduce that

$$\bar{\mathrm{P}}_\mathrm{e}(f, \varphi) \geq W(\mathcal{S}_{R_N}(Q)|\mathbf{z}^N)$$

$$- \sum_{\mathbf{y}^N \in \mathcal{S}_{R_N}(Q)} q_Q(\mathbf{y}^N) \exp\left\{ -N \cdot R_N + \sum_{i=1}^N \ln \frac{1}{\alpha_{y_i}(Q)} \right\}, \tag{230}$$

for a given $\mathbf{z}^N \in \mathcal{X}^N$ with $P_{\mathbf{z}^N} = Q$. Due to the singularity of W,

$$W(\mathcal{S}_{R_N}(Q)|\mathbf{z}^N)$$

$$= \sum_{\mathbf{y}^N} W(\mathbf{y}^N|\mathbf{z}^N) \mathbb{1}\left\{ \frac{1}{N} \sum_{i=1}^N \ln \frac{W(y_i|z_i)}{q_Q(y_i)} \leq R_N \right\} \tag{231}$$

$$\geq \epsilon - \frac{m_3(Q, W)}{\sqrt{N} V(Q, W)^{3/2}} + \frac{K(W, \epsilon, \delta, \nu)\phi(\Phi^{-1}(\epsilon))}{\sqrt{N \cdot V(Q, W)}} \tag{232}$$

$$\cdot \left(1 - \frac{K(W, \epsilon, \delta, \nu)}{2\sqrt{N \cdot V(Q, W)}\phi(\Phi^{-1}(\epsilon))} \right),$$

where the proof of (232) is similar to that of (171) and omitted for brevity.

Further, define

$$P_{XY}(x, y) := Q(x)W(y|x), \tag{233}$$

$$P_{\mathbf{X}^N \mathbf{Y}^N}(\mathbf{x}^n, \mathbf{y}^n) := \prod_{i=1}^{N} P_{XY}(x_i, y_i). \tag{234}$$

Evidently,

$$\sum_{\mathbf{y}^N \in \mathcal{S}_{R_N}(Q)} q_Q(\mathbf{y}^N) \exp\left\{-N\left[R_N - \frac{1}{N}\sum_{i=1}^{N}\ln\frac{1}{\alpha_{y_i}(Q)}\right]\right\}$$
$$\tag{235}$$

$$= \sum_{(\mathbf{x}^N, \mathbf{y}^N)} P_{\mathbf{X}^N \mathbf{Y}^N}(\mathbf{x}^N, \mathbf{y}^N) \mathbb{1}\left\{\frac{1}{N}\sum_{i=1}^{N}\ln\frac{W(y_i|x_i)}{q_Q(y_i)} \le R_N\right\}$$

$$\times \exp\left\{-N\left[R_N - \frac{1}{N}\sum_{i=1}^{N}\ln\frac{W(y_i|x_i)}{q_Q(y_i)}\right]\right\} \tag{236}$$

$$\le \frac{1}{\sqrt{2\pi N \cdot U(Q, W)}} + \frac{\tilde{m}_3(Q, W)}{\sqrt{N} U(Q, W)^{3/2}} \tag{237}$$

$$\le \frac{1}{\sqrt{N \cdot V(Q, W)}}\left(\frac{1}{\sqrt{2\pi}} + \frac{\kappa(W)}{V(Q, W)}\right), \tag{238}$$

where $U(Q, W)$ is defined in (214), and (237) follows from Lemma 2, whose application is ensured by the fact that $U(Q, W) \ge V(Q, W)$ (e.g., [11, Lemma 62]), which, along with (219), also implies (238).

By substituting (232) and (238) into (230), along with the definitions of $K(W, \epsilon, \delta, \nu)$ and $n_o(W, \epsilon, \delta, \nu)$, one can verify that

$$\bar{P}_e(f, \varphi)$$
$$> \epsilon + \frac{\max_{P \in \mathcal{S}_1(\delta, \nu)} \frac{m_3(P, W)}{V(P, W)} - \frac{m_3(Q, W)}{V(Q, W)}}{\sqrt{N \cdot V(Q, W)}} \tag{239}$$
$$\ge \epsilon, \tag{240}$$

which, in turn, implies the assertion. ∎

In order to prove the first assertion of the theorem, i.e., (33), fix some $\epsilon \in (0, \frac{1}{2})$ and assume $V_\epsilon(W) > 0$, because otherwise [31, Proposition 9] implies (33). Fix some $\delta > 0$ such that (216) holds and $\mathcal{S}_2\left(\delta, \frac{V_\epsilon(W)}{2}\right) = \emptyset$. Such a choice is possible since $V(\cdot, W)$ is continuous over $\mathcal{P}(\mathcal{X})$, as noted before. For any $P \in \mathcal{P}(\mathcal{X})$, let

$$P^*(P) := \arg\min_{Q \in \mathcal{P}_W^*} ||Q - P||_2. \tag{241}$$

Fix some $\beta_1, \beta_2 \in \mathbb{R}^+$ such that

$$I(P; W) \le C(W) - \beta_1 ||P - P^*(P)||_2^2, \tag{242}$$
$$|\sqrt{V(P, W)} - \sqrt{V(P^*(P), W)}| \le \beta_2 ||P - P^*(P)||_2, \tag{243}$$

for any $P \in \mathcal{S}_1\left(\delta, \frac{V_\epsilon(W)}{2}\right)$, whose existence is ensured by [31, Lemma 7]. In light of (242) and (III-D), for all $P \in$

$\mathcal{S}_1\left(\delta, \frac{V_\epsilon(W)}{2}\right)$ and for any $N \in \mathbb{Z}^+$,

$$N \cdot I(P; W) + \sqrt{N \cdot V(P, W)}\Phi^{-1}(\epsilon) \tag{244}$$
$$\le N \cdot C(W) + \sqrt{N \cdot V_\epsilon(W)}\Phi^{-1}(\epsilon)$$
$$\quad - \beta_1 N ||P - P^*(P)||_2^2 + \beta_2 |\Phi^{-1}(\epsilon)|\sqrt{N}||P - P^*(P)||_2 \tag{245}$$

$$\le N \cdot C(W) + \sqrt{N \cdot V_\epsilon(W)}\Phi^{-1}(\epsilon) + \frac{1}{4\beta_1}\left(\beta_2|\Phi^{-1}(\epsilon)|\right)^2, \tag{246}$$

where (246) follows from elementary calculus. Consider any $N \in \mathbb{Z}^+$ such that

$$N \ge \max\left\{N_o(W, \epsilon, \delta), \tilde{N}_o(W, \epsilon, \delta, \frac{V_\epsilon(W)}{2})\right\}, \tag{247}$$

where N_o and $\tilde{N}_o$ are given in Lemmas 12 and 14, respectively. Define

$$R_N := C(W) + \sqrt{\frac{V_\epsilon(W)}{N}}\Phi^{-1}(\epsilon)$$
$$+ \frac{1}{N}\left(\frac{1}{4\beta_1}\left(\beta_2|\Phi^{-1}(\epsilon)|\right)^2 + K(W, \epsilon, \delta, \frac{V_\epsilon(W)}{2})\right), \tag{248}$$

and consider any (N, R_N) constant-composition code (f, φ) with the common composition Q. Now, if $Q \in \mathcal{S}_3(\delta)$, then Lemma 12 implies that $\bar{P}_e(f, \varphi) > \epsilon$. Similarly, if $Q \in \mathcal{S}_1\left(\delta, \frac{V_\epsilon(W)}{2}\right)$, then Lemma 14 and (246) imply that $\bar{P}_e(f, \varphi) > \epsilon$. Since the code is arbitrary, we conclude that (33) holds.

