# Superposition Codes for Mismatched Decoding

Jonathan Scarlett
University of Cambridge
jms265@cam.ac.uk

Alfonso Martinez
Universitat Pompeu Fabra
alfonso.martinez@ieee.org

Albert Guillén i Fàbregas
ICREA & Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

***Abstract*—An achievable rate is given for discrete memoryless channels with a given (possibly suboptimal) decoding rule. The result is obtained using a refinement of the superposition coding ensemble. The rate is tight with respect to the ensemble average, and can be weakened to the LM rate of Hui and Csiszár-Körner, and to Lapidoth's rate based on parallel codebooks.**

## I. INTRODUCTION

In this paper, we consider the problem of channel coding over a discrete memoryless channel (DMC) $W(y|x)$, in which the decoder maximizes the symbol-wise product of a given decoding metric $q(x,y)$. If $q(x,y) = W(y|x)$ then the decoder is a maximum-likelihood (ML) decoder; otherwise, the decoder is said to be mismatched [1]–[6]. More precisely, the decoder estimates the message as

$$\hat{m} = \arg\max_{j} \prod_{i=1}^{n} q(x_i^{(j)}, y_i), \tag{1}$$

where $n$ is the block length, $\boldsymbol{x}^{(j)} = (x_1^{(j)}, \dots, x_n^{(j)})$ is the $j$-th codeword in the codebook, and $\boldsymbol{y} = (y_1, \dots, y_n)$ is the received vector.

An error is said to have occurred if the estimated message differs from the transmitted one. A rate $R$ is said to be achievable if, for all $\delta > 0$, there exists a sequence of codebooks with at least $\exp(n(R-\delta))$ codewords and vanishing error probability. The mismatched capacity, defined to be the supremum of all achievable rates, is unknown in general, and most existing work has focused on achievable random-coding rates. Of particular note is the LM rate [3], [5], given by (see Section I-A for notation)

$$I^{\mathrm{LM}}(Q) \triangleq \min_{\substack{\widetilde{P}_{XY} \,:\, \widetilde{P}_X = Q, \widetilde{P}_Y = P_Y \\ \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)]}} I_{\widetilde{P}}(X;Y), \tag{2}$$

where $Q$ is an arbitrary input distribution, and $P_{XY} = Q \times W$. The generalized mutual information (GMI) [6] is defined similarly, with the constraint $\widetilde{P}_X = P_X$ removed.

To our knowledge, only two improvements on the LM rate have been reported in the literature. In [4], Csiszár and Narayan show that better achievable rates can be obtained by applying the LM rate to the channel $W^{(2)}\big((y_1,y_2)|(x_1,x_2)\big) = W(y_1|x_1)W(y_2|x_2)$ with the metric $q^{(2)}\big((x_1,x_2),(y_1,x_2)\big) = q(x_1,y_1)q(x_2,y_2)$, and similarly for the $k$-th order products of $W$ and $q$. However, as $k$ increases, the required computation becomes prohibitively complex, since the LM rate is non-convex in $Q$ in general. In [1], Lapidoth gives a single-letter improvement on the LM rate using multiple-access coding techniques, in which multiple codebooks are generated in parallel.

In this paper, we obtain a new single-letter improvement on the LM rate using a refinement of superposition coding, the standard version of which is typically used in broadcast scenarios [7]. The results of this paper and the existing single-letter results in the literature can be summarized by the following list of random-coding constructions, in decreasing order of rate (see Sections II-A to II-D for further discussion):

1) Refined Superposition Coding (Theorem 1),
2) Standard Superposition Coding ((11)–(12)),
3) Expurgated Parallel Coding ([1, Theorem 4]),
4) Constant-Composition Coding (LM rate [3], [5]),
5) i.i.d. Coding (GMI [6]).

### A. Notation

The set of all probability distributions on an alphabet $\mathcal{A}$ is denoted by $\mathcal{P}(\mathcal{A})$. The set of all empirical distributions (i.e. types) corresponding to length-$n$ sequences on $\mathcal{A}$ is denoted by $\mathcal{P}_n(\mathcal{A})$. The set of all sequences of length $n$ with a given type $P_X$ is denoted by $T^n(P_X)$, and similarly for joint types. See [8, Ch. 2] for an introduction to the method of types.

The marginals of a joint distribution $P_{XY}(x,y)$ are denoted by $P_X(x)$ and $P_Y(y)$. Similarly, $P_{Y|X}(y|x)$ denotes the conditional distribution induced by $P_{XY}(x,y)$. We write $P_X = \widetilde{P}_X$ to denote element-wise equality between two probability distributions on the same alphabet. Expectation with respect to a distribution $P_X(x)$ is denoted by $\mathbb{E}_P[\cdot]$.

Given a distribution $Q(x)$ and a conditional distribution $W(y|x)$, the joint distribution $Q(x)W(y|x)$ is denoted by $Q \times W$. Information-theoretic quantities with respect to a given distribution (e.g. $P_{XY}(x,y)$) are written with a subscript (e.g. $I_P(X;Y)$). We write $[\alpha]^+ = \max(0,\alpha)$, and denote the indicator function by $\mathbb{1}\{\cdot\}$. Logarithms have base $e$, and rates are in nats except in the examples, where bits are used.

