Multiuser Random Coding Techniques for Mismatched Decoding

Jonathan Scarlett, *Member, IEEE*, Alfonso Martinez, *Senior Member, IEEE*, and Albert Guillén i Fàbregas, *Senior Member, IEEE*

***Abstract*—This paper studies multiuser random coding techniques for channel coding with a given (possibly suboptimal) decoding rule. For the mismatched discrete memoryless multiple-access channel, an error exponent is obtained that is tight with respect to the ensemble average, and positive within the interior of Lapidoth's achievable rate region. This exponent proves the ensemble tightness of the exponent of Liu and Hughes in the case of maximum-likelihood decoding. An equivalent dual form of Lapidoth's achievable rate region is given, and the latter is shown to extend immediately to channels with infinite and continuous alphabets. In the setting of single-user mismatched decoding, similar analysis techniques are applied to a refined version of superposition coding, which is shown to achieve rates at least as high as standard superposition coding for any set of random-coding parameters.**

***Index Terms*—Mismatched decoding, multiple-access channel, superposition coding, random coding, error exponents, ensemble tightness, Lagrange duality, maximum-likelihood decoding.**

I. Introduction

The mismatched decoding problem [1]–[9] seeks to characterize the performance of coded communication systems when the decoding rule is fixed and possibly suboptimal. This problem is of interest, for example, when the optimal decoding rule is infeasible due to channel uncertainty or implementation constraints. Finding a single-letter expression for the mismatched capacity (i.e. the highest achievable rate with mismatched decoding; see Section I-A for formal definitions) remains an open problem even for single-user discrete memoryless channels. The vast majority of existing works have focused on achievability results via random coding.

The most notable early works are by Hui [1] and Csiszár and Körner [2], who independently derived the achievable rate known as the LM rate, using random codes in which each codeword has a constant or nearly-constant composition. A generalization to infinite and continuous alphabets was given by Ganti *et al.* [7] using cost-constrained coding techniques, relying on a Lagrange dual formulation of the LM rate that first appeared in [4]. In general, the LM rate can be strictly smaller than the mismatched capacity [3], [6]. Motivated by the lack of converse results, the concept of ensemble tightness has been addressed in [4], [7], [8], where it has been shown that, for any DMC, the LM rate is the best rate possible for the constant-composition and cost-constrained random-coding ensembles. In [3], Csiszár and Narayan showed that better achievable rates can be obtained by applying the LM rate to the second-order product channel, and similarly for higher-order products. Random-coding error exponents for mismatched decoding were given in [8], [10], [11], and ensemble tightness was addressed in [8].

The mismatched multiple-access channel (MAC) was considered by Lapidoth [6], who obtained an achievable rate region and showed the surprising fact that the single-user LM rate can be improved by treating the single-user channel as a MAC. Thus, as well as being of independent interest, network information theory problems with mismatched decoding can also provide valuable insight into the single-user mismatched decoding problem. In recent work that developed independently of ours, Somekh-Baruch [9] gave error exponents and rate regions for the cognitive MAC (i.e. the MAC where one user knows both messages and the other only knows its own) using two multiuser coding schemes: superposition coding and random binning. When applied to single-user mismatched channels, these yield achievable rates that can improve on those by Lapidoth when certain auxiliary variables are fixed.

In this paper, we build on the work of [6] and study multiuser coding techniques for channels with mismatched decoding. Our main contributions are as follows:

1) We develop a variety of tools for studying multiuser random coding ensembles in mismatched decoding settings. Broadly speaking, our techniques permit the derivations of ensemble-tight error exponents for channels with finite input and output alphabets, as well as generalizations to continuous alphabets based on Lagrange duality analogous to those for the single-user setting mentioned above.
2) By applying our techniques to the mismatched MAC, we provide an alternative derivation of Lapidoth's rate region [6] that also yields the ensemble-tight error exponent, and the appropriate generalization to continuous alphabets. By specializing to the case of ML decoding, we

J. Scarlett was with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. He is now with the Laboratory for Information and Inference Systems, École Polytechnique Fédérale de Lausanne, CH-1015, Switzerland (e-mail: jmscarlett@gmail.com). A. Martinez is with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain (e-mail: alfonso.martinez@ieee.org). A. Guillén i Fàbregas is with the Institució Catalana de Recerca i Estudis Avançats (ICREA), the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain, and also with the Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, U.K. (e-mail: guillen@ieee.org).

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 under grant agreement 303633 and by the Spanish Ministry of Economy and Competitiveness under grants RYC-2011-08150 and TEC2012-38800-C03-03.

This work was presented in part at the 50th Annual Allerton Conference on Communication, Control and Computing 2012, and at the 2013 IEEE International Symposium on Information Theory.

prove the ensemble tightness of the exponent given in [12] for constant-composition random coding, which was previously unknown.
3) For the single-user channel, we introduce a refined version of superposition coding that yields rates at least as high as the standard version [9], [13] for any choice of parameters, with strict improvements possible when the input distribution is fixed.

To avoid overlap with [9], we have omitted the parts of our work that appeared therein; however, these can also be found in [13].

For mismatched DMCs, the results of this paper and various previous works can be summarized by the following list of random-coding constructions, in decreasing order of achievable rate:

1) Refined superposition coding (Theorems 7 and 8),
2) Standard superposition coding (Theorems 5 and 6; see [9], [13]),
3) Expurgated parallel coding [6],
4) Constant-composition or cost-constrained coding with independent codewords (LM Rate [1], [2], [7]),
5) i.i.d. coding with independent codewords (generalized mutual information [10]).

The gap between 1) and 2) can be strict for a given input distribution; no examples are known where the gap between 2) and 3) is strict; and the gaps between the remaining three can be strict even for an optimized input distribution. Numerical examples are provided in Section IV-B.

A. System Setup

Throughout the paper, we consider both the mismatched single-user channel and the mismatched multiple-access channel. Here we provide a description of each.

1) Mismatched Single-User Channel: The input and output alphabets are denoted by $\mathcal{X}$ and $\mathcal{Y}$ respectively, and the channel transition law is denoted by $W(y|x)$, thus yielding an n-letter transition law given by

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

If $\mathcal{X}$ and $\mathcal{Y}$ are finite, the channel is referred to as a discrete memoryless channel (DMC). We consider length-n block coding, in which a codebook $\mathcal{C} = \{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(M)}\}$ is known at both the encoder and decoder. The encoder takes as input a message m uniformly distributed on the set $\{1, \dots, M\}$, and transmits the corresponding codeword $\boldsymbol{x}^{(m)}$. The decoder receives the vector $\boldsymbol{y}$ at the output of the channel, and forms the estimate

$$\hat{m} = \operatorname*{arg\,max}_{j \in \{1,\dots,M\}} q^n(\boldsymbol{x}^{(j)}, \boldsymbol{y}), \tag{2}$$

where n is the length of each codeword, and $q^n(\boldsymbol{x}, \boldsymbol{y}) \triangleq \prod_{i=1}^{n} q(x_i, y_i)$. The function $q(x, y)$ is called the *decoding metric*, and is assumed to be non-negative. In the case of a tie, a codeword achieving the maximum in (2) is selected uniformly at random. In the case that $q(x, y) = W(y|x)$, the decoding rule in (2) is that of optimal maximum-likelihood (ML) decoding.

A rate R is said to be achievable if, for all $\delta > 0$, there exists a sequence of codebooks $\mathcal{C}_n$ with at least $\exp(n(R-\delta))$ codewords of length n such that $\lim_{n\to\infty} p_e(\mathcal{C}_n) = 0$ under the decoding metric q. The mismatched capacity of a given channel and metric is defined to be the supremum of all achievable rates.

An error exponent $E(R)$ is said to be achievable if there exists a sequence of codebooks $\mathcal{C}_n$ with at least $\exp(nR)$ codewords of length n such that

$$\liminf_{n\to\infty} -\frac{1}{n} \log p_e(\mathcal{C}_n) \ge E(R). \tag{3}$$

We let $\overline{p}_e(n, M)$ denote the average error probability with respect to a given random-coding ensemble that will be clear from the context. A random-coding error exponent $E_r(R)$ is said to exhibit ensemble tightness if

$$\lim_{n\to\infty} -\frac{1}{n} \log \overline{p}_e(n, e^{nR}) = E_r(R). \tag{4}$$

For all of the cases of interest in this paper, the limit will exist.

With these definitions, the above-mentioned LM rate is given as follows for an arbitrary input distribution Q:

$$I_{\mathrm{LM}}(Q) \triangleq \min_{\substack{\widetilde{P}_{XY} \,:\, \widetilde{P}_X = Q, \widetilde{P}_Y = P_Y \\ \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)]}} I_{\widetilde{P}}(X;Y), \tag{5}$$

where $P_{XY} = Q \times W$. This rate can equivalently be expressed as [4]

$$I_{\mathrm{LM}}(Q) = \sup_{s \ge 0, a(\cdot)} \mathbb{E}\left[\log \frac{q(X,Y)^s e^{a(X)}}{\mathbb{E}[q(\overline{X},Y)^s e^{a(\overline{X})} \,|\, Y]}\right], \tag{6}$$

where $(X, Y, \overline{X}) \sim Q(x)W(y|x)Q(\overline{x})$. In the terminology of [7], (5) is the primal expression and (6) is the dual expression.

2) Mismatched Multiple-Access Channel: We also consider a 2-user memoryless MAC $W(y|x_1, x_2)$ with input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ and output alphabet $\mathcal{Y}$. In the case that each alphabet is finite, the MAC is referred to as a discrete memoryless MAC (DM-MAC). The decoding metric is denoted by $q(x_1, x_2, y)$, and we write $W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \triangleq \prod_{i=1}^{n} W(y_i|x_{1,i}, x_{2,i})$ and $q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \triangleq \prod_{i=1}^{n} q(x_{1,i}, x_{2,i}, y_i)$.

Encoder $\nu = 1, 2$ takes as input a message m_ν uniformly distributed on the set $\{1, \dots, M_\nu\}$, and transmits the corresponding codeword $\boldsymbol{x}_\nu^{(m_\nu)}$ from the codebook $\mathcal{C}_\nu = \{\boldsymbol{x}_\nu^{(1)}, \dots, \boldsymbol{x}_\nu^{(M_\nu)}\}$. Given the output sequence $\boldsymbol{y}$, the decoder forms an estimate $(\hat{m}_1, \hat{m}_2)$ of the message pair, given by

$$(\hat{m}_1, \hat{m}_2) = \operatorname*{arg\,max}_{(i,j) \in \{1,\dots,M_1\} \times \{1,\dots,M_2\}} q^n(\boldsymbol{x}_1^{(i)}, \boldsymbol{x}_2^{(j)}, \boldsymbol{y}). \tag{7}$$

We assume that ties are resolved uniformly at random. Similarly to the single-user case, optimal ML decoding is recovered by setting $q(x_1, x_2, y) = W(y|x_1, x_2)$.

An error is said to have occurred if the estimate $(\hat{m}_1, \hat{m}_2)$ differs from (m_1, m_2). The error probability for a given pair of codebooks $(\mathcal{C}_1, \mathcal{C}_2)$ is denoted by $p_e(\mathcal{C}_1, \mathcal{C}_2)$, and the error probability for a given random-coding ensemble is denoted by $\overline{p}_e(n, M_1, M_2)$. We define achievable rate pairs, error exponents, and ensemble tightness analogously to the single-user setting.

