Error Exponents of the Dirty-Paper and Gel’fand–Pinsker Channels

Ran Tamir, Member, IEEE, and Neri Merhav, Life Fellow, IEEE

Abstract—We derive various error exponents for communication channels with random states, which are available non-causally at the encoder only. For both the finite-alphabet Gel’fand–Pinsker channel and its Gaussian counterpart, the dirty-paper channel, we derive random coding exponents, error exponents of the typical random codes (TRCs), and error exponents of expurgated codes. For the two channel models, we analyze some sub-optimal bin-index decoders, which turn out to be asymptotically optimal, at least for the random coding error exponent. For the dirty-paper channel, we show explicitly via a numerical example, that both the error exponent of the TRC and the expurgated exponent strictly improve upon the random coding exponent, at relatively low coding rates, which is a known fact for discrete memoryless channels without random states. We also show that at rates below capacity, the optimal values of the dirty-paper design parameter \( \alpha \) in the random coding sense and in the TRC exponent sense are different from one another, and they are both different from the optimal \( \alpha \) that is required for attaining the channel capacity. For the Gel’fand–Pinsker channel, we allow for a variable-rate random binning code construction, and prove that the previously proposed maximum penalized mutual information decoder is asymptotically optimal within a given class of decoders, at least for the random coding error exponent.

Index Terms—Dirty-paper channel, error exponent, expurgated exponent, Gel’fand–Pinsker channel, random states, side information, typical random code.

I. INTRODUCTION

Channels with random states available as side information (SI) at the encoder have been studied for more than four decades. In 1980, Gel’fand and Pinsker [11] have derived the capacity formula for this channel model with discrete alphabets, and Costa has considered the Gaussian counterpart a few years later [6], which is widely known as the dirty-paper channel (DPC). Applications and extensions of their work include computer memories with defects [12], multiuser (DPC). Applications and extensions of their work include computer memories with defects [12], multiuser

TRCs for both DMCs (which coincide with one another for identical SI versions at the encoder, the adversary, and the decoder sides subject to various cost constrains were developed by Moulin and Wang [27]. A strong converse result and a sphere-packing type bound for the Gel’fand–Pinsker channel were derived by Tyagi and Narayan [37].

Error exponents of typical random codes (TRCs) and expurgated codes for the DPC and the Gel’fand–Pinsker channel are the main theme of this work. The error exponent of the TRC [21] is defined as

\[
E_{\text{trc}}(R) = \lim_{n \to \infty} \left\{ \frac{1}{n} \ln P(C_n) \right\},
\]

where \( R \) is the coding rate, \( P(C_n) \) is the error probability of a codebook \( C_n \), and the expectation is with respect to (w.r.t.) the randomness of \( C_n \) across the ensemble of codes. As explained in [21], the error exponent of the TRC does not only strictly improve upon the random coding error exponent at low coding rates, but actually provides the exponent function around which the negative normalized logarithmic error probability concentrates.

In [3], Barg and Forney considered TRCs with independently and identically distributed codewords as well as typical linear codes, for the special case of the binary symmetric channel with maximum likelihood (ML) decoding. In [28] Nazari et al. provided bounds on the error exponents of TRCs for both DMCs (which coincide with one another for the optimal input distribution) and multiple-access channels. In a recent article by Merhav [21], an exact single-letter expression has been derived for the error exponent of typical, random, fixed-composition codes, over DMCs, and a wide class of (stochastic) decoders, collectively referred to as the generalized likelihood decoder. Later, Merhav has studied error exponents of TRCs for the colored Gaussian channel [22], typical random trellis codes [23], and derived

\[
E_{\text{trc}}(R) = \lim_{n \to \infty} \left\{ \frac{1}{n} \ln P(C_n) \right\}.
\]

1Note that this definition is different from the ordinary random-coding exponent, which is given by \( E_{\text{rc}}(R) = \lim_{n \to \infty} \left\{ -\frac{1}{n} \ln P(C_n) \right\}. \)

Manuscript received 14 May 2022; revised 8 May 2023; accepted 28 August 2023. Date of publication 13 September 2023; date of current version 22 November 2023. This work was supported by the Israel Science Foundation (ISF) under Grant 13718. The work of Ran Tamir was supported in part by the Swiss Federal Institute of Technology (ETH–Foundations of Data Science (ETH-FDS). An earlier version of this paper was presented in part at the 2023 IEEE Information Theory Workshop [DOI: 10.1109/ITW55543.2023.10161668]. (Corresponding author: Ran Tamir.)

Ran Tamir is with the Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland (e-mail: tamir@isi.ee.ethz.ch).

Neri Merhav is with the Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering, Technion–Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: merhav@ee.technion.ac.il).

Communicated by H. Tyagi, Associate Editor for Shannon Theory and Information Measures.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2023.3314210.

Digital Object Identifier 10.1109/TIT.2023.3314210.
a Lagrange–dual lower bound to the TRC exponent [24]. Recently, Tamir et al. have studied the large deviations behavior around the TRC exponent [35], error exponents of typical random Slepian–Wolf codes [33], and universal decoding for the TRC exponent [34]. More interesting concentration results of the logarithmic error probability of random codes around the error exponent of the TRC have been reported in [36]. Finally, a dual-domain lower bound to the TRC error exponent for general channels and pairwise-independent random-coding ensembles appears in [5].

For the DPC, an ordinary random binning code is analyzed and various error exponents are derived. Specifically, we provide some relatively simple expressions for the exact random coding error exponent and for the error exponent of the TRC. As in the ordinary DMC (without random states), an improvement is achieved at relatively low coding rates by code expurgation. We show also that in the DPC, the TRC exponent and the expurgated exponent strictly improve upon the random coding error exponent at some ranges of low rates. Although we implement a simple sub-optimal decoder, which seeks the closest codeword to the channel output vector, and not the optimal bin-index decoder, we show that at least for the random coding exponent, our decoder is as good as the optimal decoder. We also show that at low rates, such as rate zero, the optimal values of the design parameter \( \alpha \) in the random coding sense and in the TRC exponent sense are different, and they both differ from the optimal \( \alpha \) that is required for achieving the channel capacity.

Moving further, we turn to derive similar exponential error bounds for the Gel’fand–Pinsker channel. Also for the Gel’fand–Pinsker channel, we construct a random-binning code, but now we allow for a slightly more sophisticated code construction, which is in the spirit of [27]. For every possible SI type, we draw a different random-binning code, and select the binning rate individually, thus allowing for more degrees of freedom to improve the code performance. For completeness, we also adopt the penalized maximum mutual information (MMI) decoder proposed in [27]. The DPC and the Gel’fand–Pinsker channel are very close in spirit, and one may even consider the DPC as a special case of the Gel’fand–Pinsker channel, but still, they differ significantly by the simple fact that in the Gel’fand–Pinsker channel, we allow for a different sub-code for each SI type. As a consequence, the analyses in the two cases follow different lines and the resulting expressions in the DPC case are considerably simpler, considering the low number of parameters that should undergo optimization. In the Gel’fand–Pinsker case, on the other hand, the final expressions are given by optimization problems, the dimension of which grows with the alphabet sizes. As opposed to our analysis for the DPC, which is tight at all coding rates, here, when deriving the TRC exponent and the expurgated bound, we make a pairwise error analysis, thus providing exponent functions which are tight only at relatively low rates, which is the range where these exponent functions improve upon the random coding error exponent (and the latter is tight at the high rates). For the above described code construction of variable-rate binning (that depends on the empirical distribution of the SI), we prove that at least for the random coding error exponent, the penalized MMI decoder is actually optimal among all metrics that depend both on the joint empirical distribution of the codeword and the channel output sequence as well as on the SI possible type.4 Due to recent findings in [34], we conjecture that the penalized MMI decoder is also optimal w.r.t. the error exponents of the TRC, but leave this question for future work.

The remaining part of the paper is organized as follows. In Section II, we establish notation conventions. In Section III, we consider the DPC; in Subsection III-A, we formalize the model, review some preliminaries, and indicate the main objectives and in Subsection III-C, we provide the main results and discuss them. Section IV is devoted to the Gel’fand–Pinsker channel. In Subsection IV-A, we formalize the settings and in Subsection IV-B, we provide and discuss the results. In Section V we prove our results regarding the DPC and in the Appendices, we prove more supplementary results concerning DPC, discuss optimal bin index decoding, and prove the results concerning the Gel’fand–Pinsker model.

II. Notation Conventions

Throughout the paper, random variables will be denoted by capital letters and specific values they may take will be denoted by the corresponding lower case letters, both in the bold face font. For example, the random vector \( X = (X_1, X_2, \ldots, X_n) \) (\( n \) is a positive integer) may take a specific vector value \( x = (x_1, x_2, \ldots, x_n) \) in \( \mathbb{R}^n \). When used in the linear-algebraic context, these vectors should be thought of as column vectors, and so, when they appear with superscript \( T \), they will be transformed into row vectors by transposition. Thus, \( x^T y \) is understood as the inner product of \( x \) and \( y \). The notation \( \| x \| \) will stand for the Euclidean norm of vector \( x \). The \( n \)-dimensional hypersphere of radius \( \sqrt{n} \sigma \) will be denoted by \( S(\sqrt{n} \sigma) \) and its surface area by \( \text{Surf}(S(\sqrt{n} \sigma)) \). As customary in probability theory, we write \( X = (X_1, \ldots, X_n) \sim \mathcal{N}(0, \sigma^2 \cdot I_n) \) (\( I_n \) being the \( n \times n \) identity matrix) to denote that the probability density function of \( X \) is

\[
P_X(x) = (2\pi\sigma^2)^{-n/2} \cdot \exp \left\{ -\frac{1}{2\sigma^2} \| x \|^2 \right\}.
\]

Logarithms, here and throughout the sequel, are taken to the natural base. The probability of an event \( \mathcal{E} \) will be denoted by \( P(\mathcal{E}) \), and the expectation operator will be denoted by \( \mathbb{E}[] \). For two positive sequences \( a_n \) and \( b_n \), the notation \( a_n \asymp b_n \) will stand for equality in the exponential scale, that is, \( \lim_{n \to \infty} \frac{1}{b_n} \ln \frac{a_n}{b_n} = 0 \). Similarly, \( a_n \leq b_n \) means that \( \limsup_{n \to \infty} (1/n) \ln (a_n/b_n) \leq 0 \), and so on. The indicator function of an event \( \mathcal{E} \) will be denoted by \( 1[\mathcal{E}] \). The notation \( [x]_+ \) will stand for \( \max\{0, x\} \).

4Since the decoder has no access to the actual SI, it has to seek all codebooks pertaining to all possible SI types. Then, the penalized MMI decoder balances the fact that each codebook has a different binning rate.

---

2Recall that in the DPC, the transmitted vector is \( x = u(m, s) - \alpha s \), with \( s \) being the SI vector and \( u(m, s) \) an auxiliary codeword that depends on both the SI vector and the transmitted message \( m \in \{1, 2, \ldots, M\} \), and where \( \alpha \) is subject to optimization.

3Note that in the continuous case, the different types of \( s \) are defined according to their norm, but \( I(U; S) \) is independent of \( \| s \| \). In the Gel’fand–Pinsker channel, on the other hand, there is no such parallel property.
In the discrete case, alphabets of random variables and their realizations will be denoted by calligraphic letters. Accordingly, alphabets of random vectors and their realizations will be superscripted by their dimensions. For example, the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ may take a specific vector value $x = (x_1, \ldots, x_n)$ in $\mathcal{X}^n$, the $n$-th order Cartesian power of $\mathcal{X}$, which is the alphabet of each component of this vector. Sources and channels will be subscripted by the names of the relevant random variables/vectors and their conditionings, whenever applicable, following the standard notation conventions, e.g., $Q_X$, $Q_{Y|X}$, and so on. When there is no room for ambiguity, these subscripts will be omitted. For a generic joint distribution $Q_{XY} = \{Q_{XY}(x,y), x \in \mathcal{X}, y \in \mathcal{Y}\}$, which will often be abbreviated by $Q$, information measures will be denoted in the conventional manner, but with a subscript $Q$, that is, $H_Q(X)$ is the marginal entropy of $X$, $H_Q(X|Y)$ is the conditional entropy of $X$ given $Y$, $I_Q(X;Y) = H_Q(X) - H_Q(X|Y)$ is the mutual information between $X$ and $Y$, and similarly for other quantities. The weighted divergence between two conditional distributions (channels), say, $Q_{Y|X}$ and $W = \{W(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$, with weighting $Q_X$ is defined as

$$D(Q_{Y|X}||W|Q_X) = \sum_{x \in \mathcal{X}} Q_X(x) \sum_{y \in \mathcal{Y}} Q_{Y|X}(y|x) \ln \frac{Q_{Y|X}(y|x)}{W(y|x)}. \tag{3}$$

The empirical distribution of a sequence $x \in \mathcal{X}^n$, which will be denoted by $\hat{P}_x$, is the vector of relative frequencies, $\hat{P}_x(x)$, of each symbol $x \in \mathcal{X}$ in $x$. The type class of $x \in \mathcal{X}^n$, denoted $T^n(x)$, is the set of all vectors $x' \in \mathcal{X}^n$ with $\hat{P}_{x'} = \hat{P}_x$. When we wish to emphasize the dependence of the type class on the empirical distribution $P$, we will denote it by $T^n(P)$. The set of all types of vectors of length $n$ over $\mathcal{X}$ will be denoted by $\mathcal{P}_n(\mathcal{X})$. Information measures associated with empirical distributions will be denoted with ‘hats’ and will be subscripted by the sequences from which they are induced. Similar conventions will apply to the joint empirical distribution, the joint type class, the conditional empirical distributions and the conditional type classes associated with pairs (and multiples) of sequences of length $n$. Accordingly, $\hat{P}_{xy}$ would be the joint empirical distribution of $(x, y) = \{(x_i, y_i)\}_{i=1}^n$, $T^n(\hat{Q}_{X|Y})$ will stand for the conditional type class induced by a sequence $y$ and a relevant empirical conditional distribution $\hat{Q}_{X|Y}$, which is the set of all vectors $x \in \mathcal{X}^n$ with $\hat{P}_{xy} = Q_{X|Y} \times \hat{P}_y$, $\hat{I}_{xy}(X;Y)$ will denote the empirical mutual information induced by $x$ and $y$, and so on. Similar conventions will apply to triplets of sequences, say, $\{(x, y, z)\}$, etc. Likewise, when we wish to emphasize the dependence of empirical information measures upon a given empirical distribution given by $Q$, we denote them using the subscript $Q$, as described above.