## II. MAIN RESULT

We begin by introducing the refined superposition coding ensemble from which the achievable rate is obtained. We fix a finite alphabet $\mathcal{U}$, an input distribution $Q_{UX}$ and the rates

This work has been funded in part by the European Research Council under ERC grant agreement 259663, by the European Union's 7th Framework Programme (PEOPLE-2011-CIG) under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03.

$R_0$ and $\{R_{1u}\}_{u\in\mathcal{U}}$. We write $M_0 \triangleq e^{nR_0}$ and $M_{1u} \triangleq e^{nR_{1u}}$. We let $P_{\boldsymbol{U}}(\boldsymbol{u})$ be the uniform distribution over the type class $T^n(Q_{U,n})$, where $Q_{U,n} \in \mathcal{P}_n(\mathcal{U})$ is the most probable type under $Q_U$. For each $u \in \mathcal{U}$, we define

$$n_u \triangleq Q_{U,n}(u)n \tag{3}$$

and let $Q_{u,n_u} \in \mathcal{P}_{n_u}(\mathcal{X})$ be the most probable type under $Q_{X|U=u}$ for sequences of length $n_u$. We let $P_{\boldsymbol{X}_u}(\boldsymbol{x}_u)$ be the uniform distribution over the type class $T^{n_1}(Q_{u,n_u})$. Thus,

$$P_{\boldsymbol{U}}(\boldsymbol{u}) = \frac{1}{|T^n(Q_{U,n})|}\mathbb{1}\Big\{\boldsymbol{u} \in T^n(Q_{U,n})\Big\} \tag{4}$$

$$P_{\boldsymbol{X}_u}(\boldsymbol{x}_u) = \frac{1}{|T^{n_u}(Q_{u,n_u})|}\mathbb{1}\Big\{\boldsymbol{x}_u \in T^{n_u}(Q_{u,n_u})\Big\}. \tag{5}$$

We randomly generate the length-$n$ auxiliary codewords $\{\boldsymbol{U}^{(i)}\}_{i=1}^{M_0}$ independently according to $P_{\boldsymbol{U}}$. For each $i = 1,\dots,M_0$ and $u \in \mathcal{U}$, we further generate the length-$n_u$ partial codewords $\{\boldsymbol{X}_u^{(i,j_u)}\}_{j_u=1}^{M_{1u}}$ independently according to $P_{\boldsymbol{X}_u}$. For example, when $\mathcal{U} = \{1,2\}$ we have

$$\bigg\{\Big(\boldsymbol{U}^{(i)}, \{\boldsymbol{X}_1^{(i,j_1)}\}_{j_1=1}^{M_{11}}, \{\boldsymbol{X}_2^{(i,j_2)}\}_{j_2=1}^{M_{12}}\Big)\bigg\}_{i=1}^{M_0}$$
$$\sim \prod_{i=1}^{M_0}\bigg(P_{\boldsymbol{U}}(\boldsymbol{u}^{(i)})\prod_{j_1=1}^{M_{11}}P_{\boldsymbol{X}_1}(\boldsymbol{x}_1^{(i,j_1)})\prod_{j_2=1}^{M_{12}}P_{\boldsymbol{X}_2}(\boldsymbol{x}_2^{(i,j_2)})\bigg). \tag{6}$$

The codebook on $\mathcal{X}^n$ is indexed as $(m_0, m_{11}, \dots, m_{1|\mathcal{U}|})$. To transmit a given message, we treat $\boldsymbol{U}^{(m_0)}$ as a time-sharing sequence; at the indices where $\boldsymbol{U}^{(m_0)}$ equals $u$, we transmit the symbols of $\boldsymbol{X}_u^{(m_0,m_{1u})}$. There are $M_0\prod_u M_{1u}$ codewords, and hence the rate is $R = R_0 + \sum_u Q_{U,n}(u)R_{1u}$. We will see that in the mismatched setting, this ensemble yields higher achievable rates than the standard constant-composition superposition coding ensemble [7] in which, for all $i = 1,\dots,M_0$, the codewords $\{\boldsymbol{X}^{(i,j)}\}_{j=1}^{M_1}$ are conditionally independent given $\boldsymbol{U}^{(i)}$.

The main result of this paper is stated in the following theorem, which is proved in Section III. We define the set

$$\mathcal{T}_0(P_{UXY}) \triangleq \Big\{\widetilde{P}_{UXY} \in \mathcal{P}(\mathcal{U}\times\mathcal{X}\times\mathcal{Y}) \,:\, \widetilde{P}_{UX} = P_{UX}, \widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)]\Big\}. \tag{7}$$