In order to prove the second assertion of the theorem, i.e., (34), fix some $\epsilon \in (\frac{1}{2}, 1)$ and $\delta > 0$ such that (216) holds. Choose some $a \in \mathbb{R}^+$ that satisfies $a > \frac{2}{1-\epsilon}$ and $\nu \in \mathbb{R}^+$ such that $\nu \le \frac{1}{a}V_\epsilon(W)\left[\Phi^{-1}(\epsilon)\right]^2$. Similar to (242) and (III-D), choose $\beta_1, \beta_2 \in \mathbb{R}^+$ such that

$$I(P; W) \le C(W) - \beta_1 ||P - P^*(P)||_2^2, \tag{249}$$
$$|\sqrt{V(P, W)} - \sqrt{V(P^*(P), W)}| \le \beta_2 ||P - P^*(P)||_2, \tag{250}$$

for any $P \in \mathcal{S}_1(\delta, \nu)$. From (249) and (III-D), similar to (246), we deduce that for all $P \in \mathcal{S}_1(\delta, \nu)$ and $N \in \mathbb{Z}^+$,

$$N \cdot I(P; W) + \sqrt{N \cdot V(P, W)}\Phi^{-1}(\epsilon)$$
$$\le N \cdot C(W) + \sqrt{N \cdot V_\epsilon(W)}\Phi^{-1}(\epsilon) + \frac{1}{4\beta_1}\left(\beta_2\Phi^{-1}(\epsilon)\right)^2. \tag{251}$$

Consider any $N \in \mathbb{Z}^+$ such that

$$N \ge \max\{N_o(W, \epsilon, \delta), \tilde{N}_o(W, \epsilon, \delta, \nu)\}, \tag{252}$$

where N_o and $\tilde{N}_o$ are as given in Lemmas 12 and 14, respectively. Consider any (N, R_N) constant-composition code

(f, φ) with the common composition Q and define

$$
R_N := C(W) + \sqrt{\frac{V_\epsilon(W)}{N}} \Phi^{-1}(\epsilon)
$$
$$
+ \frac{1}{N} \left(\frac{1}{4\beta_1} \left(\beta_2 \Phi^{-1}(\epsilon) \right)^2 + K(W, \epsilon, \delta, \nu) \right.
$$
$$
\left. - \ln \left(1 - \epsilon - \frac{2}{a} \right) \right). \quad (253)
$$

If $Q \in \mathcal{S}_3(\delta)$, then $\bar{P}_e(f, \varphi) > \epsilon$ due to Lemma 12. If $Q \in \mathcal{S}_2(\delta, \nu)$, then $\bar{P}_e(f, \varphi) > \epsilon$ because of Lemma 13, which can be applied since $Q \in \mathcal{S}_2(\delta, \nu)$ guarantees that $V(P, W) < \nu$. Finally, if $Q \in \mathcal{S}_1(\delta, \nu)$, then Lemma 14, along with (251), implies that $\bar{P}_e(f, \varphi) > \epsilon$. Since the code is arbitrary, we conclude that (34) holds. ∎

IV. Discussion

A. Relation to the Minimax Converse

In the absence of feedback, one can interpret the proof of Theorem 3 in terms of the minimax converse (e.g., [39, Theorem 1]), which we illustrate next. To this end, we fix a symmetric and singular $W \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$ and note that [39, Eq. (9) and (11)] imply that for any $N \in \mathbb{Z}^+$ and $\epsilon \in (0, 1)$,

$$
\min_{P_{\mathbf{X}^N}} \max_{Q_{\mathbf{Y}^N}} \beta_{1-\epsilon}(P_{\mathbf{X}^N \mathbf{Y}^N}, P_{\mathbf{X}^N} \times Q_{\mathbf{Y}^N}) \leq \frac{1}{M^*(N, \epsilon)}, \quad (254)
$$

where

$$
P_{\mathbf{X}^N \mathbf{Y}^N}(\mathbf{x}^N, \mathbf{y}^N) := P_{\mathbf{X}^N}(\mathbf{x}^N) W(\mathbf{y}^N | \mathbf{x}^N), \quad (255)
$$
$$
(P_{\mathbf{X}^N} \times Q_{\mathbf{Y}^N})(\mathbf{x}^N, \mathbf{y}^N) := P_{\mathbf{X}^N}(\mathbf{x}^N) Q_{\mathbf{Y}^N}(\mathbf{y}^N), \quad (256)
$$

and $\beta_{1-\epsilon}(P_{\mathbf{X}^N \mathbf{Y}^N}, P_{\mathbf{X}^N} \times Q_{\mathbf{Y}^N})$ denotes the minimum probability of error under $P_{\mathbf{X}^N} \times Q_{\mathbf{Y}^N}$, subject to the constraint that the error probability under hypothesis $P_{\mathbf{X}^N \mathbf{Y}^N}$ does not exceed ϵ. Due to [39, Theorem 21], the minimum on the left side of (254) is attained by $U_{\mathcal{X}^N}$. Consider some $N \in \mathbb{Z}^+$ such that (157) holds and let R be as in (158). With these choices, we define[4]

$$
Q^*_{\mathbf{Y}^N}(\mathbf{y}^N) := \frac{e^{N \sum_y P_{\mathbf{y}^N}(y) \ln \xi_y} \mathbb{1}\left\{ \mathbf{y}^N \in \mathcal{S}(R) \right\}}{\sum_{\mathbf{b}^N} e^{N \sum_b P_{\mathbf{b}^N}(b) \ln \xi_b} \mathbb{1}\left\{ \mathbf{b}^N \in \mathcal{S}(R) \right\}}, \quad (257)
$$

where ξ_y and $\mathcal{S}(R)$ are as defined before. Evidently,

$$
Q^*_{\mathbf{Y}^N} \in \mathcal{P}(\mathcal{Y}^N). \quad (258)
$$

With a slight abuse of notation, let $\beta_{1-\epsilon}(U_{\mathcal{X}^N}, Q^*_{\mathbf{Y}^N})$ denote the value of the cost function of the optimization problem in (254) when $P_{\mathbf{X}^N} = U_{\mathcal{X}^N}$ and $Q_{\mathbf{Y}^N} = Q^*_{\mathbf{Y}^N}$. From (254),

$$
M^*(N, \epsilon) \leq \frac{1}{\beta_{1-\epsilon}(U_{\mathcal{X}^N}, Q^*_{\mathbf{Y}^N})}. \quad (259)
$$

From the Neyman-Pearson lemma (e.g., [40]), the right side of (259) is attained by a randomized threshold test with the randomization parameter $\tau \in (0, 1)$ satisfying

$$
\tau W(\mathcal{S}(R) | \mathbf{x}_o^N) = \epsilon, \quad (260)
$$

[4]The non-product distribution in (257) is inspired by [39, Eq. (168)]. In particular, if W is BEC then (257) reduces to [39, Eq. (168)].

and

$$
\beta_{1-\epsilon}(U_{\mathcal{X}^N}, Q^*_{\mathbf{Y}^N}) \quad (261)
$$
$$
= \frac{(1-\tau) W(\mathcal{S}(R) | \mathbf{x}_o^N) e^{-NR}}{\sum_{\mathbf{y}^N \in \mathcal{S}(R)} q(\mathbf{y}^N) \exp\left\{ -N \left[R - \frac{1}{N} \sum_{i=1}^N \ln \frac{1}{\alpha_{y_i}} \right] \right\}}. \quad (262)
$$

Equations (260) and (262) can be verified via elementary algebra by noticing that W is singular and symmetric. We omit the details for brevity. Finally, (170) and (171), along with (157) and (158), imply that

$$
W(\mathcal{S}(R) | \mathbf{x}_o^N)
$$
$$
- \sum_{\mathbf{y}^N \in \mathcal{S}(R)} q(\mathbf{y}^N) \exp\left\{ -N \left[R - \frac{1}{N} \sum_{i=1}^N \ln \frac{1}{\alpha_{y_i}} \right] \right\}
$$
$$
> \epsilon. \quad (263)
$$

Equations (259)–(263) imply that $M^*(N, \epsilon) < e^{NR}$, which, in turn, implies Theorem 3 in the absence of feedback.

The above interpretation of the arguments leading to (263) yield a more streamlined alternative to the one in the main text, at least for the case of no feedback. We have provided the latter because it allows for feedback and because it gives a unified method for proving converse results in the fixed-rate and fixed-error-probability regimes.

B. On Dropping the Constant-Composition Assumption

As noted before, Theorem 4 gives an $O(1)$ upper bound on the third-order term of the normal approximation for asymmetric and singular DMCs only if we consider constant-composition codes. Although this restriction is undesirable, it is quite common in converse results. Indeed, the usual proof of the converse statement of (6) involves first showing it for constant-composition codes, and then arguing that this restriction at most results in an extra $O(\ln N)$ term.

Tomamichel and Tan [31] have showed an $\ln \sqrt{N}$ upper bound on the third-order term in general by eliminating the constant-composition code restriction in the first step. This result, coupled with the existing results in the literature, gives the third-order term for a broad class of channels, which includes positive channels with positive capacity but does not include asymmetric and singular channels. The method of [31] is based on relating the channel coding problem to a binary hypothesis test by using an auxiliary output distribution, which is in the same vein as the so-called meta-converse of Polyanskiy et al. (e.g., [11, Section III.E and III.F]). As opposed to the classical applications of this idea, which use a product auxiliary output distribution and result in the aforementioned two-step procedure, the authors of [31] use an appropriately chosen non-product output distribution to dispense with the constant-composition step. However, their non-product distribution is different from the one used in the previous subsection. Investigating how to combine the analysis of [31] and the viewpoint in Section IV-A to drop the constant-composition assumption in Theorem 4 is a worthy direction for future research.