B. Notation

We use bold symbols for vectors (e.g. $\boldsymbol{x}_1$, $\boldsymbol{y}$), and denote the corresponding i-th entry using a non-bold symbol with a subscript (e.g. $x_{1,i}$, y_i). All logarithms have base e. Moreover, all rates are in units of nats except in the examples, where bits are used. We define $[c]^+ = \max\{0, c\}$, and denote the indicator function by $\mathbb{1}\{\cdot\}$.

The symbol $\sim$ means "distributed as". The set of all probability distributions on an alphabet, say $\mathcal{X}$, is denoted by $\mathcal{P}(\mathcal{X})$, and the set of all empirical distributions on a vector in $\mathcal{X}^n$ (i.e. types [14, Ch. 2], [15]) is denoted by $\mathcal{P}_n(\mathcal{X})$. Similar notations $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ and $\mathcal{P}_n(\mathcal{Y}|\mathcal{X})$ are used for conditional distributions, with the latter adopting the convention that the empirical distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$ is uniform for values of x that do not appear in $\boldsymbol{x}$. For a given $Q \in \mathcal{P}_n(\mathcal{X})$, the type class $T^n(Q)$ is defined to be the set of all sequences in $\mathcal{X}^n$ with type Q. For a given joint type $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ and sequence $\boldsymbol{x} \in T^n(P_X)$, the conditional type class $T^n_{\boldsymbol{x}}(P_{XY})$ is defined to be the set of all sequences $\boldsymbol{y}$ such that $(\boldsymbol{x}, \boldsymbol{y}) \in T^n(P_{XY})$.

The probability of an event is denoted by $\mathbb{P}[\cdot]$. The marginals of a joint distribution $P_{XY}(x, y)$ are denoted by $P_X(x)$ and $P_Y(y)$. 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 joint distribution $P_{XY}(x, y)$ is denoted by $\mathbb{E}_P[\cdot]$, or simply $\mathbb{E}[\cdot]$ when the associated probability distribution is understood from the context. Similarly, mutual information with respect to P_{XY} is written as $I_P(X;Y)$, or simply $I(X;Y)$. Given a distribution $Q(x)$ and conditional distribution $W(y|x)$, we write $Q \times W$ to denote the joint distribution defined by $Q(x)W(y|x)$.

For two positive sequences f_n and g_n, we write $f_n \doteq g_n$ if $\lim_{n\to\infty} \frac{1}{n} \log \frac{f_n}{g_n} = 0$, $f_n \,\dot{\le}\, g_n$ if $\limsup_{n\to\infty} \frac{1}{n} \log \frac{f_n}{g_n} \le 0$, and analogously for $\dot{\ge}$. We make use of the standard asymptotic notations $O(\cdot)$, $o(\cdot)$ and $\Omega(\cdot)$. When studying the MAC, we index the users as $\nu = 1, 2$, and let ν^c denote the unique index differing from ν.

II. Multiple-Access Channel

In this section, we study the mismatched multiple-access channel introduced in Section I-A. We consider random coding, in which each codeword of user $\nu = 1, 2$ is generated independently according to some distribution $P_{\boldsymbol{X}_\nu}$. We let $\boldsymbol{X}_\nu^{(i)}$ be the random variable corresponding to the i-th codeword of user ν, yielding

$$\Big(\{\boldsymbol{X}_1^{(i)}\}_{i=1}^{M_1}, \{\boldsymbol{X}_2^{(j)}\}_{i=1}^{M_2}\Big) \sim \prod_{i=1}^{M_1} P_{\boldsymbol{X}_1}(\boldsymbol{x}_1^{(i)}) \prod_{j=1}^{M_2} P_{\boldsymbol{X}_2}(\boldsymbol{x}_2^{(j)}). \tag{8}$$

We assume without loss of generality that message $(1, 1)$ is transmitted, and write $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ in place of $\boldsymbol{X}_1^{(1)}$ and $\boldsymbol{X}_2^{(1)}$. We write $\overline{\boldsymbol{X}}_1$ and $\overline{\boldsymbol{X}}_2$ to denote arbitrary codewords that are generated independently of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. The random sequence at the output of the channel is denoted by $\boldsymbol{Y}$. It follows that

$$(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}, \overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2) \sim P_{\boldsymbol{X}_1}(\boldsymbol{x}_1) P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2) \times P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2). \tag{9}$$

For clarity of exposition, we focus primarily on the case that there is no time-sharing (e.g. see [12]). In Section II-D, we discuss some of the corresponding results with time-sharing.

We study the random-coding error probability by considering the following events:

$$\text{(Type 1)} \qquad \frac{q^n(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \ge 1 \text{ for some } i \ne 1;$$

$$\text{(Type 2)} \qquad \frac{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \ge 1 \text{ for some } j \ne 1;$$

$$\text{(Type 12)} \qquad \frac{q^n(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \ge 1 \text{ for some } i \ne 1,\ j \ne 1.$$

We refer to these as error events, though they do not necessarily imply decoder errors when the inequalities hold with equality, since we have assumed that the decoder resolves ties uniformly at random.

The probabilities of the error events are denoted by $\overline{p}_{e,1}(n, M_1)$, $\overline{p}_{e,2}(n, M_2)$ and $\overline{p}_{e,12}(n, M_1, M_2)$, and the overall random-coding error probability is denoted by $\overline{p}_e(n, M_1, M_2)$. Since breaking ties as errors increases the error probability by at most a factor of two [16], we have

$$\frac{1}{2} \max\{\overline{p}_{e,1}, \overline{p}_{e,2}, \overline{p}_{e,12}\} \le \overline{p}_e \le \overline{p}_{e,1} + \overline{p}_{e,2} + \overline{p}_{e,12}. \tag{10}$$

A. Exponents and Rates for the DM-MAC

In this subsection, we study the DM-MAC using the constant-composition ensemble. For $\nu = 1, 2$, we fix $Q_\nu \in \mathcal{P}(\mathcal{X}_\nu)$ and let $P_{\boldsymbol{X}_\nu}$ be the uniform distribution on $T^n(Q_{\nu,n})$, where $Q_{\nu,n} \in \mathcal{P}_n(\mathcal{X}_\nu)$ is a type with the same support as Q_ν such that $\max_{x_\nu} |Q_{\nu,n}(x_\nu) - Q_\nu(x_\nu)| \le \frac{1}{n}$. Thus,

$$P_{\boldsymbol{X}_\nu}(\boldsymbol{x}_\nu) = \frac{1}{|T^n(Q_{\nu,n})|} \mathbb{1}\big\{\boldsymbol{x}_\nu \in T^n(Q_{\nu,n})\big\}. \tag{11}$$

Our analysis is based on the method of types [14, Ch. 2]. Throughout the section, we write $f(\boldsymbol{Q})$ to denote a quantity f that depends on Q_1 and Q_2. Similarly, we write $f(\boldsymbol{Q}_n)$ to denote a quantity that depends on $Q_{1,n}$ and $Q_{2,n}$.

1) Error Exponents: The error exponents and achievable rates are expressed in terms of the following sets ($\nu = 1, 2$):

$$\mathcal{S}(\boldsymbol{Q}) \triangleq \Big\{ P_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) \,:\, P_{X_1} = Q_1,\, P_{X_2} = Q_2 \Big\} \tag{12}$$

$$\mathcal{T}_\nu(P_{X_1 X_2 Y}) \triangleq \Big\{ \widetilde{P}_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) \,:\, \widetilde{P}_{X_\nu} = P_{X_\nu}, \widetilde{P}_{X_{\nu^c} Y} = P_{X_{\nu^c} Y}, \mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)] \ge \mathbb{E}_P[\log q(X_1, X_2, Y)] \Big\} \tag{13}$$

$$\mathcal{T}_{12}(P_{X_1 X_2 Y}) \triangleq \Big\{ \widetilde{P}_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) \,:\, \widetilde{P}_{X_1} = P_{X_1}, \widetilde{P}_{X_2} = P_{X_2}, \widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)] \ge \mathbb{E}_P[\log q(X_1, X_2, Y)] \Big\}, \tag{14}$$

where we recall that for $\nu = 1, 2$, ν^c denotes the unique element differing from ν.

Theorem 1. *For any mismatched DM-MAC, for the constant-composition ensemble in* (11) *with input distributions Q_1 and Q_2, the ensemble-tight error exponents are given as follows for $\nu = 1, 2$:*

$$\lim_{n\to\infty} -\frac{1}{n}\log \overline{p}_{e,\nu}(n, e^{nR_\nu}) = E_{r,\nu}^{\mathrm{cc}}(\boldsymbol{Q}, R_\nu) \tag{15}$$

$$\lim_{n\to\infty} -\frac{1}{n}\log \overline{p}_{e,12}(n, e^{nR_1}, e^{nR_2}) = E_{r,12}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2), \tag{16}$$

where

$$E_{r,\nu}^{\mathrm{cc}}(\boldsymbol{Q}, R_\nu) \triangleq \min_{P_{X_1X_2Y}\in\mathcal{S}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_\nu(P_{X_1X_2Y})} D(P_{X_1X_2Y}\|Q_1\times Q_2\times W) + \big[I_{\widetilde{P}}(X_\nu; X_{\nu^c}, Y) - R_\nu\big]^+ \tag{17}$$

$$E_{r,12}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) \triangleq \min_{P_{X_1X_2Y}\in\mathcal{S}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}(P_{X_1X_2Y})} D(P_{X_1X_2Y}\|Q_1\times Q_2\times W) + \Big[\max\Big\{I_{\widetilde{P}}(X_1;Y) - R_1, I_{\widetilde{P}}(X_2;Y) - R_2, D\big(\widetilde{P}_{X_1X_2Y}\|Q_1\times Q_2\times P_Y\big) - R_1 - R_2\Big\}\Big]^+. \tag{18}$$

Proof: The random-coding error probabilities $\overline{p}_{e,1}$ and $\overline{p}_{e,2}$ can be handled similarly to the single-user setting [8]. Furthermore, equivalent error exponents to (17) ($\nu = 1, 2$) were given in [17]. We therefore focus on $\overline{p}_{e,12}$, which requires a more careful analysis. We first rewrite

$$\overline{p}_{e,12} = \mathbb{E}\left[\mathbb{P}\left[\bigcup_{i\neq 1, j\neq 1}\left\{\frac{q^n(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \geq 1\right\}\middle| \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}\right]\right]. \tag{19}$$

in terms of the possible joint types of $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$ and $(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})$. To this end, we define

$$\mathcal{S}_n(\boldsymbol{Q}_n) \triangleq \Big\{P_{X_1X_2Y}\in\mathcal{P}_n(\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}) : P_{X_1} = Q_{1,n}, P_{X_2} = Q_{2,n}\Big\} \tag{20}$$

$$\mathcal{T}_{12,n}(P_{X_1X_2Y}) \triangleq \mathcal{T}_{12}(P_{X_1X_2Y}) \cap \mathcal{P}_n(\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y}). \tag{21}$$