**Theorem 1.** *For any finite set $\mathcal{U}$ and input distribution $Q_{UX}$, the rate*

$$R = R_0 + \sum_u Q_U(u)R_{1u} \tag{8}$$

*is achievable provided that $R_0$ and $\{R_{1u}\}_{u=1}^{|\mathcal{U}|}$ satisfy*

$$R_{1u} \le I^{\mathrm{LM}}(Q_{X|U=u}) \tag{9}$$

$$R_0 \le \min_{\widetilde{P}_{UXY}\in\mathcal{T}_0(Q_{UX}\times W)} I_{\widetilde{P}}(U;Y) + \bigg[\max_{\mathcal{K}\subseteq\mathcal{U}}\sum_{u\in\mathcal{K}}Q_U(u)\Big(I_{\widetilde{P}}(X;Y|U=u) - R_{1u}\Big)\bigg]^+. \tag{10}$$

By combining the techniques of [9, Section III-C] and [1, Thm. 3], it can be shown that, subject to minor technical conditions, the achievable rate of Theorem 1 is tight with respect to the ensemble average. That is, Theorem 1 gives the best possible achievable rate for the random coding ensemble under consideration. A similar statement holds for the random-coding error exponent given in the proof in Section III (cf. (30)). Furthermore, a similar analysis can be applied to the mismatched broadcast channel with degraded message sets, in which a secondary user attempts to recover only the index $m_0$.

### A. Comparison to Existing Results

The LM rate in (2) is recovered from (8)–(10) by setting $|\mathcal{U}| = 1$ and $R_0 = 0$.

Using standard constant-composition superposition coding [7], a similar (yet simpler) analysis to Section III yields the achievability of all $R = R_0 + R_1$ such that $(R_0, R_1)$ satisfy

$$R_1 \le \min_{\substack{\widetilde{P}_{UX}=Q_{UX}, \widetilde{P}_{UY}=P_{UY} \\ \mathbb{E}_{\widetilde{P}}[\log q(X,Y)]\ge\mathbb{E}_P[\log q(X,Y)]}} I_{\widetilde{P}}(X;Y|U) \tag{11}$$

$$R_0 + R_1 \le \min_{\substack{\widetilde{P}_{UXY}\in\mathcal{T}_0(Q_{UX}\times W) \\ I_{\widetilde{P}}(U;Y)\le R_0}} I_{\widetilde{P}}(U,X;Y) \tag{12}$$

This achievability result can be obtained by weakening that of Theorem 1 with the identification $R_1 = \sum_u Q_U(u)R_{1u}$. Roughly speaking, we obtain (11) by summing the $|\mathcal{U}|$ rates in (9) weighted by $Q_U(\cdot)$, identifying the corresponding $|\mathcal{U}|$ joint distributions $\widetilde{P}_{XY}$ in (2) with $\widetilde{P}_{XY|U=u}$, and relaxing the constraints on the metric in (2) to hold on average with respect to $Q_U$, rather than for each $u \in \mathcal{U}$. Furthermore, we obtain (12) from (10) by replacing the maximum over $\mathcal{K} \subseteq \mathcal{U}$ with the particular choice $\mathcal{K} = \mathcal{U}$, noting that (10) always holds when the minimizing $\widetilde{P}_{UXY}$ satisfies $I_{\widetilde{P}}(U;Y) > R_0$, and using the chain rule for mutual information.

The achievable rate of Lapidoth [1, Thm. 4] was derived by (i) fixing the finite auxiliary alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, the input distributions $Q_{X_1} \in \mathcal{P}(\mathcal{X}_1)$ and $Q_{X_2} \in \mathcal{P}(\mathcal{X}_2)$, and the function $\phi : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{X}$, (ii) coding for the mismatched multiple-access channel $W'(y|x_1,x_2) \triangleq W(y|\phi(x_1,x_2))$ with the metric $q'(x_1,x_2,y) \triangleq q(\phi(x_1,x_2),y)$ and input distributions $Q_{X_1}$ and $Q_{X_2}$, and (iii) expurgating all codeword pairs except those of a given joint type. The analysis in [1] yields the achievability of all $R' = R'_1 + R'_2$ such that $(R'_1, R'_2)$ satisfy

$$R'_1 \le \min_{\substack{\widetilde{P}_{X_1X_2}=Q_{X_1}\times Q_{X_2} \\ \widetilde{P}_{X_2Y}=P_{X_2Y} \\ \mathbb{E}_{\widetilde{P}}[\log q']\ge\mathbb{E}_P[\log q']}} I_{\widetilde{P}}(X_1;Y|X_2) \tag{13}$$

$$R'_2 \le \min_{\substack{\widetilde{P}_{X_1X_2}=Q_{X_1}\times Q_{X_2} \\ \widetilde{P}_{X_1Y}=P_{X_1Y} \\ \mathbb{E}_{\widetilde{P}}[\log q']\ge\mathbb{E}_P[\log q']}} I_{\widetilde{P}}(X_2;Y|X_1) \tag{14}$$

$$R'_1 + R'_2 \le \min_{\substack{\widetilde{P}_{X_1X_2}=Q_{X_1}\times Q_{X_2}, \widetilde{P}_Y=P_Y \\ \mathbb{E}_{\widetilde{P}}[\log q']\ge\mathbb{E}_P[\log q'] \\ I_{\widetilde{P}}(X_1;Y)\le R'_1,\, I_{\widetilde{P}}(X_2;Y)\le R'_2}} I_{\widetilde{P}}(X_1,X_2;Y), \tag{15}$$

where $P_{X_1X_2Y} = Q_{X_1} \times Q_{X_2} \times W'$, each minimization is over all $\widetilde{P}_{X_1X_2Y}$ satisfying the specified constraints, and we write $\mathbb{E}_{\widetilde{P}}[\log q']$ as a shorthand for $\mathbb{E}_{\widetilde{P}}[\log q'(X_1, X_2, Y)]$.

