Codebook Mismatch Can Be Fully Compensated by Mismatched Decoding

Neri Merhav[*] Georg Böcherer[†]

June 19, 2022

Abstract

We consider an ensemble of constant composition codes that are subsets of linear codes: while the encoder uses only the constant-composition subcode, the decoder operates as if the full linear code was used, with the motivation of simultaneously benefiting both from the probabilistic shaping of the channel input and from the linear structure of the code. We prove that the codebook mismatch can be fully compensated by using a mismatched additive decoding metric that achieves the random coding error exponent of (non-linear) constant composition codes. As the coding rate tends to the mutual information, the optimal mismatched metric approaches the maximum a posteriori probability (MAP) metric, showing that codebook mismatch with mismatched MAP metric is capacity-achieving for the optimal input assignment.

Index Terms: linear codes, constant composition codes, error exponent, channel capacity, codebook mismatch, decoding metric, mismatched decoding.

[*] N. Merhav is with the Viterbi Faculty of Electrical and Computer Engineering, Technion – Israel Institute of Technology, Haifa 3200003, Israel. Email: merhav@technion.ac.il

[†] G. Böcherer is with the Munich Research Center, Huawei Technologies Duesseldorf GmbH, Germany. Email: georg.boecherer@ieee.org

1 Introduction

As is very well known, linear codes have always been of central interest in channel coding theory, thanks to their convenient practical implementation, both at the encoder and the decoder side (see, e.g., [14, Chap. 6], [17, Part II], [22, Sections 2.9, 2.10] for major elementary textbooks, as well as a vast amount of other books and articles). The structure of linear codes, together with the additivity of the optimal channel decoding metric of certain memoryless channels, can offer reduced decoding complexity in many ways, such as in syndrome decoding (for relatively high coding rates), bounded distance decoding [16, Sect. 6.2], Chase decoding [10], and many other decoding schemes that are based on temporary hard decision, followed by a process of correction using soft information. In many cases, the resulting decoding is equivalent (or at least asymptotically so) to the optimal maximum likelihood (ML) decoding. Also, for convolutional codes, which form a special subclass of linear codes, the Viterbi decoder offers dramatic reduction in computational complexity [22] without sacrificing decoder optimality. Last but not least, polar codes, invented by Stolte [21] and Arıkan [3] and proven to be capacity-achieving by Arıkan [3], form another subclass of linear codes. Polar codes are perhaps the most attractive codes in the front line of research in coding theory today, with decoding complexity that is proportional to $n \log n$, where n is the block length.[1] In general, it would be safe to say that most of the modern practical codes are linear. For channels with a sufficient degree of symmetry, like binary-input, output-symmetric (BIOS) channels, whose capacity-achieving input distribution is uniform, it is well known that linear codes can achieve capacity, as linear codes inherently induce the uniform input distribution. Moreover, the ensemble of linear codes achieves the well known random coding error exponent at all coding rates up to channel capacity [14, Theorem 6.2.1].

An inherent limitation of linear codes, however, is that they cannot achieve capacity (and hence neither can they achieve the channel's reliability function) when the capacity-achieving input distribution is non-uniform. One way to compensate for this drawback, and to achieve capacity with linear codes nonetheless, is to extend the channel input alphabet using a many-to-one mapping that induces the desired input distribution (which is an

[1] There are, of course, additional classes of modern codes with efficient decoding schemes, like Turbo codes and low-density parity check codes, but their decodings are iterative and so, in general they are not guaranteed to be equivalent to ML decoding.

operation also known as “probabilistic shaping”), and use the linear code on symbols of the extended alphabet – see, e.g., Gallager [14, p. 208] for the details. This approach is conceptually simple, however, not very attractive from the practical point of view, because the required extended alphabet is often much larger, and so, many more coded bits need to be processed, and the inevitable consequence is increased complexity and increased power consumption.

During the years, researchers in coding theory have been pondering about the question whether it is possible to enjoy the best of both words, namely, to achieve capacity (as well as good error-rate performance at lower rates) without giving up on the above-mentioned benefits of the linear code structure and the additive metric decoding, and without paying the price of increased complexity of Gallager’s alphabet extension needed for probabilistic shaping.

In this context, the idea of probabilistic amplitude shaping (PAS) [9] was proposed to have exactly the above-mentioned features: it enables the use of linear codes with non-uniform distribution without the need of alphabet extension. The basic idea is as follows: at the transmitter, the uniformly distributed message bits are transformed into “probabilistically shaped” codewords with the desired distribution. This can be achieved efficiently using distribution matching (DM) algorithms [20]. The shaped sequences are then encoded systematically with a linear code. The systematic encoding preserves the imposed distribution of the message part, and generates in addition uniformly distributed parity bits. For input distributions that have a uniform distribution as a factor, the parity bits can be used to generate the uniform factor, and the shaped bits can be used for the non-uniform factor. For the additive white Gaussian noise (AWGN) channel with M-ary amplitude shift keying (ASK) input alphabet $\{\pm 1, \pm 3, \ldots, \pm(M-1)\}$, the uniform factor is the distribution of the symbol signs, and the non-uniform factor is the distribution of the symbol amplitudes, hence the name probabilistic amplitude shaping. In [7], the PAS encoding was extended to linear layered probabilistic shaping (LLPS), which also allows for shaped parity bits. The decoding rule for PAS [7] and LLPS [9] is essentially the maximum a posteriori (MAP) decoding rule, which takes into account the input shaping. The schemes proposed in [7], [9] suggest that in principle, linear codes can be used for channels with non-uniform input distribution, by encoding into those codewords that have the required distribution.

For instance, suppose that for a channel with input distribution P_X, only codewords of type P_X are used. Thus, the set of codewords that are actually transmitted forms a constant composition code [11, Chapter 10], yet the decoder decodes as it would even if the entire (linear) codebook was used. Therefore, one can think of this structure as a linear extension of the constant composition code. This introduces a *codebook mismatch* between the set of transmitted codewords (e.g., a constant composition code) and the set of hypotheses for decoding (e.g., the linear code).

Several previous works have discussed codebook mismatch, linear codes, and additive metrics for analyzing probabilistic shaping schemes. In [1, Section I], codebook mismatch was discussed and judged to be suboptimal. The works [4], [5, Chapter 7], [7], and [8], model codebook mismatch by considering a random code generated according to a uniform distribution as the extended codebook used for decoding. In [4], a discrete memoryless channel (DMC) is considered and a typicality decoder is used. Discrete-input, continuous-output channels and additive decoding metrics are considered in [5], [7], [8], and achievable rates for successful encoding and successful decoding are analyzed separately. The encoding error analysis by typicality in [5, Section 7.3] is simplified in [7, Appendix A] by using a simple counting argument. In [8], encoding and decoding rates are combined to an achievable rate without further proof. The work [5, Chapter 10] derives an error exponent for PAS and shows that PAS using constant composition DM (CCDM) [18] and MAP decoding on a random linear extension code achieves the mutual information of discrete-input, continuous-output memoryless channels. This result implies that PAS achieves capacity for a number of practically relevant channels, including the AWGN channel with ASK input. Note that the related works [2] and [15] do not model the codebook mismatch in their analysis. In [2], a joint source channel coding scenario is considered, while in [15], decoding is performed on the set of shaped sequences and no extension code is considered for decoding.

It is this background that motivates the study of codebook mismatch with linear extension codes and additive decoding metrics that we conduct in this work. Instead of considering the PAS configuration, we examine the following, simpler, and more general setup, whose focus is on harnessing linear codes for constant-composition coding: Consider the ensemble of linear codes, where the encoder uses only a subset of the codebook that forms a constant-composition code (corresponding to a certain type class), whereas the decoder uses an arbitrary additive decoding metric (e.g., the ML or MAP decoding metric) and decodes the same way as if the full linear code was used, ignoring the fact that non-typical codewords are actually never used by the encoder. The motivation for the decoder not to discard the non-typical codewords is in order to maintain the linear structure of the code, along with its benefits, as described earlier. The main questions that we concern ourselves with are the following:

1. For a given discrete memoryless channel (DMC) and a given decoding metric, what is the random coding error exponent associated with this setting?
2. Considering the fact that, due to the codebook mismatch, the decoder is sub-optimal, can we choose an alternative additive decoding metric that would compensate for the codebook mismatch?

In response to these two questions, we first derive the exact error exponent for linear codes under code mismatch for a given, additive decoding metric, and then show that by optimizing this decoding metric, we can improve the random coding exponent so as to coincide with that of general (non-linear) constant composition codes [11, Theorem 10.2], which means, among other things, that capacity is achieved. More specifically, regarding item no. 1 above, we show that the error exponent of any given additive metric cannot be larger than the random coding exponent of the ensemble of fixed composition codes, but on the other hand, with regard to item no. 2, we fully characterize the optimal metric that achieves this upper bound. We further show that as rate approaches mutual information, the optimal metric becomes the MAP metric, implying that MAP decoding on the linear extension code achieves capacity, given that the constant composition is the capacity-achieving input distribution.