III. ERROR EXPONENTS OF DIRTY-PAPER CODING

A. Setup, Preliminaries, and Objectives

Consider the DPC,

$$\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{Z}, \tag{4}$$

where $\mathbf{X} = (X_1, \ldots, X_n)$ is the channel input vector, whose power is limited according to $\|\mathbf{X}\|^2 \leq nP$. $\mathbf{S} = (S_1, \ldots, S_n) \sim \mathcal{N}(0, \mathbf{Q}_I)$, is the random interference signal vector, and $\mathbf{Z} = (Z_1, \ldots, Z_n) \sim \mathcal{N}(0, \sigma_Z^2 \cdot \mathbf{I}_n)$ is the additive Gaussian noise, which is independent of $\mathbf{X}$ and $\mathbf{S}$. Here, $Q > 0$ and $\sigma_Z^2 > 0$ are the interference variance and the noise variance, respectively.

Consider the following mechanism of random codebook selection. Let the coding rate, $R \geq 0$, be given. Let $\alpha > 0$ be a design parameter (to be optimized later), and define

$$\rho_0 \triangleq \sqrt{\frac{\alpha^2 Q}{P + \alpha^2 Q}} = \sqrt{\frac{\alpha^2 Q}{W}}, \tag{5}$$

where $W \triangleq P + \alpha^2 Q$. Next, define

$$I_{US} \triangleq \frac{1}{2} \ln \left( \frac{1}{1 - \rho_0^2} \right) = \frac{1}{2} \ln \left( \frac{W}{P} \right). \tag{6}$$

Generate a random codebook, $C$, of $M = e^{n(R + I_{US} + \Delta)}$ ($\Delta > 0$ being arbitrarily small) codewords, $u_k$, $k = 0, 1, \ldots, M - 1$, by independent random selection of each codeword under the uniform distribution over the surface of an $n$-dimensional hypersphere of radius $\sqrt{nW}$, centered at the origin. Next, divide $C$ into $M_1 = e^{nR}$ bins, $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{M_1 - 1}$, each of size $M_2 = M/M_1 = e^{n(I_{US} + \Delta)}$.

Let $\epsilon > 0$ be arbitrarily small. Given a message $m \in \{0, 1, \ldots, M_1 - 1\}$ and a realization, $s$ of $\mathbf{S}$, the encoder seeks, within bin number $m$, a codeword, $u_k$, $k \in \{M_2m, M_2m + 1, \ldots, M_2(m + 1) - 1\}$, such that

$$\alpha_s \|s\|^2 \leq u_k^T s \leq \alpha_s (1 + \epsilon) \|s\|^2. \tag{7}$$

The coefficient $\alpha_s$ in (7) is defined as

$$\alpha_s \triangleq \alpha \cdot \sqrt{\frac{nQ}{\|s\|^2}}. \tag{8}$$

Let the set $\mathcal{T}(U|s)$ be defined as

$$\{u : \|u\|^2 = nW, \alpha_s \|s\|^2 \leq u^T s \leq \alpha_s (1 + \epsilon) \|s\|^2\} \tag{9}$$

The criterion (7) is equivalent to seeking a codeword $u_k$ whose empirical correlation coefficient with $s$, $\hat{\rho}(u_k, s) \triangleq u_k^T s/\|u_k\| \cdot \|s\|$, is nearly $\rho_0$. At this point, it is important to note that for fixed $s \in \mathbb{R}^n$ and for $U$ uniformly distributed over a hypersphere, the probability that the empirical correlation coefficient $\hat{\rho}(U, s)$ is around $\rho_0$ decays exponentially fast with an exponent given by $I_{US} = -\frac{1}{2} \ln (1 - \rho_0^2)$. Now, since the bin size, $M_2$, is exponentially larger than $e^{nI_{US}}$ (by $\Delta > 0$), the probability of not finding even one codeword in bin number $m$, which satisfies this condition, decays double exponentially with $n$. We therefore safely neglect this event of encoding failure. We define the encoder, more precisely, as follows: suppose that for the given $s$ and $m$ observed by the encoder, there are $K_n(s, m) \geq 1$ codewords within the bin $\mathcal{C}_m$ that satisfy (7). Then, the encoder randomly selects one of them under the uniform distribution, to be the transmitted codeword, $u_k$, i.e.,

$$\mathbb{P}(u_k = u|s, m) = \frac{1}{K_n(s, m)}, \quad \forall u \in \mathcal{C}_m \cap \mathcal{T}(U|s). \tag{10}$$

5Note that the optimal $\alpha$ here may not be the same as the optimal $\alpha$ that achieves the capacity [6].
Upon randomly selecting $u_k$, the encoder transmits
\begin{equation}
    x = u_k - \alpha_s \cdot s.
\end{equation}

Note that
\begin{align}
    \|x\|^2 &= \|u_k\|^2 - 2\alpha_s u_k^T s + \alpha_s^2 \|s\|^2 \\
    &\leq nW - 2\alpha_s^2 \|s\|^2 + \alpha_s^2 \|s\|^2 \\
    &= nW - n\alpha^2 Q \\
    &= nP,
\end{align}

and so, the power constraint is met.

The receiver implements the MMI decoder, which in the Gaussian-quadratic case considered here, amounts to seeking the codeword, $u_k$, $k \in \{0, 1, \ldots, M - 1\}$, that maximizes the squared (or, equivalently, the absolute) empirical correlation, $\hat{\rho}^2(u_k, y)$, i.e.,
\begin{equation}
    k = \arg \max_k \hat{\rho}^2(u_k, y) = \arg \max_k \frac{(u_k^T y)^2}{\|u_k\|^2 \cdot \|y\|^2} = \arg \max_k \frac{(u_k^T y)^2}{\|u_k\|^2 \cdot \|y\|^2},
\end{equation}

and finally, decoding $m$ according to the bin to which $u_k$ belongs, that is,
\begin{equation}
    \hat{m} = \left\lfloor \frac{k}{M_2} \right\rfloor.
\end{equation}

This two–stage decoding process is asymptotically optimal, at least in the random-coding sense, as it follows by a combination of two arguments. First, we argue by the end of Subsection V-A ahead that MMI decoding is asymptotically optimal, provided that one wants to fully decode $u_k$ (and not only the bin), and second, we prove in Appendix A that a simple ML decoder and the optimal bin index decoder are asymptotically equivalent, at least in the random coding sense.

For a given code, let $P_t(C_n)$ be the probability of error in the bin index decoding. The random coding error exponent is defined as
\begin{equation}
    E_t(R) = \lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E} \left[ P_t(C_n) \right],
\end{equation}

and the error exponent of the TRC is defined by
\begin{equation}
    E_m(R) = \lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E} \left[ P_{\text{trc}}(C_n) \right].
\end{equation}

Our first objective is to derive exact single-letter expressions for (16) and (17).

Another objective is to prove the existence of a sequence of codes $\mathcal{C} = \{C_n\}_{n=1}^{\infty}$, whose error exponent is strictly higher than $E_t(R)$ and $E_m(R)$, at least at low coding rates, and obtain a single–letter expression that lower bounds the following limit
\begin{equation}
    E_{\alpha}(\mathcal{C}) = \lim_{n \to \infty} \text{inf}_{m} \frac{1}{n} \ln \max_{m} P_{\text{trc}}(C_n),
\end{equation}

where $P_{\text{trc}}(C_n)$ is the conditional error probability, given that message $m$ was transmitted.

In this section, we analyze, study and discuss the random coding exponent, the typical code exponent and the expurgated exponent associated with this model, and we examine them, not only as functions of the rate $R$, but also as functions of $\alpha$ and $Q$. In particular, we show that the optimal value of $\alpha$ for rates below capacity may be different from the optimal $\alpha$ that is needed to achieve capacity, namely, $\alpha^* = P/(P + \sigma_s^2)$ [6, eq. (7)].

For the sake of convenience and simplicity, we initially assume that the random interference vector, $S$, is distributed uniformly across the surface of the $n$-dimensional hypersphere of radius $\sqrt{nQ}$ (that is, $\|S\|^2 = nQ$ with probability one) and return to the Gaussian model later (by weighing the various radii according to the Gaussian divergence between $Q$ and $Q$), as part of our main theorem in this section. Recall that the encoder depends on $S$ only via its empirical correlations with the various codewords, $\{u_k\}$, and these empirical correlations are scale–invariant. The decoder, on the other hand, has no access to $S$ anyway.

B. On the Various Error Exponents

A few words concerning the different error exponents studied in this paper are in order. The random coding error exponent defined in (16) provides the asymptotic decay rate of the average probability of error. While at relatively high coding rates, the random coding error exponent provides the error exponent of the best code, this is not the case at relatively low coding rates. At relatively low coding rates, the average error probability is being dominated by the relatively poor codes in the ensemble [21], thus $E_t(R)$ is not the error exponent of the typical code in the ensemble. Hence, we also define in (17) the error exponent of the TRC, which captures the true exponential behavior of the error probability of a randomly drawn codes. At least for DMCs, it has been proved in [35] that the exponential rate of decay of the error probability of a randomly chosen code concentrates around $E_m(R)$ at all coding rates. Still, the error exponent of the TRC describes the behavior of the normalized logarithmic error probability for a randomly drawn code, and we are also interested in the question of existence of exceptionally good codes, whose error exponents are strictly higher than the error exponent of the TRC, at least at relatively low coding rates. For this reason, we also define the error exponent of the expurgated code in (18), which is given as a function of a specific sequence of (good) codes (in comparison to the former two error exponents, which are related to expectations over the ensemble of codes). While these three error exponents have been extensively studied in various communication scenarios over the past few decades (e.g., [28] study these error exponents for DMCs and multiple-access channels), in the current work we broaden the scope of research and study them for the DPC and also for the Gel’fand–Pinsker channel.

One may also consider a more general encoding scheme, where $x$ is randomly selected from the conditional type class
\begin{equation}
    \left\{ x : \|x\|^2 = nP, \ x^T u = n\sqrt{P\rho_z u}, \ x^T s = n\sqrt{T\rho_s u} \right\},
\end{equation}

where the parameters $\rho_z u$ and $\rho_s u$ are subjected to optimization. Note that this more general coding is much closer in spirit to the coding strategy for the Gel’fand–Pinsker channel, as given in Subsection IV-A, where the transmitted codeword is drawn from the conditional type class given $u, s$.

Due to the fact that it coincides with the sphere packing bound, which serves as an upper bound on the reliability function.

In the random coding error exponent, we average the error probability directly, while in the TRC exponent, we average the logarithmic error probability.
C. Statement of the Main Results

Let $P$, $Q$, and $\alpha \in [0, 1]$ be given. Define the following quantities:

$$z = 1 - e^{-(R + I_{\text{trc}})} = 1 - \frac{Pe^{-2R}}{W} \tag{19}$$

$$a = \alpha \sqrt{Q} (\sqrt{Q - \alpha \sqrt{Q}}) + W, \tag{20}$$

and for two given reals, $\rho$ and $\varrho$, both in $(-1, 1)$, define also

$$b = \sqrt{W} (\sqrt{Q - \alpha \sqrt{Q}} \varrho + \rho W). \tag{21}$$

For two given positive reals, $p$ and $q (p \leq q)$, define

$$\mu(p, q) = 1 + (q - p)z \tag{22}$$

$$\Delta(p, q) = (\mu(p, q) + p) [\mu(p, q) - q] + \rho^2 pq. \tag{23}$$

Next, define

$$E(R, Q, \alpha, \rho, \varrho, p, q) = \frac{1}{2} \ln \mu(p, q) + \frac{1}{2 \sigma^2_\rho \mu(p, q)} [q - p)z [W + \hat{Q} - \alpha^2 Q] + \mu(p, q)(p(a^2 - \varrho^2) - pq(a^2 - 2pab + b^2)]}{W \cdot \Delta(p, q)} \tag{24}$$

$$E(R, Q, \alpha, \rho, \varrho) = \sup_{(p, q): 0 \leq p \leq q, \Delta(p, q) > 0} E(R, \hat{Q}, \alpha, \rho, \varrho, p, q, p). \tag{25}$$

Define also $\mathcal{P}(\rho, \rho_0)$ as the set of values of $\rho$ within $(-1, 1)$ such that the correlation matrix

$$\begin{pmatrix} 1 & \rho & \rho_0 \\ \rho & 1 & \varrho \\ \rho_0 & \varrho & 1 \end{pmatrix}$$

would be positive semi-definite, and let

$$E(R, \hat{Q}, \alpha, \rho) = \inf_{\varrho \in \mathcal{P}(\rho, \rho_0)} \left\{ E(R, \hat{Q}, \alpha, \rho, \varrho) + L(\rho, \varrho) \right\} + \frac{1}{2} \ln (1 - \rho^2) - \frac{1}{2} \ln \frac{W}{P}, \tag{27}$$

where

$$L(\rho, \varrho) = -\frac{1}{2} \ln (1 - \rho^2 - \varrho^2 - \rho_0^2 + 2\rho \rho_0). \tag{28}$$

Finally, letting

$$E_0(R, \hat{Q}, \alpha, \rho) \triangleq E(R, \hat{Q}, \alpha, \rho) + \frac{1}{2} \ln \frac{1 - \rho^2}{1 - \rho^2} - \frac{1}{2} \ln \frac{W}{P} - R$$

$$= \inf_{\varrho \in \mathcal{P}(\rho, \rho_0)} \left\{ E(R, \hat{Q}, \alpha, \rho, \varrho) + L(\rho, \varrho) \right\} - \ln \frac{W}{P} - R; \tag{29}$$

we have the following result, the proof of which is given in Section V.