Roughly speaking, $\mathcal{S}_n$ is the set of possible joint types of $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})$, and $\mathcal{T}_{12,n}(P_{X_1X_2Y})$ is the set of types of $(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})$ that lead to decoding errors when $(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) \in T^n(P_{X_1X_2Y})$. The constraints on P_{X_ν} and $\widetilde{P}_{X_\nu}$ arise from the fact that we are using constant-composition random coding, and the constraint $\mathbb{E}_{\widetilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)]$ holds if and only if $q^n(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$ for $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \in T^n(P_{X_1X_2Y})$ and $(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1X_2Y})$. Fixing $P_{X_1X_2Y} \in \mathcal{S}_n(\boldsymbol{Q}_n)$ and letting $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$ be an arbitrary triplet of sequences such that $(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}) \in T^n(P_{X_1X_2Y})$, it follows that the event in (19) can be written as

$$\bigcup_{i\neq 1, j\neq 1} \bigcup_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12,n}} \Big\{(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y}) \in T^n(\widetilde{P}_{X_1X_2Y})\Big\}. \tag{22}$$

Expanding the probability and expectation in (19) in terms of types, substituting (22), and interchanging the order of the unions, we obtain

$$\overline{p}_{e,12} = \sum_{P_{X_1X_2Y}\in\mathcal{S}_n(\boldsymbol{Q}_n)} \mathbb{P}\big[(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) \in T^n(P_{X_1X_2Y})\big] \times \mathbb{P}\Bigg[\bigcup_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12,n}(P_{X_1X_2Y})} \bigcup_{i\neq 1, j\neq 1} \Big\{(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1X_2Y})\Big\}\Bigg] \tag{23}$$

$$\doteq \max_{P_{X_1X_2Y}\in\mathcal{S}_n(\boldsymbol{Q}_n)} \mathbb{P}\big[(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}) \in T^n(P_{X_1X_2Y})\big] \times \max_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12,n}(P_{X_1X_2Y})} \mathbb{P}\Bigg[\bigcup_{i\neq 1, j\neq 1} \Big\{(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{y}) \in T^n(\widetilde{P}_{X_1X_2Y})\Big\}\Bigg], \tag{24}$$

where $\boldsymbol{y}$ is an arbitrary element of $T^n(P_Y)$ (hence depending *implicitly* on $P_{X_1X_2Y}$), and (24) follows from the union bound and since the number of joint types is polynomial in n.

By a standard property of types [14, Ch. 2], the exponent of the first probability in (24) is given by $D(P_{X_1X_2Y}\|Q_1\times Q_2\times W)$, so it only remains to determine the exponential behavior of the second probability. To this end, we make use of Lemma 2 in Appendix A with $Z_1(i) = \boldsymbol{X}_1^{(i)}$, $Z_2(j) = \boldsymbol{X}_2^{(j)}$, $\mathcal{A} = T_{\boldsymbol{y}}^n(\widetilde{P}_{X_1X_2Y})$, $\mathcal{A}_1 = T_{\boldsymbol{y}}^n(\widetilde{P}_{X_1Y})$ and $\mathcal{A}_2 = T_{\boldsymbol{y}}^n(\widetilde{P}_{X_2Y})$. Using (A.10)–(A.11) and standard properties of types [14, Ch. 2], it follows that the second probability in (24) has an exponent of

$$\Big[\max\Big\{I_{\widetilde{P}}(X_1;Y) - R_1, I_{\widetilde{P}}(X_2;Y) - R_2, D\big(\widetilde{P}_{X_1X_2Y}\|Q_1\times Q_2\times P_Y\big) - R_1 - R_2\Big\}\Big]^+. \tag{25}$$

Upon substituting (25) into (24), it only remains to replace the sets $\mathcal{S}_n$ and $\mathcal{T}_{12,n}$ by $\mathcal{S}$ and $\mathcal{T}_{12}$ respectively. This is seen to be valid since the underlying objective function is continuous in $\widetilde{P}_{X_1X_2Y}$, and since any joint distribution has a corresponding joint type which is within $\frac{1}{n}$ in each value of the probability mass function. See the discussion around [18, Eq. (30)] for the analogous continuity argument in the single-user setting. ∎

Theorem 1 and (10) reveal that the overall ensemble-tight error exponent is given by

$$E_r^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2) \triangleq \min\Big\{E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1), E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2), E_{r,12}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2)\Big\}. \tag{26}$$

The proof of Theorem 1 made use of the refined union bound given in Lemma 2. If we had instead used the standard truncated union bound in (A.1), we would have obtained the weaker type-12 exponent

$$E_{r,12}^{\mathrm{cc}'}(\boldsymbol{Q},R_1,R_2) \triangleq \min_{P_{X_1X_2Y}\in\mathcal{S}(\boldsymbol{Q})} \min_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}(P_{X_1X_2Y})} D(P_{X_1X_2Y}\|Q_1\times Q_2\times W) + \big[D(\widetilde{P}_{X_1X_2Y}\|Q_1\times Q_2\times P_Y) - (R_1+R_2)\big]^+, \quad (27)$$

which coincides with an achievable exponent given in [17].

2) Achievable Rate Region: The following theorem is a direct consequence of Theorem 1, and provides an alternative proof of Lapidoth's ensemble-tight achievable rate region [6].

Theorem 2. *The overall error exponent* $E_r^{\mathrm{cc}}(\boldsymbol{Q},R_1,R_2)$ *in* (26) *is positive for all rate pairs* (R_1,R_2) *in the interior of* $\mathcal{R}_{\mathrm{LM}}(\boldsymbol{Q})$*, defined to be the set of all rate pairs* (R_1,R_2) *satisfying the following for* $\nu=1,2$*:*

$$R_\nu \le \min_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_\nu(Q_1\times Q_2\times W)} I_{\widetilde{P}}(X_\nu;X_{\nu^c},Y) \quad (28)$$

$$R_1+R_2 \le \min_{\substack{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}(Q_1\times Q_2\times W)\\ I_{\widetilde{P}}(X_1;Y)\le R_1,\, I_{\widetilde{P}}(X_2;Y)\le R_2}} D(\widetilde{P}_{X_1X_2Y}\|Q_1\times Q_2\times P_Y). \quad (29)$$

Proof: The conditions in (28)–(29) are obtained from (17)–(18) respectively. Focusing on (29), we see that the objective in (18) is always positive when $D(P_{X_1X_2Y}\|Q_1\times Q_2\times W)>0$, $I_{\widetilde{P}}(X_1;Y)>R_1$ or $I_{\widetilde{P}}(X_2;Y)>R_2$. Moreover, by a similar argument to [3, Lemma 1], the right-hand side of (18), with only the second minimization kept, is continuous as a function of $P_{X_1X_2Y}$ when restricted to distributions with the same support as $Q_1\times Q_2\times W$. Hence, we may substitute $Q_1\times Q_2\times W$ for $P_{X_1X_2Y}$ (thus forcing the first divergence to zero) and introduce the constraints $I_{\widetilde{P}}(X_1;Y)\le R_1$ and $I_{\widetilde{P}}(X_2;Y)\le R_2$ to obtain the condition in (29). ■

Using a time-sharing argument [6], [19] (see also Section II-D), it follows from Theorem 2 that we can achieve any rate pair in the convex hull of $\bigcup_{\boldsymbol{Q}}\mathcal{R}_{\mathrm{LM}}(\boldsymbol{Q})$, where the union is over all distributions Q_1 and Q_2 on $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively.

Using a similar argument to the proof of Theorem 2, we see that (27) yields the rate condition

$$R_1+R_2 \le \min_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}(Q_1\times Q_2\times W)} D(\widetilde{P}_{X_1X_2Y}\|Q_1\times Q_2\times P_Y). \quad (30)$$

In Section IV-A, we compare (18) and (29) with the weaker expressions in (27) and (30).

B. Exponents and Rates for General Alphabets

In this section, we present equivalent dual expressions for the rates given in Theorem 2, and extend them to the memoryless MAC with general alphabets. While we focus on rates for brevity, dual expressions and continuous-alphabet generalizations for the exponents in Theorem 1 can be obtained similarly; see [13, Sec. 4.2] for details.

We use the cost-constrained ensemble [8], [11], defined as follows. We fix $Q_1\in\mathcal{P}(\mathcal{X}_1)$ and $Q_2\in\mathcal{P}(\mathcal{X}_2)$, and choose

$$P_{\boldsymbol{X}_\nu}(\boldsymbol{x}_\nu)=\frac{1}{\mu_{\nu,n}}\prod_{i=1}^n Q_\nu(x_{\nu,i})\mathbb{1}\big\{\boldsymbol{x}_\nu\in\mathcal{D}_{\nu,n}\big\} \quad (31)$$

for $\nu=1,2$, where $\mu_{\nu,n}$ is a normalizing constant, and

$$\mathcal{D}_{\nu,n}\triangleq\left\{\boldsymbol{x}_\nu\,:\,\left|\frac{1}{n}\sum_{i=1}^n a_{\nu,l}(x_{\nu,i})-\phi_{\nu,l}\right|\le\frac{\delta}{n},\; l=1,\dots,L_\nu\right\}, \quad (32)$$

where $\{a_{\nu,l}\}_{l=1}^{L_\nu}$ are auxiliary cost functions, δ is a positive constant, and $\phi_{\nu,l}\triangleq\mathbb{E}_{Q_\nu}[a_{\nu,l}(X_\nu)]$. Thus, the codewords for user ν are constrained to satisfy L_ν cost constraints in which the empirical mean of $a_{\nu,l}(\cdot)$ is close to the true mean. We allow each of the parameters to be optimized, including the cost functions. The case $L_\nu=0$ should be understood as corresponding to the case that $\mathcal{D}_{\nu,n}$ contains all $\boldsymbol{x}_\nu$ sequences, thus recovering the i.i.d. distribution studied in [20]. In the case of finite input alphabets, the constant-composition ensemble can also be recovered by setting $L_\nu=|\mathcal{X}_\nu|$ and letting each auxiliary cost function be the indicator function of its argument equaling a given input symbol [8].

The cost-constrained ensemble has primarily been used with $L_\nu=1$ [11], [21], but the inclusion of multiple cost functions has proven beneficial in the mismatched single-user setting [8], [22]. We will see that the use of multiple costs is beneficial for both the matched and mismatched MAC. We note that *system* costs (as opposed to the *auxiliary* costs used here) can easily be handled (e.g. see [8, Sec. VII], [22]), but in this paper we assume for simplicity that the channel is unconstrained.

The following proposition from [8] will be useful.

Proposition 1. [8, Prop. 1] *For* $\nu=1,2$*, fix the input distribution* Q_ν *along with* L_ν *and the auxiliary cost functions* $\{a_{\nu,l}\}_{l=1}^{L_\nu}$*. Then* $\mu_{\nu,n}=\Omega(n^{-L_\nu/2})$ *provided that* $\mathbb{E}_{Q_\nu}[a_{\nu,l}(X_\nu)^2]<\infty$ *for* $l=1,\dots,L_\nu$*.*

The main result of this subsection is the following theorem.