The remainder of this work is organized as follows. In Section 2, we establish the notation conventions. In Section 3, we formalize the problem setting and spell out the objectives of this work. In Section 4, we present the main theorems and discuss them. Finally, in Section 5, we prove the theorems.

2 Notation

Throughout the paper, random variables will be denoted by capital letters, specific values they may take will be denoted by the corresponding lower case letters, and their alphabets will be denoted by calligraphic letters. Random vectors and their realizations will be denoted, respectively, by capital letters and the corresponding lower case letters, both in the bold face font. Their alphabets will be superscripted by their dimensions. For example, the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, ($n$ – positive integer) may take a specific vector value $\boldsymbol{x} = (x_1, \ldots, x_n)$ 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 denoted by the letter P, Q, and W, sometimes subscripted by the names of the relevant random variables/vectors and their conditionings, if applicable, following the standard notation conventions, e.g., P_X, $Q_{Y|X}$, and so on. When there is no room for ambiguity, these subscripts will be omitted. The probability of an event $\mathcal{E}$ will be denoted by $\Pr\{\mathcal{E}\}$, and the expectation operator will be denoted by $\boldsymbol{E}\{\cdot\}$. For two positive sequences a_n and b_n, the notation $a_n \doteq b_n$ will stand for equality in the exponential scale, that is, $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$. Similarly, $a_n \overset{\cdot}{\leq} b_n$ means that $\limsup_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} \leq 0$, and so on. The indicator function of an event $\mathcal{E}$ will be denoted by $\mathcal{I}\{E\}$. The notation $[x]_+$ will stand for $\max\{0, x\}$. Logarithms will be defined to the base 2, unless specified otherwise.

The empirical distribution of a sequence $\boldsymbol{x} \in \mathcal{X}^n$, which will be denoted by $\hat{P}_{\boldsymbol{x}}$, is the vector of relative frequencies $\hat{P}_{\boldsymbol{x}}(x)$ of each symbol $x \in \mathcal{X}$ in $\boldsymbol{x}$. The type class of $\boldsymbol{x} \in \mathcal{X}^n$, denoted $\mathcal{T}(\boldsymbol{x})$, is the set of all vectors $\boldsymbol{x}'$ with $\hat{P}_{\boldsymbol{x}'} = \hat{P}_{\boldsymbol{x}}$. When we wish to emphasize the dependence of the type class on the empirical distribution, say P, we will denote it by $\mathcal{T}(P)$. Information measures associated with empirical distributions will be denoted with 'hats' and will be subscripted by the sequences from which they are induced. For example, the entropy associated with $\hat{P}_{\boldsymbol{x}}$, which is the empirical entropy of $\boldsymbol{x}$, will be denoted by $\hat{H}_{\boldsymbol{x}}(X)$. An alternative notation, following the conventions described in the previous paragraph, is $H(\hat{P}_{\boldsymbol{x}})$. 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, $\hat{P}_{\boldsymbol{xy}}$ would be the joint empirical distribution of $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_i, y_i)\}_{i=1}^n$, $\mathcal{T}(\boldsymbol{x}, \boldsymbol{y})$ or $\mathcal{T}(\hat{P}_{\boldsymbol{xy}})$ will denote the joint type class of $(\boldsymbol{x}, \boldsymbol{y})$, $\mathcal{T}(\boldsymbol{x}|\boldsymbol{y})$ will stand for the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$, $\hat{H}_{\boldsymbol{xy}}(X, Y)$ will designate the empirical joint entropy of $\boldsymbol{x}$ and $\boldsymbol{y}$, $\hat{H}_{\boldsymbol{xy}}(X|Y)$ will be the empirical conditional entropy, $\hat{I}_{\boldsymbol{xy}}(X; Y)$ will denote empirical mutual information, and so on.

Given a fixed probability assignment, P_X, of a random variable X, and given a generic conditional distribution $Q_{Y|X}$, we denote the induced information measures using the conventional notation rules of the information theory literature, but with the subscript Q. For example, $I_Q(X; Y)$, $H_Q(Y)$, $H_Q(Y|X)$, and $H_Q(X|Y)$ will denote, respectively, the mutual information between X and Y, the marginal entropy of Y, the conditional entropy of Y given X and the conditional entropy of X given Y, all induced by $P_X \times Q_{Y|X}$. Likewise, given P_X and $Q_{Y|X}$, the induced conditional distribution of X given Y will be denoted by $Q_{X|Y}$. The same notation conventions will apply whenever other auxiliary random variables will be involved, such as $X' \sim P_X$. The weighted Kullback-Leibler divergence between two conditional distributions, say, $Q_{Y|X}$ and $W = \{W(y|x),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$, is defined as

$$D(Q_{Y|X} \| W | P_X) = \sum_{x \in \mathcal{X}} P_X(x) \sum_{y \in \mathcal{Y}} Q_{Y|X}(y|x) \log \frac{Q_{Y|X}(y|x)}{W(y|x)}. \tag{1}$$

3 Problem Setting

We consider coded communication via a discrete memoryless channel (DMC) with a finite input alphabet $\mathcal{X}$, a finite output alphabet $\mathcal{Y}$, and a single–letter transition probability matrix, $W = \{W(y|x),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$. When the channel is fed by an input vector $\boldsymbol{X} = \boldsymbol{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$, it outputs a random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n) \in \mathcal{Y}^n$, according to the conditional probability distribution,

$$\Pr\{\boldsymbol{Y} = \boldsymbol{y} | \boldsymbol{X} = \boldsymbol{x}\} = W(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} W(y_i|x_i). \tag{2}$$

Without essential loss of generality, we assume the cardinality of $\mathcal{X}$ to be a power of two, i.e., $|\mathcal{X}| = 2^m$ for some positive integer m. In the absence of this property, one can always formally extend $\mathcal{X}$ to be of the size of $\exp_2(\lceil \log_2 |\mathcal{X}| \rceil)$, by adding to $\mathcal{X}$ some dummy input symbols that are never actually used. We adopt this assumption in order to allow the restriction to binary linear codes, and thereby simplify the notation and the derivations.

The transmitter is assumed to employ a binary linear code of block length nm and code dimension $k = nmr_{\text{fec}}$, where $0 < r_{\text{fec}} \leq 1$.

Remark 1. For r_{fec}, the subscript FEC stands for forward error correction and emphasizes following [7] that r_{fec} is the code rate of the employed binary linear FEC code. As we will see below, the effective rate R at which information is transmitted over the channel also depends on the type P_X of the constant composition, i.e., it is not determined by the FEC code rate r_{fec} alone.

The encoding mechanism is as follows. An information message, $w \in \{0, 1, 2, \ldots, 2^k - 1\}$, with a binary representation denoted by $\boldsymbol{b}(w) \in \{0, 1\}^k$, is mapped into a codeword $\boldsymbol{c}(w)$

according to

$$\boldsymbol{c}(w) = \boldsymbol{b}(w) \cdot G + \boldsymbol{v}, \tag{3}$$

where $\boldsymbol{G} \in \{0,1\}^{k \times nm}$, $\boldsymbol{v} \in \{0,1\}^{nm}$, and where the entries of $\boldsymbol{G}$ and $\boldsymbol{v}$ are selected independently at random according to the uniform distribution over $\{0,1\}$. The binary codeword $\boldsymbol{c}(w)$ is mapped into a channel input vector $\boldsymbol{x}(w) \in \mathcal{X}^n$ via a labeling function $\phi\colon \{0,1\}^m \to \mathcal{X}$ that indexes the symbols in the channel alphabet $\mathcal{X}$ by m bits, i.e.,

$$\boldsymbol{c}(w) \to \boldsymbol{x}(w) = \phi(c_1(w) \dots c_m(w))\phi(c_{m+1}(w) \dots c_{2m}(w)) \dots \phi(c_{(n-1)m+1}(w) \dots c_{nm}(w)), \tag{4}$$

$c_i(w)$, $i = 0, \dots, nm-1$, being the components of $\boldsymbol{c}(w)$. We denote the codebook

$$\mathcal{C} = \{\boldsymbol{x}(w),\ w \in \{0, 1, \dots, 2^k - 1\}\}. \tag{5}$$

In contrast to the traditional setting, where all codewords of the linear code are used, in this work, we consider the case where only a subset of the codebook is used, namely, codewords, $\{\boldsymbol{x}(w)\}$, which belong to a given type class, $\mathcal{T}(P_X)$, where P_X is a certain empirical distribution over $\mathcal{X}$. Since $|\mathcal{T}(P_X)| \doteq 2^{nH(P_X)}$, and since the code partitions the Hamming space, $\{0,1\}^{nm}$, into $2^{nm(1-r_{\text{fec}})}$ disjoint cosets, each of size $2^{nmr_{\text{fec}}}$, then there must be at least one coset that includes at least

$$\frac{|\mathcal{T}(P_X)|}{2^{nm(1-r_{\text{fec}})}} \doteq \frac{2^{nH(P_X)}}{2^{nm(1-r_{\text{fec}})}} = 2^{n[H(P_X) - m(1-r_{\text{fec}})]}$$

codewords in $\mathcal{T}(P_x)$. Our encoder will use (a subset of) such a coset, henceforth denoted $\mathcal{C}'$, to encode information at the rate,

$$R = H(P_X) - m(1 - r_{\text{fec}}), \tag{6}$$

where we keep in mind the well known fact that the probability of error does not depend on the coset representative. Thus, the effective coding rate, R, associated with $\mathcal{C}'$, is always less than or equal to mr_{fec}, with equality iff P_X is the uniform distribution over $\mathcal{X}$.