C. Limitation in the Error Exponents Regime

One might conjecture that by following the same program used to prove Theorem 4, one could prove the following lower bound for asymmetric and singular channels

$$\liminf_{N \to \infty} \frac{\bar{P}_{e,c}(N,R)}{\frac{1}{\sqrt{N}}e^{-N E_{SP}(R)}} \geq K(R,W), \qquad (264)$$

where $K(R,W)$ is a positive constant that depends on R and W. However, a proof of (264) seems to be more involved than its counterpart in the normal approximation regime, i.e., Theorem 4. The main technical difficulty is proving the continuity properties of $E_{SP}(R, \cdot)$ that are required to distinguish between the "good types", for which $E_{SP}(R,Q) \approx E_{SP}(R)$ and hence one can use a result like Lemma 14 to deduce an $\Omega(\frac{1}{\sqrt{N}})$ sub-exponential term directly, and the "bad types", for which $E_{SP}(R,Q)$ is bounded away from $E_{SP}(R)$ and hence one can utilize this inferiority of the exponent to deduce an $\Omega(\frac{1}{\sqrt{N}})$ sub-exponential term. Indeed, justifications of these continuity properties appear to be quite intricate. For an analogous upper bound, see Honda [15], [16].

APPENDIX A
PROOF OF PROPOSITION 3

(i) Thanks to the symmetry of the channel, $\tilde{E}_{SP}(R) = \tilde{E}_{SP}(R,U_{\mathcal{X}})$ (e.g., [9, p. 145]). Moreover, due to the facts that $E_{SP}(R) = \tilde{E}_{SP}(R)$ and $E_{SP}(R,P) \geq \tilde{E}_{SP}(R,P)$ for all $P \in \mathcal{P}(\mathcal{X})$, which have been noted before, we conclude that $E_{SP}(R) = E_{SP}(R,U_{\mathcal{X}})$.

(ii) Fix any $\rho \in \mathbb{R}_+$ and consider the following convex program

$$\min_{Q \in \mathcal{P}(\mathcal{X})} \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} Q(x) W(y|x)^{1/(1+\rho)} \right)^{1+\rho}, \quad (265)$$

whose convexity is verified in [9, Theorem 5.6.5]. Next, we recall the necessary and sufficient conditions for any $Q \in \mathcal{P}(\mathcal{X})$ to attain the minimum in (265), due to [9, Theorem 5.6.5], for all $x \in \mathcal{X}$,

$$\sum_{y \in \mathcal{Y}} W(y|x)^{1/(1+\rho)} \left(\sum_{z \in \mathcal{X}} Q(z) W(y|z)^{1/(1+\rho)} \right)^{\rho}$$

$$\geq \sum_{y \in \mathcal{Y}} \left(\sum_{z \in \mathcal{X}} Q(z) W(y|z)^{1/(1+\rho)} \right)^{1+\rho}, \quad (266)$$

with equality if $Q(x) > 0$. Thanks to the symmetry of the channel, $U_{\mathcal{X}}$ is an optimizer of (265) (e.g., [9, p. 145]) and hence (266) implies (60).

(iii) We first note the following, which is an easy consequence of elementary convex optimization arguments (e.g., [10, Ex. 2.5.23])

$$E_{SP}(R,U_{\mathcal{X}}) = \max_{\rho \geq 0} \min_{q \in \mathcal{P}(\mathcal{Y})} \left\{ -\rho R - (1+\rho) \right.$$

$$\left. \cdot \sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) \ln \sum_{y \in \mathcal{Y}} W(y|x)^{1/(1+\rho)} q(y)^{\rho/(1+\rho)} \right\}. \quad (267)$$

Due to [13, Propositions 1 and 2], (267) has a unique saddle-point. Further, [13, Proposition 3] ensures that ρ_R is the $\mathbb{R}_+$ component of this saddle-point. Owing to the properties of saddle-points (e.g., [41, Lemma 36.2]) ρ_R attains the maximum in (267), and the fact that $E_{SP}(R) = E_{SP}(R,U_{\mathcal{X}}) > 0$ ensures its positivity. Hence,

$$E_{SP}(R,U_{\mathcal{X}}) \qquad (268)$$

$$= \min_{q \in \mathcal{P}(\mathcal{Y})} \left\{ -\rho_R R - (1+\rho_R) \right. \qquad (269)$$

$$\left. \cdot \sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) \ln \sum_{y \in \mathcal{Y}} W(y|x)^{1/(1+\rho_R)} q(y)^{\rho_R/(1+\rho_R)} \right\}$$

$$\leq -\rho_R R - (1+\rho_R) \qquad (270)$$

$$\cdot \sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) \ln \sum_{y \in \mathcal{Y}} W(y|x)^{1/(1+\rho_R)} q_R(y)^{\rho_R/(1+\rho_R)}.$$

Next, we claim that

$$E_{SP}(R,U_{\mathcal{X}}) \leq -\rho_R R + E_o(\rho_R, U_{\mathcal{X}}) \leq \tilde{E}_{SP}(R,U_{\mathcal{X}}). \qquad (271)$$

In light of the definition of $\tilde{E}_{SP}(R,U_{\mathcal{X}})$, i.e., (17), the second inequality is evident, so we prove the first one. To this end,

$$\ln \sum_{y \in \mathcal{Y}} W(y|x)^{\frac{1}{1+\rho_R}} q_R(y)^{\frac{\rho_R}{1+\rho_R}} \qquad (272)$$

$$= \ln \sum_{y \in \mathcal{Y}} \frac{W(y|x)^{\frac{1}{1+\rho_R}} \left(\sum_{z \in \mathcal{X}} U_{\mathcal{X}}(z) W(y|z)^{\frac{1}{1+\rho_R}} \right)^{\rho_R}}{\left(\sum_{b \in \mathcal{Y}} \left(\sum_{a \in \mathcal{X}} U_{\mathcal{X}}(a) W(b|a)^{\frac{1}{1+\rho_R}} \right)^{1+\rho_R} \right)^{\frac{\rho_R}{1+\rho_R}}} \qquad (273)$$

$$= \ln \frac{\sum_{y \in \mathcal{Y}} \left(\sum_{z \in \mathcal{X}} U_{\mathcal{X}}(z) W(y|z)^{\frac{1}{1+\rho_R}} \right)^{1+\rho_R}}{\left(\sum_{b \in \mathcal{Y}} \left(\sum_{a \in \mathcal{X}} U_{\mathcal{X}}(a) W(b|a)^{\frac{1}{1+\rho_R}} \right)^{1+\rho_R} \right)^{\frac{\rho_R}{1+\rho_R}}} \qquad (274)$$

$$= -\frac{1}{1+\rho_R} E_o(\rho_R, U_{\mathcal{X}}), \qquad (275)$$

where (273) follows from the definition of q_R, (274) follows from the second assertion of this proposition, and (275) follows from the definition of $E_o(\cdot, \cdot)$. Plugging (275) into (270) implies the first inequality of (271).

In light of the first assertion of this proposition, i.e., (59), (271) implies that ρ_R attains the maximum in the definition of $\tilde{E}_{SP}(R,U_{\mathcal{X}})$.

(iv) Equation (271) and the first assertion of this proposition ensure that q_R attains the minimum in (269). Hence, by recalling the definition of a saddle-point (e.g., [41, p. 380]), in order to conclude the proof, it suffices to show that ρ_R attains the supremum in the following

optimization problem:

$$\sup_{\rho \in \mathbb{R}_+} \left\{ -\rho R - (1+\rho) \cdot \right.$$
$$\left. \sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) \ln \sum_{y \in \mathcal{Y}} W(y|x)^{1/(1+\rho)} q_R(y)^{\rho/(1+\rho)} \right\}. \tag{276}$$

To this end, for any $\rho \in \mathbb{R}_+$, temporarily define

$$q_\rho(y) := \frac{\left(\sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) W(y|x)^{1/(1+\rho)} \right)^{1+\rho}}{\sum_{b \in \mathcal{Y}} \left(\sum_{a \in \mathcal{X}} U_{\mathcal{X}}(a) W(b|a)^{1/(1+\rho)} \right)^{1+\rho}}, \tag{277}$$

$$V_\rho(y|x) := \frac{W(y|x)^{1/(1+\rho)} q_\rho(y)^{\rho/(1+\rho)}}{\sum_{b \in \mathcal{Y}} W(b|x)^{1/(1+\rho)} q_\rho(b)^{\rho/(1+\rho)}}. \tag{278}$$