Theorem 1: Consider the setting described in Subsection III-A. The error exponents $E_c(R)$ and $E_{w}(R)$ associated with dirty-paper coding and MMI decoding are given by:

$$E_c(R, \hat{Q}, \alpha) = \inf_{\rho \in (-1, 1)} E_0(R, \hat{Q}, \alpha, \rho), \tag{30}$$

$$E_{w}(R, \hat{Q}, \alpha, \rho) = \inf_{\rho^2 < 1 - (P/W)^2 e^{-2R}} E_0(R, \hat{Q}, \alpha, \rho), \tag{31}$$

respectively. In addition, there exists a sequence of codes, $\mathcal{C} = \{C_n\}_{n=1}^\infty$, such that the error exponent $E_c(\mathcal{C})$ associated with dirty-paper coding and MMI decoding is lower-bounded by:

$$E_c(R, \hat{Q}, \alpha) \leq \inf_{\rho^2 < 1 - (P/W)^2 e^{-2R}} E_0(R, \hat{Q}, \alpha, \rho). \tag{32}$$

If $S$ is distributed uniformly over the surface of a hypersphere of radius $\sqrt{n}Q$, take $Q = Q$ in all above expressions, and then maximize over the design parameter $\alpha \in [0, 1]$ for each error exponent, to obtain optimal performance. Alternatively, if $S \sim N(0, Q \cdot I_n)$, first minimize over $\hat{Q}$ one of the corresponding error exponents plus the additional term

$$D_o(\hat{Q}||Q) = \frac{1}{2} \left[ \frac{\hat{Q}}{Q} - \ln \left( \frac{\hat{Q}}{Q} \right) - 1 \right], \tag{33}$$

which is the KL divergence between two Gaussian distributions (with zero mean) and variances $Q, \hat{Q}$, and finally, maximize again over $\alpha$.

We can simplify the bounds significantly by confining the inner-most optimization (25) to $p = q$, corresponding to a relatively simple union-bound analysis of pairwise error probabilities, which is tight at some range of low rates. This is relevant since the interesting range where the TRC error exponent differs from the random coding error exponent is the range of low rates anyway. Accordingly, the choice $p = q$ yields:

$$E(R, \hat{Q}, \alpha, \rho, \varrho, q, q) = \frac{q(a^2 - b^2) - q^2(a^2 - 2pab + b^2)}{2W \sigma^2_q (1 - q^2(1 - \rho^2))} \tag{34}$$

to be maximized over $q \in [0, 1/\sqrt{1 - \rho^2}]$ which can be carried out explicitly, yielding

$$E(R, \hat{Q}, \alpha, \rho, \varrho) \geq \frac{1}{2W \sigma^2_q \sqrt{1 - \rho^2}} \cdot T \left( a^2 - b^2; \frac{a^2 - 2pab + b^2}{\sqrt{1 - \rho^2}} \right), \tag{35}$$

where the function $T(\cdot, \cdot)$ is defined as

$$T(g, h) \triangleq \sup_{0 \leq \tau < 1} g\tau - h \tau^2 \left( \frac{1}{1 - \tau^2} \right) \begin{cases} g \leq 0 \\
\frac{g^2}{2(h + \sqrt{h^2 - g^2})} & h \geq 0 \\\n\infty & g \geq 0 \end{cases} \tag{36}$$

The random coding exponent, $E_c(R)$, the typical–code error exponent, $E_w(R)$, and the expurgated error exponent, $E_{w,a}(R)$, are presented in Figure 1 for the DPC with the parameters $P = 10$ and $Q = \sigma^2_{Q} = 1$, and with optimized $\alpha$ in the interval $[0, 1]$. As can be seen in Figure 1, $E_{w,a}(R)$ and $E_{w}(R)$ coincide at $R = 0$, but $E_{w}(R) < E_{w,a}(R)$ at relatively low positive coding rates. At relatively high rates, the three exponent functions coincide. Figure 2 presents zero-rate error exponents as a function of $Q$ for the DPC with the parameters $P = 10$ and $\sigma^2_{Q} = 1$, and with optimized $\alpha$ in the interval $[0, 1]$. The two exponents $E_{w,a}(0)$ and $E_{w}(0)$ coincide. Note that the random coding exponent is almost a constant, independently of
The probability law of the DMC is specified by $P$ and memoryless with known probability mass function also different from the optimal, capacity–achieving alphabet. For a given type $Q$ where $T_{\text{class}} Q$ for any conditional type (to be optimized for every $X$ with finite input, state and output alphabets is $\alpha Q$ as a function of $\alpha$ might be positive. Figure 3 presents zero–rate error exponents $Q = X + 10 P$. Fig. 1. Low–rate error exponent functions for the DPC with the parameters, $P = 10$, $Q = \sigma^2_2 = 1$, and with optimized $\alpha$ in the interval $[0, 1]$.

$Q$, in agreement with [15]. It turns out that for larger values of $Q$, $E_1(0)$ is slightly increasing with $Q$, which may seem counter–intuitive. However, note that for $Q > 0$, even if we select $\alpha = 1$ (which is sub-optimal), we completely cancel the interference (yielding $Y = (U – S) + S + Z = U + Z$, yet the effective transmission power (the power of $U$) is $P + Q$ (rather than $P$), although the transmission power of $X = U – S$ is still $P$. On the other hand, there is the effective rate reduction of $I(U; S) = 1/2 \ln(1 + 2P)$ (which does not exist when $Q = 0$). Still, the overall effect of increasing $Q$ might be positive. Figure 3 presents zero–rate error exponents as a function of $\alpha$ for the parameters $P = 10$ and $Q = \sigma^2_2 = 1$. The optimal values of $\alpha$ are different from each other: $\alpha^*_Q \approx 0.90$, $\alpha^*_R = \alpha^*_S \approx 0.38$. The latter two are also different from the optimal, capacity–achieving $\alpha$, which is $\alpha^*_{\text{capacity}} = P/(P + \sigma^2_2) = 10/11 = 0.9091$.

IV. ERROR EXPONENTS OF THE GEL’FAND–PINSKER CHANNEL

A. Setting and Objectives

Consider a state dependent DMC $W : \mathcal{X} \times \mathcal{S} \rightarrow \mathcal{Y}$ with finite input, state and output alphabets $\mathcal{X}$, $\mathcal{S}$ and $\mathcal{Y}$, respectively. The $S$-valued state process $\{S_t\}_{t=1}^\infty$ is stationary and memoryless with known probability mass function $P_S$. The probability law of the DMC is specified by

$$W(y|x, s) = \prod_{i=1}^n W(y_i|x_i, s_i),$$

where $x \in \mathcal{X}^n$, $s \in \mathcal{S}^n$, and $y \in \mathcal{Y}^n$.

Let the coding rate $R$ be given. Let $\mathcal{U}$ be a finite auxiliary alphabet. For a given type $Q_S \in \mathcal{P}_n(S)$, let $Q_{U|S}(Q_S)$ be any conditional type (to be optimized for every $Q_S$). Furthermore, for any $Q_S$, let $Q_{X|SU}(Q_S)$ be another conditional type (to be optimized as well). Let $\varepsilon > 0$ be given. For $Q = Q_{U|S} \times Q_S$, let $R(Q_S) = I_Q(S; U) + \varepsilon$. For any $Q_S \in \mathcal{P}_n(S)$, draw $\epsilon^N(R + R(Q_S))$ independent $n$-vectors, uniformly over the type class $T(Q_U)$, where $Q_U$ is the $U$-marginal of $Q_{U|S} \times Q_S$. Partition these codewords into $M = \epsilon^N(M)$ bins, each one of size $M(Q_S) = \epsilon^N(R(Q_S))$. Let us denote those bins as

$$B(Q_S, m) = \{u_{Q_S, m, 1}, u_{Q_S, m, 2}, \ldots, u_{Q_S, m, M(Q_S)}\},$$

where $Q_S \in \mathcal{P}_n(S)$ and $m \in \{1, 2, \ldots, M\}$. Note that this random-binning code construction is similar to the one in Subsection III-A, but here, we allow the binning rate to depend on the SI type, thus allowing for more degrees of freedom to improve the overall performance.

For a given message $m \in \{1, 2, \ldots, M\}$ and a state vector $s \in \mathcal{S}^n$, we choose the vector $u$ from the set $C(m, s) = B(P_s, m) \cap T(Q_{U|S}|s)$ with equal probabilities. The probability that $C(m, s)$ is empty decays double-exponentially fast, hence we neglect this possible error event. For a given $s \in \mathcal{S}^n$ and a selected vector $u$, we draw the codeword $x$ according to the uniform distribution over the conditional type class $T(Q_{X|SU}|s, u)$ and transmit it over the channel.

This fact follows from Lemma 1 in Subsection V-A.
Following the same rationale as in Subsection III-A, also here we implement a sub-optimal two-stage decoder, which first decodes for the \( u \) codeword according to some metric that depends on the joint empirical distribution with the received vector \( y \), and afterwards, provides only the bin index that corresponds to the decoded \( u \). More precisely, the decoder observes the received vector \( y \) and decodes for a bin index using the following generalized deterministic decoder:

\[
\hat{m}(y) = \arg\max_{m \in \{1,2,\ldots,M\}} \left\{ \max_{Q_S \in \mathcal{P}_n(S)} \max_{u \in \mathcal{B}(Q_S,m)} G(\hat{P}_{uy}, Q_S) \right\},
\]

(39)

where \( \hat{P}_{uy} \) is the empirical distribution of \((u, y)\), and \( G(\cdot, \cdot) \) is a given continuous, real-valued functional of this empirical distribution and of another distribution \( Q_S \in \mathcal{P}_n(S) \).

For a given codebook, the probability of error is given by \( P_e(C_n) \)

\[
P_e(C_n) = \frac{1}{M} \sum_{m=1}^{M} \sum_{x \in S^n} P(s) \sum_{y \in Y^n} W(y|x, s) \cdot \mathbb{1}\{\hat{m}(y) \neq m\}.
\]

(40)

Our objectives here are identical to those defined in Subsection III-A.

B. Main Results

In order to present the random coding error exponent, define the exponent function \( E_c(R) \)

\[
E_c(R) = \min_{Q_S} \max_{Q_Y} \min_{Q_{UX} X} D(Q_{UXS} || P_{UXS}) + D(Q_Y | SUX) || W_Y \{X|S \times Q_S \} + [I_Q(U; Y) - I_Q(U; S) - R]^+.
\]

(41)

Then, we have the following result, which is proved in Appendix C.

**Theorem 2:** Consider the setting described in Subsection IV-A.

1) It holds that

\[
\lim_{n \to \infty} - \frac{1}{n} \ln \mathbb{E}[P_e(C_n)] = E_c(R).
\]

(42)

2) The universal decoder

\[
\hat{m} = \arg\max_{m \in \{1,2,\ldots,M\}} \max_{Q_S \in \mathcal{P}_n(S)} \max_{u \in \mathcal{B}(Q_S,m)} [I_{uy}(U; Y) - R(Q_S)]
\]

(43)

achieves \( E_c(R) \).

**Discussion** An expression similar to (41) can be found in two previous works. First in [30], where the coding technique is slightly different than ours and composed by two steps. First, the empirical distribution of the state sequence is transmitted to the receiver, and only then, the message is encoded similarly as in Subsection IV-A, by using a codebook, which is optimally designed for the empirical statistics of \( s \). Since the decoder knows in which codebook to look at, it uses an ordinary MMI decoder to decode for a bin index. The final expression in [30] is given by the minimum between (41) and another expression, which is related to the first decoding phase of transmitting the statistics of \( s \). Hence, our exponential error bound is at least as tight as the one in [30]. The expression in (41) appears also in [27], where a similar codebook generation has already been encountered. Although the universal decoder (43) has already been proposed in [27], its optimality has not been proved before.

We continue by presenting a numerical example of the exact error exponent given in Theorem 2. Let \( S = U = X = \mathcal{Y} = \{0, 1\} \) and \( \mathbb{P}(S = 0) = 1 - \mathbb{P}(S = 1) = p \in [0, 1] \). It is important to note that choosing \( |U| = 2 \) may not be optimal, but we make this choice in order to keep the computational complexity relatively low. Regarding the state-dependent channel, consider the following. If \( S = 0 \), then \( Y = X \) with probability one, i.e., the channel is clean. Otherwise, if \( S = 1 \), the channel is stuck at 0: \( \mathbb{P}(Y = 0|X = 0) = 1 - \mathbb{P}(Y = 1|X = 0) = 1 \) and \( \mathbb{P}(Y = 0|X = 1) = 1 - \mathbb{P}(Y = 1|X = 1) = 1 \). The capacity of this channel is given by

\[
C_{ap} = \max_{Q_{UX}, Q_{Y,S}} [I(U; Y) - I(U; S)] = p \quad \text{[bits]},
\]

(44)

which follows as a minor modification of [8, pp. 178-180, Example 7.3].