At the receiver side, a metric decoder is used. The decoding metric is a function $U : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$, which induces the following decoding rule:

$$\hat{w}(\boldsymbol{y}) = \arg\max_{w \in \{0,1,\dots,2^k-1\}} U(\boldsymbol{x}(w), \boldsymbol{y}), \tag{7}$$

where

$$U(\boldsymbol{x}(w),\boldsymbol{y})=\prod_{i=1}^{n}U(x_i(w),y_i), \tag{8}$$

$x_i(w)$, $i=1,\ldots,n$, being the components of $\boldsymbol{x}(w)$. Note that for practical reasons (discussed in the Introduction), the decoder examines *all codewords of the original linear code,* $\mathcal{C}$, not only those of the subcode, $\mathcal{C}'$, of P_X-typical codewords. In other words, $\boldsymbol{x}(\hat{w}(\boldsymbol{y}))$ can be any codeword in $\mathcal{C}$, not necessarily in $\mathcal{C}'$.

The probability of error for a given $\boldsymbol{w}$ (with $\boldsymbol{x}(w)\in\mathcal{C}'$) and a given code $\mathcal{C}$, is defined as

$$P_{\mathrm{e}|w}(\mathcal{C})=\Pr\{\hat{w}(\boldsymbol{Y})\neq w\}, \tag{9}$$

where the randomness of $\boldsymbol{Y}$ is due to the channel only. The average error probability is defined as

$$\bar{P}_{\mathrm{e}}=\boldsymbol{E}\left\{\frac{1}{2^{nR}}\sum_{\{w:\ \boldsymbol{x}(w)\in\mathcal{C}'\}}\bar{P}_{\mathrm{e}|w}(\mathcal{C})\right\}, \tag{10}$$

where the expectation is with respect to the randomness of $(\boldsymbol{G},\boldsymbol{v})$. The random coding error exponent is defined as

$$E_{\mathrm{r}}(R)=\lim_{n\to\infty}\left\{-\frac{\log\bar{P}_{\mathrm{e}}}{n}\right\}, \tag{11}$$

provided that the limit exists. In the sequel, whenever we need to emphasize the dependence of the random coding error exponent upon the decoding metric, U, we will denote it by $E_{\mathrm{r}}(R,U)$.

Observe that our setting exhibits a situation of *codebook mismatch*: While the encoder uses only the subcode, $\mathcal{C}'$, the decoder acts as if the entire larger code, $\mathcal{C}$, was fully used, without taking advantage of the knowledge that $\boldsymbol{x}(w)$ must be in $\mathcal{C}'$. This is done in order to avoid ruining the linear structure of the code, which is useful for fast decoding. Consequently, the decoder is suboptimal even if its decoding metric is the maximum likelihood metric, $U(x,y)=U_{\mathrm{ML}}(x,y)\stackrel{\triangle}{=}W(y|x)$. The question that we study, in this work, is whether $U_{\mathrm{ML}}(x,y)$ can be replaced by another decoding metric, $U(x,y)$, that would compensate for the codebook mismatch.

Our main result in this work is in answering this question affirmatively. To this end, we first derive a single–letter formula for $E_{\mathrm{r}}(R,U)$, for a given, arbitrary decoding metric, U, and then we demonstrate that by maximizing $E_{\mathrm{r}}(R,U)$ w.r.t. U, we can significantly improve over $E_{\mathrm{r}}(R, U_{\mathrm{ML}})$ as well as over $E_{\mathrm{r}}(R, U_{\mathrm{MAP}})$, where $U_{\mathrm{MAP}}(x,y) \stackrel{\triangle}{=} P_X(x)W(y|x)$. The comparison to U_{MAP} is relevant because, in spite of the mismatch, it still achieves coding rates arbitrarily close to the 'capacity' associated with P_X, namely, the mutual information induced by $P_X \times W$. Moreover, our main result is in proving that, on the one hand, $E_{\mathrm{r}}(R, U)$ cannot exceed the random coding exponent of (non-linear) fixed composition codes [11, Theorem 10.2]:

$$E_{\mathrm{r}}^{\mathrm{cc}}(R) = \min_{Q_{Y|X}} \left\{ D(Q_{Y|X} \| W | P_X) + [I_Q(X;Y) - R]_+ \right\}. \tag{12}$$

but on the other hand, we characterize the optimal decoding metric, U_*, and show that it achieves $E_{\mathrm{r}}^{\mathrm{cc}}(R)$.

4 Main Results

Our main result is in the following theorem, whose proof can be found in Appendix A.

Theorem 1. *Consider the setting formulated in Section 3.*

1. *For a given decoding metric U,*

$$E_{r}(R, U) = \max_{0 \le \rho \le 1} \sup_{\theta \ge 0} \left\{ -\sum_{x \in \mathcal{X}} P_X(x) \log \left(\sum_{y \in \mathcal{Y}} \frac{W(y|x)}{[U_\theta(x|y)]^\rho} \right) + \rho [H(P_X) - R] \right\}, \tag{13}$$

where

$$U_\theta(x|y) \stackrel{\triangle}{=} \frac{[U(x,y)]^\theta}{\sum_{x' \in \mathcal{X}} [U(x', y)]^\theta}. \tag{14}$$

2. *For every metric U, $E_{r}(R, U) \le E_{r}^{cc}(R)$.*

3. *For every given $\rho \in [0,1]$, assume that there exists a vector $Z = Z_\rho \stackrel{\triangle}{=} \{Z_\rho(x),\ x \in \mathcal{X}\}$, with strictly positive components, that satisfies the system of simultaneous equations,*

$$Z_\rho(x) = \sum_y [W(y|x)]^{1/(1+\rho)} \left[\sum_{x'} \frac{P_X(x')[W(y|x')]^{1/(1+\rho)}}{Z(x')} \right]^\rho, \qquad \forall\ x \in \mathcal{X}, \tag{15}$$

and define

$$U(x|y,\rho) = \frac{P_X(x)[W(y|x]^{1/(1+\rho)}/Z_\rho(x)}{\sum_{x'} P_X(x')[W(y|x')]^{1/(1+\rho)}/Z_\rho(x')}, \tag{16}$$

and

$$U_\theta(x|y,\rho) = \frac{[U(x|y,\rho)]^\theta}{\sum_{x'} [U(x'|y,\rho)]^\theta}. \tag{17}$$

Finally, let

$$\rho_\star = \textit{arg max}_{\rho\in[0,1]} \sup_{\theta\geq 0}\left\{-\sum_{x\in\mathcal{X}} P_X(x)\log\left(\sum_{y\in\mathcal{Y}}\frac{W(y|x)}{[U_\theta(x|y,\rho)]^\rho}\right)+\rho[H(P_X)-R]\right\}. \tag{18}$$

Then, the metric $U_\star \overset{\triangle}{=} \{U(x|y,\rho_\star),\ x\in\mathcal{X},\ y\in\mathcal{Y}\}$ *achieves* $E_r^{cc}(R)$.

Remark 2. The expression of $E_{\rm r}(R,U)$ does not seem to lend itself to closed form derivation of U^* using traditional optimization techniques, and therefore, the proof of the third part of the theorem will be based on a chain of inequalities relating $E_{\rm r}(R,U)$ and $E_{\rm r}^{\rm cc}(R)$ and examining the conditions under which the inequalities become equalities. As can be seen, the optimal metric, $U_\star$, that maximizes $E_{\rm r}(R,U)$, depends, in quite a complicated manner, not only on the input assignment, P_X, and the channel, W, but also on the coding rate, R, via ρ^*.

The remaining part of this section is devoted to a discussion on Theorem 1.

4.1 Numerical Example of Error Exponents

In Figure 1, we compare the error exponent for constant composition codes, achieved by $U_\star$, to those associated with $U_{\rm MAP}$ and $U_{\rm ML}$. The example considered refers to a quantized Gaussian channel with input distribution

$$P_X(-3)=P_X(3)=0.05,\quad P_X(-1)=P_X(1)=0.45. \tag{19}$$

The channel output is quantized to 4 levels and the noise variance is chosen such that the mutual information between input X and the quantized output is 0.5 bits. The equivalent channel matrix is

$$W=\begin{bmatrix}0.8036 & 0.1964 & 0.0052 & 0.0000\\ 0.1912 & 0.6072 & 0.1912 & 0.0052\\ 0.0052 & 0.1912 & 0.6072 & 0.1912\\ 0.0000 & 0.0052 & 0.1964 & 0.8036\end{bmatrix} \tag{20}$$

where the ith column is a distribution on the output alphabet, given that the ith input symbol was transmitted.