**Proposition 1.** *After the optimization of $\mathcal{U}$ and $Q_{UX}$, the achievable rate $R = \max_{R_0,R_1} R_0 + R_1$ described by (11)–(12) is at least as high as the achievable rate $R' = \max_{R'_1,R'_2} R'_1 + R'_2$ described by (13)–(15) with optimized parameters $\mathcal{X}_1$, $\mathcal{X}_2$, $Q_{X_1}$, $Q_{X_2}$ and $\phi(x_1, x_2)$.*

*Proof:* We fix the alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, the distributions $Q_{X_1}(x_1)$ and $Q_{X_2}(x_2)$, and the function $\phi(\cdot,\cdot)$ arbitrarily. We consider the superposition coding parameters $U = X_2$ and

$$Q_{UX}(u, x) = \sum_{x_1} Q_{X_1}(x_1)Q_{X_2}(x_2)\mathbb{1}\{x = \phi(x_1, x_2)\}. \tag{16}$$

Since the $(U, X)$ marginal of $\widetilde{P}_{UXY}$ is constrained to be equal to $Q_{UX}$ in both (11) and (12), we can equivalently rewrite each optimization as a minimization over $\widetilde{P}_{Y|UX}$. The corresponding objectives and constraints depend on $\widetilde{P}_{Y|UX}(y|u, x)$ only through $\widetilde{P}_{Y|X_1X}(y|x_1, \phi(x_1, x_2))$, which is a conditional distribution on $\mathcal{Y}$ given $\mathcal{X}_1$ and $\mathcal{X}_2$. Thus, the bounds can be weakened by taking the minimizations over all distributions on $\mathcal{Y}$ given $\mathcal{X}_1$ and $\mathcal{X}_2$ satisfying similar constraints to (11)–(12). With some simple algebra and by renaming $R_1 = R'_1$ and $R_0 = R'_2$, we obtain the achievability of $(R'_1, R'_2)$ satisfying (13) and (15) with the constraint $I_{\widetilde{P}}(X_1; Y) \le R'_1$ removed. Repeating these steps with $U = X_1$, $R_1 = R'_2$, $R_0 = R'_1$ and

$$Q_{UX}(u, x) = \sum_{x_2} Q_{X_1}(x_1)Q_{X_2}(x_2)\mathbb{1}\{x = \phi(x_1, x_2)\}, \tag{17}$$

we obtain the achievability of $(R'_1, R'_2)$ satisfying (14) and (15) with the constraint $I_{\widetilde{P}}(X_2; Y) \le R'_2$ removed.

Finally, we make use of [10, Lemma 1], which states that $R_1 + R_2$ satisfies the inequality in (15) if and only if $R_1 + R_2$ satisfies at least one of two similar inequalities, each of the same form as (15) with one constraint $I_{\widetilde{P}}(X_\nu; Y) \le R_\nu$ ($\nu = 1, 2$) removed. It follows that the union of the above two derived regions contains the region in (13)–(15), and hence the former yields a sum rate at least as high as the latter. ■

*B. Discussion*

The benefits of both parallel and superposition coding arise from the dependence among the randomly generated codewords. Under parallel coding without expurgation [1], one generates the codewords $\{\boldsymbol{X}_1^{(i)}\}_{i=1}^{M_1}$ and $\{\boldsymbol{X}_2^{(j)}\}_{j=1}^{M_2}$ independently. One can picture the overall codewords as being arranged in an $M_1 \times M_2$ grid, where the $(i, j)$-th entry is a deterministic function of $(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)})$. In this case, every codeword in row $i$ depends on $\boldsymbol{X}_1^{(i)}$, and every codeword in column $j$ depends on $\boldsymbol{X}_2^{(j)}$. Similarly, under superposition coding, one can arrange the codewords in an $M_0 \times M_1$ grid in which every codeword in the $i$-th row depends on $\boldsymbol{U}^{(i)}$.

The dependence among both rows and columns under parallel coding yields two constraints of the form $I_{\widetilde{P}}(X_\nu; Y) \le R_\nu$ ($\nu = 1, 2$) in (15), whereas (12) has just one analogous constraint. However, from [10, Lemma 1], at most one of the two constraints affects the minimization for any given rate pair.

Superposition coding allows one to specify a joint composition of $(U, X)$, yielding the constraint $\widetilde{P}_{UX} = Q_{UX}$ in (11)–(12). On the other hand, one cannot specify the joint composition of $(X_1, X_2)$ under parallel coding. However, using the expurgation step of [1], one recovers a codebook in which the joint composition is fixed.