We plot the exponent function itself in Figure 4 and find that its functional behavior is similar to that of the random coding error exponent in a DMC without random states; it decreases in an affine fashion for relatively low coding rates, and in a strictly convex fashion at the high coding rates. The reliability of this state-dependent channel obviously depends on the probability of \( S = 0 \); the higher this probability is, the percentage of time the channel is clean is bigger, and the channel is more reliable, although the decoder is ignorant of the state realization. As can be seen in Figure 4, where we compare between \( p = 0.7 \), \( p = 0.5 \), and \( p = 0.3 \), the reliability of the channel is highest at \( p = 0.7 \) and lowest at \( p = 0.3 \).
Following the studies in [3], [21], and [28] on TRCs in ordinary channel coding, we claim that also in the Gel’fand–Pinsker channel, the random coding error exponent, whose exact value is given in Theorem 2, does not yield the true exponential behavior of the error probability of a randomly chosen code, since it is dominated by the relatively bad codes in the ensemble, rather than the channel noise, at least at low coding rates. Due to the definition of the TRC exponent, the derivation of a single-letter expression is not as easy as in ordinary random coding (for example, see the proof in [21, Section 5]), since the expectations over the randomness of the ensemble and over the randomness of the channel cannot be switched, which is one of the first steps in random coding analysis. Let us choose the universal decoding metric \( g(Q_{U|X}, Q_S) = I_Q(U;Y) - R(Q_S) \). In order to present a lower bound on the error exponent of the TRC, define the expressions

\[
E_0(Q_{UU^*|XS}, Q_S) = \min_{\{Q_{UU^*|XS} : g(Q_{UU^*|XS}) \geq g(Q_{U|X}, Q_S)\}} D(Q_{Y|XS} \| W_{Y|XS} (Q_{UU^*|XS})),
\]

\[
E_1(Q_{UU^*S}, Q_S^r) = \min_{\{Q_{UU^*S} : Q_{XS} = Q_{XS}\}} [I_Q(U^*;X|US) + E_0(Q_{UU^*S} \times \hat{Q}_{UX} | Q_{UU^*S}, Q_S^r)],
\]

and,

\[
E_2(Q_{UU^*S}, Q_S^r) = \min_{\{Q_{UU^*S} : Q_S = Q_S\}} [I_Q(S;U^*|U) + E_1(Q_{UU^*} \times \hat{Q}_{S|UU^*} | Q_S^r)].
\]

Denote the set \( \mathcal{Q}(Q_U, Q_{U^*}) = \{ \hat{Q}_{U^*} : \hat{Q}_U = Q_U, \hat{Q}_{U^*} = Q_{U^*} \} \) and define the exponent function

\[
E_m(R) = \max_{\{Q_{UXS}\}} \min_{Q_S\neq Q_{S'}} \min_{\{Q_{UU^*S} : I_Q(U^*;U) \leq 2R + R(Q_{S}) + R(Q_{S'})\}} [D(Q_S \| P_S) + E_2(\hat{Q}_{UU^*}, Q_{S'}) + I_Q(U;U') - R - R(Q_{S'})].
\]

Then, our second result is the following theorem, which is proved in Appendix D.

**Theorem 3:** Consider the setting described in Subsection IV-A. It holds that

\[
\lim_{n \to \infty} -\frac{1}{n} \mathbb{E} \ln P_i(C_n) \geq E_m(R).
\]

Since our analysis in Appendix D amounts only to pairwise error events, the resulting exponent function is only a lower bound in general and is not tight at relatively high coding rates. Note that at high coding rates, the random coding error exponent of (41) provides the exact value of the error exponent of the TRC.

In ordinary channel coding, the random coding error exponent, and the error exponent of the TRC are both improved at relatively low coding rates by code expurgation. As we already seen in Figure 1 above, this fact is also true for the DPC. Upon using similar techniques as in [34, Appendix A], which is an error exponent derivation of an expurgated code in ordinary channel coding, we are able to derive a bound which is tighter than \( E(R) \) and \( E_m(R) \), at least at low coding rates. Let us define the exponent function

\[
E_n(R) = \max_{\{Q_{UXS}\}} \min_{Q_S, Q_{S'}} \min_{\{Q_{UU^*S} : I_Q(U^*;U') \leq R + R(Q_S) + R(Q_{S'})\}} [D(Q_S \| P_S) + E_2(\hat{Q}_{UU^*}, Q_{S'}) + I_Q(U;U') - R - R(Q_{S'})].
\]

Then, our third result is the following theorem, which is proved in Appendix E.

**Theorem 4:** Consider the setting described in Subsection IV-A. There exists a sequence of constant composition codes, \( \{C_n, n = 1, 2, \ldots\} \), such that

\[
\liminf_{n \to \infty} -\frac{1}{n} \ln \max_m P_{i,m}(C_n) \geq E_n(R).
\]

V. PROOF OF THEOREM 1

A. Preparatory Steps

Before moving on to the actual derivations of the various error exponent functions, two basic preparatory steps will prove useful. The first is associated with certain helpful properties that a randomly chosen code in the ensemble satisfies with an overwhelmingly high probability, and the second is associated with the structure of the joint distribution of the various variables in the coded communication system under discussion.

1) Two Properties of Good Codes: Recall the following definitions. First, for a given \( m \in \{0, 1, \ldots, M - 1\} \) and \( s \in \mathcal{S}(\sqrt{n}Q) \), we have defined \( K_n(s, m) \) to be

\[
K_n(s, m) = |C_m \cap \mathcal{T}(U|s)|,
\]

where \( \mathcal{T}(U|s) \) is defined by

\[
\{ u : \| u \|_2 = nW, \alpha_s\| s \|^2 \leq u^T s \leq \alpha_s(1 + \epsilon)\| s \|^2 \}.
\]

Second, for a given \( u \in \mathbb{R}^n \) and \( y \in \mathbb{R}^n \), the empirical correlation coefficient between \( u \) and \( y \) is defined by

\[
\hat{\rho}(u, y) = \frac{u^T y}{\| u \| \cdot \| y \|}.
\]

There are two desired properties that are associated with a randomly chosen code in our ensemble, with an extremely high probability, and we henceforth refer to codes with those properties as good codes. Roughly speaking, a good code has the following two properties at the same time: (i) \( K_n(s, m) \) is close to \( e^{n\Delta} \) for all \( s \) and \( m \), and (ii) For every given \( y \), there exists at least one codeword, \( u_k \), such that \( \hat{\rho}^2(u_k, y) \) is essentially at least as large as \( 1 - \exp[-2(R + I_{us})] \). These two features will be pivotal in our analysis.

To state properties (i) and (ii) more formally, we assert the following lemma, whose proofs appear in Appendix B.

**Lemma 1:** Let \( \Delta > 0 \) be an arbitrarily small constant and denote by \( \mathcal{Q}_n \) the ensemble of codes with \( M = e^{n(R + I_{us}) + \Delta} \) codewords. Let \( \hat{Q} > 0 \) be fixed as well. Let \( \epsilon > 0 \) be...
arbitrarily small and let $G_0^n \subseteq \mathcal{C}^n$ be the subset of all codes such that
\[
e^{-n\Delta} \leq K_u(s, m) \leq e^{n(\Delta + \epsilon)} \tag{55}\]
for all $m \in \{0, 1, \ldots, M - 1\}$ and $s \in \mathcal{S}(\sqrt{nQ}) \triangleq \{s : \|s\|^2 = nQ\}$. Then, a randomly selected code falls in $G_0^n$ with probability that tends to double-exponentially as a function of $n$.

**Lemma 2**: Let $\epsilon > 0$ be arbitrarily small and let $G^n_2$ be the subset of all codes such that for all $k, k' \in \{0, 1, \ldots, M - 1\}$, $k' \neq k$,
\[
\max_{k \neq k'} P^2(u_k, y) \geq 1 - \exp[-2(R + I_{os} - \epsilon)], \tag{56}\]
for all $y \in \mathbb{R}^n$. Then, a randomly selected code falls in $G^n_2$ with probability that tends to unity double-exponentially as a function of $n$.

Finally, a good code is in $G^n = G^n_0 \cap G^n_2$, which still holds a probability that tends to unity double-exponentially rapidly as $n$ tends to infinity.

2) **On the Structure of the Joint Probability Distribution of $(k, s, y)$**: Since our decoder, first decodes the index $k$ of the auxiliary codeword, $u_k (0 \leq k \leq M - 1)$ and then decodes the message to be the index $m$ of the bin to which $k$ belongs (see Equations (14), (15)), an obvious upper bound to the probability of error in estimating $m$ is the probability of error in estimating $k$, since the latter also counts errors associated with incorrect values of $k$ that still belong to the correct bin, $m$. But the gap is negligible since the number of incorrect codewords in the correct bin is only $M_2 - 1$, which is of a smaller exponential order than $M - 1$. We therefore henceforth approach the error probability analysis in the same way as if we had to assess the probability of error in decoding $k$. To this end, it is instrumental to characterize the joint distribution of $k$ with other entities that play a role in the coded communication system, namely, $s$ and $y$.

We first observe that for a good code, the probability distribution of the index $k$ is essentially uniform. To see why this is true, consider first the probability of $k$ given the message $m$, where $m = \lfloor k/M_2 \rfloor$,
\[
P(k|m) = \int_{\mathbb{R}^n} P(s) \cdot e^{-n\Delta} \mathbb{1}_{\{u_k \in C_m \cap T(U|s)\}} \text{ ds} = \int_{T(S|u_k)} P(s) e^{-n\Delta} \text{ ds} = e^{-n\Delta} \int_{T(S|u_k)} \frac{\text{ ds}}{\text{Vol}(T(S|\sqrt{nQ}))}, \tag{57}\]
where $T(S|u_k) = \{s : u_k \in T(U|s)\}$. Similarly as in the proof of Lemma 2, the last integral is easily seen to be of the exponential order of $\exp \left[\frac{\alpha}{2} \ln(1 - \rho_0^2)\right] = e^{-nI_{us}}$, and so,
\[
P(k|m) = e^{-n(I_{us} + \Delta)}, \quad \text{whenever } m = \left\lfloor \frac{k}{M_2} \right\rfloor \tag{58}\]
and therefore, for any $k = 0, 1, \ldots, M - 1$,
\[
P(k) = \mathbb{P} \left\{ m = \left\lfloor \frac{k}{M_2} \right\rfloor \right\} \cdot P(k|m) = \frac{1}{M_1} \cdot \frac{1}{M_2} = \frac{1}{M}, \tag{59}\]
where we have assumed that the prior distribution over the various messages, $\{m\}$, is uniform.

For a good code, the joint distribution of $(k, s, y)$ is given by
\[
P(k, s, y) = P(k) P(s|u_k) P(y|u_k, s), \tag{60}\]
where $P(y|u_k, s) = N (u_k + (1 - \alpha_k) s, \sigma_s^2 I_n)$ is the additive white Gaussian noise channel, with input $x + s = u_k + (1 - \alpha_k) s$ and noise variance, $\sigma_s^2$.

\[
P(s|u_k) = \frac{P(s) \cdot P(u_k|s)}{\int_{\mathbb{R}^n} P(s') \cdot P(u_k|s') \text{ ds'}} = \frac{P(s) \cdot e^{-n\Delta} \mathbb{1}_{\{u_k \in C_m \cap T(U|s')\}}}{\int_{\mathbb{R}^n} P(s') e^{-n\Delta} \mathbb{1}_{\{u_k \in C_m \cap T(U|s')\}} \text{ ds'}} = \frac{1}{\int_{T(S|u_k)} \frac{1}{\text{Vol}(T(S|\sqrt{nQ}))} \text{ ds'}} = \frac{1}{\text{Vol}(T(S|u_k))}, \tag{61}\]
where $\text{Vol}(T(S|u_k))$ is the volume of $T(S|u_k)$.

As a consequence of these relations, we can now present the error probability (in $k$) as follows.
\[
P(C) = \frac{1}{M} \sum_{k=0}^{M-1} \int_{T(S|u_k)} \frac{\text{ ds}}{\text{Vol}(T(S|u_k))} \int_{\mathcal{N}_k^1} P(y|u_k, s) \text{ dy} = \frac{1}{M} \sum_{k=0}^{M-1} \int_{\mathcal{N}_k^1} \left[ \int_{T(S|u_k)} \frac{P(y|u_k, s)}{\text{Vol}(T(S|u_k))} \text{ ds} \right] \text{ dy} = \frac{1}{M} \sum_{k=0}^{M-1} \int_{\mathcal{N}_k^1} P(y|u_k) \text{ dy}, \tag{62}\]
where $\mathcal{N}_k^1$ is the complement to the decision region in favor of $u_k$. This is an ordinary expression of the probability of error, associated with a codebook $C = \{u_k\}$ over the effective channel
\[
P(y|u_k) = \int_{T(S|u_k)} \frac{P(y|u_k, s) \text{ ds}}{\text{Vol}(T(S|u_k))}, \tag{63}\]
According to this effective channel, $y$ depends on $u_k$ only via the empirical correlation coefficient, $\hat{\rho}(u_k, y)$. Therefore, the optimal decoding metric also depends on $(u_k, y)$ only via $\hat{\rho}(u_k, y)$ (in other words, $\hat{\rho}(u_k, y)$ serves as sufficient statistics). It follows from the results of [18] (see in particular, Example 2 therein) that at least for the random coding error exponent, the MMI decoder is asymptotically optimal in the sense of achieving the same error exponent as the maximum likelihood decoder.