Theorem 3. *The region* $\mathcal{R}_{\mathrm{LM}}(\boldsymbol{Q})$ *in* (28)–(29) *can be expressed as the set of rate pairs* (R_1,R_2) *satisfying*

$$R_1\le\sup_{s\ge0,a_1(\cdot)}\mathbb{E}\left[\log\frac{q(X_1,X_2,Y)^s e^{a_1(X_1)}}{\mathbb{E}\big[q(\overline{X}_1,X_2,Y)^s e^{a_1(\overline{X}_1)}\,|\,X_2,Y\big]}\right] \quad (33)$$

$$R_2\le\sup_{s\ge0,a_2(\cdot)}\mathbb{E}\left[\log\frac{q(X_1,X_2,Y)^s e^{a_2(X_2)}}{\mathbb{E}\big[q(X_1,\overline{X}_2,Y)^s e^{a_2(\overline{X}_2)}\,|\,X_1,Y\big]}\right], \quad (34)$$

and at least one of

$$R_1\le\sup_{\rho_2\in[0,1],s\ge0,a_1(\cdot),a_2(\cdot)} -\rho_2R_2 + \mathbb{E}\left[\log\frac{\big(q(X_1,X_2,Y)^s e^{a_2(X_2)}\big)^{\rho_2}e^{a_1(X_1)}}{\mathbb{E}\Big[\Big(\mathbb{E}\big[q(\overline{X}_1,\overline{X}_2,Y)^s e^{a_2(\overline{X}_2)}\,\big|\,\overline{X}_1\big]\Big)^{\rho_2}e^{a_1(\overline{X}_1)}\Big|Y\Big]}\right] \quad (35)$$

$$R_2\le\sup_{\rho_1\in[0,1],s\ge0,a_1(\cdot),a_2(\cdot)} -\rho_1R_1 + \mathbb{E}\left[\log\frac{\big(q(X_1,X_2,Y)^s e^{a_1(X_1)}\big)^{\rho_1}e^{a_2(X_2)}}{\mathbb{E}\Big[\Big(\mathbb{E}\big[q(\overline{X}_1,\overline{X}_2,Y)^s e^{a_1(\overline{X}_1)}\,\big|\,\overline{X}_2\big]\Big)^{\rho_1}e^{a_2(\overline{X}_2)}\Big|Y\Big]}\right], \quad (36)$$

where $(X_1, X_2, Y, \overline{X}_1, \overline{X}_2)$ *is distributed as* $Q_1(x_1)Q_2(x_2)W(y|x_1, x_2)Q_1(\overline{x}_1)Q_2(\overline{x}_2)$.

Moreover, this region is achievable for any memoryless MAC (possibly having infinite or continuous alphabets) and any pair (Q_1, Q_2), *where each supremum is subject to* $\mathbb{E}_{Q_\nu}[a_\nu(X_\nu)^2] < \infty$ $(\nu = 1, 2)$. *Any point in the region can be achieved using cost-constrained coding with* $L_1 = L_2 = 3$.

Proof: The equivalence of this rate region to (28)–(29) is proved in Appendix C. Here we prove the second claim of the theorem by providing a direct derivation.

The key initial step is to obtain the following non-asymptotic bound on the type-12 error event, holding for any codeword distributions $P_{\boldsymbol{X}_1}$ and $P_{\boldsymbol{X}_2}$:

$$\overline{p}_{e,12}(n, M_1, M_2) \le \min_{\nu=1,2} \mathrm{rcu}_{12,\nu}(n, M_1, M_2), \tag{37}$$

where for $\nu = 1, 2$ we define

$$\mathrm{rcu}_{12,\nu}(n, M_1, M_2) \triangleq \mathbb{E}\Bigg[\min\bigg\{1, (M_\nu - 1)\mathbb{E}\bigg[\min\bigg\{1, (M_{\nu^c} - 1) \mathbb{P}\bigg[\frac{q^n(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \ge 1 \,\bigg|\, \overline{\boldsymbol{X}}_\nu\bigg]\bigg\} \,\bigg|\, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}\bigg]\bigg\}\Bigg]. \tag{38}$$

To prove this, we first write

$$\overline{p}_{e,12} = \mathbb{P}\Bigg[\bigcup_{i\neq 1, j\neq 1}\bigg\{\frac{q^n(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \ge 1\bigg\}\Bigg] \tag{39}$$

$$= \mathbb{E}\Bigg[\mathbb{P}\Bigg[\bigcup_{i\neq 1, j\neq 1}\bigg\{\frac{q^n(\boldsymbol{X}_1^{(i)}, \boldsymbol{X}_2^{(j)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y})} \ge 1\bigg\}\,\Bigg|\, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y}\Bigg]\Bigg]. \tag{40}$$

We obtain the above-mentioned bounds by applying Lemma 1 in Appendix A to the union in (40) (with $Z_1(i) = \boldsymbol{X}_1^{(i)}$ and $Z_2(j) = \boldsymbol{X}_2^{(j)}$), and then writing $\min\{1, \alpha, \beta\} \le \min\{1, \alpha\}$ and $\min\{1, \alpha, \beta\} \le \min\{1, \beta\}$.

Define $Q_\nu^n(\boldsymbol{x}_\nu) \triangleq \prod_{i=1}^n Q_\nu(x_{\nu,i})$ for $\nu = 1, 2$. Expanding (38) and applying Markov's inequality and $\min\{1, \alpha\} \le \alpha^\rho$ $(0 \le \rho \le 1)$, we obtain[1]

$$\mathrm{rcu}_{12,1}(n, M_1) \le \sum_{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}} P_{\boldsymbol{X}_1}(\boldsymbol{x}_1)P_{\boldsymbol{X}_2}(\boldsymbol{x}_2)W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2)\Bigg(M_1\sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) \times\bigg(M_2\frac{\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2)q^n(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y})^s}{q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})^s}\bigg)^{\rho_2}\Bigg)^{\rho_1} \tag{41}$$

for any $\rho_1 \in [0, 1]$, $\rho_2 \in [0, 1]$ and $s \ge 0$. For $\nu = 1, 2$, we let $a_\nu(x)$ be one of the three cost functions in the ensemble, and we define $a_\nu^n(\boldsymbol{x}_\nu) \triangleq \sum_{i=1}^n a_\nu(x_{\nu,i})$ and $\phi_\nu \triangleq \mathbb{E}_{Q_\nu}[a_\nu(X_\nu)]$. In accordance with the theorem statement, we assume that $\mathbb{E}_{Q_\nu}[a_\nu(X_\nu)]^2 < \infty$, so that Proposition 1 holds. Using the bounds on the cost functions in (32), we can weaken (41) to

$$\mathrm{rcu}_{12,1}(n, M_1) \le e^{2\delta(\rho_1+\rho_1\rho_2+1)} \times \sum_{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}} P_{\boldsymbol{X}_1}(\boldsymbol{x}_1)P_{\boldsymbol{X}_2}(\boldsymbol{x}_2)W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2)\Bigg(M_1\sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) \times\bigg(M_2\frac{\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2)q^n(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{y})^s e^{a_2^n(\overline{\boldsymbol{x}}_2)}}{q^n(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})^s e^{a_2^n(\boldsymbol{x}_2)}}\bigg)^{\rho_2}\frac{e^{a_1^n(\overline{\boldsymbol{x}}_1)}}{e^{a_1^n(\boldsymbol{x}_1)}}\Bigg)^{\rho_1}. \tag{42}$$

We upper bound (42) by substituting (31) and replacing the summations over $\mathcal{D}_{\nu,n}$ by summations over all sequences on $\mathcal{X}_\nu^n$. Writing the resulting terms (e.g. $W^n(\boldsymbol{y}|\boldsymbol{x}_1, \boldsymbol{x}_2)$) as a product from 1 to n and taking the supremum over (s, ρ_1, ρ_2) and the cost functions, we obtain a bound whose exponent is

$$\max_{\rho_1\in[0,1], \rho_2\in[0,1]} E_{0,12,1}^{\mathrm{cost}}(\boldsymbol{Q}, \rho_1, \rho_2) - \rho_1(R_1 + \rho_2 R_2), \tag{43}$$

where

$$E_{0,12,1}^{\mathrm{cost}}(\boldsymbol{Q}, \rho_1, \rho_2) \triangleq \sup_{s\ge 0, a_1(\cdot), a_2(\cdot)} -\log\mathbb{E}\Bigg[\Bigg(\mathbb{E}\Bigg[\bigg(\frac{\mathbb{E}\big[q(\overline{X}_1, \overline{X}_2, Y)^s e^{a_2(\overline{X}_2)} \,|\, \overline{X}_1\big]}{q(X_1, X_2, Y)^s e^{a_2(X_2)}}\bigg)^{\rho_2} \times \frac{e^{a_1(\overline{X}_1)}}{e^{a_1(X_1)}} \,\Bigg|\, X_1, X_2, Y\Bigg]\Bigg)^{\rho_1}\Bigg] \tag{44}$$

We obtain the condition in (35) by taking the derivative of $E_{0,12,1}^{\mathrm{cc}}$ at zero, analogously to the proof of Theorem 3. We obtain (36) analogously by starting with $\mathrm{rcu}_{12,2}$ in place of $\mathrm{rcu}_{12,1}$, and we obtain (33)–(34) via a simpler analysis following the standard single-user setting [8].

Finally, we note that $L_1 = L_2 = 3$ suffices due to the fact that the cost functions used in deriving (35)–(36) may coincide, since the theorem statement only requires one of the two to hold. ∎

Theorem 3 extends Lapidoth's MAC rate region to general alphabets, analogously to the extension of the single-user LM rate to general alphabet by Ganti *et al.* [7]. Compared to the single-user setting, the extension is non-trivial, requiring refined union bounds, as well as a technique for handling the two additional in constraints in (29) one at a time, thus leading to two type-12 conditions in (35)–(36).

C. Matched MAC Error Exponent

Here we apply our results to the setting of ML decoding, where $q(x_1, x_2, y) = W(y|x_1, x_2)$. The best known exponent for the constant-composition ensemble was derived by Liu and Hughes [12], and was shown to yield a strict improvement over Gallager's exponent for the i.i.d. ensemble [20] even after the optimization of the input distributions.

We have seen that for a general decoding metric, the overall error exponent E_r^{cc} given in (26) may be reduced when $E_{r,12}^{\mathrm{cc}\prime}$ in (27) is used in place of $E_{r,12}^{\mathrm{cc}}$. The following result shows that the resulting expressions are in fact identical in the matched case.

[1] In the case of continuous alphabets, the summations should be replaced by integrals as necessary.

Theorem 4. *Under ML decoding (i.e. $q(x_1, x_2, y) = W(y|x_1, x_2)$), we have for any input distributions (Q_1, Q_2) and rates (R_1, R_2) that*

$$\min\left\{E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1), E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2), E_{r,12}^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2)\right\} = \min\left\{E_{r,1}^{\mathrm{cc}}(\boldsymbol{Q}, R_1), E_{r,2}^{\mathrm{cc}}(\boldsymbol{Q}, R_2), E_{r,12}^{\mathrm{cc}'}(\boldsymbol{Q}, R_1, R_2)\right\}. \tag{45}$$

Thus, both the left-hand side and right-hand side of (45) *equal the overall ensemble-tight error exponent.*

Proof: See Appendix C. ∎

While it is possible that $E_{r,12}^{\mathrm{cc}} > E_{r,12}^{\mathrm{cc}'}$ under ML decoding, Theorem 4 shows that this never occurs in the region where $E_{r,12}^{\mathrm{cc}}$ achieves the minimum in (26). Thus, combining Theorem 4 and Theorem 1, we conclude that the exponent given in [12] is ensemble-tight for the constant-composition ensemble under ML decoding.