We note that the gaps are considerable.

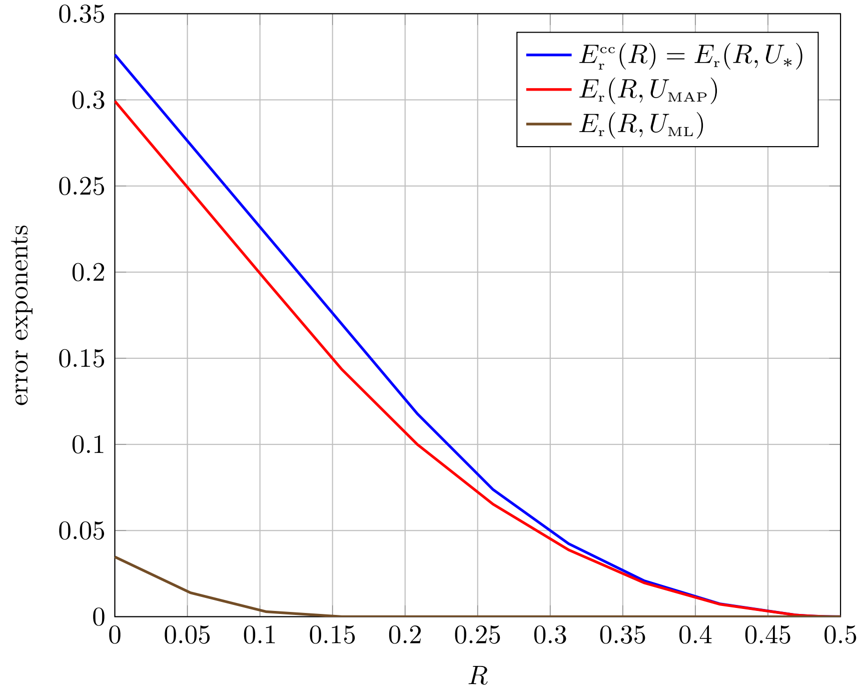

Figure 1: Comparison of the random coding exponents associated with the ML metric, U_{ML}, the MAP metric, U_{MAP}, and the random coding exponent of the ensemble of non-linear constant composition codes (12), which is achieved by $U_\star$.

4.2 The MAP Metric Achieves $I(X;Y)$

We argue that a necessary and sufficient condition for a metric $\{U(x|y),\ x \in \mathcal{X},\ y \in \mathcal{Y}\}$ to achieve the mutual information rate, $I(X;Y)$ (induced by P_X and W), is $U(x|y) = P_{X|Y}(x|y)$ (or an equivalent metric), where $P_{X|Y}$ is the posterior distribution induced by $P_X \times W$. Indeed, let $R = I(X;Y) - \epsilon$ be given, where $\epsilon > 0$ is arbitrarily small. Then, for $E_{\mathrm{r}}(I(X;Y) - \epsilon, U)$ to be strictly positive, there must exist $\rho \in [0,1]$ and $\theta \geq 0$ such that

$$-\sum_x P_X(x) \log\left(\sum_y \frac{W(y|x)}{[U_\theta(x|y)]^\rho}\right) + \rho[H(X|Y) + \epsilon] > 0. \tag{21}$$

Let $\theta > 0$ be such that there exists $\rho \in [0,1]$ for which this condition holds true. Now, for this given θ, consider the function,

$$F(\rho) \stackrel{\triangle}{=} -\sum_x P_X(x) \log\left(\sum_y \frac{W(y|x)}{[U_\theta(x|y)]^\rho}\right) + \rho[H(X|Y) + \epsilon]. \tag{22}$$

It is easy to see that $F(0) = 0$ and that $F(\rho)$ is concave, which means that the derivative, $F'(\rho)$, is monotonically non-increasing. Therefore a necessary and sufficient condition for the existence of $\rho \in [0,1]$ such that $F(\rho) > 0$ (i.e., $\max_{\rho \in [0,1]} F(\rho) > 0$) is that $F'(0) > 0$.

Now,

$$\begin{aligned}
F'(0) &= -\sum_x P_X(x)\cdot\frac{\sum_y W(y|x)[U_\theta(x|y)]^{-0}\log[1/U_\theta(x|y)]}{\sum_y W(y|x)[U_\theta(x|y)]^{-0}}+H(X|Y)+\epsilon &(23)\\
&= \sum_x P_X(x)\sum_y W(y|x)\log U_\theta(x|y)+H(X|Y)+\epsilon &(24)\\
&= -D(P_{X|Y}\|U_\theta|P_Y)+\epsilon. &(25)
\end{aligned}$$

By the arbitrariness of $\epsilon > 0$, it follows that in order that $F'(0) > 0$ for all $\epsilon > 0$, the divergence term must vanish, and so, we must have that $U_\theta(x|y) = P_{X|Y}(x|y)$ for all $x \in \mathcal{X}$ and for all y such that $P_Y(y) > 0$.

Note that the optimal metric U_* agrees with the MAP metric when $R = I(X;Y)$, which corresponds to $\rho \to 0$, because in this case, the power $1/(1+\rho)$ tends to unity and $Z_\rho(x)$ tends to 1 for all x.

5 Proof of Theorem 1

Beginning from the first part of the theorem, let $\boldsymbol{x}$ and $\boldsymbol{y}$ be the transmitted codeword and the received channel output vector, respectively. Let $P_{\mathrm{e}}(\boldsymbol{x},\boldsymbol{y})$ denote the expected error probability given that $\boldsymbol{x}$ was transmitted and $\boldsymbol{y}$ was received, where the expectation is with respect to the randomness of all other codewords in $\mathcal{C}$. Then,

$$\begin{aligned}
\bar{P}_{\mathrm{e}}(\boldsymbol{x},\boldsymbol{y}) &= \Pr\left[\bigcup_{\boldsymbol{x}'\in\mathcal{C}\setminus\{\boldsymbol{x}\}}\left\{U(\boldsymbol{x}',\boldsymbol{y})\ge U(\boldsymbol{x},\boldsymbol{y})\right\}\right]\\
&\le \min\left\{1, 2^{k-1}\sum_{\{\boldsymbol{x}':\ U(\boldsymbol{x}',\boldsymbol{y})\ge U(\boldsymbol{x},\boldsymbol{y})\}}2^{-nm}\right\}\\
&\le \min\left\{1, |\mathcal{M}(\boldsymbol{x},\boldsymbol{y})|2^{k-nm}\right\}, &(26)
\end{aligned}$$

where

$$\mathcal{M}(\boldsymbol{x},\boldsymbol{y}) = \{\boldsymbol{x}':\ U(\boldsymbol{x}',\boldsymbol{y})\ge U(\boldsymbol{x},\boldsymbol{y})\}, \tag{27}$$

and where we have used the (truncated) union bound, the union being taken over all 2^{k-1} pairwise error events, where an incorrect codeword $\boldsymbol{x}'$, randomly drawn under the uniform distribution over $\{0,1\}^{nm}$, happens to have a metric score, $U(\boldsymbol{x}',\boldsymbol{y})$, that exceeds the one of the transmitted codeword, $U(\boldsymbol{x},\boldsymbol{y})$. In Appendix A, we show that this truncated union bound is exponentially tight in spite of the fact that the codewords of a random linear code are not mutually independent. This is done by deriving a lower bound of the same exponential order.