The first of the two definitions is an abuse of notation in that it conflicts with the definition of q_R in (57). But the conflict is temporary and the subscript will successfully distinguish the two. We proceed by noting that

$$\sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) V_\rho(y|x) = \tag{279}$$
$$\sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) \frac{W(y|x)^{\frac{1}{1+\rho}} \left[\sum_{z \in \mathcal{X}} U_{\mathcal{X}}(z) W(y|z)^{\frac{1}{1+\rho}} \right]^\rho}{\sum_{b \in \mathcal{Y}} W(b|x)^{\frac{1}{1+\rho}} \left[\sum_{a \in \mathcal{X}} U_{\mathcal{X}}(a) W(b|a)^{\frac{1}{1+\rho}} \right]^\rho}$$

so that

$$\sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) V_\rho(y|x) \tag{280}$$
$$= \frac{\sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) W(y|x)^{\frac{1}{1+\rho}} \left[\sum_{z \in \mathcal{X}} U_{\mathcal{X}}(z) W(y|z)^{\frac{1}{1+\rho}} \right]^\rho}{\sum_{b \in \mathcal{Y}} \left[\sum_{a \in \mathcal{X}} U_{\mathcal{X}}(a) W(b|a)^{\frac{1}{1+\rho}} \right]^{1+\rho}} \tag{281}$$
$$= q_\rho(y), \tag{282}$$

where (279) follows by substituting (277) into (278), (281) follows from (60), which is verified in item (ii) of this proposition, and (282) follows from the definition of q_ρ, i.e., (277). Note that

$$I(U_{\mathcal{X}}; V_\rho) = D(V_\rho \| q_\rho | U_{\mathcal{X}}), \tag{283}$$

which is a direct consequence of the definitions of the mutual information and relative entropy and (282). Next, we note that for any $\rho \in \mathbb{R}_+$,

$$-\frac{1}{1+\rho} (D(V_\rho \| W | U_{\mathcal{X}}) + \rho I(U_{\mathcal{X}}; V_\rho))$$
$$= \sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) \ln \sum_{y \in \mathcal{Y}} W(y|x)^{1/(1+\rho)} q_\rho(y)^{\rho/(1+\rho)}. \tag{284}$$

To see (284), first observe that

$$D(V_\rho \| W | U_{\mathcal{X}}) = \sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) \sum_{y \in \mathcal{Y}} V_\rho(y|x)$$
$$\cdot \left\{ \frac{\rho}{(1+\rho)} \ln \frac{q_\rho(y)}{W(y|x)} \right.$$
$$\left. - \ln \sum_{b \in \mathcal{Y}} W(b|x)^{1/(1+\rho)} q_\rho(b)^{\rho/(1+\rho)} \right\}, \tag{285}$$

which is a direct consequence of the definition of $V_\rho(y|x)$, i.e., (278). Further, (278), coupled with (283), implies that

$$\rho I(U_{\mathcal{X}}; V_\rho) = \rho \left[\sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) \sum_{y \in \mathcal{Y}} V_\rho(y|x) \cdot \right.$$
$$\left\{ \frac{1}{(1+\rho)} \ln \frac{W(y|x)}{q_\rho(y)} \right.$$
$$\left. \left. - \ln \sum_{b \in \mathcal{Y}} W(b|x)^{1/(1+\rho)} q_\rho(b)^{\rho/(1+\rho)} \right\} \right]. \tag{286}$$

Equations (285) and (286) imply (284). We continue with the following assertion:

Lemma 15:

$$E_{SP}(R, U_{\mathcal{X}}) = -\rho_R R + D(V_{\rho_R} \| W | U_{\mathcal{X}}) + \rho_R I(U_{\mathcal{X}}; V_{\rho_R}), \tag{287}$$

and V_{ρ_R} is a minimizer for $E_{SP}(R, U_{\mathcal{X}})$.

Proof: First, note that

$$E_{SP}(R, U_{\mathcal{X}}) = \max_{\rho \in \mathbb{R}_+} \left\{ -\rho R \right.$$
$$\left. + \min_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} [D(V \| W | U_{\mathcal{X}}) + \rho I(U_{\mathcal{X}}; V)] \right\}, \tag{288}$$

which is verified in [10, Ex. 2.5.23]. By the subdifferential characterization of Lagrange multipliers (e.g., [41, Theorem 29.1]), ρ_R is the unique maximizer in (288), and hence

$$E_{SP}(R, U_{\mathcal{X}}) = -\rho_R R$$
$$+ \min_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \{D(V \| W | U_{\mathcal{X}}) + \rho_R I(U_{\mathcal{X}}; V)\}. \tag{289}$$

Now, for any $\rho \in \mathbb{R}_+$,

$$D(V_\rho \| W | U_{\mathcal{X}}) + \rho I(U_{\mathcal{X}}; V_\rho) \tag{290}$$
$$= -\ln \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} U_{\mathcal{X}}(x) W(y|x)^{1/(1+\rho)} \right)^{1+\rho} \tag{291}$$
$$= E_o(\rho, U_{\mathcal{X}}), \tag{292}$$

which follows from routine computations once we employ (60) on the right side of (284) along with the definition of q_ρ, i.e., (277). Also, for any $\rho \in \mathbb{R}_+$,

$$\min_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} [D(V \| W | U_{\mathcal{X}}) + \rho I(U_{\mathcal{X}}; V)] \geq E_o(\rho, U_{\mathcal{X}}), \tag{293}$$

which follows from routine convex analysis arguments (e.g., [10, Ex. 2.5.23]). Equations (292) and (293), along with the strict convexity of $D(\cdot\|W|U_\mathcal{X})$, which is an immediate consequence of the strict convexity of the function $\mathbb{R}_+ \ni x \mapsto x \ln x$, imply that V_{ρ_R} is the unique minimizer in (289), which, in turn, establishes (287). Since V_{ρ_R} is the unique minimizer in (289), it must also be primal optimal (e.g., [41, Theorem 28.1]), i.e., it must be a minimizer of $E_{SP}(R, U_\mathcal{X})$. ∎

In order to conclude the proof, consider

$$e_{SP}(R, R) = \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X}): \, D(V\|q_R|U_\mathcal{X}) \leq R} D(V\|W|U_\mathcal{X}) \tag{294}$$

from (74). By noting the fact that V_{ρ_R} is a minimizer of $E_{SP}(R, U_\mathcal{X})$, which is verified in Lemma 15, along with (283), we have

$$I(U_\mathcal{X}; V_{\rho_R}) = D(V_{\rho_R}\|q_R|U_\mathcal{X}) \leq R, \tag{295}$$

which, in turn, implies that

$$e_{SP}(R, R) \leq E_{SP}(R, U_\mathcal{X}). \tag{296}$$

Further,

$$e_{SP}(R, R)$$
$$\geq \sup_{\rho \in \mathbb{R}_+} \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \{D(V\|W|U_\mathcal{X}) \tag{297}$$
$$+ \rho \left[D(V\|q_R|U_\mathcal{X}) - R\right]\}$$
$$\geq \inf_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \{D(V\|W|U_\mathcal{X}) \tag{298}$$
$$+ \rho_R \left[D(V\|q_R|U_\mathcal{X}) - R\right]\}$$
$$= D(V_{\rho_R}\|W|U_\mathcal{X}) + \rho_R \left[D(V_{\rho_R}\|q_R|U_\mathcal{X}) - R\right], \tag{299}$$
$$= -\rho_R R + D(V_{\rho_R}\|W|U_\mathcal{X}) + \rho_R I(U_\mathcal{X}; V_{\rho_R}) \tag{300}$$
$$= E_{SP}(R, U_\mathcal{X}), \tag{301}$$

where (299) follows by solving the convex program in (298), (300) follows from (283), and (301) is (287). Hence, (296), (297) and (301) imply that

$$E_{SP}(R, U_\mathcal{X}) \tag{302}$$
$$= e_{SP}(R, R) \tag{303}$$
$$= \max_{\rho \in \mathbb{R}_+} \min_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \{D(V\|W|U_\mathcal{X}) \tag{304}$$
$$+ \rho \left[D(V\|q_R|U_\mathcal{X}) - R\right]\}$$
$$= \max_{\rho \in \mathbb{R}_+} \left\{ -\rho R - (1 + \rho) \sum_{x \in \mathcal{X}} U_\mathcal{X}(x) \tag{305} \right.$$
$$\left. \cdot \ln \sum_{y \in \mathcal{Y}} W(y|x)^{1/(1+\rho)} q_R(y)^{\rho/(1+\rho)} \right\}$$
$$\geq -\rho_R R \tag{306}$$
$$- (1 + \rho_R) \sum_{x \in \mathcal{X}} U_\mathcal{X}(x) \ln \sum_{y \in \mathcal{Y}} W(y|x)^{\frac{1}{1+\rho_R}} q_R(y)^{\frac{\rho_R}{1+\rho_R}}$$
$$= E_{SP}(R, U_\mathcal{X}), \tag{307}$$

where (305) follows by solving the convex program in (304), and (307) follows from (284) and (287). Hence, we conclude that ρ_R attains the supremum in (276). ∎