In [13, Sec. 4.2.4], [23], we show that the error exponent of [12] admits a dual form resembling the i.i.d. exponent of Gallager [20], but with additional optimization parameters $a_1(\cdot)$ and $a_2(\cdot)$ that are functions of the input alphabets $\mathcal{X}_1$ and $\boldsymbol{X}_2$. As usual, this dual form can also be derived directly via the cost-constrained ensemble, with the analysis remaining valid for infinite and continuous alphabets.

D. Time-Sharing

Thus far, we have focused on the standard random coding ensemble described by (8), where the codewords are independent. It is well-known that even in the matched case, the union of the resulting achievable rate regions over all (Q_1, Q_2) may be non-convex, and time-sharing is needed to achieve the rest of the capacity region [24]. There are two distinct ways of doing so: (i) With explicit time-sharing, one splits the block of length n into two or more smaller blocks, and uses separate codebooks within each block; (ii) With coded time-sharing, one still generates a single codebook, but the codewords are *conditionally* independent given some time-sharing sequence $\boldsymbol{U}$ on a time-sharing alphabet $\mathcal{U}$. In particular, in the case of constant-composition random coding, one may let $\boldsymbol{U}$ be uniform on a type class corresponding to $Q_U \in \mathcal{P}(\mathcal{U})$, and let each $\boldsymbol{X}_\nu$ be uniform on a conditional type class corresponding to $Q_\nu \in \mathcal{P}(\mathcal{X}_\nu\,|\,\mathcal{U})$.

While both of these schemes yield the entire capacity region in the matched case [19, Ch. 4], the coded time-sharing approach is generally preferable in terms of exponents [12]. Intuitively, this is because explicit time-sharing shortens the effective block length, thus diminishing the exponent.

Surprisingly, however, explicit time-sharing can outperform coded time-sharing in the mismatched case, even in terms of the achievable rate region. This is most easily understood via the dual-domain expressions, and for concreteness we consider the case $|\mathcal{U}| = 2$ with $Q_U = (\lambda, 1-\lambda)$. Let $I_1(\boldsymbol{Q}, s)$ denote the right-hand side of (33) with a fixed value of s in place of the supremum. Using explicit time-sharing with two different input distribution pairs $\boldsymbol{Q}^{(1)}$ and $\boldsymbol{Q}^{(2)}$, the condition corresponding to (33) is given by

$$R_1 \le \lambda \sup_{s\ge 0} I_1(\boldsymbol{Q}^{(1)}, s) + \big(1-\lambda\big) \sup_{s\ge 0} I_1(\boldsymbol{Q}^{(2)}, s), \tag{46}$$

whereas coded time-sharing only permits

$$R_1 \le \sup_{s\ge 0} \Big(\lambda I_1(\boldsymbol{Q}^{(1)}, s) + \big(1-\lambda\big) I_1(\boldsymbol{Q}^{(2)}, s)\Big). \tag{47}$$

These are obtained using similar arguments to the case without time-sharing; see [13, Sec. 4.2.5] for further details. Similar observations apply for the other rate conditions, including the parameters ρ_1 and ρ_2 in (35)–(36).

It is evident from (46) (and the other analogous rate conditions) that explicit time-sharing between two points can be used to obtain any pair (R_1, R_2) on the line connecting two achievable pairs corresponding to $\boldsymbol{Q}^{(1)}$ and $\boldsymbol{Q}^{(2)}$. On the other hand, the same is only true for coded time-sharing if there exists a single parameter s simultaneously maximizing both terms in the objective function of (47) (and similarly for the other rate conditions), which is not the case in general.

Building on this insight, in the following section, we compare two forms of superposition coding for single-user channels. The standard version can be viewed as analogous to coded time-sharing, whereas the refined version can be viewed as analogous to explicit time-sharing. As a result, the latter can lead to higher achievable rates.

III. Superposition Coding

In this section, we turn to the single-user mismatched channel introduced in Section I-A1, and consider multiuser coding schemes that can improve on standard schemes with independent codewords. Some numerical examples are given in Section IV.

A. Standard Superposition Coding

We first discuss a standard form of superposition coding that has had extensive application in degraded broadcast channels [25]–[27] and other network information theory problems [19]. This ensemble was studied in the context of mismatched decoding in [9], [13], so we do not repeat the details here.

The parameters of the ensemble are an auxiliary alphabet $\mathcal{U}$, an auxiliary codeword distribution $P_{\boldsymbol{U}}$, and a conditional codeword distribution $P_{\boldsymbol{X}|\boldsymbol{U}}$. We fix two rates R_0 and R_1. An auxiliary codebook $\{\boldsymbol{U}^{(i)}\}_{i=1}^{M_0}$ with $M_0 \triangleq \lfloor e^{nR_0} \rfloor$ codewords is generated at random, with each auxiliary codeword independently distributed according to $P_{\boldsymbol{U}}$. For each $i = 1, \dots, M_0$, a codebook $\{\boldsymbol{X}^{(i,j)}\}_{j=1}^{M_1}$ with $M_1 \triangleq \lfloor e^{nR_1} \rfloor$ codewords is generated at random, with each codeword conditionally independently distributed according to $P_{\boldsymbol{X}|\boldsymbol{U}}$. The message m at the input to the encoder is indexed as (m_0, m_1), and for any such pair, the corresponding codeword is $\boldsymbol{X}^{(m_0,m_1)}$.

The following achievable rate for DMCs is obtained using constant-composition coding with some input distribution $Q_{UX} \in \mathcal{P}(\mathcal{U}\times\mathcal{X})$, in which $P_{\boldsymbol{U}}$ is the uniform distribution on a type class corresponding to Q_U, and $P_{\boldsymbol{X}|\boldsymbol{U}}$ is the uniform distribution on a conditional type class corresponding to $Q_{X|U}$. We define the sets

$$\mathcal{S}(Q_{UX}) \triangleq \big\{P_{UXY} \in \mathcal{P}(\mathcal{U}\times\mathcal{X}\times\mathcal{Y}) \,:\, P_{UX} = Q_{UX}\big\} \tag{48}$$

$$\mathcal{T}_0(P_{UXY}) \triangleq \Big\{ \widetilde{P}_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}) \,:\, \widetilde{P}_{UX} = P_{UX}, \\ \widetilde{P}_Y = P_Y, \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \geq \mathbb{E}_P[\log q(X,Y)] \Big\} \quad (49)$$

$$\mathcal{T}_1(P_{UXY}) \triangleq \Big\{ \widetilde{P}_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}) \,:\, \widetilde{P}_{UX} = P_{UX}, \\ \widetilde{P}_{UY} = P_{UY}, \mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \geq \mathbb{E}_P[\log q(X,Y)] \Big\}. \quad (50)$$

Theorem 5. [9], [13] *Suppose that W is a DMC. For any finite auxiliary alphabet $\mathcal{U}$, and input distribution $Q_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$, the rate*

$$R = R_0 + R_1 \quad (51)$$

is achievable provided that (R_0, R_1) satisfy

$$R_1 \leq \min_{\widetilde{P}_{UXY} \in \mathcal{T}_1(Q_{UX} \times W)} I_{\widetilde{P}}(X;Y|U) \quad (52)$$

$$R_0 + R_1 \leq \min_{\substack{\widetilde{P}_{UXY} \in \mathcal{T}_0(Q_{UX} \times W) \\ I_{\widetilde{P}}(U;Y) \leq R_0}} I_{\widetilde{P}}(U,X;Y). \quad (53)$$

This rate is also known to be tight with respect to the ensemble average [9], [13]. It is known to be at least as high as Lapidoth's expurgated parallel coding rate [6], though it is not known whether the improvement can be strict.

Using similar steps to those in the previous section, one can obtain the following equivalent dual form, which also remains valid in the case of continuous alphabets [13, Sec. 5.2.2].

Theorem 6. [13] *The achievable rate conditions in* (52)–(53) *can be expressed as*

$$R_1 \leq \sup_{s \geq 0, a(\cdot,\cdot)} \mathbb{E}\left[\log \frac{q(X,Y)^s e^{a(U,X)}}{\mathbb{E}[q(\widetilde{X},Y)^s e^{a(U,\widetilde{X})} \,|\, U, Y]} \right] \quad (54)$$

$$R_0 \leq \sup_{\rho_1 \in [0,1], s \geq 0, a(\cdot,\cdot)} -\rho_1 R_1 \\ + \mathbb{E}\left[\log \frac{\big(q(X,Y)^s e^{a(U,X)}\big)^{\rho_1}}{\mathbb{E}\Big[\Big(\mathbb{E}\big[q(\overline{X},Y)^s e^{a(\overline{U},\overline{X})} \,\big|\, \overline{U}\big]\Big)^{\rho_1} \,\Big|\, Y\Big]} \right], \quad (55)$$

where $(U, X, Y, \widetilde{X}, \overline{U}, \overline{X})$ is distributed as $Q_{UX}(u,x)W(y|x)Q_{X|U}(\widetilde{x}|u)Q_{UX}(\overline{u},\overline{x})$.

We observe that superposition coding has some similarity to the coded time-sharing ensemble discussed in Section II-D, in that both involve generating codewords $\boldsymbol{x}$ conditionally on auxiliary sequences $\boldsymbol{u}$ according to the uniform distribution on a type class. We saw in Section II-D that better rates are in fact achieved by explicit time-sharing, in which one splits the block length into sub-blocks and codes individually on each one. We now apply this approach to superposition coding, yielding a refined ensemble that can lead to higher achievable rates than the standard version.

B. Refined Superposition Coding

The ensemble is defined as follows. We fix a finite alphabet $\mathcal{U}$, an input distribution $Q_U \in \mathcal{P}(\mathcal{U})$ and the rates R_0 and $\{R_{1u}\}_{u \in \mathcal{U}}$. We write $M_0 \triangleq \lfloor e^{nR_0} \rfloor$ and $M_{1u} \triangleq \lfloor e^{nR_{1u}} \rfloor$. We let $P_{\boldsymbol{U}}(\boldsymbol{u})$ be the uniform distribution on the type class

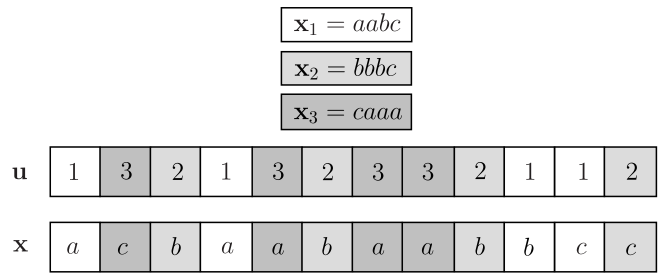

Figure 1. The construction of the codeword from the auxiliary sequence $\boldsymbol{u}$ and the partial codewords $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ for refined SC. Here we have $\mathcal{U} = \{1,2,3\}$, $\mathcal{X} = \{a,b,c\}$, $n_1 = n_2 = n_3 = 4$, and $n = 12$.