Since $U(\boldsymbol{x}', \boldsymbol{y})$ depends on $(\boldsymbol{x}', \boldsymbol{y})$ only via their joint type, we can assess the cardinality of $M(\boldsymbol{x}, \boldsymbol{y})$ by the method of types [11] as

$$
\begin{aligned}
|\mathcal{M}(\boldsymbol{x}, \boldsymbol{y})| &= \sum_{\{\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y}):\ U(\boldsymbol{x}', \boldsymbol{y}) \geq U(\boldsymbol{x}, \boldsymbol{y})\}} |\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y})| \\
&\doteq \sum_{\{\mathcal{T}(\boldsymbol{x}'|\boldsymbol{y}):\ \log U(\boldsymbol{x}', \boldsymbol{y}) \geq \log U(\boldsymbol{x}, \boldsymbol{y})\}} \exp_2 \left\{ n \hat{H}_{\boldsymbol{x}'\boldsymbol{y}}(X'|Y) \right\} \\
&= \exp_2 \left\{ n \max_{Q_{X'|Y} \in \mathcal{E}(Q_{XY})} H_Q(X'|Y) \right\}. \qquad (28)
\end{aligned}
$$

where

$$
\mathcal{E}(Q_{XY}) = \left\{ Q_{X'|Y} :\ \sum_{x,y} Q_{X'Y}(x,y) \log U(x,y) \geq \sum_{x,y} Q_{XY}(x,y) \log U(x,y) \right\}. \qquad (29)
$$

It follows that

$$
\begin{aligned}
\bar{P}_{\mathrm{e}}(\boldsymbol{x}, \boldsymbol{y}) &\doteq \min \left\{ 1, \exp_2 \left[k - nm + n \cdot \max_{Q_{X'|Y} \in \mathcal{E}(Q_{XY})} H_Q(X'|Y) \right] \right\} \\
&= \exp_2 \left\{ -n \left[m(1 - r_{\mathrm{fec}}) - \max_{Q_{X'|Y} \in \mathcal{E}(Q_{XY})} H_Q(X'|Y) \right]_+ \right\} \\
&= \exp_2 \left\{ -n \left[H(P_X) - R - \max_{Q_{X'|Y} \in \mathcal{E}(Q_{XY})} H_Q(X'|Y) \right]_+ \right\}, \qquad (30)
\end{aligned}
$$

where in the last equality we have used the relation (6). Averaging over the randomness of $\boldsymbol{Y}$, we get

$$
\begin{aligned}
\bar{P}_{\mathrm{e}}(\boldsymbol{x}) &= \sum_{\boldsymbol{y} \in \mathcal{Y}^n} W(\boldsymbol{y}|\boldsymbol{x}) P_{\mathrm{e}}(\boldsymbol{x}, \boldsymbol{y}) \\
&\doteq \sum_{\mathcal{T}(\boldsymbol{y}|\boldsymbol{x})} |\mathcal{T}(\boldsymbol{y}|\boldsymbol{x})| \cdot \exp_2 \left\{ -n \left[H(P_X) - R - \max_{Q_{X'|Y} \in \mathcal{E}(Q_{XY})} H_Q(X'|Y) \right]_+ \right\} \\
&\doteq \exp_2 \Bigg\{ -n \min_{Q_{Y|X}} \Bigg(D(Q_{Y|X} \| W | P_X) + \\
&\qquad \left[H(P_X) - R - \max_{Q_{X'|Y} \in \mathcal{E}(Q_{XY})} H_Q(X'|Y) \right]_+ \Bigg) \Bigg\}, \qquad (31)
\end{aligned}
$$

and since this expression depends on $\boldsymbol{x}$ only via its type class $\hat{P}_{\boldsymbol{x}} = P_X$, then the same formula holds also for the average error probability, which includes also the expectation w.r.t. the randomness of the transmitted codeword, $\boldsymbol{x}$, i.e.,

$$
\bar{P}_{\mathrm{e}} \doteq \exp_2 \Bigg\{ -n \min_{Q_{Y|X}} \Bigg(D(Q_{Y|X} \| W | P_X) +
$$

$$\left[H(P_X) - R - \max_{Q_{X'|Y} \in \mathcal{E}(Q_{XY})} H_Q(X'|Y)\right]_+\Bigg)\Bigg\}. \tag{32}$$

It should be pointed out that this expression of the average error probability is very similar to the one obtained for the ensemble of non-linear constant composition codes for a given decoding metric, U. There is one important difference, however: In the case of non-linear fixed composition codes, the definition of the set $\mathcal{E}(Q_{XY})$ should include the additional constraint that $\sum_{y \in \mathcal{Y}} Q_Y(y) Q_{X'|Y}(x|y) = P_X(x)$ for every $x \in \mathcal{X}$, whereas here, this constraint is absent.

We next derive the Lagrange-dual to this expression. Beginning from the inner maximization, we have

$$\begin{aligned}
&\max_{Q_{X'|Y} \in \mathcal{E}(Q_{XY})} H_Q(X'|Y) \\
&= \max_{Q_{X'|Y}} \inf_{\theta \geq 0} \Bigg\{ H_Q(X'|Y) + \theta \sum_y Q_Y(y) \Bigg[\sum_x Q_{X'|Y}(x|y) \log U(x,y) - \\
&\quad \sum_x Q_{X|Y}(x|y) \log U(x,y) \Bigg]\Bigg\}.
\end{aligned} \tag{33}$$

Moving on to the outer minimization, we obtain

$$\begin{aligned}
E_{\mathrm{r}}(R,U) &= \min_{Q_{Y|X}} \left(D(Q_{Y|X} \| W | P_X) + \left[H(P_X) - R - \max_{Q_{X'|Y} \in \mathcal{E}(Q_{XY})} H_Q(X'|Y) \right]_+ \right) \\
&= \min_{Q_{Y|X}} \max_{0 \leq \rho \leq 1} \left(D(Q_{Y|X} \| W | P_X) + \rho \left[H(P_X) - R - \max_{Q_{X'|Y} \in \mathcal{E}(Q_{XY})} H_Q(X'|Y) \right] \right) \\
&= \min_{Q_{Y|X}} \max_{0 \leq \rho \leq 1} \Bigg(D(Q_{Y|X} \| W | P_X) + \rho \Bigg[H(P_X) - R - \max_{Q_{X'|Y}} \inf_{\theta \geq 0} \Bigg\{ H_Q(X'|Y) + \\
&\quad \theta \sum_y Q_Y(y) \Bigg[\sum_x Q_{X'|Y}(x|y) \log U(x,y) - \sum_x Q_{X|Y}(x|y) \log U(x,y) \Bigg] \Bigg\} \Bigg] \Bigg) \\
&= \min_{Q_{Y|X}} \max_{0 \leq \rho \leq 1} \min_{Q_{X'|Y}} \sup_{\theta \geq 0} \Bigg(D(Q_{Y|X} \| W | P_X) + \rho [H(P_X) - R] - \rho H_Q(X'|Y) - \\
&\quad \rho\theta \sum_y Q_Y(y) \Bigg[\sum_x Q_{X'|Y}(x|y) \log U(x,y) - \sum_x Q_{X|Y}(x|y) \log U(x,y) \Bigg] \Bigg) \\
&\overset{\text{(a)}}{=} \min_{Q_{Y|X}} \max_{0 \leq \rho \leq 1} \min_{Q_{X'|Y}} \sup_{\theta \geq 0} \Bigg(D(Q_{Y|X} \| W | P_X) + \rho [H(P_X) - R] - \rho H_Q(X'|Y) - \\
&\quad \hat{\theta} \sum_y Q_Y(y) \Bigg[\sum_x Q_{X'|Y}(x|y) \log U(x,y) - \sum_x Q_{X|Y}(x|y) \log U(x,y) \Bigg] \Bigg) \\
&\overset{\text{(b)}}{=} \min_{Q_{Y|X}} \max_{0 \leq \rho \leq 1} \sup_{\hat{\theta} \geq 0} \min_{Q_{X'|Y}} \Bigg(D(Q_{Y|X} \| W | P_X) + \rho [H(P_X) - R] - \rho H_Q(X'|Y) - \\
&\quad \hat{\theta} \sum_y Q_Y(y) \Bigg[\sum_x Q_{X'|Y}(x|y) \log U(x,y) - \sum_x Q_{X|Y}(x|y) \log U(x,y) \Bigg] \Bigg),
\end{aligned} \tag{34}$$

where in (a) we have defined $\hat{\theta} = \rho\theta$ and in (b) we have used the fact that the objective is convex in $Q_{X'|Y}$ and affine in θ. Since the objective function is affine in $(\rho, \hat{\theta})$, then after inner-most minimization over $Q_{X'|Y}$ it becomes concave in $(\rho, \hat{\theta})$. The inner most minimization amounts to the maximization

$$\begin{aligned}
&\max_{Q_{X'|Y}} \left\{ \rho H_Q(X'|Y) + \hat{\theta} \sum_y Q_Y(y) \sum_x Q_{X'|Y}(x|y) \log U(x,y) \right\} \\
&= \rho \max_{Q_{X'|Y}} \sum_y Q_Y(y) \sum_x Q_{X'|Y}(x|y) \log \frac{[U(x,y)]^{\hat{\theta}/\rho}}{Q_{X'|Y}(x|y)} \\
&= \rho \cdot \sum_y Q_Y(y) \log \left(\sum_x [U(x,y)]^{\hat{\theta}/\rho} \right) \\
&\stackrel{\triangle}{=} \rho \cdot \sum_y Q_Y(y) \log Z(y, \hat{\theta}/\rho). \qquad (35)
\end{aligned}$$