Let

$$\hat{S}_N := \sum_{n=1}^{N} \frac{Z_n}{N}, \tag{308}$$

and μ_N (resp. $\tilde{\mu}_N$) denote the law of $\hat{S}_N$ when Z_n are independent with laws ν_n (resp. $\tilde{\nu}_n$). Let

$$W_N := \sum_{n=1}^{N} \frac{T_n}{\sqrt{m_{2,N}}}, \tag{309}$$

where T_n and $m_{2,N}$ are defined right before the statement of the lemma. Via routine change of measure arguments (e.g., [42, p. 111]), one can check that

$$\mu_N\left([c, \infty)\right) = e^{-N\Lambda_N^*(c)} \int_0^\infty e^{-x\eta\sqrt{m_{2,N}}} dF_N(x) \tag{310}$$
$$= e^{-N\Lambda_N^*(c)} \int_0^\infty e^{-t} \left[F_N\left(\frac{t}{\psi_N}\right) - F_N(0)\right] dt, \tag{311}$$

where F_N is the distribution of W_N when Z_n are independent with laws $\tilde{\nu}_n$, $\psi_N := \eta\sqrt{m_{2,N}}$ and (311) follows from an application of the integration by parts. To deduce (69), first note that for any $t \in \mathbb{R}_+$

$$F_N\left(\frac{t}{\psi_N}\right) - F_N(0) \geq \Phi\left(\frac{t}{\psi_N}\right) - \Phi(0) - \frac{2m_{3,N}}{m_{2,N}^{3/2}} \tag{312}$$
$$\geq t\frac{\phi(0)}{\psi_N} - t^2\frac{1}{\psi_N^2 2\sqrt{2\pi e}} - \frac{2m_{3,N}}{m_{2,N}^{3/2}}, \tag{313}$$

where (312) follows from the Berry-Esseen theorem (e.g., [43, Theorem III.1]), and (313) follows from a power series approximation, coupled with the observation that $\phi'(\cdot) \geq -\frac{1}{\sqrt{2\pi e}}$ on $\mathbb{R}_+$. Using (313), we deduce that

$$\int_0^\infty e^{-t} \left[F_N\left(\frac{t}{\psi_N}\right) - F_N(0)\right] dt \tag{314}$$
$$\geq \int_{at_N}^\infty e^{-t} \left[F_N\left(\frac{t}{\psi_N}\right) - F_N(0)\right] dt \tag{315}$$
$$\geq \int_{at_N}^\infty e^{-t} \left[\frac{t\left(1 - \frac{1}{a}\right)}{\eta\sqrt{2\pi m_{2,N}}} - \frac{t^2}{\psi_N^2 2\sqrt{2\pi e}}\right] dt. \tag{316}$$

By carrying out the integration on the right side of (316) (e.g., [13, Eq. (221), (222)]), we conclude that (69) holds. ∎

Define $S_N := \sum_{n=1}^{N} Z_n$ and let F_N denote the distribution function of S_N. For convenience, let $B_N(r)$ denote the left

side of (72) and $m_{1,N} := \sum_{n=1}^{N} \mathrm{E}[Z_n]$. We have

$$B_N(r)$$

$$= \mathrm{e}^{-r} \int_{-\infty}^{r} \mathrm{e}^z \mathrm{d}F_N(z) \tag{317}$$

$$= F_N(r) - \int_{-\infty}^{r} \mathrm{e}^{(z-r)} F_N(z) \mathrm{d}z \tag{318}$$

$$= \int_0^\infty \mathrm{e}^{-x} \left[F_N(r) - F_N(r-x) \right] \mathrm{d}x \tag{319}$$

$$\leq \int_0^\infty \mathrm{e}^{-x} \left\{ \int_{\frac{r-m_{1,N}}{\sqrt{m_{2,N}}} - \frac{x}{\sqrt{m_{2,N}}}}^{\frac{r-m_{1,N}}{\sqrt{m_{2,N}}}} \frac{\mathrm{e}^{-\frac{a^2}{2}}}{\sqrt{2\pi}} \mathrm{d}a + c \frac{m_{3,N}}{m_{2,N}^{3/2}} \right\} \mathrm{d}x \tag{320}$$

$$\leq \frac{1}{\sqrt{2\pi m_{2,N}}} + c \frac{m_{3,N}}{m_{2,N}^{3/2}}, \tag{321}$$

where (318) follows from integration by parts, (320) follows from the Berry-Esseen Theorem[5] and $c = 2$ (resp. $c = 1$) if the random variables are independent (resp. i.i.d.).   ∎

APPENDIX D
PROOF OF LEMMA 3

We begin by recalling the fact that (ρ_R, q_R) is the unique saddle-point of the right side of (61), which is shown in Proposition 3(iv), and hence we are in a position to invoke the results proven in [13] throughout the proof.

(i) This assertion is a direct consequence of [13, Lemma 3(ii)].

(ii) The claim follows from [13, Theorem 2]. It was also shown earlier as part of the proof of Proposition 3(iv) (see (304)).

(iii) First, note that given any $r \in (\mathrm{D}(W_R\|q_R|U_{\mathcal{X}}), R]$,

$$e_{\mathrm{SP}}(r, R) = \max_{\rho \in \mathbb{R}_+} \min_{V \in \mathcal{P}(\mathcal{Y}|\mathcal{X})} \{ \mathrm{D}(V\|W|U_{\mathcal{X}}) \tag{322}$$
$$+ \rho \left(\mathrm{D}(V\|q_R|U_{\mathcal{X}}) - r \right) \}$$
$$= \max_{\rho \in \mathbb{R}_+} \left\{ -\rho r - (1+\rho)\Lambda \left(\tfrac{\rho}{1+\rho} \right) \right\}, \tag{323}$$

where (322) follows since the convex program $e_{\mathrm{SP}}(r, R)$, i.e., (74), has zero duality gap, thanks to the fact that Slater's condition (e.g., [41, Corollary 28.2.1]) holds, which is a direct consequence of the first assertion of this lemma, and (323) follows by solving the convex program on the right side of (322).

The proof of the assertion goes by contradiction. Assume that there exists $\lambda_\mathrm{o} \in [0, 1)$ with $\Lambda''(\lambda_\mathrm{o}) = 0$. From (85) and (86), this is equivalent to

$$W(y|x_\mathrm{o}) = q_R(y)\mathrm{e}^{-\Lambda'(\lambda_\mathrm{o})}, \, \forall\, y \in \mathrm{supp}(W(\cdot|x_\mathrm{o})). \tag{324}$$

Further, (323) and (324), along with the definition of $\Lambda(\cdot)$, imply that

$$e_{\mathrm{SP}}(R, R) = \max_{\rho \in \mathbb{R}_+} -\rho \left[R + \Lambda'(\lambda_\mathrm{o}) \right]. \tag{325}$$

Since $e_{\mathrm{SP}}(R, R) = \mathrm{E}_{\mathrm{SP}}(R)$, which is shown in the second assertion of this lemma, (325) implies that either $\mathrm{E}_{\mathrm{SP}}(R) = 0$, which contradicts the fact that $\mathrm{E}_{\mathrm{SP}}(R) > 0$ (e.g., [9, p. 158]), or $\mathrm{E}_{\mathrm{SP}}(R) = \infty$, which contradicts the fact that $R > R_\infty$. Hence, we conclude that for all $\lambda \in [0, 1)$, $\Lambda''(\lambda) > 0$.

(iv) For notational convenience, let

$$e_\mathrm{o}(\rho, R) := -(1+\rho)\Lambda \left(\tfrac{\rho}{1+\rho} \right). \tag{326}$$

Hence, (323) reads

$$e_{\mathrm{SP}}(r, R) = \max_{\rho \in \mathbb{R}_+} \{ e_\mathrm{o}(\rho, R) - \rho r \}. \tag{327}$$

$e_{\mathrm{SP}}(\cdot, R)$ is differentiable owing to [13, Corollary 2], and hence we conclude that $s_{(\cdot)}$ is well-defined. Since differentiable convex functions of one variable are continuously differentiable, the second assertion follows. To verify the last two assertions, observe that (327) is the Lagrangian dual of the convex program $e_{\mathrm{SP}}(r, R)$, i.e., (74), which is established in (322) and (323). Hence, we can use the subdifferential characterization of the Lagrange multipliers (e.g., [41, Theorem 29.1]) to deduce that the set of optimizers in (327) coincides with the negative of the subdifferential of $e_{\mathrm{SP}}(\cdot, R)$ at r, i.e., $\rho \in \mathbb{R}_+$ maximizes (327) if and only if

$$\rho \in -\partial e_{\mathrm{SP}}(\cdot, R)(r). \tag{328}$$

Since $e_{\mathrm{SP}}(\cdot, R)$ is differentiable at r, $-\partial e_{\mathrm{SP}}(\cdot, R)(r) = \{s_r\}$ and hence s_r uniquely attains the maximum in (327). Further, since $e_{\mathrm{SP}}(r, R) \geq e_{\mathrm{SP}}(R, R) = \mathrm{E}_{\mathrm{SP}}(R) > 0$, we have $s_r \in \mathbb{R}^+$.