**B. Derivation**

Considering the MMI decoder, which is equivalent to the maximum absolute correlation decoder, that chooses the index
$k$ with maximum $|u_k^T y|$, we have:

$$P(C_n) \leq \frac{1}{M} \sum_{k=0}^{M-1} \sum_{k \neq k'} \int_{R^n} P(s) ds \int_{R^n} P(y | u_k, s) \times \left\{ |u_k^T y| \geq \max \{ |u_k^T y|, |u_k^T y| \} \right\} dy. \tag{64}$$

Our good codes (in $G_n$) satisfy

$$\max_{k \neq k'} |u_k^T y| \geq \sqrt{n(P + \sigma^2 Q)} \|y\|^2 \sqrt{1 - e^{-2(R_1 - \epsilon)}}, \tag{65}$$

for every $k, k', y$, where $R_1 = R + I_{30}$. Therefore, we can safely further bound the error probability of every $C_n \in G_n$ according to

$$P(C_n) \leq \frac{1}{M} \sum_{k=0}^{M-1} \sum_{k \neq k'} \int_{R^n} P(s) ds \int_{R^n} P(y | u_k, s) \times \left\{ |u_k^T y| \geq \max \{ |u_k^T y|, |u_k^T y| \} \right\} dy \tag{66}$$

$$= \frac{1}{M} \sum_{k=0}^{M-1} \sum_{k \neq k'} \int_{R^n} P(y | u_k, s) \times \left\{ |u_k^T y| \geq \max \{ |u_k^T y|, |u_k^T y| \} \right\} dy \tag{67}$$

where (67) is due to the simple identity $\{a \geq \max\{b, c\} = \{a \geq b, a \geq c\}$. For the sake of notational simplicity, we henceforth use the following alternative notation: $u = u_k, v = u_k', and w = u + (1 - \alpha) s$. Now,

$$I(s) \triangleq \frac{(2\pi \sigma^2)^{-n/2}}{2} \int_{R^n} dy \cdot \exp \left\{ -\frac{1}{2\sigma^2} \|y - w\|^2 \right\} \times \left\{ |u^T y| \geq |u^T y|, |v^T y| \geq \sqrt{nW \|y\|^2} \right\} \tag{68}$$

$$\leq \exp\left\{ -\frac{\|w\|^2/2\sigma^2}{(2\pi \sigma^2)^n/2} \right\} \frac{\Gamma}{2\sigma^2} \exp \left\{ \frac{t}{2} y^T (vW - uu^T) y + \frac{1}{2} y^T (v^T - nW \cdot I_n) y \right\} \tag{69}$$

$$= \exp\left\{ -\frac{\|w\|^2/2\sigma^2}{(2\pi \sigma^2)^n/2} \right\} \frac{\Gamma}{2\sigma^2} \exp \left\{ \frac{t}{2} y^T (vW - uu^T) y + \frac{1}{2} y^T \Gamma y \right\}, \tag{70}$$

where (68) follows from the definitions of $W$ and $\zeta$, $\theta > 0$ and $\zeta > 0$ in (69) are Chernoff parameters, and $\Gamma$ in (70) is defined by

$$\Gamma = \left( \frac{1}{\sigma^2} + \zeta nW \right) \cdot I_n + \theta uu^T T - (\theta + \zeta) vW T$$

$$= \frac{1}{\sigma^2} \left[ (1 + \zeta n^2 zW) \cdot I_n + \sigma^2 \theta uu^T T - \sigma^2 (\theta + \zeta) vW T \right]. \tag{71}$$

We now use the following identity

$$\frac{1}{(2\pi)^n/2 |\Lambda|^{1/2}} \int_{R^n} dy \exp \left\{ y^T \omega - \frac{1}{2} y^T \Lambda^{-1} y \right\} = \exp \left\{ \frac{1}{2} \omega^T \Lambda \omega \right\}, \tag{72}$$

and so, continuing from (70),

$$I(s) \leq \exp\left\{ -\frac{\|w\|^2/2\sigma^2}{(2\pi \sigma^2)^n/2} \right\} \frac{\Gamma}{2\sigma^2} \exp \left\{ \frac{1}{2} \omega^T \Gamma \omega \right\} \tag{73}$$

$$= \exp\left\{ -\frac{\|w\|^2/2\sigma^2}{(2\pi \sigma^2)^n/2} \right\} \frac{\Gamma}{2\sigma^2} \exp \left\{ \frac{1}{2} \omega^T (1 + \zeta n^2 zW) I_n + \sigma^2 \theta uu^T T - \sigma^2 (\theta + \zeta) vW T \right\} \tag{74}$$

$$= \exp\left\{ -\frac{\|w\|^2/2\sigma^2}{(2\pi \sigma^2)^n/2} \right\} \frac{\Gamma}{2\sigma^2} \exp \left\{ \frac{1}{2} \omega^T (1 + \zeta n^2 zW) I_n - \sigma^2 (\theta + \zeta) vW T \right\} \tag{75}$$

where (75) follows by substituting (71), and in (76) we have denoted $\mu = 1 + \zeta n^2 zW, t = \sigma^2 \theta, and \sigma = \sigma^2 (\theta + \zeta)$. We now have to find both the inverse and the determinant of $\mu I_n + t uu^T - rvv^T$, where we use the fact that $u^T u = v^T v = nW$ and $u^T v = nW$. Beginning from the inverse, we invoke the matrix inversion lemma, asserting that for given matrices, $A, B, C$, and $D$, of dimensions, $n \times n, n \times k, k \times k$, and $k \times n$, respectively ($k \leq n$),

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}. \tag{77}$$

Setting $k = 2, A = \mu I_n, B = [tu \ rv], C = I_2, D = [u - v]^T, p = nWt and q = nWr$, we arrive at

$$\langle \mu I_n + t uu^T - rvv^T \rangle^{-1}$$

$$= \frac{1}{\mu I_n}$$

$$- t(\mu - q) uu^T T - r(\mu + p) vv^T T + \frac{pq}{{\mu}(\mu - q) + \rho^2 pq}$$

$$- \frac{1}{\mu I_n}$$

$$- \frac{t uu^T T - rvv^T T - \rho(uu^T T + vv^T T)}{{\mu}(\mu + p)(\mu - q) + \rho^2 pq}, \tag{78}$$

where we have used the fact that $pr = qt$. To find the determinant of $\mu I_n + t uu^T - rvv^T$, we find the eigenvalues and take their product. First, observe that all $n - 2$ linearly independent vectors that are orthogonal to both $u$ and $v$ are eigenvectors pertaining to the eigenvalue $\mu$. The two remaining eigenvalues correspond to two linearly independent vectors that lie in the subspace spanned by $u$ and
v. Let us denote \( r = au + bv \). Then for \( r \) to be an eigenvector,
\[
(\mu I_n + t\mu u^T - r vv^T)r
= (\mu I_n + t\mu u^T - r vv^T)(au + bv)
= \mu au + \mu bv + ap\mu + bopp\mu - bpq = [\mu a + p(a + \rho b)]u + [\mu b - q(\rho a + b)]v.
\]

The resulting eigenvalues are therefore the eigenvalues of the \( 2 \times 2 \) matrix
\[
\begin{pmatrix}
\mu + p & pp \\
-pq & \mu - q
\end{pmatrix},
\]
which are readily found to be \( \lambda = \mu + (p - q)/2 \pm \sqrt{(p + q)^2 - 4\rho^2pq}/2 \). Thus,
\[
|\mu I_n + t\mu u^T - r vv^T|^{1/2} = \mu^{n/2 - 1} \sqrt{\mu^2 + \mu(p - q) - pq(1 - \rho^2)}.
\]

Now, substituting (79) and (84) back into (76) yields that
\[
I(s) \leq \exp \left\{ - \frac{1}{2} \ln[1 + (q - p)z] + \frac{1}{2\sigma_2^2[1 + (q - p)z]} \right[ (q - p)zW + (1 - \alpha^2)Q
\]
\[+ \frac{1}{2\sigma_2^2[1 + (q - p)z]} \left[ (q - p)(pa^2 - qB^2) - pq(\alpha^2 - 2\rho ab + b^2) \right] + \frac{1}{nW}(1 + p(1 - z) + qz) \right] \left[ 1 - p(1 - z) + qz \right] + \rho^2pq) \right) \}
\]
\[
\triangleq e^{-nE(p,q,\hat{\sigma}_s^2,\rho,\varepsilon)},
\]

where the dependence of \( E(p,q,\hat{\sigma}_s^2,\rho,\varepsilon) \) on \( \varepsilon \) is through \( b \), and where \( E(p,q,\hat{\sigma}_s^2,\rho,\varepsilon) \) should be maximized over the set \( A = \{(p,q): q \geq p \geq 0, \}
\[1 + p(1 - z) + qz][1 - p(1 - z)] + \rho^2pq > 0 \}
\]
resulting in
\[
E(\hat{\sigma}_s^2,\rho,\varepsilon) = \sup_{(p,q) \in A} E(p,q,\hat{\sigma}_s^2,\rho,\varepsilon).
\]

As for the simpler bound in (34)-(36), we now choose \( q = p \) (amounting to simple union-bound analysis), which simplifies to
\[
E(q,q,\hat{\sigma}_s^2,\rho,\varepsilon) = \frac{q(a^2 - b^2) - q\hat{\sigma}_s^2(2\rho ab + b^2)}{2\sigma_2^2W[1 - q^2(1 - \rho^2)]},
\]

and the supremum over \( q \) should be taken in the range \( q \in [0,1/\sqrt{1 - \rho^2}] \). Alternatively, defining \( \tau = q\sqrt{1 - \rho^2} \), then (101) is equivalent to
\[
E(\tau,\hat{\sigma}_s^2,\rho,\varepsilon) = \frac{1}{2\sigma_2^2W\sqrt{1 - \rho^2}} \left[ g\tau - h\tau^2 \right],
\]
which should be maximized over \( \tau \in [0,1] \), where \( g = a^2 - b^2 \) and
\[
h = \frac{a^2 - 2\rho ab + b^2}{\sqrt{1 - \rho^2}}.
\]

Then,
\[
T(g,h) \triangleq \sup_{0 \leq \tau \leq 1} \frac{g\tau - h\tau^2}{1 - \tau^2}
\]
\[
= \begin{cases} 
0 & g \leq 0 \\
\frac{0}{2(h + \sqrt{h^2 - g^2})} & h \geq g > 0 \\
\infty & g > h > 0.
\end{cases}
\]

It follows that
\[
I(s) \leq \exp \left\{ - \frac{n}{2W\sigma_2^2\sqrt{1 - \rho^2}} \cdot T \left( a^2 - b^2, \frac{a^2 - 2\rho ab + b^2}{\sqrt{1 - \rho^2}} \right) \right\},
\]

Note that for a given \( \hat{Q} = \|s\|^2/n \), and a given correlation coefficient, \( \rho \), between \( u \) and \( v \), the variable \( a \) is a deterministic constant and \( b \) depends only on the empirical correlation
coefficient, $\varrho$. Let us then denote the simplified exponent for given $\hat{\sigma}_s^2$, $\rho$, and $\varrho$, by

$$E_{aw}(\hat{\sigma}_s^2, \rho, \varrho) = \frac{1}{2W}\frac{1}{\sigma_s^2} T_{\sigma_s^2/1} - \frac{1}{2} \ln \left(1 + \frac{\alpha^2Q}{P}\right) - R$$

(106)

The probability that $s$ would have an empirical correlation coefficient, $\varrho$, with $\nu$ given that it has empirical correlation $\rho_0$ with $u$ is exponentially

$$P(T(s|u,v)|s \in T(s|u)) = \frac{\text{Vol}\{T(s|u,v)\}}{\text{Vol}\{T(s|u)\}} = \frac{(2\pi \sigma_s^2)^n}{(2\pi \epsilon \sigma_s^2)^n} e^{-nZ(\rho, \varrho)}$$

(107)

The probability of the event $\{\nu_1^2 u_k^2/(\nu \nu') \approx \rho_{uu'}\}$ is about $e^{2n(\rho + I_{\rho\nu})}$ codeword pairs in $c_n$ and since the probability of the event $\{\nu_1^2 u_k^2/(\nu \nu') \approx \rho_{uu'}\}$ is about $e^{2n(\rho + I_{\rho\nu})}$, the typical value of the number $M(\rho_{uu'})$ is of the exponential order of $\exp[2n(\rho + I_{\rho\nu})^2]$, the typical value of the number $M(\rho_{uu'})$ is of the exponential order of $\exp[2n(\rho + I_{\rho\nu})^2]$, which is of the exponential order of $\exp[2n(\rho + I_{\rho\nu})^2]$, and is zero otherwise. More details on this point can be found in [21] and [22]. One can arrive at the constraint in (121) by performing a standard expurgation process (like the one provided in [34, Appendix A]) according to the conditional error probabilities.

**APPENDIX A**

**BIN INDEX DECODING**

In principle, both in the DPC and the Gel’fand–Pinsker channel models defined in Subsections III-A and IV-A, respectively, an optimal bin index decoder should be implemented, in order to minimize the probability of error. Such an optimal decoder compares between ‘metrics’ that depend on the whole set of $u$’s in each bin. Therefore, this optimal bin index decoder is relatively hard to implement, and moreover, it is quite complicated to analyze. Considering the papers [2] and [19], where it has been proved that specific sub-optimal bin index decoders attain the same random coding exponent as the optimal bin index decoder, it is reasonable to suspect that also in the current work, one may lose nothing, at least in the random coding exponent sense, when using a two-stage
sub-optimal decoder, like the one in (14) and (15). In the lines to follow, we show by relatively simple arguments that this is indeed the case here.