$T^n(Q_{U,n})$, where $Q_{U,n}$ is a type with the same support as Q_U such that $\max_u |Q_{U,n}(u) - Q_U(u)| \leq \frac{1}{n}$. We set

$$P_{\boldsymbol{U}}(\boldsymbol{u}) = \frac{1}{|T^n(Q_{U,n})|} \mathbb{1}\Big\{ \boldsymbol{u} \in T^n(Q_{U,n}) \Big\} \quad (56)$$

and generate the length-n auxiliary codewords $\{\boldsymbol{U}^{(i)}\}_{i=1}^{M_0}$ independently according to $P_{\boldsymbol{U}}$. The difference compared to standard superposition coding is that the codewords are not generated conditionally independently given $\boldsymbol{U}^{(i)}$. Instead, we generate a number of partial codewords, and construct the length-n codeword by placing the entries of a partial codeword in the indices where $\boldsymbol{U}$ takes a particular value.

More precisely, for each $u \in \mathcal{U}$, we define

$$n_u \triangleq Q_{U,n}(u) n \quad (57)$$

and fix a partial codeword distribution $P_{\boldsymbol{X}_u} \in \mathcal{P}(\mathcal{X}^{n_u})$. For each $i = 1, \dots, M_0$ and $u \in \mathcal{U}$, we 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

$$\bigg\{ \Big(\boldsymbol{U}^{(i)}, \{\boldsymbol{X}_1^{(i,j_1)}\}_{j_1=1}^{M_{11}}, \{\boldsymbol{X}_2^{(i,j_2)}\}_{j_2=1}^{M_{12}} \Big) \bigg\}_{i=1}^{M_0} \\ \sim \prod_{i=1}^{M_0} \bigg(P_{\boldsymbol{U}}(\boldsymbol{u}^{(i)}) \prod_{j_1=1}^{M_{11}} P_{\boldsymbol{X}_1}(\boldsymbol{x}_1^{(i,j_1)}) \prod_{j_2=1}^{M_{12}} P_{\boldsymbol{X}_2}(\boldsymbol{x}_2^{(i,j_2)}) \bigg). \quad (58)$$

The message m at the encoder 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 $M = M_0 \prod_u M_{1u}$ codewords, and hence the rate is $R = R_0 + \sum_u Q_{U,n}(u) R_{1u}$. An example of the construction of the codeword $\boldsymbol{x}$ from the auxiliary sequence $\boldsymbol{u}$ and partial codewords $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ is shown in Figure 1, where we have $\mathcal{U} = \{1,2,3\}$ and $\mathcal{X} = \{a,b,c\}$.

While our main result is stated for an arbitrary finite alphabet $\mathcal{U}$, the analysis will be presented for $\mathcal{U} = \{1,2\}$ for clarity. We proceed by presenting several definitions for this specific choice. 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 we write

$$\boldsymbol{X}^{(i,j_1,j_2)} \triangleq \Xi(\boldsymbol{U}^{(i)}, \boldsymbol{X}_1^{(i,j_1)}, \boldsymbol{X}_2^{(i,j_2)}). \quad (59)$$

We let $\boldsymbol{y}_u(\boldsymbol{u})$ denote the subsequence of $\boldsymbol{y}$ corresponding to the indices where $\boldsymbol{u}$ equals u, and similarly for $\boldsymbol{Y}_u(\boldsymbol{u})$.

We assume without loss of generality that $(m_0, m_1, m_2) = (1, 1, 1)$. We let $\boldsymbol{U}$, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ and $\boldsymbol{X}$ be the codewords corresponding to $(1, 1, 1)$, yielding $\boldsymbol{X} = \Xi(\boldsymbol{U}, \boldsymbol{X}_1, \boldsymbol{X}_2)$. We 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. It follows that $\overline{\boldsymbol{X}}^{(j_1,j_2)} = \Xi(\overline{\boldsymbol{U}}, \overline{\boldsymbol{X}}_1^{(j_1)}, \overline{\boldsymbol{X}}_2^{(j_2)})$.

Upon receiving a realization $\boldsymbol{y}$ of the output sequence $\boldsymbol{Y}$, the decoder forms the estimate

$$
\begin{aligned}
&(\hat{m}_0, \hat{m}_1, \hat{m}_2) \\
&\quad = \arg\max_{(i,j_1,j_2)} q^n(\boldsymbol{x}^{(i,j_1,j_2)}, \boldsymbol{y}) && (60)\\
&\quad = \arg\max_{(i,j_1,j_2)} q^{n_1}\big(\boldsymbol{x}_1^{(i,j_1)}, \boldsymbol{y}_1(\boldsymbol{u}^{(i)})\big) q^{n_2}\big(\boldsymbol{x}_2^{(i,j_2)}, \boldsymbol{y}_2(\boldsymbol{u}^{(i)})\big), && (61)
\end{aligned}
$$

where the objective in (61) follows by separating the indices where $u = 1$ from those where $u = 2$. By writing the objective in this form, we see 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 $q^{n_2}(\boldsymbol{x}_2^{(i,j_2)}, \boldsymbol{y}_2(\boldsymbol{u}^{(i)}))$. We thus consider three error events:

(Type 0) $\quad \dfrac{q^n(\boldsymbol{X}^{(i,j_1,j_2)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1$ for some $i \neq 1$, j_1, j_2;

(Type 1) $\quad \dfrac{q^{n_1}(\overline{\boldsymbol{X}}_1^{(1,j_1)}, \boldsymbol{Y}_1(\boldsymbol{U}))}{q^{n_1}(\boldsymbol{X}_1, \boldsymbol{Y}_1(\boldsymbol{U}))} \geq 1$ for some $j_1 \neq 1$;

(Type 2) $\quad \dfrac{q^{n_2}(\overline{\boldsymbol{X}}_2^{(1,j_2)}, \boldsymbol{Y}_2(\boldsymbol{U}))}{q^{n_2}(\boldsymbol{X}_2, \boldsymbol{Y}_2(\boldsymbol{U}))} \geq 1$ for some $j_2 \neq 1$.

The corresponding probabilities are denoted by $\overline{p}_{e,0}(n, M_0, M_{11}, M_{12})$, $\overline{p}_{e,1}(n, M_{11})$ and $\overline{p}_{e,2}(n, M_{12})$ respectively. Analogously to (10), the overall random-coding error probability $\overline{p}_e(n, M_0, M_{11}, M_{12})$ satisfies

$$
\frac{1}{2}\max\{\overline{p}_{e,0}, \overline{p}_{e,1}, \overline{p}_{e,2}\} \leq \overline{p}_e \leq \overline{p}_{e,0} + \overline{p}_{e,1} + \overline{p}_{e,2}. \quad (62)
$$

While our analysis of the error probability will yield non-asymptotic bounds and error exponents as intermediate steps, we focus on the resulting achievable rates for clarity.

C. Rates for DMCs

In this subsection, we assume that the channel is a DMC. We fix a joint distribution Q_{UX}, and let $Q_{UX,n}$ be a corresponding type with $\max_{u,x} |Q_{UX,n}(u,x) - Q_{UX}(u,x)| \leq \frac{1}{n}$. We let $P_{\boldsymbol{X}_u}$ be the uniform distribution on the type class $T^{n_u}\big(Q_{X|U,n}(\cdot|u)\big)$, yielding

$$
\begin{aligned}
&P_{\boldsymbol{X}_u}(\boldsymbol{x}_u) \\
&\quad = \frac{1}{\big|T^{n_u}\big(Q_{X|U,n}(\cdot|u)\big)\big|} \mathbb{1}\Big\{\boldsymbol{x}_u \in T^{n_u}\big(Q_{X|U,n}(\cdot|u)\big)\Big\}. \quad (63)
\end{aligned}
$$

Combining this with (56), we have by symmetry that each pair $(\boldsymbol{U}^{(i)}, \boldsymbol{X}^{(i,j_1,j_2)})$ is uniformly distributed on $T^n(Q_{UX})$.

The main result of this section is stated in the following theorem, which makes use of the LM rate defined in (5) and the set $\mathcal{T}_0$ defined in (49).

Theorem 7. *For any finite set $\mathcal{U}$ and input distribution Q_{UX}, the rate*

$$
R = R_0 + \sum_u Q_U(u) R_{1u} \quad (64)
$$

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

$$
R_{1u} \leq I_{\mathrm{LM}}\big(Q_{X|U}(\cdot|u)\big), \quad u \in \mathcal{U} \quad (65)
$$

$$
\begin{aligned}
R_0 \leq &\min_{\widetilde{P}_{UXY} \in \mathcal{T}_0(Q_{UX} \times W)} I_{\widetilde{P}}(U;Y) + \\
&\left[\max_{\mathcal{K} \subseteq \mathcal{U}, \mathcal{K} \neq \emptyset} \sum_{u \in \mathcal{K}} Q_U(u)\Big(I_{\widetilde{P}}(X;Y|U=u) - R_{1u}\Big)\right]^+. \quad (66)
\end{aligned}
$$

Proof: As mentioned above, the proof is presented only for $\mathcal{U} = \{1, 2\}$; the same arguments apply in the general case. Observe that the type-1 error event corresponds to the error event for the standard constant-composition ensemble with rate R_{11}, length $n_1 = nQ_U(1)$, input distribution $Q_{X|U}(\cdot|1)$, and ties treated as errors. 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 LM rate derivation [1], [2], yielding (65).

The error probability for the type-0 event is given by

$$
\overline{p}_{e,0} = \mathbb{P}\left[\bigcup_{i \neq 1}\bigcup_{j_1, j_2}\left\{\frac{q^n(\boldsymbol{X}^{(i,j_1,j_2)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1\right\}\right], \quad (67)
$$

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

$$
\begin{aligned}
&\overline{p}_{e,0} = c_0 \mathbb{E}\Bigg[\min\Bigg\{1, (M_0 - 1) \\
&\times \mathbb{E}\Bigg[\mathbb{P}\Bigg[\bigcup_{j_1, j_2}\left\{\frac{q^n(\overline{\boldsymbol{X}}^{(j_1,j_2)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1\right\} \Bigg| \overline{\boldsymbol{U}}\Bigg] \Bigg| \boldsymbol{U}, \boldsymbol{X}, \boldsymbol{Y}\Bigg]\Bigg\}\Bigg], \quad (68)
\end{aligned}
$$

where $c_0 \in [\frac{1}{2}, 1]$, since for independent events the truncated union bound is tight to within a factor of $\frac{1}{2}$ [28, Lemma A.2]. 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

$$
\frac{q^n(\overline{\boldsymbol{X}}^{(j_1,j_2)}, \boldsymbol{Y})}{q^n(\boldsymbol{X}, \boldsymbol{Y})} \geq 1 \quad (69)
$$

can be written as

$$
\widetilde{P}_{UXY} \in \mathcal{T}_{0,n}(P_{UXY}) \triangleq \mathcal{T}_0(P_{UXY}) \cap \mathcal{P}_n(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}), \quad (70)
$$

where $\mathcal{T}_0$ is defined in (49). 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 $\mathbb{E}_{\widetilde{P}}[\log q(X,Y)] \geq$

$\mathbb{E}_P[\log q(X,Y)]$ coincides with the condition in (69). Thus, expanding (68) in terms of types, we obtain

$$\overline{p}_{e,0} = c_0 \sum_{P_{UXY}} \mathbb{P}\Big[(\boldsymbol{U},\boldsymbol{X},\boldsymbol{Y}) \in T^n(P_{UXY})\Big]$$
$$\times \min\Bigg\{1, (M_0-1) \sum_{\widetilde{P}_{UXY} \in \mathcal{T}_{0,n}(P_{UXY})} \mathbb{P}\Big[(\overline{\boldsymbol{U}},\boldsymbol{y}) \in T^n(\widetilde{P}_{UY})\Big]$$
$$\times \mathbb{P}\Big[\bigcup_{j_1,j_2} \Big\{\big(\overline{\boldsymbol{u}},\overline{\boldsymbol{X}}^{(j_1,j_2)},\boldsymbol{y}\big) \in T^n(\widetilde{P}_{UXY})\Big\}\Big]\Bigg\}, \quad (71)$$

where we write $(\overline{\boldsymbol{u}},\boldsymbol{y})$ to denote an arbitrary pair such that $\boldsymbol{y} \in T^n(P_Y)$ and $(\overline{\boldsymbol{u}},\boldsymbol{y}) \in T^n(\widetilde{P}_{UY})$; note that these sequences *implicitly* depend on P_{UXY} and $\widetilde{P}_{UXY}$.