Moreover, via direct differentiation, one can verify that

$$\frac{\partial^2}{\partial \rho^2} \left[-\rho r + e_\mathrm{o}(\rho, R) \right] = \frac{\partial^2 e_\mathrm{o}(\rho, R)}{\partial \rho^2} \tag{329}$$

$$= -\frac{\Lambda'' \left(\frac{\rho}{1+\rho} \right)}{(1+\rho)^3} \tag{330}$$

$$< 0, \tag{331}$$

where (331) follows from the third assertion of this lemma. As a direct consequence of (331), we conclude that s_r is the unique positive real number satisfying

$$r = \left. \frac{\partial e_\mathrm{o}(\rho, R)}{\partial \rho} \right|_{\rho = s_r}. \tag{332}$$

This observation, coupled with (331) and the inverse function theorem, further implies that s_r is strictly decreasing in r.

(v) Since $\Lambda(\cdot)$ is a convex function (e.g., [42, Lemma 2.2.5(a)]), $\lambda[e_{\mathrm{SP}}(r, R) - r] - \Lambda(\lambda)$ is a concave function of λ and hence a sufficient condition for $\lambda_\mathrm{o} \in \mathbb{R}$ to attain $\Lambda^*(e_{\mathrm{SP}}(r, R) - r)$ is

$$\Lambda'(\lambda_\mathrm{o}) = e_{\mathrm{SP}}(r, R) - r. \tag{333}$$

As noted above, s_r is the unique positive real number satisfying $r = \left. \frac{\partial e_\mathrm{o}(\rho, R)}{\rho} \right|_{\rho = s_r}$, hence, an elementary calculation implies that

$$r = -\Lambda \left(\tfrac{s_r}{1+s_r} \right) - \tfrac{1}{(1+s_r)}\Lambda' \left(\tfrac{s_r}{1+s_r} \right), \tag{334}$$

and hence

$$e_{\mathrm{SP}}(r, R) = \frac{s_r}{(1+s_r)}\Lambda'\left(\frac{s_r}{1+s_r}\right) - \Lambda\left(\frac{s_r}{1+s_r}\right). \quad (335)$$

Equations (334) and (335) imply that

$$\Lambda'\left(\frac{s_r}{1+s_r}\right) = e_{\mathrm{SP}}(r, R) - r. \quad (336)$$

Equation (336) ensures that $\frac{s_r}{1+s_r}$ attains $\Lambda^*(e_{\mathrm{SP}}(r, R) - r)$ and hence

$$\Lambda^*(e_{\mathrm{SP}}(r, R) - r) = \frac{s_r}{(1+s_r)}[e_{\mathrm{SP}}(r, R) - r] - \Lambda\left(\frac{s_r}{1+s_r}\right) \quad (337)$$

$$= e_{\mathrm{SP}}(r, R), \quad (338)$$

where (338) follows by substituting (336) into (335). Finally, let

$$\eta_r := \frac{s_r}{1+s_r}, \quad (339)$$

and note that $\eta_r \in \mathbb{R}^+$, since $s_r \in \mathbb{R}^+$. Hence, (336) implies the existence of a real number in $(0, 1)$, namely η_r, with

$$\Lambda'(\eta_r) = e_{\mathrm{SP}}(r, R) - r. \quad (340)$$

To verify the uniqueness, it suffices to note that $e_{\mathrm{SP}}(\cdot, R) - (\cdot)$ is strictly decreasing, along with the third assertion of this lemma and the inverse function theorem.

(vi) From the proof of part (iv) we know that s_R is the unique ρ that achieves the maximum in

$$\max_{\rho \geq 0} \{e_{\mathrm{o}}(\rho, R) - \rho R\} \quad (341)$$

$$= \max_{\rho \geq 0} \left\{-\rho R - (1+\rho)\Lambda\left(\frac{\rho}{1+\rho}\right)\right\} \quad (342)$$

$$= \max_{\rho \geq 0} \left\{ - \rho R \right. \quad (343)$$

$$\left. - (1+\rho)\ln\sum_{y \in \mathcal{Y}} q_R(y)^{\rho/(1+\rho)} W(y|x_{\mathrm{o}})^{1/(1+\rho)}\right\}. \quad (344)$$

But by Proposition 3(iv) and the symmetry of the channel, ρ_R achieves the maximum in (344). The conclusion follows. ∎

APPENDIX E
PROOF OF LEMMA 7

The proof follows from essentially the same arguments given in [13, Section III.E]. We provide an outline for completeness.

Since $\Lambda(\cdot)$ is smooth (by [42, Ex. 2.2.24]) and strictly convex over $(0, 1)$ (by Lemma 3(iii)), by [41, Corollary 23.5.1] and the inverse function theorem we have that $\Lambda^*(\cdot)$ is twice differentiable over the domain

$$(-D(W\|q_R|U_{\mathcal{X}}), D(W_R\|W|U_{\mathcal{X}}))$$

and

$$\Lambda^{*\prime}(e_{\mathrm{SP}}(r, R) - r) = \eta_r, \quad (345)$$

$$\Lambda^{*\prime\prime}(e_{\mathrm{SP}}(r, R) - r) = \frac{1}{\Lambda''(\eta_r)}, \quad (346)$$

for any $r \in [\bar{R}, R]$. Via calculations similar to the ones leading to [13, Eq. (92)], one can verify that

$$\Lambda^*(e_{\mathrm{SP}}(R_N, R) - R_N)$$
$$= \Lambda^*(e_{\mathrm{SP}}(R, R) - R) + \varepsilon_N \eta_R \quad (347)$$
$$\quad + (e_{\mathrm{SP}}(R_N, R) - e_{\mathrm{SP}}(R, R))\eta_R$$
$$\quad + \frac{\Lambda^{*\prime\prime}(\bar{x})}{2}[e_{\mathrm{SP}}(R_N, R) - R_N - e_{\mathrm{SP}}(R, R) + R]^2,$$

for some $\bar{x} \in (e_{\mathrm{SP}}(R, R) - R, e_{\mathrm{SP}}(R_N, R) - R_N)$. Note that since $e_{\mathrm{SP}}(\cdot, R) - (\cdot)$ is strictly decreasing from the definition of $e_{\mathrm{SP}}(\cdot, \cdot)$ in (74) and $R_N < R$, $e_{\mathrm{SP}}(R, R) - R < e_{\mathrm{SP}}(R_N, R) - R_N$. Using Lemma 3(iv) and (v), along with the definition of ε_N, (347) further implies that

$$e_{\mathrm{SP}}(R_N, R) = e_{\mathrm{SP}}(R, R) + \varepsilon_N s_R \quad (348)$$
$$\quad + \varepsilon_N^2 (1 + s_R)\frac{\Lambda^{*\prime\prime}(\bar{x})}{2}$$
$$\quad \cdot \left(1 + \frac{1}{\varepsilon_N}[e_{\mathrm{SP}}(R_N, R) - e_{\mathrm{SP}}(R, R)]\right)^2.$$

By using (346), along with the fact that $e_{\mathrm{SP}}(\cdot, R) - (\cdot)$ is strictly decreasing and continuous over $[\bar{R}, R]$, which is again a consequence of the definition of $e_{\mathrm{SP}}(\cdot, R)$, i.e., (74), we deduce that

$$\Lambda^{*\prime\prime}(\bar{x}) \leq \frac{1}{m_{2,\min}} \in \mathbb{R}^+. \quad (349)$$

Now Lemma 3(iv) and (vi) imply that

$$s_R = \rho_R = |E'_{\mathrm{SP}}(R)|. \quad (350)$$

Finally, via a first-order power series approximation, along with Lemma 3(iv) and (v), one can verify that

$$\left(1 + \frac{1}{\varepsilon_N}[e_{\mathrm{SP}}(R_N, R) - e_{\mathrm{SP}}(R, R)]\right)^2 \leq (1 + s_{\bar{R}})^2. \quad (351)$$