Consider a random codebook of size $M_0 = e^{nR_0}$, partitioned into $M = e^{n(R_0 - R_B)}$ bins, each of size $M_B = e^{nR_B}$. Let $m \in \{0, 1, \ldots, M_0 - 1\}$ be the index of the transmitted codeword, and let $\mu = \lfloor m/M_B \rfloor \in \{0, 1, \ldots, M - 1\}$ be the index of its bin, $C_\mu$. Consider a sub-optimal, two-step decoder that decodes $m$ by first decoding $\mu$, using an optimal bin index decoder, and then decoding $m$ by ML decoding within the sub-code pertaining to the decoded bin. Then, on the one hand,

$$\mathbb{P}\{\hat{m} \neq m\} \geq e^{-nE_c(R_0)}. \quad \text{(A.1)}$$

On the other hand,

$$\mathbb{P}\{\hat{m} \neq m\} = \mathbb{P}\{\hat{m} \neq m, \hat{\mu} \neq \mu\} + \mathbb{P}\{\hat{m} \neq m, \hat{\mu} = \mu\} \leq \mathbb{P}\{\hat{\mu} \neq \mu\} + \mathbb{P}\{\arg \max_{u_m \in C_\mu} p(y|u_m') \neq m, \hat{\mu} = \mu\} \leq \mathbb{P}\{\hat{\mu} \neq \mu\} + e^{-nE_c(R_B)}. \quad \text{(A.2)}$$

and so, it follows from (A.1) and (A.5) that

$$\mathbb{P}\{\hat{\mu} \neq \mu\} \leq e^{-nE_c(R_B)} - e^{-nE_c(R_0)}. \quad \text{(A.3)}$$

Now, let $\hat{m}$ be the ordinary ML-decoded version of $m$ and let $\hat{\mu} = \lfloor \hat{m}/M_B \rfloor$ be a sub-optimal bin index decoder that relies simply on $\hat{m}$. Then, as a matching upper bound to (A.7) we have

$$\mathbb{P}\{\hat{\mu} \neq \mu\} \leq \mathbb{P}\{\hat{m} \neq m\} \leq e^{-nE_c(R_0)}. \quad \text{(A.4)}$$

We conclude that

$$\mathbb{P}\{\hat{\mu} \neq \mu\} \leq \mathbb{P}\{\hat{m} \neq m\}, \quad \text{(A.5)}$$

i.e., at least in the random coding sense, the suggested sub-optimal bin index decoder is as good as the optimal bin index decoder.

**APPENDIX B**

**Proofs of Lemmas 1 and 2**

**A. Proof of Lemma 1**

Consider first a given $m$ and $s \in S(\sqrt{nQ})$. Observe that $K_n(s,m) = |C_m \cap T(U)|$ is a binomial random variable with $|C_m| = M_B = e^{n(1 - \Delta)}$ trials and probability of success of the exponential order of $e^{-nK_m}$, and therefore $K_n(m,s)$ concentrates double-exponentially rapidly around $e^{n\Delta}$. More precisely, similarly as in [20, Eqs. (24)–(25)], given $0 < \epsilon \ll \Delta$, we have:

$$\mathbb{P}\{K_n(s,m) > e^{n(\Delta + \epsilon)}\} \leq \exp\{-(\epsilon n - 1)e^{n\epsilon}\} \quad \text{(B.1)}$$

$$\mathbb{P}\{K_n(s,m) < e^{n(\Delta - \epsilon)}\} \leq \exp\{-(1 - (n\epsilon + 1)e^{-n\epsilon})e^{n\epsilon}\}. \quad \text{(B.2)}$$

Let $G_n^s$ be the collection of codes for which $e^{n(\Delta - \epsilon)} \leq K_n(s,m) \leq e^{n(\Delta + \epsilon)}$ for all $m \in \{0, 1, \ldots, M - 1\}$ and $s \in S(\sqrt{nQ})$. We now show that a randomly selected code falls in $G_n^s$ with a probability that tends to one double-exponentially. To this end, consider a fine quantization of the vectors in $S_n(\sqrt{nQ})$ in the following manner: every $s \in S_n(\sqrt{nQ})$ can be represented as follows. Let $\theta_i \in [0, 2\pi), i = 1, 2, \ldots, n - 1$, and let the components of $s$ be given by

$$s_1 = \sqrt{nQ} \sin \theta_1, \quad \text{(B.3)}$$

$$s_2 = \sqrt{nQ} \cos \theta_2 \sin \theta_2, \quad \text{(B.4)}$$

$$s_3 = \sqrt{nQ} \cos \theta_3 \cos \theta_2 \sin \theta_3, \quad \text{(B.5)}$$

$$\vdots$$

$$s_{n-1} = \sqrt{nQ} \cos \theta_1 \cdots \cos \theta_{n-2} \sin \theta_{n-1}, \quad \text{(B.6)}$$

$$s_n = \sqrt{nQ} \cos \theta_1 \cdots \cos \theta_{n-2} \cos \theta_{n-1}. \quad \text{(B.7)}$$

Now, let us quantize each $\theta_i$ uniformly using $n^2$ points with a quantization step-size of $2\pi/n^2$. Thus, the total number of quantization points within the hypercube $[0, 2\pi)^n$ is $n^{2(n-1)}$. Given $s \in S_n(\sqrt{nQ})$, we extract $\theta_1, \ldots, \theta_{n-1}$, and then quantize each $\theta_i$ to its nearest quantization value, $\hat{\theta}_i$. Note that $|\sin \theta_i - \sin \hat{\theta}_i| \leq |\theta_i - \hat{\theta}_i| \leq \pi/n^2$ since the absolute value of the derivative of the sinusoidal function is bounded by 1.

The same comment applies also to $|\cos \theta_i - \cos \hat{\theta}_i|$. According to the above representation, we can think of each component $g_{ik}, k = 1, 2, \ldots, n$, as being given by $\sqrt{nQ} \prod_{i=1}^n g_i(k_{i,\theta})$, where each $g_i(k_{i,\theta})$ is either $\sin \theta_i$, or $\cos \theta_i$, or it is identical to 1. Clearly, it is also true that $|g_i(\theta_i) - g_i(\hat{\theta}_i)| \leq |\theta_i - \hat{\theta}_i| \leq \pi/n^2$, and that $|g_i(\theta_i)|$ and $|g_i(\hat{\theta}_i)|$ are upper bounded by 1. The quantization error in $s_\ell, \ell = 1, 2, \ldots, n$, is therefore upper bounded by

$$|s_\ell - \hat{s}_\ell| \leq \sqrt{nQ} \sum_{j=1}^n \left| \prod_{i=1}^j g_i(\theta_i) \right| \left| \prod_{i=j+1}^n g_i(\theta_i) \right| \quad \text{(B.8)}$$

$$\leq \sqrt{nQ} \sum_{j=1}^n \prod_{i=1}^j g_i(\hat{\theta}_i) \prod_{i=j+1}^n g_i(\theta_i) \quad \text{(B.9)}$$

$$\leq \sqrt{nQ} \sum_{j=1}^n \prod_{i=1}^j g_i(\hat{\theta}_i) \prod_{i=j+1}^n [g_i(\theta_i) - g_i(\hat{\theta}_i)] \quad \text{(B.10)}$$

$$\leq \sqrt{nQ} \sum_{j=1}^n \prod_{i=1}^j g_i(\hat{\theta}_i) \prod_{i=j+1}^n [g_i(\theta_i) - g_i(\hat{\theta}_i)] \quad \text{(B.11)}$$

$$\left| g_i(\theta_i) - g_i(\hat{\theta}_i) \right| \quad \text{(B.12)}$$
\[ \leq \sqrt{nQ} \sum_{j=1}^{n} \frac{\pi}{n^2} \]

\[ = \frac{\pi \sqrt{Q}}{\sqrt{n}}, \]  \hspace{1cm} (B.13)

and so, \( \| s - \hat{s} \| \leq \pi \sqrt{Q} \). Since the large-deviations bounds on \( K_n(s, m) \) (Equations (B.1) and (B.2)), decay double-exponentially with \( n \), while \( M \) is only exponential and the number of quantization points is only \( n^{2(n-1)} \), it follows by the union bound that with probability that tends to one double-exponentially, these bounds hold simultaneously for all \( m \) and all quantization points on the hypersphere. Given that this happens, then for any point on the hypersphere surface, we have

\[ u^T s = u^T [s + (s - \hat{s})] \]

\[ = u^T \hat{s} + u^T (s - \hat{s}) \]

\[ \geq \alpha_n \hat{s}^T \hat{s} - \| u \| \| s - \hat{s} \| \]

\[ \geq \alpha_n \hat{s}^T \hat{s} - \sqrt{nW} \cdot \pi \sqrt{Q} \]

\[ = n\alpha \sqrt{QQ - \pi \sqrt{nWQ}}, \]  \hspace{1cm} (B.17)

where (B.17) follows by assuming that \( u \) is a codeword in \( C_m \cap \mathcal{T}(U|\hat{s}) \). The second term in (B.19), which scales with \( \sqrt{n} \), is asymptotically negligible relative to the first one, which is linear in \( n \). It follows that every codeword in \( C_m \cap \mathcal{T}(U|\hat{s}) \) is also in \( C_m \cap \mathcal{T}(U|s) \), provided that \( s \) is quantized to \( \hat{s} \), and provided that \( \mathcal{T}(U|s) \) is slightly expanded by reducing the threshold, \( \alpha_n \| s \|^2 \), within a relative amount that remains bounded as \( n \) grows without bound.

**Appendix C**

**Proof of Theorem 2**

A. Step 1: Error Exponent for a General Decoding Metric

Given \( m \in \{1, 2, \ldots, M\} \), \( S = s, U = u \in \mathcal{C}(m, s) \), \( X = x \in \mathcal{T}(Q_X|SU|s, u) \), and \( Y = y \), the probability of error is given by

\[ P_e(m, s, u, x, y) = \mathbb{P}\left\{ \bigcup_{Q_{S'}} \bigcup_{m' \neq m} \bigcup_{k=1}^{M(Q_{S'})} G(\hat{P}_{U_{Q_{S'}}, m', s, y}, Q_{S'}) \geq G(\hat{P}_{uy}, \hat{P}_s) \right\} \]

\[ \leq \max \left\{ \mathbb{P}\left\{ \rho^2(U_{\hat{k}}, y) < 1 - \exp\{-2(R + I_{us} - \epsilon)\} \right\} \right\} \]

\[ = \left[ 1 - \exp\{-n(R + I_{us} - \epsilon)\} \right]^M \]

\[ = \left[ 1 - \exp\{-n(R + I_{us} - \epsilon)\} \right]^{\exp\{n[R + I_{us} + \Delta]\}} \]

\[ = \exp\left\{ - \exp\{n[R + I_{us} + \Delta]\} \exp\{-n(R + I_{us} - \epsilon)\} \right\}, \]  \hspace{1cm} (B.22)

which decays double-exponentially as \( n \to \infty \). Applying the union bound over all pairs \( \{k, k'\} \), introduces an exponential factor of the order of \( M^2 \), which leaves the probability of the union double-exponentially small. As for the union over \( \{y\} \), it is sufficient to consider one hypersphere surface since the empirical correlation coefficient is invariant to scaling of \( y \). Consider again, quantization of the hypersphere of \( y \), since the large-deviations factor of the order of \( M^2 \). This property applies to every quantization point at the same time, since the factor of \( n^{2(n-1)} \) still leaves the probability of the union double-exponentially small. Finally, we pass to the continuum of \( \{y\} \) in the hypersphere, in the same way as in the proof of Lemma 1, by using the fact that the quantization error is small and hence affects the empirical correlation coefficient by a vanishingly small amount as \( n \) grows without bound.

**Appendix C**

**Proof of Theorem 2**

A. Step 1: Error Exponent for a General Decoding Metric

Given \( m \in \{1, 2, \ldots, M\} \), \( S = s, U = u \in \mathcal{C}(m, s) \), \( X = x \in \mathcal{T}(Q_X|SU|s, u) \), and \( Y = y \), the probability of error is given by

\[ P_e(m, s, u, x, y) = \mathbb{P}\left\{ \bigcup_{Q_{S'}} \bigcup_{m' \neq m} \bigcup_{k=1}^{M(Q_{S'})} G(\hat{P}_{U_{Q_{S'}}, m', s, y}, Q_{S'}) \geq G(\hat{P}_{uy}, \hat{P}_s) \right\} \]

\[ \leq \max \left\{ \mathbb{P}\left\{ \rho^2(U_{\hat{k}}, y) < 1 - \exp\{-2(R + I_{us} - \epsilon)\} \right\} \right\} \]

\[ = \left[ 1 - \exp\{-n(R + I_{us} - \epsilon)\} \right]^M \]

\[ = \left[ 1 - \exp\{-n(R + I_{us} - \epsilon)\} \right]^{\exp\{n[R + I_{us} + \Delta]\}} \]

\[ = \exp\left\{ - \exp\{n[R + I_{us} + \Delta]\} \exp\{-n(R + I_{us} - \epsilon)\} \right\}, \]  \hspace{1cm} (B.22)

which decays double-exponentially as \( n \to \infty \). Applying the union bound over all pairs \( \{k, k'\} \), introduces an exponential factor of the order of \( M^2 \), which leaves the probability of the union double-exponentially small. As for the union over \( \{y\} \), it is sufficient to consider one hypersphere surface since the empirical correlation coefficient is invariant to scaling of \( y \). Consider again, quantization of the hypersphere of \( y \), since the large-deviations factor of the order of \( M^2 \). This property applies to every quantization point at the same time, since the factor of \( n^{2(n-1)} \) still leaves the probability of the union double-exponentially small. Finally, we pass to the continuum of \( \{y\} \) in the hypersphere, in the same way as in the proof of Lemma 1, by using the fact that the quantization error is small and hence affects the empirical correlation coefficient by a vanishingly small amount as \( n \) grows without bound.
\[ P(m, s, u, x, y) = \min \left\{ 1, \sum_{Q_{s'}} \sum_{m' \neq m} \sum_{k=1}^{M} \exp \left\{ -n A(\hat{P}_{uy}, \hat{P}_s) \right\} \right\} \]  
\[ = \exp \left\{ -nA(\hat{P}_{uy}, \hat{P}_s) \right\} \]  
where,  
\[ A(\hat{P}_{uy}, \hat{P}_s) = \min_{\{Q_{u'y}: G(Q_{u'y} \times P_u, Q_{s'}) \geq G(\hat{P}_{uy}, \hat{P}_s)\}} \]  
\[ I_Q(U'; Y) \]  
Note that the expression in (C.7) does not depend on \( m' \neq m \) and \( k \in \{1, 2, \ldots, M(Q_{s'})\} \), hence, substituting (C.7) back into (C.2) yields that  
\[ P(m, s, u, x, y) = \min \left\{ 1, \sum_{Q_{s'}} \sum_{m' \neq m} \sum_{k=1}^{M} \exp \left\{ -n A(\hat{P}_{uy}, \hat{P}_s) \right\} \right\} \]  
\[ = \exp \left\{ -nA(\hat{P}_{uy}, \hat{P}_s) \right\} \]  