On substituting this back into the expression of $E_{\mathrm{r}}(R,U)$, we have

$$\begin{aligned}
E_{\mathrm{r}}(R,U) &= \min_{Q_{Y|X}} \max_{0 \le \rho \le 1} \sup_{\hat{\theta} \ge 0} \Bigg(D(Q_{Y|X} \| W | P_X) + \rho[H(P_X) - R] - \\
&\quad \rho \cdot \sum_{x,y} P_X(x) Q_{Y|X}(y|x) \log Z(y, \hat{\theta}/\rho) + \hat{\theta} \sum_x P_X(x) Q_{Y|X}(y|x) \log U(x,y) \Bigg) \\
&\stackrel{\text{(a)}}{=} \max_{0 \le \rho \le 1} \sup_{\hat{\theta} \ge 0} \min_{Q_{Y|X}} \Bigg(\sum_{x,y} P_X(x) Q_{Y|X}(y|x) \log \frac{Q_{Y|X}(y|x)}{W(y|x)} + \rho[H(P_X) - R] - \\
&\quad \rho \cdot \sum_{x,y} P_X(x) Q_{Y|X}(y|x) \log Z(y, \hat{\theta}/\rho) + \hat{\theta} \sum_x P_X(x) Q_{Y|X}(y|x) \log U(x,y) \Bigg) \\
&= \max_{0 \le \rho \le 1} \sup_{\hat{\theta} \ge 0} \min_{Q_{Y|X}} \left(\sum_{x,y} P_X(x) Q_{Y|X}(y|x) \log \frac{Q_{Y|X}(y|x) [U(x,y)]^{\hat{\theta}}}{W(y|x) [Z(y, \hat{\theta}/\rho)]^{\rho}} + \rho[H(P_X) - R] \right) \\
&= \max_{0 \le \rho \le 1} \sup_{\hat{\theta} \ge 0} \left(- \sum_x P_X(x) \log \left(\sum_y \frac{W(y|x) Z(y, \hat{\theta}/\rho)]^{\rho}}{[U(x,y)]^{\hat{\theta}}} \right) + \rho[H(P_X) - R] \right) \\
&\stackrel{\text{(b)}}{=} \max_{0 \le \rho \le 1} \sup_{\theta \ge 0} \left[- \sum_x P_X(x) \log \left(\sum_y \frac{W(y|x)}{[U_\theta(x|y)]^{\rho}} \right) + \rho[H(P_X) - R] \right], \qquad (36)
\end{aligned}$$

where in (a) we have used the fact that the objective is convex in $Q_{X'|Y}$ and concave in $(\rho, \hat{\theta})$, and in (b) we returned to the original optimization parameter, $\theta = \hat{\theta}/\rho$. This completes the proof of part 1 of the theorem.

Moving on to part 2 of the theorem, in Appendix B, we prove that the following (Lagrange-dual) expression may serve as an alternative representation of $E_{\mathrm{r}}^{\mathrm{cc}}(R)$:

$$E_{\mathrm{r}}^{\mathrm{cc}}(R) = \min_V \max_{0 \le \rho \le 1} \left\{ -(1+\rho) \sum_x P_X(x) \log \left[\sum_y (W(y|x)[V(y)]^{\rho})^{1/(1+\rho)} \right] - \rho R \right\}, \qquad (37)$$

where the minimization is over all probability assignments, $V=\{V(y)\}$ over $\mathcal{Y}$.

Next, consider the following chain of equalities and inequalities:

$$\begin{aligned}
&\sum_x P_X(x)\log\left(\sum_y \frac{W(y|x)}{[U_\theta(x|y)]^\rho}\right)\\
=&\sum_x P_X(x)\log\left(\sum_y \tilde{W}(y|x)\frac{W(y|x)}{[U_\theta(x|y)]^\rho\tilde{W}(y|x)}\right)\\
\overset{\text{(a)}}{\geq}&\max_{\tilde{W}}\sum_x P_X(x)\sum_y \tilde{W}(y|x)\log\left(\frac{W(y|x)}{[U_\theta(x|y)]^\rho\tilde{W}(y|x)}\right)\\
=&\max_{\tilde{W}}\left\{\sum_x P_X(x)\sum_y \tilde{W}(y|x)\log\left(\frac{W(y|x)}{\tilde{W}(y|x)}\right)-\rho\sum_{x,y}P_X(x)\tilde{W}(y|x)\log U_\theta(x|y)\right\}\\
\overset{\text{(b)}}{\geq}&\max_{\tilde{W}}\left\{\sum_x P_X(x)\sum_y \tilde{W}(y|x)\log\left(\frac{W(y|x)}{\tilde{W}(y|x)}\right)+\rho\tilde{H}(X|Y)\right\}\\
=&\max_{\tilde{W}}\left\{\sum_x P_X(x)\sum_y \tilde{W}(y|x)\log\left(\frac{W(y|x)}{\tilde{W}(y|x)}\right)+\rho[H(X)+\tilde{H}(Y|X)-\tilde{H}(Y)]\right\}\\
\overset{\text{(c)}}{=}&\max_{\tilde{W}}\max_V\sum_x P_X(x)\sum_y \tilde{W}(y|x)\Bigg[\log\left(\frac{W(y|x)}{\tilde{W}(y|x)}\right)+\\
&\rho H(X)-\rho\log\tilde{W}(y|x)+\rho\log V(y)\Bigg]\\
=&\max_V\max_{\tilde{W}}\sum_x P_X(x)\sum_y \tilde{W}(y|x)\left[\log\left(\frac{W(y|x)[V(y)]^\rho}{[\tilde{W}(y|x)]^{1+\rho}}\right)+\rho H(X)\right]\\
=&\max_V\max_{\tilde{W}}\sum_x P_X(x)\sum_y \tilde{W}(y|x)\left[(1+\rho)\log\left(\frac{(W(y|x)[V(y)]^\rho)^{1/(1+\rho)}}{\tilde{W}(y|x)}\right)+\rho H(X)\right]\\
=&\max_V\left\{(1+\rho)\sum_x P_X(x)\log\left(\sum_y (W(y|x)[V(y)]^\rho)^{1/(1+\rho)}\right)+\rho H(X)\right\}
\end{aligned}\tag{38}$$

where in (a) we have used Jensen's inequality, in (b) and onward, $\tilde{H}(Y|X)$ and $\tilde{H}(Y)$ refer to entropies induced by $P_X\times\tilde{W}$, and in (c), V is a probability distribution on $\mathcal{Y}$. Consequently,

$$\begin{aligned}
&-\sum_x P_X(x)\log\left(\sum_y \frac{W(y|x)}{[U_\theta(x|y)]^\rho}\right)+\rho[H(X)-R]\\
\leq&\min_V\left\{-(1+\rho)\sum_x P_X(x)\log\left(\sum_y (W(y|x)[V(y)]^\rho)^{1/(1+\rho)}\right)-\rho H(X)+\rho[H(X)-R)\right\}\\
=&\min_V\left\{-(1+\rho)\sum_x P_X(x)\log\left(\sum_y (W(y|x)[V(y)]^\rho)^{1/(1+\rho)}\right)-\rho R\right\}
\end{aligned}\tag{39}$$

and after maximizing both sides over $\rho\in[0,1]$ and $\theta\geq 0$, we get $E_{\mathrm{r}}(R,U)\leq E_{\mathrm{r}}^{\mathrm{cc}}(R)$, in view of eq. (37). This completes the proof of part 2 of the theorem.

To prove part 3 of the theorem, let us examine the conditions under which the inequalities in (38) become equalities: The first inequality, which is an application of Jensen's inequality, is met with equality if the expression,

$$\frac{W(y|x)}{[U_\theta(x|y)]^\rho \tilde{W}(y|x)}$$

is independent of $\boldsymbol{y}$, though it is still allowed to depend on x. In other words, $U_\theta(x|y)$ must satisfy

$$\frac{W(y|x)}{\tilde{W}(y|x)U_\theta(x|y)^\rho} = K(x), \tag{40}$$

for some function $K(x)$. The second inequality becomes equality if

$$U_\theta(x|y) = \frac{P_X(x)\tilde{W}(y|x)}{\sum_{x'} P(x')\tilde{W}(y|x')}. \tag{41}$$

Now, the optimal $\tilde{W}$, that achieves the last line of (38), is given by

$$\tilde{W}(y|x) = \frac{[W(y|x)]^{1/(1+\rho)}[V(y)]^{\rho/(1+\rho)}}{\sum_{y'}[W(y'|x)]^{1/(1+\rho)}[V(y')]^{\rho/(1+\rho)}} \stackrel{\triangle}{=} \frac{[W(y|x)]^{1/(1+\rho)}[V(y)]^{\rho/(1+\rho)}}{Z(x)}. \tag{42}$$

We argue that the following metric satisfies both requirements.

$$U_\theta(x|y) = \frac{P_X(x)[W(y|x)]^{1/(1+\rho)}/Z(x)}{\sum_{x'} P_X(x')[W(y|x')]^{1/(1+\rho)}/Z(x')} \stackrel{\triangle}{=} \frac{P_X(x)[W(y|x)]^{1/(1+\rho)}/Z(x)}{\zeta(y)}. \tag{43}$$

Indeed,

$$\begin{aligned} & \frac{W(y|x)}{\tilde{W}(y|x)[U_\theta(x|y)]^\rho} \\ = & \frac{W(y|x)Z(x)\zeta^\rho(y)Z^\rho(x)}{[W(y|x)]^{1/(1+\rho)}[V(y)]^{\rho/(1+\rho)}P_X^\rho(x)[W(y|x)]^{\rho/(1+\rho)}} && (44) \\ = & \frac{[Z(x)]^{1+\rho}\zeta^\rho(y)}{[V(y)]^{\rho/(1+\rho)}P_X^\rho(x)} && (45) \end{aligned}$$