Similarly to the discussion following (61), 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 $\big(\overline{\boldsymbol{X}}_u^{(j_u)},\boldsymbol{y}_u(\overline{\boldsymbol{u}})\big) \in T^{n_u}(\widetilde{P}_{XY|U}(\cdot,\cdot|u))$ for $u=1,2$. Thus, applying Lemma 2 in Appendix A with $Z_1(j_1) = \boldsymbol{X}_1^{(j_1)}$, $Z_2(j_2) = \boldsymbol{X}_2^{(j_2)}$, $\mathcal{A}_1 = T^{n_1}_{\boldsymbol{y}_1(\overline{\boldsymbol{u}})}(\widetilde{P}_{XY|U}(\cdot,\cdot|1))$, $\mathcal{A}_2 = T^{n_2}_{\boldsymbol{y}_2(\overline{\boldsymbol{u}})}(\widetilde{P}_{XY|U}(\cdot,\cdot|2))$, and $\mathcal{A} = \big\{(\boldsymbol{X}_1,\boldsymbol{X}_2) : \boldsymbol{X}_u \in T^{n_u}_{\boldsymbol{y}_u(\overline{\boldsymbol{u}})}(\widetilde{P}_{XY|U}(\cdot,\cdot|u)),\, u=1,2\big\}$, we obtain

$$\mathbb{P}\Big[\bigcup_{j_1,j_2} \Big\{\big(\overline{\boldsymbol{u}},\overline{\boldsymbol{X}}^{(j_1,j_2)},\boldsymbol{y}\big) \in T^n(\widetilde{P}_{UXY})\Big\}\Big] = \quad (72)$$
$$\zeta_0' \min\Bigg\{1, \min_{u=1,2} M_{1u}\mathbb{P}\Big[\big(\overline{\boldsymbol{X}}_u,\boldsymbol{y}_u(\overline{\boldsymbol{u}})\big) \in T^{n_u}\big(\widetilde{P}_{XY|U}(\cdot,\cdot|u)\big)\Big],$$
$$M_{11}M_{12}\mathbb{P}\Big[\bigcap_{u=1,2} \Big\{\big(\overline{\boldsymbol{X}}_u,\boldsymbol{y}_u(\overline{\boldsymbol{u}})\big) \in T^{n_u}\big(\widetilde{P}_{XY|U}(\cdot,\cdot|u)\big)\Big\}\Big]\Bigg\}, \quad (73)$$

where $\zeta_0' \in [\frac{1}{4},1]$. This is a minimization of four terms corresponding to the four subsets of $\{1,2\}$.

Substituting (73) into (71) and applying standard properties of types [14, Ch. 2], we obtain

$$\lim_{n\to\infty} -\frac{1}{n}\log\overline{p}_{e,0} = \min_{P_{UXY}\,:\,P_{UX}=Q_{UX}}$$
$$\min_{\widetilde{P}_{UXY}\in\mathcal{T}_0(P_{UXY})} D(P_{UXY}\|Q_{UX}\times W) + \Bigg[I_{\widetilde{P}}(U;Y)+$$
$$\Bigg[\max_{\mathcal{K}\subseteq\mathcal{U},\mathcal{K}\neq\emptyset} \sum_{u\in\mathcal{K}} Q_U(u)\Big(I_{\widetilde{P}}(X;Y|U=u)-R_{1u}\Big)\Bigg]^+ - R_0\Bigg]^+, \quad (74)$$

where we have replaced the minimizations over types by minimizations over all distributions in the same way as the proof of Theorem 1. By a similar argument to [2, Lemma 1], the right-hand side of (74), with only the second minimization kept, is continuous as a function of P_{UXY} when restricted to distributions whose support is the same as that of $Q_{UX}\times W$. It follows that the right-hand side of (74) is positive whenever (66) holds with strict inequality. ■

The proof of Theorem 7 gives an exponentially tight analysis yielding the exponent in (74). This does not prove that the resulting rate is ensemble-tight, since a subexponential decay of the error probability to zero is possible in principle. However, the changes required to prove the tightness of the rate are minimal. We saw that each condition in (64) corresponds to an error event with independent constant-composition codewords and a reduced block length, and hence it follows from existing analyses [4], [6] that $\overline{p}_{e,1}\to 1$ when R_{11} fails this condition, and analogously for $\overline{p}_{e,2}$ and R_{12}. To see that $\overline{p}_{e,0}\to 1$ when (66) fails, we let $\mathcal{E}_i$ be the event that $q^n(\boldsymbol{X}^{(i,j_1,j_2)},\boldsymbol{Y}) \geq q^n(\boldsymbol{X},\boldsymbol{Y})$ for some (j_1,j_2), let $I_0(P_{UXY})$ denote the right-hand side of (66) with P_{UXY} in place of $Q_1\times Q_2\times W$, and write

$$\overline{p}_{e,0} = \mathbb{P}\Big[\bigcup_{i\neq 1}\mathcal{E}_i\Big] \quad (75)$$
$$= \sum_{P_{UXY}} \mathbb{P}[(\boldsymbol{U},\boldsymbol{X},\boldsymbol{Y})\in T^n(P_{UXY})] \times \Big(1-\big(1-\mathbb{P}[\mathcal{E}_2\,|\,P_{UXY}]\big)^{M_0-1}\Big) \quad (76)$$
$$\geq \sum_{P_{UXY}} \mathbb{P}[(\boldsymbol{U},\boldsymbol{X},\boldsymbol{Y})\in T^n(P_{UXY})] \times \Big(1-\big(1-p_0(n)e^{-nI_0(P_{UXY})}\big)^{M_0-1}\Big), \quad (77)$$

where (76) follows since the events $\mathcal{E}_i$ are conditionally i.i.d. given that $(\boldsymbol{U},\boldsymbol{X},\boldsymbol{Y})$ has a given joint type P_{UXY}, and (77) holds for some subexponential factor $p_0(n)$ by (74). Next, we observe from the law of large numbers that the joint type of $(\boldsymbol{U},\boldsymbol{X},\boldsymbol{Y})$ approaches $Q_1\times Q_2\times W$ with high probability as $n\to\infty$. Moreover, by the same argument as that of the LM rate [3, Lemma 1], $I_0(P_{UXY})$ is continuous in P_{UXY}. Combining these observations, we readily obtain from (77) that $\overline{p}_{e,0}\to 1$ if $R_0 > I_0(Q_1\times Q_2\times W)$, as desired.

D. Comparison to Standard Superposition Coding

In this subsection, we show that the conditions in (65)–(66) can be weakened to (52)–(53) upon identifying

$$R_1 = \sum_u Q_U(u)R_{1u}. \quad (78)$$

Proposition 2. *For any finite auxiliary alphabet $\mathcal{U}$ and input distribution Q_{UX}, the rate $\max_{R_0,R_{11},\dots,R_{1|\mathcal{U}|}} R_0 + \sum_u Q_U(u)R_{1u}$ resulting from Theorem 7 is at least as high as the rate $\max_{R_0,R_1} R_0+R_1$ resulting from Theorem 5.*

Proof: We begin by weakening (66) to (53). We lower bound the right-hand side of (66) by replacing the maximum over $\mathcal{K}$ by the particular choice $\mathcal{K}=\mathcal{U}$, yielding

$$R_0 \leq \min_{\widetilde{P}_{UXY}\in\mathcal{T}_0(Q_{UX}\times W)} I_{\widetilde{P}}(U;Y) + \Big[I_{\widetilde{P}}(X;Y|U)-R_1\Big]^+, \quad (79)$$

where we have used (78) and the definition of conditional mutual information. We can weaken (79) to (53) using the chain rule for mutual information, and noting that (79) is always satisfied when the minimizing $\widetilde{P}_{UXY}$ satisfies $I_{\widetilde{P}}(U;Y) > R_0$.

Next, we show that highest value of R_1 permitted by the $|\mathcal{U}|$ conditions in (65), denoted by R_1^*, can be lower bounded by the right-hand side of (52). From (78) and (65), we have

$$R_1^* = \sum_u Q_U(u) I_{\widetilde{P}^*}(X;Y|U=u), \quad (80)$$

where $\widetilde{P}^*_{XY|U}(\cdot,\cdot|u)$ is the distribution that achieves the minimum in (5) under $Q_{X|U}(\cdot|u)$. Defining the joint distribution $\widetilde{P}^*_{UXY}$ accordingly with $\widetilde{P}^*_U = Q_U$, we can write (80) as

$$R_1^* = I_{\widetilde{P}^*}(X;Y|U). \tag{81}$$

Therefore, we can lower bound R_1^* by the right-hand side of (52) provided that $\widetilde{P}^*_{UXY} \in \mathcal{T}_1(Q_{UX}\times W)$. The constraints $\widetilde{P}^*_{UX} = Q_{UX}$ and $\widetilde{P}^*_{UY} = P_{UY}$ in (50) are satisfied since we have chosen $\widetilde{P}^*_U = Q_U$, and since the constraints in (5) imply $\widetilde{P}^*_{X|U}(\cdot|u) = Q_{X|U}(\cdot|u)$ and $\widetilde{P}^*_{Y|U}(\cdot|u) = P_{Y|U}(\cdot|u)$ for all $u\in\mathcal{U}$. The constraint $\mathbb{E}_{\widetilde{P}^*}[\log q(X,Y)] \ge \mathbb{E}_P[\log q(X,Y)]$ is satisfied since, from (5), we have $\mathbb{E}_{\widetilde{P}^*}[\log q(X,Y)\,|\,U=u] \ge \mathbb{E}_P[\log q(X,Y)\,|\,U=u]$ for all $u\in\mathcal{U}$. ∎

Intuitively, one can think of the gain of the refined superposition coding ensemble as being due to a stronger dependence among the codewords. For standard SC, the codewords $\{\boldsymbol{X}^{(i,j)}\}_{j=1}^{M_1}$ are conditionally independent given $\boldsymbol{U}^{(i)}$, whereas for refined superposition coding this is generally not the case. The additional structure leads to further constraints in the minimizations, and maxima over more terms in the objective functions, both leading to higher overall rates.

It should be noted, however, that the exponents for standard superposition coding may be higher, particularly at low to moderate rates. In particular, we noted in the proof of Theorem 7 that the type-1 and type-2 error events are equivalent to a single-user channel, but the corresponding block lengths are only n_1 and n_2. Thus, if either $Q_U(1)$ or $Q_U(2)$ is close to zero, the corresponding exponent is small.

Finally, we recall that the standard superposition coding rate is at least as high as Lapidoth's expurgated parallel coding rate [9], though no example of strict improvement is known.