Next, given \( m \in \{1, 2, \ldots, M\} \), \( S = s, U = u \in C(m, s) \), and \( X = x \in T(Q_{X|SU}|s, u) \), we average over the randomness of the channel output vector \( Y \):  
\[ P(m, s, u, x) = \sum_{y \in Y} W(y|x, s) \exp \left\{ -n E_0(R, \hat{P}_s, \hat{P}_{uy}) \right\} \]  
\[ = \sum_{Q_{Y|SU} \in T(Q_{Y|SU}|S, U, x)} \sum_{y \in Y} W(y|x, s) \]  
\[ \times \exp \left\{ -n E_0(R, \hat{P}_s, \hat{P}_{uy}) \right\} \]  
\[ = \sum_{Q_{Y|SU}} \exp \left\{ -n D(Q_{Y|SU}||W_{Y|XS}|\hat{P}_{sux}) \right\} \]  
\[ \times \exp \left\{ -n E_0(R, \hat{P}_s, \hat{P}_{UY}) \right\} \]  
\[ = \exp \left\{ -n \min_{Q_{Y|SU}} \left\{ D(Q_{Y|SU}||W_{Y|XS}|\hat{P}_{sux}) \right\} \right\} \]  
\[ + E_0(R, \hat{P}_s, \hat{P}_{UY}) \]  
\[ \triangleq \exp \left\{ -n \cdot E_1(R, \hat{P}_s, Q_{UX|S}) \right\} \]  
(C.17)  

Note that (C.15) depends on \( s, u, x \) only via their joint empirical distribution, hence, averaging over the randomness of \( U \in B(\hat{P}_s, m) \cap T(Q_{U'|S}|s) \) and \( X \in T(Q_{X|SU}|s, u) \) yields  
\[ P(m, s) = \exp \left\{ -n \cdot \min_{Q_{Y|SU}} \left\{ D(Q_{Y|SU}||W_{Y|XS}|Q_{UX|S} \times \hat{P}_s) \right\} \right\} \]  
\[ + E_0(R, \hat{P}_s, Q_{UY}) \]  
\[ \triangleq \exp \left\{ -n \cdot E_1(R, \hat{P}_s, Q_{UX|S}) \right\} \]  
(C.18)  

The exponent function in (C.17) depends on \( Q_{UX|S} \), which are design parameters of the system that can be chosen to minimize the average error probability. Thus, minimizing the average probability of error w.r.t. \( Q_{UX|S} \), we arrive at  
\[ P(m, s) = \exp \left\{ -n \cdot \max_{Q_{UX|S}} E_1(R, \hat{P}_s, Q_{UX|S}) \right\} \]  
(C.18)
APPENDIX D
PROOF OF THEOREM 3

Assume that the matrix \( \{Q_{UX|S}\} \) is fixed. Let us choose the universal decoding metric \( g(Q_{UY}, Q_S) = I_Q(U; Y) - R(Q_S) \). For a given codebook, the probability of error is given by

\[
P_e(C_n) = \sum_{Q_S} \mathbb{P}[S \in T(Q_S)] \frac{1}{M \cdot M(Q_S)} \sum_{m=1}^{M(Q_S)} \sum_{\ell=1}^{M(Q_S)} \sum_{s \in T(Q_S)} P(s | u_{m, \ell})
\]

\[
\times \sum_{x \in T(Q_{X|US}|u_{m, \ell}, s)} \frac{1}{T(Q_{X|US}|u_{m, \ell}, s)}
\]

\[
\times \sum_{y \in Y^n} W(y|x, s) \{ \bigcup_{Q_{SR} \neq m} \bigcup_{\ell=1}^{M(Q_{SR})} \left\{ \exp\{ng(\hat{P}_{um, \ell y}, Q_{SR})\} \geq \exp\{ng(\hat{P}_{um, \ell y}, Q_S)\} \right\} \}
\]

We upper-bound the inner-most summation over \( y \) by\(^{10}\):

\[
\sum_{y \in Y^n} W(y|x, s) \left\{ \bigcup_{Q_{SR} \neq m} \bigcup_{\ell=1}^{M(Q_{SR})} \left\{ \exp\{ng(\hat{P}_{um, \ell y}, Q_{SR})\} \geq \exp\{ng(\hat{P}_{um, \ell y}, Q_S)\} \right\} \right\}
\]

\[
\leq \sum_{y \in Y^n} W(y|x, s) \sum_{Q_{SR} \neq m} \sum_{\ell=1}^{M(Q_{SR})} \inf_{\mu \geq 0} \left( \frac{\exp\{ng(\hat{P}_{um, \ell y}, Q_{SR})\}}{\exp\{ng(\hat{P}_{um, \ell y}, Q_S)\}} \right)^\mu
\]

\[
= \sum_{Q_{SR} \neq m} \sum_{\ell=1}^{M(Q_{SR})} \sum_{y \in Y^n} W(y|x, s) \inf_{\mu \geq 0} \left( \frac{\exp\{ng(\hat{P}_{um, \ell y}, Q_{SR})\}}{\exp\{ng(\hat{P}_{um, \ell y}, Q_S)\}} \right)^\mu
\]

where the inner-most sum over \( y \in Y^n \) is assessed by the method of types as

\[
\sum_{y \in Y^n} W(y|x, s) \cdot \exp\left\{ -n \cdot \sup_{\mu \geq 0} \left\{ \mu \cdot \{ g(\hat{P}_{um, \ell y}, Q_{SR}) - g(\hat{P}_{um, \ell y}, Q_S) \} \right\} \right\}
\]

\[
= \sum_{Q_{Y|UXS} \neq s} \sum_{y \in T(Q_{Y|UXS}|u_{m, \ell, s}, x)} W(y|x, s) \times \exp\left\{ -n \cdot \Psi(\hat{P}_{um, \ell y}, Q_{SR}, \hat{P}_{um, \ell y}, Q_S) \right\}
\]

\[
= \sum_{Q_{Y|UXS} \neq s} \sum_{y \in T(Q_{Y|UXS}|u_{m, \ell, s}, x)} e^{nE_Q[\ln W(Y|X, S)]} \times \exp\left\{ -n \cdot \Psi(\hat{P}_{um, \ell y}, Q_{SR}, \hat{P}_{um, \ell y}, Q_S) \right\}
\]

\[
= \sum_{Q_{Y|UXS} \neq s} e^{-nD(Q_{Y|UXS}|W_{Y|XS}|u_{m, \ell, s}, x)} \times \exp\left\{ -n \cdot \Psi(\hat{P}_{um, \ell y}, Q_{SR}, \hat{P}_{um, \ell y}, Q_S) \right\}
\]

\[
= \exp\left\{ -nE_0(\hat{P}_{um, \ell y}, u_{m, \ell, s}, x, Q_{SR}) \right\}
\]

Substituting it back into (D.1) yields that

\[
P_e(C_n)
\]

\[
\leq \sum_{Q_S} \mathbb{P}[S \in T(Q_S)] \frac{1}{M \cdot M(Q_S)} \sum_{m=1}^{M(Q_S)} \sum_{\ell=1}^{M(Q_S)} \sum_{s \in T(Q_{X|US}|u_{m, \ell}, s)} \frac{1}{T(Q_{X|US}|u_{m, \ell}, s)}
\]

\[
\times \sum_{Q_{SR} \neq m} \sum_{\ell=1}^{M(Q_{SR})} \sum_{y \in Y^n} W(y|x, s) \inf_{\mu \geq 0} \left( \frac{\exp\{ng(\hat{P}_{um, \ell y}, Q_{SR})\}}{\exp\{ng(\hat{P}_{um, \ell y}, Q_S)\}} \right)^\mu
\]

\[
= \sum_{Q_S} \mathbb{P}[S \in T(Q_S)] \frac{1}{M \cdot M(Q_S)} \sum_{m=1}^{M(Q_S)} \sum_{\ell=1}^{M(Q_S)} \sum_{s \in T(Q_{X|US}|u_{m, \ell}, s)} \frac{1}{T(Q_{X|US}|u_{m, \ell}, s)}
\]

\[
\times \exp\left\{ -nE_0(\hat{P}_{um, \ell y}, u_{m, \ell, s}, x, Q_{SR}) \right\}
\]

\[
= \exp\left\{ -nE_0(\hat{P}_{um, \ell y}, u_{m, \ell, s}, x, Q_{SR}) \right\}
\]

The inner-most sum over \( x \) is assessed by the method of types as

\[
\sum_{x \in T(Q_{X|US}|u_{m, \ell}, s)} \frac{1}{T(Q_{X|US}|u_{m, \ell}, s)} \times \exp\left\{ -nE_0(\hat{P}_{um, \ell y}, u_{m, \ell, s}, x, Q_{SR}) \right\}
\]

\[
= \sum_{Q_{X|US} = Q_{X|US}} \sum_{x \in T(Q_{X|US}|u_{m, \ell, s}, x, s)} \frac{1}{T(Q_{X|US}|u_{m, \ell, s})} \exp\left\{ -nE_0(\hat{P}_{um, \ell y}, u_{m, \ell, s}, x, s, Q_{SR}) \right\}
\]

\[
= \exp\left\{ -nE_0(\hat{P}_{um, \ell y}, u_{m, \ell, s}, x, s, Q_{SR}) \right\}
\]
\[
= \sum_{\{Q_{S|UU'}: Q_S = Q_S(U)\}} \mathbb{P}[S \in T(Q_S)] \frac{1}{M \cdot M(Q_S)} \sum_{m=1}^{M(Q_S)} \sum_{\ell=1}^{M(Q_{S'})} \\
\exp \left\{-nE_0(\hat{P}_{u_m,\ell,u_{m',\ell'},s} \times \hat{Q}_{S|UU',Q_S'}) \right\}
\]
\[
\sum_{\{Q_{S|UU': Q_S = Q_S(U)\}} \mathbb{P}[S \in T(Q_S)] \frac{1}{M \cdot M(Q_S)} \sum_{m=1}^{M(Q_S)} \sum_{\ell=1}^{M(Q_{S'})} \\
\exp \left\{-nE_0(\hat{P}_{u_m,\ell,u_{m',\ell'},s} \times \hat{Q}_{S|UU',Q_S'}) \right\}
\]
\[
\exp \left\{-n \cdot \min_{\{Q_{S|UU': Q_S = Q_S(U)\}}} [I_Q(S;U'|U) + E_1(\hat{P}_{u_m,\ell,u_{m',\ell'},s} \times \hat{Q}_{S|UU',Q_S'})] \right\}
\]
\[
= \exp \left\{-nE_1(\hat{P}_{u_m,\ell,u_{m',\ell'},s} \times \hat{Q}_{S|UU',Q_S'}) \right\}
\]

Substituting it back into (D.19) yields that
\[
P_C(C_n)
\leq \sum_{Q_S Q_{S'}} e^{-nD(Q_S||P_S)} e^{-n(R+R(Q_S))}
\sum_{m=1}^{M(Q_S)} \sum_{\ell=1}^{M(Q_{S'})} \sum_{m' \neq m} \sum_{\ell' \neq \ell}
\exp \left\{-nE_2(\hat{P}_{u_m,\ell,u_{m',\ell'},s} \times \hat{Q}_{S|UU',Q_S'}) \right\}
\]
\[
\sum_{Q_S Q_{S'}} e^{-nD(Q_S||P_S)} e^{-n(R+R(Q_S))}
\sum_{m=1}^{M(Q_S)} \sum_{\ell=1}^{M(Q_{S'})} \sum_{m' \neq m} \sum_{\ell' \neq \ell}
\exp \left\{-nE_2(\hat{P}_{u_m,\ell,u_{m',\ell'},s} \times \hat{Q}_{S|UU',Q_S'}) \right\}
\]
\[
\sum_{Q_S Q_{S'}} e^{-nD(Q_S||P_S)} e^{-n(R+R(Q_S))}
\sum_{m=1}^{M(Q_S)} \sum_{\ell=1}^{M(Q_{S'})} \sum_{m' \neq m} \sum_{\ell' \neq \ell}
\exp \left\{-nE_2(\hat{P}_{u_m,\ell,u_{m',\ell'},s} \times \hat{Q}_{S|UU',Q_S'}) \right\}
\]
\[
\sum_{Q_S Q_{S'}} e^{-nD(Q_S||P_S)} e^{-n(R+R(Q_S))}
\sum_{m=1}^{M(Q_S)} \sum_{\ell=1}^{M(Q_{S'})} \sum_{m' \neq m} \sum_{\ell' \neq \ell}
\exp \left\{-nE_2(\hat{P}_{u_m,\ell,u_{m',\ell'},s} \times \hat{Q}_{S|UU',Q_S'}) \right\}
\]

Again, the inner-most sum is assessed by the method of types as
\[
= \sum_{s \in T(Q_S)} P(s|u_{m,\ell}) \exp \left\{-nE_1(\hat{P}_{u_m,\ell,u_{m',\ell'},s} \times \hat{Q}_{S|UU',Q_S'}) \right\}
\]

Now, for any \(\rho > 1\),
\[
\mathbb{E} \left[ P_C(C_n)^{1/\rho} \right] \leq \sum_{Q_S Q_{S'}} e^{-nD(Q_S||P_S)/\rho} e^{-n(R+R(Q_S))/\rho}
\sum_{Q_{UU'} \in Q(U,U')} \mathbb{E} \left[ N(\hat{Q}_{UU'})^{1/\rho} \right]
\exp \left\{-nE_2(\hat{P}_{u_m,\ell,u_{m',\ell'},s} \times \hat{Q}_{S|UU',Q_S'}) \right\}
\]

The expectation in (D.30) is upper-bounded as [21]
\[
\mathbb{E} \left[ N(\hat{Q}_{UU'})^{1/\rho} \right] \leq \begin{cases} 
\exp \{n[2R + R(Q_S) + R(Q_S') - I_Q(U;U')]/\rho \} & \text{if } 2R + R(Q_S) + R(Q_S') \geq I_Q(U;U') \\
\exp \{n[2R + R(Q_S) + R(Q_S') - I_Q(U;U')]/\rho \} & \text{if } 2R + R(Q_S) + R(Q_S') < I_Q(U;U')
\end{cases}
\]
and so,
\[
(E \left[N(\tilde{Q}_{UU'})^{1/\rho}\right])^\rho \\
\leq \exp\{n(2R + R(Q_S) + R(Q_{S'}) - I_D(U; U') + \rho[I_Q(U; U') - 2R - R(Q_S) - R(Q_{S'})]) + D(Q_S\|P_S) + E_2(\tilde{Q}_{UU'}, Q_{S'}) + I_Q(U; U') - R - R(Q_{S'})]\}
\]
(D.32)

which completes the proof of Theorem 3.