Now, observe that the maximizing V in (38) is given by

$$V(y) = \sum_x P_X(x)\tilde{W}(y|x) = \sum_x P_X(x)\cdot\frac{[W(y|x)]^{1/(1+\rho)}[V(y)]^{\rho/(1+\rho)}}{Z(x)} \tag{46}$$

or, equivalently, by dividing both sides by $[V(y)]^{\rho/(1+\rho)}$,

$$[V(y)]^{1/(1+\rho)} = \sum_x P_X(x)\cdot\frac{[W(y|x)]^{1/(1+\rho)}}{Z(x)}, \tag{47}$$

and so, raising both sides to the power of ρ, we get

$$[V(y)]^{\rho/(1+\rho)} = \left(\sum_x P_X(x)\cdot\frac{[W(y|x)]^{1/(1+\rho)}}{Z(x)}\right)^\rho = \zeta^\rho(y), \tag{48}$$

so the ratio $\zeta^{\rho}(y)/[V(y)]^{\rho/(1+\rho)} = 1$, which, together with (44), implies that

$$\frac{W(y|x)}{\tilde{W}(y|x)[U_{\theta}(x|y)]^{\rho}} = \frac{[Z(x)]^{1+\rho}}{P_X^{\rho}(x)}, \tag{49}$$

which is independent of y, as required.

As for the second requirement,

$$\begin{aligned} U_{\theta}(x|y) &= \frac{P_X(x)\tilde{W}(y|x)}{\sum_{x'} P_X(x')\tilde{W}(y|x')} \\ &= \frac{P_X(x)[W(y|x)]^{1/(1+\rho)}[V(y)]^{\rho/(1+\rho)}/Z(x)}{\sum_{x'} P_X(x')[W(y|x')]^{1/(1+\rho)}[V(y)]^{\rho/(1+\rho)}/Z(x')} \\ &= \frac{P_X(x)[W(y|x)]^{1/(1+\rho)}/Z(x)}{\sum_{x'} P_X(x')[W(y|x')]^{1/(1+\rho)}/Z(x')}. \end{aligned} \tag{50}$$

Finally, the relationship between $Z(x)$ and $V(y)$ is as follows: on the one hand, by definition,

$$Z(x) = \sum_y [W(y|x)]^{1/(1+\rho)}[V(y)]^{\rho/(1+\rho)}, \tag{51}$$

and on the other hand, we saw that

$$V(y) = \left[\sum_x \frac{P_X(x)[W(y|x)]^{1/(1+\rho)}}{Z(x)}\right]^{1+\rho}. \tag{52}$$

On substituting the second relation into the first one, we end up with following system of equations in the vector $Z = \{Z(x),\ x \in \mathcal{X}\}$:

$$Z(x) = \sum_y [W(y|x)]^{1/(1+\rho)} \left[\sum_{x'} \frac{P_X(x')[W(y|x')]^{1/(1+\rho)}}{Z(x')}\right]^{\rho}, \quad \forall\ x \in \mathcal{X} \tag{53}$$

In summary, assuming that there exists a solution to this set of equations, the metric

$$U_{\star}(x|y,\rho) = \frac{P(x)[W(y|x]^{1/(1+\rho)}/Z(x)}{\sum_{x'} P(x')[W(y|x')]^{1/(1+\rho)}/Z(x')}. \tag{54}$$

saturates all inequalities in (38) at the same time, and hence, after optimization over ρ, achieves $E_{\mathrm{r}}^{\mathrm{cc}}(R)$.

6 Conclusions

In this work, we have shown that the code mismatch of using for a constant composition code a linear extension code for decoding can be fully compensated by using a mismatched additive decoding metric. That is, the error exponent [11, Theorem 10.2] is achieved and in particular, the capacity of any DMC can be achieved by decoding a linear code with an additive MAP decoding metric.

Interesting direction for future research are:

1. In this work, we have considered DMCs and we used the method of types in our theoretical derivations. Can the presented results be generalized to continuous output channels?

2. For finite length, constant composition codes may not be optimal and minimum cost DM [19] may be preferable. A theoretic analysis may find a replacement of constant composition codes that is optimal in the finite length regime.

3. In this work, decoding was our main focus. Future work may include practical encoding aspects, for instance, by assuming the PAS [9] architecture or the LLPS [6] generalization.

Appendix A

In this appendix, we show that the first inequality in eq. (26) is exponentially tight. This is done by deriving a matching lower bound of the same exponential order as the upper bound.

Owing to the results of Domb, Zamir and Feder [13], we begin with the following observation. Consider three distinct messages w, w', and $\tilde{w}$. First, if their binary representations, $\boldsymbol{b}(w)$, $\boldsymbol{b}(w')$, and $\boldsymbol{b}(\tilde{w})$ are linearly independent, the three corresponding codewords are statistically mutually independent by the independently sampled rows of $\boldsymbol{G}$. If the binary representations are linearly dependent, i.e., $\boldsymbol{b}(\tilde{w}) = \boldsymbol{b}(w) \oplus \boldsymbol{b}(w')$, then let us define

$$\boldsymbol{a}(w) = \boldsymbol{b}(w) \cdot \boldsymbol{G}, \quad \boldsymbol{a}(w') = \boldsymbol{b}(w') \cdot \boldsymbol{G}, \tag{A.1}$$

and observe that $\boldsymbol{a}(w)$, $\boldsymbol{a}(w')$, and $\boldsymbol{v}$ are statistically mutually independent. The corresponding codewords are given by

$$\boldsymbol{c}(w) = \boldsymbol{a}(w) \oplus \boldsymbol{v} \tag{A.2}$$

$$\boldsymbol{c}(w') = \boldsymbol{a}(w') \oplus \boldsymbol{v} \tag{A.3}$$

$$\boldsymbol{c}(\tilde{w}) = \boldsymbol{a}(w) \oplus \boldsymbol{a}(w') \oplus \boldsymbol{v}, \tag{A.4}$$

and therefore, the inverse transformation is given by

$$\boldsymbol{a}(w) = \boldsymbol{c}(w') \oplus \boldsymbol{c}(\tilde{w}) \tag{A.5}$$

$$\boldsymbol{a}(w') = \boldsymbol{c}(w) \oplus \boldsymbol{c}(\tilde{w}) \tag{A.6}$$

$$\boldsymbol{v} \quad = \quad \boldsymbol{c}(w) \oplus \boldsymbol{c}(w') \oplus \boldsymbol{c}(\tilde{w}). \tag{A.7}$$

Since the transformation between these two triples of vectors is one-to-one, and every triple $(\boldsymbol{a}(w), \boldsymbol{a}(w'), \boldsymbol{v})$ has probability 2^{-3nm}, then the same is true for every triple $(\boldsymbol{c}(w), \boldsymbol{c}(w'), \boldsymbol{c}(\tilde{w}))$, which means that the three codewords are mutually independent and each one of them is uniformly distributed across $\{0,1\}^{nm}$.

Consider the pairwise error events,

$$A_{w'} = \left\{U(\boldsymbol{x}(w'), \boldsymbol{y}) \geq U(\boldsymbol{x}(w), \boldsymbol{y})\right\}. \tag{A.8}$$

Since the codewords are pairwise independent, message w' is assigned to any particular codeword with probability 2^{-nm} and the probability of event $A_{w'}$ is

$$\Pr(A_{w'}) = |\mathcal{M}(\boldsymbol{x}(w), \boldsymbol{y})| \cdot 2^{-nm} \stackrel{\triangle}{=} \alpha. \tag{A.9}$$

Since the codewords are triple-wise independent, we can use de Caen's lower bound [12] to the probability of a union, and obtain

$$\begin{aligned}
\bar{P}_{\mathrm{e}}(\boldsymbol{x}(w), \boldsymbol{y}) &= \Pr\Big(\bigcup_{w' \neq w} A_{w'}\Big) && \text{(A.10)}\\
&\geq \sum_{w' \neq w} \frac{[\Pr(A_{w'})]^2}{\sum_{\tilde{w} \neq w} \Pr(A_{w'} \cap A_{\tilde{w}})} && \text{(A.11)}\\
&= \sum_{w' \neq w} \frac{[\Pr(A_{w'})]^2}{\Pr(A_{w'}) + \sum_{\tilde{w} \neq w, w'} \Pr(A_{w'} \cap A_{\tilde{w}})} && \text{(A.12)}\\
&= \sum_{w' \neq w} \frac{\alpha^2}{\alpha + \sum_{\tilde{w} \neq w, w'} \Pr(A_{w'} \cap A_{\tilde{w}})}. && \text{(A.13)}
\end{aligned}$$