Assembling (348)–(351), along with the fact that $E_{\mathrm{SP}}(R) = e_{\mathrm{SP}}(R, R)$, which is shown in Lemma 3(ii), we conclude that (122) holds. ∎

APPENDIX F
PROOF OF LEMMA 8

(i) Let $\{\mathcal{Y}_l\}_{l=1}^L$ be a partition of the columns of W mentioned in Definition 1, whose choice is immaterial in what follows. Since each column is a permutation of every other column for any sub-channel defined by this partition, $q(y)$ is the same for any $y \in \mathcal{Y}_l$, which, in turn, allows us to particularize Proposition 1(ii) with $Q_Y \leftarrow q$ to deduce that for any $r \in \mathbb{R}^+$,

$$\sum_{\mathbf{y}^N \in \mathcal{S}(r)} W(y_1|\psi_1)\prod_{n=2}^N W(y_n|\psi_n(\mathbf{y}^{n-1}))$$
$$= W\left\{\mathcal{S}(r)|\mathbf{x}_{\mathrm{o}}^N\right\}. \quad (352)$$

(ii) Define

$$\Lambda(\lambda) := \ln \mathrm{E}_{W(\cdot|x_{\mathrm{o}})}\left[e^{\lambda \ln \frac{q(Y)}{W(Y|x_{\mathrm{o}})}}\right] \quad (353)$$

$$= \ln \sum_{y \in \mathrm{supp}(W(\cdot|x_{\mathrm{o}}))} W(y|x_{\mathrm{o}})^{1-\lambda} q(y)^\lambda. \quad (354)$$

The singularity of W, along with (126), implies that

$$\Lambda(\lambda) = \ln \sum_{y \in \text{supp}(W(\cdot|x_o))} \xi_y \alpha_y^\lambda. \qquad (355)$$

Observe that for any $\lambda \in \mathbb{R}_+$,

$$\sum_{y \in \text{supp}(W(\cdot|x_o))} \xi_y \alpha_y^\lambda \qquad (356)$$

$$= \sum_{y \in \text{supp}(W(\cdot|x_o))} W(y|x_o)^{\frac{1}{1+\lambda}}. \qquad (357)$$

$$\left(\sum_{z \in \mathcal{X}: W(y|z) > 0} U_\mathcal{X}(z) W(y|x_o)^{\frac{1}{1+\lambda}} \right)^\lambda \qquad (358)$$

$$= \sum_{y \in \text{supp}(W(\cdot|x_o))} W(y|x_o)^{\frac{1}{1+\lambda}}. \qquad (359)$$

$$\left(\sum_{z \in \mathcal{X}} U_\mathcal{X}(z) W(y|z)^{\frac{1}{1+\lambda}} \right)^\lambda \qquad (360)$$

$$= \sum_{y \in \mathcal{Y}} \left(\sum_{z \in \mathcal{X}} U_\mathcal{X}(z) W(y|z)^{\frac{1}{1+\lambda}} \right)^{1+\lambda} \qquad (361)$$

$$= \sum_{y \in \mathcal{Y}} \xi_y \alpha_y^{1+\lambda} \qquad (362)$$

where (358), (360), and (362) follow from the singularity of the channel, and (361) follows from Proposition 3(ii). In light of (355) and (362), we note that for any $\lambda \in \mathbb{R}_+$,

$$\Lambda(\lambda) = \ln \sum_{y \in \mathcal{Y}} \xi_y \alpha_y^{1+\lambda} \qquad (363)$$

$$= -\text{E}_o(\lambda, U_\mathcal{X}), \qquad (364)$$

where $\text{E}_o(\cdot, \cdot)$ is defined in (21) and (364) follows from an elementary calculation that uses the singularity of the channel. The identity (364) enables us to relate the large deviations rate function

$$\Lambda^*(-R) := \sup_{\lambda \in \mathbb{R}} \{-\lambda R - \Lambda(\lambda)\} \qquad (365)$$

to $\text{E}_{\text{SP}}(R)$, our target exponent, and hence is the crucial step of the proof (see (386) to (390) to follow). Note that it depends crucially on the singularity of the channel. Continuing with the proof, one can check that

$$\Lambda'(\lambda) = \sum_y \frac{\xi_y \alpha_y^\lambda}{\sum_b \delta_b \alpha_b^\lambda} \ln \alpha_y, \qquad (366)$$

$$\Lambda''(\lambda) = \sum_y \frac{\xi_y \alpha_y^\lambda}{\sum_b \delta_b \alpha_b^\lambda} \left(\ln \alpha_y - \Lambda'(\lambda) \right)^2, \qquad (367)$$

$$\geq 0, \qquad (368)$$

for any $\lambda \in \mathbb{R}_+$, where all sums are over the set $\text{supp}(W(\cdot|x_o))$. Further, define

$$m_3(\lambda) := \sum_y \frac{\xi_y \alpha_y^\lambda}{\sum_b \delta_b \alpha_b^\lambda} \left| \ln \alpha_y - \Lambda'(\lambda) \right|^3, \qquad (369)$$

where again both sums are over the set $\text{supp}(W(\cdot|x_o))$. Evidently, $\Lambda'(\cdot), \Lambda''(\cdot)$ and $m_3(\cdot)$ are bounded and continuous over $\mathbb{R}_+$. Next, we prove that

$$\forall \lambda \in \mathbb{R}_+, \ \Lambda''(\lambda) > 0. \qquad (370)$$

In order to see (370), first note that

$$\Lambda''(\lambda) \geq 0, \forall \lambda \in \mathbb{R}_+, \qquad (371)$$

due to (368). Assume there exists $\lambda_o \in \mathbb{R}_+$ with $\Lambda''(\lambda_o) = 0$. This, however, implies that $R_{\text{cr}} = C(W)$, owing to (364), [9, Theorem 5.6.3], Remark 1(i) and the fact that $U_\mathcal{X}$ is a capacity-achieving input distribution for W, which yields a contradiction.
For any $r \in (R_\infty, R]$, let

$$\rho_r := -\left. \frac{\partial \text{E}_{\text{SP}}(a, U_\mathcal{X})}{\partial a} \right|_{a=r}, \qquad (372)$$

which is a well-defined mapping owing to [13, Proposition 3]. Further, observe that for any $r \in (R_\infty, R]$,

$$-r = \Lambda'(\rho_r), \qquad (373)$$

which is evident in light of

$$r = \left. \frac{\partial \text{E}_o(\rho, U_\mathcal{X})}{\partial \rho} \right|_{\rho=\rho_r} \qquad (374)$$

$$= -\Lambda'(\rho_r), \qquad (375)$$

where (374) follows by recalling the fact that ρ_r attains $\tilde{\text{E}}_{\text{SP}}(r, U_\mathcal{X})$, which is shown in Proposition 3(iii), and (375) follows from (364). Moreover, since ρ_r attains $\tilde{\text{E}}_{\text{SP}}(r, U_\mathcal{X})$ and for any $r \in (R_\infty, R]$,

$$\tilde{\text{E}}_{\text{SP}}(r, U_\mathcal{X}) \geq \tilde{\text{E}}_{\text{SP}}(R, U_\mathcal{X}) = \tilde{\text{E}}_{\text{SP}}(R) > 0, \qquad (376)$$

we deduce that $\rho_r \in \mathbb{R}^+$. Further, (370), (373) and the inverse function theorem ensure that $\rho_{(\cdot)}$ is strictly decreasing over $(R_\infty, R]$.
To conclude the proof, we fix some $a > 1$ and define

$$t_{\max} := a 2 \sqrt{2\pi} \rho_{\bar{R}} \max_{\lambda \in [0, \rho_{\bar{R}}]} \frac{m_3(\lambda)}{\Lambda''(\lambda)}, \qquad (377)$$

$$m_{2,\min} := \min_{\lambda \in [0, \rho_{\bar{R}}]} \Lambda''(\lambda), \qquad (378)$$

$$m_{2,\max} := \max_{\lambda \in [0, \rho_{\bar{R}}]} \Lambda''(\lambda), \qquad (379)$$

where $\bar{R} = \frac{R+R_\infty}{2}$, as defined before. Clearly, all of the above are well-defined and positive quantities. For convenience, let

$$\frac{e^{-t_{\max}} \left(1 - \frac{1}{a} \right)}{\rho_{\bar{R}} 2 \sqrt{2\pi m_{2,\max}}} =: k_o \in \mathbb{R}^+. \qquad (380)$$