While the ability to choose the joint distribution $Q_{UX}$ in (11)–(12) may appear to give more freedom than the ability to choose $Q_{X_1}$ and $Q_{X_2}$ in (13)–(15), it can be shown that any joint distribution of $(X_1, X)$ (or $(X_2, X)$) can be induced in the latter setting with the additional freedom in choosing $\phi(\cdot,\cdot)$. This observation suggests that for many (and possibly all) channels and decoding metrics the converse to Proposition 1 holds true, i.e. that the achievable rates of standard superposition and expurgated parallel coding coincide after the full optimization of the parameters. On the other hand, we believe that the local optimization of the former rate over $(\mathcal{U}, Q_{UX})$ has a lower computational complexity than the local optimization of the latter rate over $(\mathcal{X}_1, \mathcal{X}_2, Q_{X_1}, Q_{X_2}, \phi)$. Since computational complexity generally prohibits the global optimization of the random-coding parameters, the ability to perform such local optimizations is of great interest.

Finally, the advantage of the refined superposition coding ensemble over that of standard superposition coding arises from a stronger dependence among the codewords in a given row. Specifically, unlike the former ensemble, the latter ensemble has codewords in each row which are conditionally independent given given $\boldsymbol{U}^{(i)}$.

*C. Example 1: Sum Channel*

Given two channels $(W_1, W_2)$ respectively defined on the alphabets $(\mathcal{X}_1, \mathcal{Y}_1)$ and $(\mathcal{X}_2, \mathcal{Y}_2)$, the *sum channel* is defined to be the channel $W(y|x)$ with $|\mathcal{X}| = |\mathcal{X}_1| + |\mathcal{X}_2|$ and $|\mathcal{Y}| = |\mathcal{Y}_1| + |\mathcal{Y}_2|$ such that one of the two subchannels is used on each transmission [11]. One can similarly combine two metrics $q_1(x_1, y_1)$ and $q_2(x_2, y_2)$ to form a *sum metric* $q(x, y)$.

Let $(W, q)$ be the sum channel and metric associated with $(W_1, q_1)$ and $(W_2, q_2)$, and let $\hat{Q}_1$ and $\hat{Q}_2$ be the distributions which maximize the LM rate in (2) on the respective subchannels. We set $\mathcal{U} = \{1, 2\}$, $Q_{X|U=1} = (\hat{Q}_1, \mathbf{0})$ and $Q_{X|U=2} = (\mathbf{0}, \hat{Q}_2)$, where $\mathbf{0}$ denotes the zero vector. We leave $Q_U$ to be specified.

Combining the constraints $\widetilde{P}_{UX} = Q_{UX}$ and $\mathbb{E}_{\widetilde{P}}[\log q(X, Y)] \ge \mathbb{E}_P[\log q(X, Y)]$ in (7), it is straightforward to show that the minimizing $\widetilde{P}_{UXY}(u, x, y)$ in (10) only takes on non-zero values for $(u, x, y)$ such that (i) $u = 1$, $x \in \mathcal{X}_1$ and $y \in \mathcal{Y}_1$, or (ii) $u = 2$, $x \in \mathcal{X}_2$ and $y \in \mathcal{Y}_2$, where we assume without loss of generality that the subchannel alphabets are disjoint (i.e. $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$ and $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$). It follows that $U$ is a deterministic function of $Y$ under the minimizing $\widetilde{P}_{UXY}$, and hence

$$I_{\widetilde{P}}(U; Y) = H(Q_U) - H_{\widetilde{P}}(U|Y) = H(Q_U). \tag{18}$$

Therefore, the right-hand side of (10) is lower bounded by $H(Q_U)$. Using (8) and performing a simple optimization of $Q_U$, it follows that the rate $\log\left(e^{I_1^{\mathrm{LM}}(\hat{Q}_1)} + e^{I_2^{\mathrm{LM}}(\hat{Q}_2)}\right)$ is achievable, where $I_\nu^{\mathrm{LM}}$ is the LM rate for subchannel $\nu$. An analogous result has been proved in the setting of matched decoding using the known formula for channel capacity [11]. It should be noted that the LM rate of $(W, q)$ can be strictly less than $\log\left(e^{I_1^{\mathrm{LM}}(\hat{Q}_1)} + e^{I_2^{\mathrm{LM}}(\hat{Q}_2)}\right)$ even for simple examples (e.g. binary symmetric subchannels).

### D. Example 2

We consider the channel and decoding metric described by the entries of the matrices

$$\boldsymbol{W} = \begin{bmatrix} 0.99 & 0.01 & 0 & 0 \\ 0.01 & 0.99 & 0 & 0 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{bmatrix} \tag{19}$$

$$\boldsymbol{q} = \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0.05 & 0.15 & 1 & 0.05 \\ 0.15 & 0.05 & 0.5 & 1 \end{bmatrix}. \tag{20}$$

We have intentionally chosen a highly asymmetric channel and metric, since such examples generally yield larger gaps between the various achievable rates. Using an exhaustive search to three decimal places, we find the optimized LM rate to be $R_{\mathrm{LM}}^* = 1.111$ bits/use, which is achieved by the input distribution $Q_X^* = (0.403, 0.418, 0, 0.179)$.