Next, we consider the term $\Pr(A_{w'} \cap A_{\tilde{w}})$. Here, the three channel codewords $\boldsymbol{x}(w)$, $\boldsymbol{x}(w')$, and $\boldsymbol{x}(\tilde{w})$ are involved, with the three messages being pairwise distinct. By the triple-wise independence, $\boldsymbol{x}(w')$ and $\boldsymbol{x}(\tilde{w})$ are conditionally independent given $\boldsymbol{x}(w)$, and so,

$$\Pr(A_{w'} \cap A_{\tilde{w}}) = \Pr(A_{w'}) \cdot \Pr(A_{\tilde{w}}) = \alpha^2. \tag{A.14}$$

Continuing with (A.13), we have

$$\begin{aligned}
\bar{P}_{\mathrm{e}}(\boldsymbol{x}(w), \boldsymbol{y}) &\geq \sum_{w' \neq w} \frac{\alpha^2}{\alpha + (2^k - 2)\alpha^2} && \text{(A.15)}\\
&= \frac{(2^k - 1)\alpha}{1 + (2^k - 2)\alpha} && \text{(A.16)}
\end{aligned}$$

$$\geq \frac{(2^k-1)\alpha}{2\cdot\max\{(2^k-2)\alpha,1\}} \tag{A.17}$$

$$\geq \frac{(2^k-2)\alpha}{2\cdot\max\{(2^k-2)\alpha,1\}} \tag{A.18}$$

$$= \min\left\{\frac{1}{2},\frac{(2^k-2)\alpha}{2}\right\} \tag{A.19}$$

$$\doteq \min\left\{1,|\mathcal{M}(\boldsymbol{x}(w),\boldsymbol{y})|2^k 2^{-nm}\right\}, \tag{A.20}$$

which is of the same exponential order as the truncated union bound in (26).

Appendix B

In this appendix, we prove the alternative form of the random coding error exponent for constant composition codes.

$$\begin{aligned}
&\min_{Q_{Y|X}}\left\{D(Q_{Y|X}\|W|P_X)+[I_Q(X;Y)-R]_+\right\}\\
=\;&\min_{Q_{Y|X}}\max_{0\leq\rho\leq1}\left\{D(Q_{Y|X}\|W|P_X)+\rho[H_Q(Y)-H_Q(Y|X)-R]\right\}\\
=\;&\min_{Q_{Y|X}}\max_{0\leq\rho\leq1}\left\{\sum_{x,y}P_X(x)Q_{Y|X}(y|x)\left[\log\frac{Q_{Y|X}(y|x)}{W(y|x)}+\rho\log\frac{1}{Q_Y(y)}+\right.\right.\\
&\left.\left.\rho\log Q_{Y|X}(y|x)\right]-\rho R\right\}\\
\overset{\text{(a)}}{=}\;&\min_{Q_{Y|X}}\max_{0\leq\rho\leq1}\min_{V}\left\{\sum_{x,y}P_X(x)Q_{Y|X}(y|x)\left[\log\frac{Q_{Y|X}(y|x)}{W(y|x)}+\rho\log\frac{1}{V(y)}+\right.\right.\\
&\left.\left.\rho\log Q_{Y|X}(y|x)\right]-\rho R\right\}\\
\overset{\text{(b)}}{=}\;&\min_{Q_{Y|X}}\min_{V}\max_{0\leq\rho\leq1}\left\{\sum_{x,y}P_X(x)Q_{Y|X}(y|x)\left[\log\frac{Q_{Y|X}(y|x)}{W(y|x)}+\rho\log\frac{1}{V(y)}+\right.\right.\\
&\left.\rho\log Q_{Y|X}(y|x))-\rho R\right\}\\
=\;&\min_{V}\min_{Q_{Y|X}}\max_{0\leq\rho\leq1}\left\{\sum_{x,y}P_X(x)Q_{Y|X}(y|x)\left[\log\frac{Q_{Y|X}(y|x)}{W(y|x)}+\rho\log\frac{1}{V(y)}+\right.\right.\\
&\left.\left.\rho\log Q_{Y|X}(y|x)\right]-\rho R\right\}\\
\overset{\text{(c)}}{=}\;&\min_{V}\max_{0\leq\rho\leq1}\min_{Q_{Y|X}}\left\{\sum_{x,y}P_X(x)Q_{Y|X}(y|x)\log\frac{[Q_{Y|X}(y|x)]^{1+\rho}}{W(y|x)[V(y)]^\rho}-\rho R\right\}\\
=\;&\min_{V}\max_{0\leq\rho\leq1}\min_{Q_{Y|X}}\left\{(1+\rho)\sum_{x,y}P_X(x)Q_{Y|X}(y|x)\log\frac{Q_{Y|X}(y|x)}{(W(y|x)[V(y)]^\rho)^{1/(1+\rho)}}-\rho R\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \min_{V} \max_{0\le\rho\le1} \left\{-(1+\rho)\sum_{x} P_X(x)\log\left[\sum_{y}(W(y|x)[V(y)]^{\rho})^{1/(1+\rho)}\right] - \rho R\right\} \\
&= \min_{V} \max_{0\le\rho\le1} \Bigg\{ -\sum_{x} P_X(x)\log\left[\sum_{y}\left(\frac{W(y|x)[V(y)]^{\rho}}{[P_X(x)]^{\rho}}\right)^{1/(1+\rho)}\right]^{1+\rho} + \\
&\quad \rho[H(P_X) - R]\Bigg\}, \qquad \text{(B.1)}
\end{aligned}
$$

where the inner-most minimization in (a) is over all probability distributions $\{V(y)\}$ on $\mathcal{Y}$, (b) holds because the objective is convex in V and concave in ρ, and similarly, (c) is because the objective is convex in $Q_{Y|X}$ and concave in ρ.

References

[1] S. Achtenberg, and D. Raphaeli, “Theoretic Shaping Bounds for Single Letter Constraints and Mismatched Decoding,” arXiv preprint, 2013. [Online]. Available: https://arxiv.org/abs/1308.5938v1.

[2] R. A. Amjad, “Information rates and error exponents for probabilistic amplitude shaping,” in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2018.

[3] E. Arikan, “Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[4] G. Böcherer, “Achievable Rates for Shaped Bit-Metric Decoding,” arXiv preprint, 2016. [Online]. Available: https://arxiv.org/pdf/1410.8075v6.

[5] G. Böcherer, “Principles of Coded Modulation,” Habilitation thesis, Technische Universität München, 2018.

[6] G. Böcherer, D. Lentner, A. Cirino, and F. Steiner, “Probabilistic parity shaping for linear codes,” *Oberpfaffenhofen Workshop on High Throughput Coding*, Oberpfaffenhofen, Germany, 2019. [Online]. Available: https://arxiv.org/abs/1902.10648v1.

[7] G. Böcherer, P. Schulte, and F. Steiner, “Probabilistic shaping and forward error correction for fiber-optic communication systems,” *J. Lightw. Technol.*, vol. 37, no. 2, pp. 230-244, January 2019.

[8] G. Böcherer, P. Schulte, and F. Steiner, “Probabilistic Shaping: A Random Coding Experiment,” in *Proc. Int. Zurich Seminar Commun.*, Zurich, Switzerland, 2020.

[9] G. Böcherer, F. Steiner, and P. Schulte, “Bandwidth efficient and rate-matched low-density parity-check coded modulation,” *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651-4665, December 2015.

[10] D. Chase, “Class of algorithms for decoding block codes with channel measurement information,” *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 170–182, January 1972.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Cambridge University Press, 2011.

[12] D. de Caen, “A lower bound on the probability of a union,” *Discrete mathematics*, vol. 169, no. 1-3, pp. 217-220, 1997.

[13] Y. Domb, R. Zamir, and M. Feder, “The random coding bound is tight for the average linear code or lattice,” *IEEE Trans. Inform. Theory*, vol. 62, no. 1, pp. 121-130, January 2016.

[14] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., New York 1968.

[15] Y. C. Gültekin, A. Alvarado, and F. M. Willems, “Achievable information rates for probabilistic amplitude shaping: An alternative approach via random sign-coding arguments,” *Entropy*, vol. 22, no. 7, 2020.

[16] S. Lin and D. J. Costello, *Error Control Coding*, second edition, Prentice Hall, 2004.

[17] R. J. McEliece, *The Theory of Information and Coding*, Cambridge University Press, New York, 1984.

[18] P. Schulte and G. Böcherer, “Constant Composition Distribution Matching,” *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430434, January 2016.

[19] P. Schulte and F. Steiner, “Divergence-optimal fixed-to-fixed length distribution matching with shell mapping,” IEEE Wireless Commun. Letters, vol. 8, no. 2, pp. 620623, Apr. 2019.

[20] P. Schulte, “Algorithms for Distribution Matching,” Ph.D. thesis, Technische Universität München, 2020.

[21] N. Stolte, “Rekursive Codes mit der Plotkin-Konstruktion und ihre Decodierung,” Ph.D. thesis, TU Darmstadt, 2002.

[22] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, New York, 1979.