Let $N \in \mathbb{Z}^+$ be sufficiently large such that

$$R_N \geq \bar{R}, \qquad (381)$$

$$\frac{1 + (1 + t_{\max})^2}{\rho_{\bar{R}} \left(1 - \frac{1}{a} \right) 2 \sqrt{e N m_{2,\min}}} \leq 1/2, \qquad (382)$$

and note that

$$W\left\{\mathcal{S}(R_N)|\mathbf{x}_o^N\right\} \tag{383}$$

$$\geq k_o\left(1 + a2\sqrt{2\pi}\rho_{R_N}\frac{m_3(\rho_{R_N})}{\Lambda''(\rho_{R_N})}\right)\frac{1}{\sqrt{N}}e^{-N\Lambda^*(-R_N)} \tag{384}$$

$$\geq \frac{k_o}{\sqrt{N}}e^{-N\Lambda^*(-R_N)}, \tag{385}$$

where (384) follows from Lemma 1, which is applicable thanks to (370) and (373), along with (381) and (382). Since $\rho_{(\cdot)} \in \mathbb{R}^+$ is strictly decreasing and $\Lambda(\cdot)$ is convex, (373) implies that

$$\Lambda^*(-R_N) = \max_{0\leq\lambda\leq\rho_{\bar{R}}}\left\{-\lambda\left(R-\frac{k}{N}\right)-\Lambda(\lambda)\right\} \tag{386}$$

$$\leq \frac{k\rho_{\bar{R}}}{N} + \max_{0\leq\lambda\leq\rho_{\bar{R}}}\left\{-\lambda R - \Lambda(\lambda)\right\} \tag{387}$$

$$\leq \frac{k\rho_{\bar{R}}}{N} + \sup_{\lambda\in\mathbb{R}_+}\left\{-\lambda R - \Lambda(\lambda)\right\} \tag{388}$$

$$= \frac{k\rho_{\bar{R}}}{N} + \sup_{\lambda\in\mathbb{R}_+}\left\{-\lambda R + \mathrm{E}_o(\lambda, U_\mathcal{X})\right\} \tag{389}$$

$$= \frac{k\rho_{\bar{R}}}{N} + \mathrm{E}_{SP}(R), \tag{390}$$

where (389) follows from (364) and (390) follows from Proposition 3(i). By substituting (390) into (385), we deduce the assertion. ∎

ACKNOWLEDGMENT

REFERENCES

[1] A. Feinstein, "A new basic theorem of information theory," *Trans. IRE Prof. Group Inf. Theory*, vol. 4, no. 4, pp. 2–22, Sep. 1954.

[2] P. Elias, "Coding for two noisy channels," in *Proc. Inf. Theory*, 3rd ed. London, U.K.: Butterworths Scientific Press, 1955, pp. 61–76. [Online]. Available: https://link.springer.com/content/pdf/10.1038/176773a0.pdf

[3] C. E. Shannon, "Certain results in coding theory for noisy channels," *Inf. Control*, vol. 1, no. 1, pp. 6–25, Sep. 1957.

[4] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications*. New York, NY, USA: Wiley, 1961.

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

[6] R. 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.

[7] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels," *Inf. Control*, vol. 10, pp. 65–103, Jan. 1967.

[8] E. A. Haroutunian, "Estimates of the error exponents for the semi-continuous memoryless channel," (in Russian), *Probl. Per. Inf.*, vol. 4, no. 4, pp. 37–48, 1968.

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

[10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York, NY, USA: Academic, 1981.

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

[12] F. Alajaji and P.-N. Chen, *An Introduction to Single-User Information Theory*. Singapore: Springer, 2018.

[13] Y. Altuğ and A. B. Wagner, "Refinement of the sphere-packing bound: Asymmetric channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1592–1615, Mar. 2014.

[14] 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.

[15] J. Honda, "Exact asymptotics for the random coding error probability," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2015, pp. 91–95.

[16] J. Honda, "Exact asymptotics of random coding error probability for general memoryless channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1844–1848.

[17] J. Scarlett, A. Martinez, and A. Guillén i Fabregas, "A derivation of the asymptotic random-coding prefactor," in *Proc. Ann. Allerton Conf. Commun., Control, Comput.*, Oct. 2013, pp. 956–961.

[18] J. Scarlett, A. Martinez, and A. Guillen i Fabregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2647–2666, May 2014.

[19] J. Font-Segura, A. Martinez, and A. G. i Fabregas, "Asymptotics of the random coding union bound," in *Proc. Int. Symp. Inf. Theory Appl. (ISITA)*, Oct. 2018, pp. 125–129.

[20] B. Nakiboglu, "The sphere packing bound via Augustin's method," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 816–840, Feb. 2019.

[21] B. Nakiboglu, "The sphere packing bound for DSPCs with feedback a la augustin," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7456–7467, Nov. 2019.

[22] Y. Altuğ and A. B. Wagner, "Feedback can improve the second-order coding performance in discrete memoryless channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2014, pp. 2361–2365.

[23] N. V. Shende and A. B. Wagner, "On very noisy channels with feedback," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2017, pp. 852–859.

[24] N. V. Shende, A. B. Wagner, and Y. Altug, "When does feedback improve the second-order coding rate in discrete memoryless channels?" in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1485–1489.

[25] A. B. Wagner, N. V. Shende, and Y. Altuğ, "A new method for employing feedback to improve coding performance," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, pp. 6660–6681, Nov. 2020.

[26] R. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 405–417, Jul. 1974.

[27] 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.

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

[29] R. R. Bahadur and R. R. Rao, "On deviations of the sample mean," *Ann. Math. Statist.*, vol. 31, no. 4, pp. 1015–1027, Dec. 1960.

[30] E. A. Haroutunian, "Lower bound for error probability in channels with feedback," *Problems Inf. Transmiss.*, vol. 13, no. 2, pp. 107–114, 1977.

[31] 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.

[32] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, USA, Nov. 2010.

[33] J. Font-Segura, G. Vazquez-Vilar, A. Martinez, A. Guillen i Fabregas, and A. Lancho, "Saddlepoint approximations of lower and upper bounds to the error probability in channel coding," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2018, pp. 1–6.

[34] 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.

[35] P. Billingsley, *Probability and Measure*, 3rd ed. Hoboken, NJ, USA: Wiley, 1995.

[36] Y. Altuğ, "Moderate deviations and exact asymptotics in channel coding," Ph.D. dissertation, Dept. Elect. Comput. Eng., Cornell Univ., Ithaca, NY, USA, Aug. 2013.

[37] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," in *Proc. Trans. 3rd Prague Conf. Inf. Theory*. Prague, Czech Republic: Czechoslovak Academy of Sciences, 1962, pp. 689–723.

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

[39] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.

[40] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypothesis," *Philos. Trans. Roy. Soc. London. Ser. A*, vol. 231, pp. 289–337, 1933.

[41] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1970.

[42] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed. New York, NY, USA: Springer, 1998.

[43] C.-G. Esseen, "Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law," *Acta Math.*, vol. 77, no. 1, pp. 1–125, 1945.

[44] V. Y. Korolev and I. G. Shevtsova, "A new moment-type estimate of convergence rate in the Lyapunov theorem," *Theory Probab. Appl.*, vol. 55, no. 3, pp. 505–509, Jan. 2011.

Yücel Altuğ received the B.S. and M.S. degrees in electrical and electronics engineering from Boğaziçi University, Turkey, in 2006 and 2008, respectively, and the Ph.D. degree in electrical and computer engineering from Cornell University, in 2013, where he has been awarded the ECE Director's Ph.D. Thesis Research Award. After post-doctoral appointments at Cornell University and Princeton University, he is currently a Lead Data Scientist at Natera Inc. His research interests include Shannon theory, feedback communications, and stochastic modeling and algorithm design for next-generation DNA sequencing and genetic testing.

Aaron B. Wagner (Fellow, IEEE) received the B.S. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1999, and the M.S. and Ph.D. degrees in electrical engineering and computer sciences from the University of California, Berkeley, CA, USA, in 2002 and 2005, respectively. From 2005 to 2006, he was a Post-Doctoral Research Associate with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, and a Visiting Assistant Professor with the School of Electrical and Computer Engineering, Cornell University. Since 2006, he has been with the School of Electrical and Computer Engineering at Cornell, where he is currently a Professor. He has received the NSF CAREER Award, the David J. Sakrison Memorial Prize from the U.C. Berkeley EECS Department, the Bernard Friedman Memorial Prize in Applied Mathematics from the U.C. Berkeley Department of Mathematics, the James L. Massey Research and Teaching Award for Young Scholars from the IEEE Information Theory Society, and teaching awards at the department, college, and university level at Cornell. He was a Distinguished Lecturer for the IEEE Information Theory Society (2018–2019).