The global optimization of (8)–(10) over $\mathcal{U}$ and $Q_{UX}$ is difficult. Setting $|\mathcal{U}| = 2$ and applying local optimization techniques using a number of starting points, we obtained an achievable rate of $R^* = 1.313$ bits/use, with $Q_U = (0.698, 0.302)$, $Q_{X|U=1} = (0.5, 0.5, 0, 0)$ and $Q_{X|U=2} = (0, 0, 0.528, 0.472)$. We denote the corresponding input distribution by $Q_{UX}^{(1)}$.

Applying similar techniques to the standard superposition coding rate in (11)–(12), we obtained an achievable rate of $R_{\mathrm{SC}}^* = 1.236$ bits/use, with $Q_U = (0.830, 0.170)$, $Q_{X|U=1} = (0.435, 0.450, 0.115, 0)$ and $Q_{X|U=2} = (0, 0, 0, 1)$. We denote the corresponding input distribution by $Q_{UX}^{(2)}$.

The achievable rates for this example are summarized in Table I, where $Q_{UX}^{(\mathrm{LM})}$ denotes the distribution in which $U$ is deterministic and the $X$-marginal maximizes the LM rate. While the achievable rate of Theorem 1 coincides with that of (11)–(12) under $Q_{UX}^{(2)}$, the former is significantly higher under $Q_{UX}^{(1)}$. Both types of superposition coding yield a strict improvement over the LM rate.

Our parameters may not be globally optimal, and thus we cannot conclude from this example that Theorem 1 yields a strict improvement over (11)–(12) (and hence over (13)–(15)) after optimizing $\mathcal{U}$ and $Q_{UX}$. However, since the optimization of the input distribution can be a non-convex problem even for the LM rate, finding the global optimum is computationally infeasible in general. Thus, improvements for fixed $|\mathcal{U}|$ and $Q_{UX}$ are of great interest.

Table I
ACHIEVABLE RATES FOR THE MISMATCHED CHANNEL IN (20)-(21).

| Input Distribution | Refined SC | Standard SC |
|---|---|---|
| $Q_{UX}^{(1)}$ | 1.313 | 1.060 |
| $Q_{UX}^{(2)}$ | 1.236 | 1.236 |
| $Q_{UX}^{(\mathrm{LM})}$ | 1.111 | 1.111 |

## III. Proof of Theorem 1

For clarity of exposition, we present the proof in the case that $\mathcal{U} = \{1, 2\}$. The same arguments apply to the general case. We let $\Xi(\boldsymbol{u}, \boldsymbol{x}_1, \boldsymbol{x}_2)$ denote the function for constructing the length-$n$ codeword from the auxiliary sequence and partial codewords, and write

$$\boldsymbol{X}^{(i,j_1,j_2)} \triangleq \Xi(\boldsymbol{U}^{(i)}, \boldsymbol{X}_1^{(i,j_1)}, \boldsymbol{X}_2^{(i,j_2)}). \tag{21}$$

For the remainder of the proof, we let $\boldsymbol{y}_u(\boldsymbol{u})$ denote the subsequence of $\boldsymbol{y}$ corresponding to the indices where $\boldsymbol{u}$ equals $u$. Furthermore, we define $q^n(\boldsymbol{x}, \boldsymbol{y}) \triangleq \prod_{i=1}^n q(x_i, y_i)$ and $W^n(\boldsymbol{y}|\boldsymbol{x}) \triangleq \prod_{i=1}^n W(y_i|x_i)$.

Upon receiving $\boldsymbol{y}$, the decoder forms the estimate

$$(\hat{m}_0, \hat{m}_1, \hat{m}_2) = \operatorname*{arg\,max}_{(i,j_1,j_2)} q^n(\boldsymbol{x}^{(i,j_1,j_2)}, \boldsymbol{y}) \tag{22}$$

$$= \operatorname*{arg\,max}_{(i,j_1,j_2)} q^{n_1}\big(\boldsymbol{x}_1^{(i,j_1)}, \boldsymbol{y}_1(\boldsymbol{u}^{(i)})\big) q^{n_2}\big(\boldsymbol{x}_2^{(i,j_2)}, \boldsymbol{y}_2(\boldsymbol{u}^{(i)})\big), \tag{23}$$

where the objective in (23) follows by separating the indices where $u = 1$ from those where $u = 2$. By writing the objective in this form, it is easily seen that for any given $i$, the pair $(j_1, j_2)$ with the highest metric is the one for which $j_1$ maximizes $q^{n_1}(\boldsymbol{x}_1^{(i,j_1)}, \boldsymbol{y}_1(\boldsymbol{u}^{(i)}))$ and $j_2$ maximizes $q^{n_2}(\boldsymbol{x}_2^{(i,j_2)}, \boldsymbol{y}_2(\boldsymbol{u}^{(i)}))$. Therefore, we can split the error event into three (not necessarily disjoint) events:

*(Type 0)* $\hat{m}_0 \neq m_0$
*(Type 1)* $\hat{m}_0 = m_0$ and $\hat{m}_{11} \neq m_{11}$
*(Type 2)* $\hat{m}_0 = m_0$ and $\hat{m}_{12} \neq m_{12}$

We denote the corresponding random-coding error probabilities by $\overline{p}_{e,0}$, $\overline{p}_{e,1}$ and $\overline{p}_{e,2}$ respectively.