**APPENDIX E**

**PROOF OF THEOREM 4**

Assuming that message \(m\) was transmitted, the probability of error, for a given code \(C_n\), is given by
\[
P_{d|m}(C_n) = \sum_{Q_S} \mathbb{P}[S \in T(Q_S)] \frac{1}{M(Q_S)} \sum_{\ell=1}^{M(Q_S)} \sum_{s \in T(Q_S)} P(s|u_{m,\ell}) \times \sum_{y \in \mathbb{Y}^n} \mathbb{W}(y|x, s) \left\{ \bigcup_{Q_{S'}} \bigcup_{U \neq m} \bigcup_{\ell' = 1}^{M(Q_{S'})} \{G(\hat{P}_{u_{m',\ell'}, y}, Q_{S'}) \geq G(\hat{P}_{u_{m,\ell}, y}, Q_S)\} \right\}.
\]
(E.1)

Since the whole difference relative to the total probability of error given in (D.1) is in the averaging over the set of bins (which is of size \(M\)), we may proceed using the same initial steps as in the proof of Theorem 3, and arrive at (upon using the result in (D.26))
\[
P_{d|m}(C_n) \leq \sum_{Q_S} \sum_{Q_{S'}} \mathbb{P}[S \in T(Q_S)] \frac{1}{M(Q_S)} \sum_{\ell=1}^{M(Q_S)} \sum_{m' \neq m} \sum_{\ell' = 1}^{M(Q_{S'})} \exp\{-nE_2(\hat{P}_{u_{m,\ell}, u_{m',\ell'}, Q_{S'}})\}
\]
(E.2)
\[
\quad \times \sum_{Q_{U'}} \sum_{U \neq m} \sum_{\ell' = 1}^{M(Q_{U'})} \hat{N}(\tilde{Q}_{UU'}) \exp\{-nE_2(\tilde{Q}_{UU'}, Q_{S'})\},
\]
(E.4)

where \(Q(\tilde{Q}_{UU'}): \tilde{Q}_U = Q_U, \tilde{Q}_{U'} = Q_{U'}\) and
\[
\hat{N}(\tilde{Q}_{UU'}) \triangleq \sum_{\ell=1}^{M(Q_S)} \sum_{m' \neq m} \sum_{\ell' = 1}^{M(Q_{S'})} \mathbb{1}\left\{ (u_{m,\ell}, u_{m',\ell'}) \in T(\tilde{Q}_{UU'}) \right\}.
\]
(E.5)

Now, for any \(\rho > 1\),
\[
E\left[P_{d|m}(C_n)^{1/\rho}\right] \leq \sum_{Q_S} \sum_{Q_{S'}} e^{-nD(Q_S\|P_S)/\rho} e^{-nR(Q_S)/\rho}
\]
Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
\begin{align*}
&\times \sum_{\bar{Q}_{UU'} \in \mathcal{Q}(Q_U, Q_{U'})} \mathbb{E} \left[ \hat{N}(\bar{Q}_{UU'})^{1/\rho} \right] \\
&\quad \times \exp \left\{ -n E_2(\bar{Q}_{UU'}, Q_{S'}) / \rho \right\}. \quad (E.6)
\end{align*}

The expectation in (E.6) is upper-bounded as \cite{21}
\begin{align*}
\mathbb{E} \left[ \hat{N}(\bar{Q}_{UU'})^{1/\rho} \right] &\leq \left\{ \begin{array}{ll}
\exp \left\{ n \left[ R + R(Q_S) + R(Q_{S'}) - I_Q(U; U') \right] / \rho \right\} & \text{if } R + R(Q_S) + R(Q_{S'}) \geq I_Q(U; U') \\
\exp \left\{ n \left[ R + R(Q_S) + R(Q_{S'}) - I_Q(U; U') \right] \right\} & \text{if } R + R(Q_S) + R(Q_{S'}) < I_Q(U; U') \end{array} \right.
\end{align*}
\begin{align*}
&\leq \left\{ \begin{array}{ll}
\exp \left\{ n \left[ R + R(Q_S) + R(Q_{S'}) - I_Q(U; U') \right] / \rho \right\} \\
\exp \left\{ n \left[ R + R(Q_S) + R(Q_{S'}) - I_Q(U; U') \right] \right\}
\end{array} \right.
\end{align*}
\begin{align*}
&\triangleq \exp \left\{ n G(\bar{Q}_{UU'}, R, R(Q_S), R(Q_{S'}), \rho) \right\}. \quad (E.7)
\end{align*}

Substituting (E.9) back into (E.6) yields a bound which we shall shortly denote by \( \Psi(R, \rho) \).

According to Markov’s inequality, we get
\begin{align*}
P \left\{ \frac{1}{M} \sum_{m=1}^{M} P_{\text{lim}}(C_n)^{1/\rho} > 2\Psi(R, \rho) \right\} \leq \frac{1}{2}, \quad \text{(E.10)}
\end{align*}

which means that there exists a code with
\begin{align*}
\frac{1}{M} \sum_{m=1}^{M} P_{\text{lim}}(C_n)^{1/\rho} \leq \Psi(R, \rho). \quad \text{(E.11)}
\end{align*}

We conclude that there exists a code \( C'_n \) with \( M/2 \) bins for which
\begin{align*}
\max_m P_{\text{lim}}(C'_n)^{1/\rho} \leq 4\Psi(R, \rho), \quad \text{(E.12)}
\end{align*}

and so
\begin{align*}
\max_{m} P_{\text{lim}}(C'_n) &\leq \left( \sum_{Q_{U'}} \sum_{Q_{S'}} \sum_{\bar{Q}_{UU'} \in \mathcal{Q}(Q_U, Q_{U'})} \exp \left\{ \frac{n D(Q_{S'} \| P_S) R(Q_S)}{\rho} \right\} e^{-n R(Q_{S'}) / \rho} \right)^{1/\rho} \quad \text{(E.13)}
\end{align*}
\begin{align*}
&\leq \left( \sum_{Q_{U'}} \sum_{Q_{S'}} \sum_{\bar{Q}_{UU'} \in \mathcal{Q}(Q_U, Q_{U'})} \exp \left\{ n \rho G(\bar{Q}_{UU'}, R, R(Q_S), R(Q_{S'}), \rho) \right\} \right) \quad \text{(E.14)}
\end{align*}
\begin{align*}
&\leq \exp \left\{ -n \min_{Q_{U'}, Q_{S'}} \min_{\bar{Q}_{UU'} \in \mathcal{Q}(Q_U, Q_{U'})} \left\{ E_2(\bar{Q}_{UU'}, Q_{S'}) \right\} \right. \\
&\quad - \rho G(\bar{Q}_{UU'}, R, R(Q_S), R(Q_{S'}), \rho) \\
&\quad + D(Q_S \| P_S) + R(Q_S) \bigg\}, \quad \text{(E.15)}
\end{align*}

thus,
\begin{align*}
\liminf_{n \to \infty} - \frac{1}{n} \ln \max_{m} P_{\text{lim}}(C'_n) &\geq \min_{Q_S, Q_{S'}} \min_{\bar{Q}_{UU'} \in \mathcal{Q}(Q_U, Q_{U'})} \left\{ E_2(\bar{Q}_{UU'}, Q_{S'}) \right\} \\
&\quad - \rho G(\bar{Q}_{UU'}, R, R(Q_S), R(Q_{S'}), \rho) \\
&\quad + D(Q_S \| P_S) + R(Q_S). \quad \text{(E.16)}
\end{align*}

Since the inequality in (E.16) holds for every \( \rho \geq 1 \), the negative exponential rate of the maximal probability of error can be bounded as
\begin{align*}
\liminf_{n \to \infty} - \frac{1}{n} \ln \max_{m} P_{\text{lim}}(C'_n) &\geq \sup_{\rho \geq 1} \min_{Q_S, Q_{S'}} \min_{\bar{Q}_{UU'} \in \mathcal{Q}(Q_U, Q_{U'})} \left\{ E_2(\bar{Q}_{UU'}, Q_{S'}) \right\} \\
&\quad - \rho G(\bar{Q}_{UU'}, R, R(Q_S), R(Q_{S'}), \rho) \\
&\quad + D(Q_S \| P_S) + R(Q_S). \quad \text{(E.17)}
\end{align*}
\begin{align*}
&\geq \liminf_{n \to \infty} - \frac{1}{n} \ln \max_{m} P_{\text{lim}}(C'_n) \\
&\quad \geq \min_{\rho \geq \rho_G} \min_{Q_S, Q_{S'}} \min_{\bar{Q}_{UU'} \in \mathcal{Q}(Q_U, Q_{U'})} \left\{ E_2(\bar{Q}_{UU'}, Q_{S'}) \right\} \\
&\quad - \rho G(\bar{Q}_{UU'}, R, R(Q_S), R(Q_{S'}), \rho) \\
&\quad + D(Q_S \| P_S) + R(Q_S). \quad \text{(E.18)}
\end{align*}

At this point, we shall proceed along the same steps as in \cite[Eq.s (A.26)-(A.37)]{34}, and conclude that
\begin{align*}
\liminf_{n \to \infty} - \frac{1}{n} \ln \max_{m} P_{\text{lim}}(C'_n) &\geq \min_{Q_S, Q_{S'}} \min_{\bar{Q}_{UU'} \in \mathcal{Q}(Q_U, Q_{U'})} \left\{ E_2(\bar{Q}_{UU'}, Q_{S'}) \right\} \\
&\quad - \rho G(\bar{Q}_{UU'}, R, R(Q_S), R(Q_{S'}), \rho) \\
&\quad + D(Q_S \| P_S) + R(Q_S). \quad \text{(E.19)}
\end{align*}
\begin{align*}
\liminf_{n \to \infty} - \frac{1}{n} \ln \max_{m} P_{\text{lim}}(C'_n) &\geq \min_{Q_S, Q_{S'}} \min_{\bar{Q}_{UU'} \in \mathcal{Q}(Q_U, Q_{U'})} \left\{ E_2(\bar{Q}_{UU'}, Q_{S'}) \right\} \\
&\quad - \rho G(\bar{Q}_{UU'}, R, R(Q_S), R(Q_{S'}), \rho) \\
&\quad + D(Q_S \| P_S) + E_2(\bar{Q}_{UU'}, Q_{S'}) + I_{\bar{Q}(U^'; U')} - R - R(Q_{S'}) \bigg\}, \quad \text{(E.20)}
\end{align*}

As a final step, we maximize the exponent function over the design parameters of the code:
\begin{align*}
\liminf_{n \to \infty} - \frac{1}{n} \ln \max_{m} P_{\text{lim}}(C'_n) &\geq \max_{Q_{UXI}} \min_{Q_S, Q_{S'}} \min_{\bar{Q}_{UU'} \in \mathcal{Q}(Q_U, Q_{U'})} \left\{ E_2(\bar{Q}_{UU'}, Q_{S'}) \right\} \\
&\quad - \rho G(\bar{Q}_{UU'}, R, R(Q_S), R(Q_{S'}), \rho) \\
&\quad + D(Q_S \| P_S) + E_2(\bar{Q}_{UU'}, Q_{S'}) + I_{\bar{Q}(U^'; U')} - R - R(Q_{S'}) \bigg\}, \quad \text{(E.21)}
\end{align*}
which completes the proof of Theorem 4.

\textbf{REFERENCES}
\begin{enumerate}
\item E. Arian and N. Merhav, “Guessing subject to distortion,” IEEE Trans. Inf. Theory, vol. 44, no. 3, pp. 1041–1056, May 1998.
\item R. Averbuch and N. Merhav, “Exact random coding exponents and universal decoders for the asymmetric broadcast channel,” IEEE Trans. Inf. Theory, vol. 64, no. 7, pp. 5070–5086, Jul. 2018.
\item A. Barg and G. D. Forney, Jr., “Random codes: Minimum distances and error exponents,” IEEE Trans. Inf. Theory, vol. 48, no. 9, pp. 2568–2573, Sep. 2002.
\item R. J. Barron, B. Chen, and G. W. Wornell, “The duality between information embedding and source coding with side information and some applications,” IEEE Trans. Inf. Theory, vol. 49, no. 5, pp. 1159–1180, May 2003.
\item G. Cocco, A. Guillén i Farrés, and J. Font-Segura, “Typical error exponents: A dual domain derivation,” IEEE Trans. Inf. Theory, vol. 69, no. 2, pp. 776–793, Feb. 2023.
\end{enumerate}