From the construction of the random-coding ensemble, the type-1 error probability $\overline{p}_{e,1}$ is precisely that of the single-user constant-composition ensemble with rate $R_{11}$, length $n_1 = nQ_U(1)$, and input distribution $Q_{X|U=1}$. A similar statement holds for the type-2 error probability $\overline{p}_{e,2}$, and the analysis for these error events is identical to the derivation of the LM rate [3], [5], yielding the rate conditions in (9). For the remainder of this section, we analyze the type-0 event.

We assume without loss of generality that $(m_0, m_1, m_2) = (1, 1, 1)$. We let $\boldsymbol{U}$ and $\boldsymbol{X}$ be the codewords corresponding to $(1, 1, 1)$, and let $\overline{\boldsymbol{U}}$, $\overline{\boldsymbol{X}}_1$ and $\overline{\boldsymbol{X}}_2$ be the codewords corresponding to an arbitrary message with $m_0 \neq 1$. For the index $i$ corresponding to $\overline{\boldsymbol{U}}$, we write $\overline{\boldsymbol{X}}_1^{(j_1)}$, $\overline{\boldsymbol{X}}_2^{(j_2)}$ and $\overline{\boldsymbol{X}}^{(j_1,j_2)}$ in place of $\boldsymbol{X}_1^{(i,j_1)}$, $\boldsymbol{X}_2^{(i,j_2)}$ and $\boldsymbol{X}^{(i,j_1,j_2)}$ respectively. Thus, from (21), we have $\overline{\boldsymbol{X}}^{(j_1,j_2)} = \Xi(\overline{\boldsymbol{U}}, \overline{\boldsymbol{X}}_1^{(j_1)}, \overline{\boldsymbol{X}}_2^{(j_2)})$.

The error probability for the type-0 event satisfies

$$\overline{p}_{e,0} \le \mathbb{P}\left[\bigcup_{i\neq 1}\bigcup_{j_1,j_2}\left\{\frac{q^n(\boldsymbol{X}^{(i,j_1,j_2)},\boldsymbol{Y})}{q^n(\boldsymbol{X},\boldsymbol{Y})}\ge 1\right\}\right], \tag{24}$$

where $(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x})\sim W^n(\cdot|\boldsymbol{x})$. Writing the probability as an expectation given $(\boldsymbol{U},\boldsymbol{X},\boldsymbol{Y})$ and applying the truncated union bound, we obtain

$$\overline{p}_{e,0} \le \mathbb{E}\Bigg[\min\Bigg\{1,(M_0-1) \times\mathbb{E}\Bigg[\mathbb{P}\Bigg[\bigcup_{j_1,j_2}\left\{\frac{q^n(\overline{\boldsymbol{X}}^{(j_1,j_2)},\boldsymbol{Y})}{q^n(\boldsymbol{X},\boldsymbol{Y})}\ge 1\right\}\Bigg|\overline{\boldsymbol{U}}\Bigg]\Bigg|\boldsymbol{U},\boldsymbol{X},\boldsymbol{Y}\Bigg]\Bigg\}\Bigg], \tag{25}$$

where we have written the probability of the union over $j_1$ and $j_2$ as an expectation given $\overline{\boldsymbol{U}}$.

Let the joint types of $(\boldsymbol{U},\boldsymbol{X},\boldsymbol{Y})$ and $(\overline{\boldsymbol{U}},\overline{\boldsymbol{X}}^{(j_1,j_2)},\boldsymbol{Y})$ be denoted by $P_{UXY}$ and $\widetilde{P}_{UXY}$ respectively. We claim that

$$\frac{q^n(\overline{\boldsymbol{X}}^{(j_1,j_2)},\boldsymbol{Y})}{q^n(\boldsymbol{X},\boldsymbol{Y})}\ge 1 \tag{26}$$

if and only if

$$\widetilde{P}_{UXY}\in\mathcal{T}_{0,n}(P_{UXY})\triangleq\mathcal{T}_0(P_{UXY})\cap\mathcal{P}_n(\mathcal{U}\times\mathcal{X}\times\mathcal{Y}), \tag{27}$$

where $\mathcal{T}_0(P_{UXY})$ is defined in (7). The constraint $\widetilde{P}_{UX}=P_{UX}$ follows from the construction of the random coding ensemble, $\widetilde{P}_Y=P_Y$ follows since $(\boldsymbol{U},\boldsymbol{X},\boldsymbol{Y})$ and $(\overline{\boldsymbol{U}},\overline{\boldsymbol{X}}^{(j_1,j_2)},\boldsymbol{Y})$ share the same $\boldsymbol{Y}$ sequence, and $\mathbb{E}_{\widetilde{P}}[\log q(X,Y)]\ge\mathbb{E}_P[\log q(X,Y)]$ coincides with the condition in (26). Thus, expanding (25) in terms of types yields

$$\overline{p}_{e,0}\le\sum_{P_{UXY}}\mathbb{P}\Big[(\boldsymbol{U},\boldsymbol{X},\boldsymbol{Y})\in T^n(P_{UXY})\Big]\min\Bigg\{1, (M_0-1)\sum_{\widetilde{P}_{UXY}\in\mathcal{T}_{0,n}(P_{UXY})}\mathbb{P}\Big[(\overline{\boldsymbol{U}},\boldsymbol{y})\in T^n(\widetilde{P}_{UY})\Big] \times\mathbb{P}\Bigg[\bigcup_{j_1,j_2}\Big\{\big(\overline{\boldsymbol{u}},\overline{\boldsymbol{X}}^{(j_1,j_2)},\boldsymbol{y}\big)\in T^n(\widetilde{P}_{UXY})\Big\}\Bigg]\Bigg\}, \tag{28}$$

where we write $(\overline{\boldsymbol{u}},\boldsymbol{y})$ to denote an arbitrary pair such that $\boldsymbol{y}\in T^n(P_Y)$ and $(\overline{\boldsymbol{u}},\boldsymbol{y})\in T^n(\widetilde{P}_{UY})$. The dependence of these sequences on $P_Y$ and $\widetilde{P}_{UY}$ is kept implicit.

Using a similar argument to the discussion following (23), we observe that $\big(\overline{\boldsymbol{u}},\overline{\boldsymbol{X}}^{(j_1,j_2)},\boldsymbol{y}\big)\in T^n(\widetilde{P}_{UXY})$ if and only if $\big(\overline{\boldsymbol{X}}_u^{(j_u)},\boldsymbol{y}_u(\overline{\boldsymbol{u}})\big)\in T^{n_u}(\widetilde{P}_{XY|U=u})$ for $u=1,2$. Thus, for any subset $\mathcal{K}$ of $\mathcal{U}$, we can upper bound the probability that $\big(\overline{\boldsymbol{u}},\overline{\boldsymbol{X}}^{(j_1,j_2)},\boldsymbol{y}\big)\in T^n(\widetilde{P}_{UXY})$ by the probability that $\big(\overline{\boldsymbol{X}}_u^{(j_u)},\boldsymbol{y}_u(\overline{\boldsymbol{u}})\big)\in T^{n_u}(\widetilde{P}_{XY|U=u})$ for all $u\in\mathcal{K}$. By further upper bounding via the union bound, we obtain

$$\mathbb{P}\Big[\bigcup_{j_1,j_2}\Big\{\big(\overline{\boldsymbol{u}},\overline{\boldsymbol{X}}^{(j_1,j_2)},\boldsymbol{y}\big)\in T^n(\widetilde{P}_{UXY})\Big\}\Big]\le\min\Big\{1, \min_{u=1,2}M_{1u}\mathbb{P}\Big[\big(\overline{\boldsymbol{X}}_u,\boldsymbol{y}_u(\overline{\boldsymbol{u}})\big)\in T^{n_u}(\widetilde{P}_{XY|U=u})\Big], M_{11}M_{12}\mathbb{P}\Big[\bigcap_{u=1,2}\Big\{\big(\overline{\boldsymbol{X}}_u,\boldsymbol{y}_u(\overline{\boldsymbol{u}})\big)\in T^{n_u}(\widetilde{P}_{XY|U=u})\Big\}\Big]\Big\}, \tag{29}$$

where the four terms in the minimization correspond to the four subsets of $\{1,2\}$.

Substituting (29) into (28), applying the property of types in [8, Ex. 2.3(b)] multiple times, and using the fact that the number of joint types is polynomial in $n$, we obtain

$$\lim_{n\to\infty}-\frac{1}{n}\log\overline{p}_{e,0} \ge\min_{P_{UXY}}\min_{\widetilde{P}_{UXY}\in\mathcal{T}_0(P_{UXY})}D(P_{UXY}\|Q_{UX}\times W)+\Bigg[I_{\widetilde{P}}(U;Y) +\Bigg[\max_{\mathcal{K}\subseteq\mathcal{U}}\sum_{u\in\mathcal{K}}Q_U(u)\Big(I_{\widetilde{P}}(X;Y|U=u)-R_{1u}\Big)\Bigg]^+-R_0\Bigg]^+, \tag{30}$$

where the minimization over $P_{UXY}$ is subject to $P_{UX}=Q_{UX}$. Taking $P_{UXY}\to Q_{UX}\times W$, we obtain that the right-hand side of (30) is positive whenever (10) holds with strict inequality, thus concluding the proof.

One can show that (30) holds with equality by following the steps of [9, Section III-C]. Thus, the analysis presented in this section is exponentially tight for the given ensemble.

## References


[1] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1439–1452, Sep. 1996.

[2] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[3] J. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, MIT, 1983.

[4] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 35–43, Jan. 1995.

[5] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.

[6] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *Arch. Elek. Über.*, vol. 47, no. 4, pp. 228–239, 1993.

[7] J. Körner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 670–679, Nov. 1980.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[9] J. Scarlett and A. Guillén i Fàbregas, "An achievable error exponent for the mismatched multiple-access channel," in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, Oct. 2012.

[10] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "The mismatched multiple-access channel: General alphabets," in *IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, 2013.

[11] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423, July and Oct. 1948.