E. Dual Expressions and General Alphabets

In this subsection, we present a dual expression for the rate given in Theorem 7 in the case that $|\mathcal{U}|=2$, as well as extending the result to general alphabets $\mathcal{X}$ and $\mathcal{Y}$.

With $\mathcal{U}=\{1,2\}$, the condition in (66) is given by

$$\begin{aligned} R_0 \le \min_{\widetilde{P}_{UXY}\in\mathcal{T}_0(Q_{UX}\times W)} & I_{\widetilde{P}}(U;Y) \\ &+ \Big[\max\Big\{Q_U(1)\big(I_{\widetilde{P}}(X;Y|U=1)-R_{11}\big), \\ Q_U(2)\big(I_{\widetilde{P}}(X;Y|U=2)-R_{12}\big), & I_{\widetilde{P}}(X;Y|U)-R_1\Big\}\Big]^+, \end{aligned} \tag{82}$$

where

$$R_1 \triangleq \sum_u Q_U(u)R_{1u}. \tag{83}$$

Since the right-hand side of (65) is the LM rate, we can use the dual expression in (6). The main result of this subsection gives a dual expression for (82), and extends its validity to memoryless MACs with infinite or continuous alphabets.

We again use cost-constrained random coding. We consider the ensemble given in (58), with $P_{\boldsymbol{X}_u}$ given by

$$P_{\boldsymbol{X}_u}(\boldsymbol{x}_u) = \frac{1}{\mu_{u,n_u}}\prod_{i=1}^{n_u} Q_{X|U}(x_{u,i}|u_i)\mathbb{1}\big\{\boldsymbol{x}_u\in\mathcal{D}_{u,n_u}\big\}, \tag{84}$$

where

$$\mathcal{D}_{u,n_u} \triangleq \Bigg\{\boldsymbol{x}_u : \Bigg|\frac{1}{n_u}\sum_{i=1}^{n_u} a_{u,l}(x_{u,i}) - \phi_{u,l}\Bigg| \le \frac{\delta}{n_u},\ l=1,\dots,L_u\Bigg\} \tag{85}$$

$$\phi_{u,l} \triangleq \mathbb{E}_{Q_u}\big[a_{u,l}(X_u)\,|\,U=u\big], \tag{86}$$

and where μ_{u,n_u}, $\{a_{u,l}\}$ and δ are defined analogously to (32), and n_u is defined in (57).

Theorem 8. *The condition in* (82) *holds if and only if the following holds for at least one of $u=1,2$:*

$$R_0 \le \sup_{s\ge0,\rho_1\in[0,1],\rho_2\in[0,1],a(\cdot,\cdot)} -\sum_{u'=1,2}\rho_u(u')Q_U(u')R_{1u'} + \mathbb{E}\left[\log\frac{\big(q(X,Y)^{s_u(U)}e^{a(U,X)}\big)^{\rho_u(U)}}{\mathbb{E}\Big[\Big(\mathbb{E}\big[q(\overline{X},Y)^{s_u(\overline{U})}e^{a(\overline{U},\overline{X})}\,|\,\overline{U}\big]\Big)^{\rho_u(\overline{U})}\,\Big|\,Y\Big]}\right] \tag{87}$$

where

$$\rho_1(1)=\rho_1,\ \rho_1(2)=\rho_1\rho_2,\ s_1(1)=\rho_2 s,\ s_1(2)=s \tag{88}$$

$$\rho_2(1)=\rho_1\rho_2,\ \rho_2(2)=\rho_2,\ s_2(1)=s,\ s_2(2)=\rho_1 s \tag{89}$$

and $(U,X,Y,\overline{U},\overline{X}) \sim Q_{UX}(u,x)W(y|x)Q_{UX}(\overline{u},\overline{x})$.

Moreover, for any mismatched memoryless channel (possibly having infinite or continuous alphabets) and input distribution Q_{UX} ($\mathcal{U}=1,2$), the rate $R=R_0+\sum_{u=1,2}Q_U(u)R_{1u}$ is achievable for any triplet (R_0,R_{11},R_{12}) satisfying (65) *(with I_{LM} defined in* (6)*) and* (87) *for at least one of $u=1,2$. The supremum in* (6) *is subject to $\mathbb{E}_Q[a(X)^2]<\infty$, and that in* (87) *is subject to $\mathbb{E}_Q[a(U,X)^2]<\infty$. Furthermore, the rate is achievable using cost-constrained coding in* (84) *with $L_1=L_2=2$.*

Proof: Both the proof of the primal-dual equivalence is and the direct derivation of (87) are given in Appendix D. The choice $L_1=L_2=2$ suffices since for $u=1,2$, one cost is required for (65) and another for (87). It suffices to let the cost functions for (87) with $u=1$ and $u=2$ coincide, since the theorem only requires that one of the two hold. ∎

The condition in (87) bears a strong resemblance to the standard superposition coding condition in (55); the latter can be recovered by setting $\rho_2=1$ in the condition with $u=1$, or or $\rho_1=1$ in the condition with $u=2$.

IV. Numerical Examples

A. Error Exponent for the Multiple-Access Channel

We revisit the parallel BSC example given by Lapidoth [6], consisting of binary inputs $\mathcal{X}_1=\mathcal{X}_2=\{0,1\}$ and a pair of binary outputs $\mathcal{Y}=\{0,1\}^2$. The output is given by $Y=(Y_1,Y_2)$, where for $\nu=1,2$, Y_ν is generated by passing X_ν through a binary symmetric channel (BSC) with some crossover probability $\delta_\nu<0.5$. The mismatched decoder assumes that both crossover probabilities are equal to $\delta<0.5$. The decoder assumes that both crossover probabilities

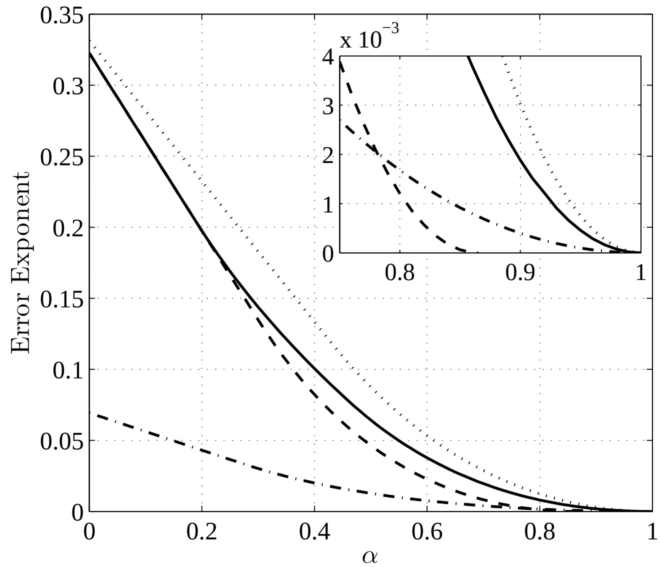

Figure 2. Error exponents $E_{r,1}^{\mathrm{cc}}$ (dotted), $E_{r,2}^{\mathrm{cc}}$ (dash-dot), $E_{r,12}^{\mathrm{cc}}$ (solid) and $E_{r,12}^{\mathrm{cc}'}$ (dashed) for the parallel channel using $\delta_1 = 0.05$, $\delta_2 = 0.25$ and equiprobable input distributions on $\{0,1\}$. The rate pair is given by $(R_1, R_2) = (\alpha C_1, \alpha C_2)$.

are equal. The corresponding decoding rule is equivalent to minimizing sum of t_1 and t_2, where t_ν is the number of bit flips from the input sequence $\boldsymbol{x}_\nu$ to the output sequence $\boldsymbol{y}_\nu$. As noted in [6], this decision rule is in fact equivalent to ML.

We let both Q_1 and Q_2 be equiprobable on $\{0,1\}$. With this choice, it was shown in [6] that the right-hand side of (30) is no greater than

$$2\left(1 - H_2\left(\frac{\delta_1+\delta_2}{2}\right)\right) \text{ bits/use,} \tag{90}$$

where $H_2(\cdot)$ is the binary entropy function in bits. In fact, this is the same rate that would be obtained by considering the corresponding *single-user* channel with $X = (X_1, X_2)$, and applying the LM rate with a uniform distribution on the quaternary input alphabet [6].

On the other hand, the refined condition in (29) can be used to prove the achievability of *any* (R_1, R_2) within the rectangle with corners $(0,0)$ and (C_1, C_2), where $C_\nu \triangleq 1 - H_2(\delta_\nu)$ [6]. This implies that the mismatched capacity region coincides with the (matched) capacity region.

We evaluate the error exponents using the optimization software YALMIP [29]. Figure 2 plots each of the exponents as a function of α, where the rate pair is $(R_1, R_2) = (\alpha C_1, \alpha C_2)$. While the overall error exponent $E_r^{\mathrm{cc}}(\boldsymbol{Q}, R_1, R_2)$ in (26) is unchanged at low to moderate values of α when $E_{r,12}^{\mathrm{cc}'}$ in (27) is used in place of $E_{r,12}^{\mathrm{cc}}$, this is not true for high values of α. Furthermore, consistent with the preceding discussion, $E_{r,12}^{\mathrm{cc}'}$ is non-zero only for $\alpha < 0.865$, whereas $E_{r,12}^{\mathrm{cc}}$ is positive for all $\alpha < 1$. The fact that $E_{r,12}^{\mathrm{cc}}$ and $E_{r,12}^{\mathrm{cc}'}$ coincide at low values of α is consistent with [17, Cor. 5], which states that $E_{r,12}^{\mathrm{cc}'}$ is ensemble-tight at low rates.

B. Achievable Rates for Single-User Channels

In this subsection, we provide examples comparing the two versions of superposition coding and the LM rate. We do not explicitly give values for Lapidoth's rate [6], since for each example given, we found it to coincide with the superposition coding rate (see Theorem 5).

1) Sum Channel: We first consider a sum-channel analog of the parallel-channel example given in Section IV-A. 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 [30]. One can similarly combine two metrics $q_1(x_1, y_1)$ and $q_2(x_2, y_2)$ to form a sum metric $q(x,y)$. Assuming without loss of generality that $\mathcal{X}_1$ and $\mathcal{X}_2$ are disjoint and $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint, we have

$$q(x,y) = \begin{cases} q_1(x_1,y_1) & x_1 \in \mathcal{X}_1 \text{ and } y_1 \in \mathcal{Y}_1 \\ q_2(x_2,y_2) & x_2 \in \mathcal{X}_2 \text{ and } y_2 \in \mathcal{Y}_2 \\ 0 & \text{otherwise,} \end{cases} \tag{91}$$

and similarly for $W(y|x)$. Let $\hat{Q}_1$ and $\hat{Q}_2$ be the distributions that maximize the LM rate in (5) on the respective subchannels. We set $\mathcal{U} = \{1,2\}$, $Q_{X|U}(\cdot|1) = (\hat{Q}_1, \mathbf{0})$ and $Q_{X|U}(\cdot|2) = (\mathbf{0}, \hat{Q}_2)$, 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 (49), we find that the minimizing $\widetilde{P}_{UXY}$ in (66) only has 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 $y \in \mathcal{Y}_2$. It follows that U is a deterministic function of Y under the minimizing $\widetilde{P}_{UXY}$, and hence

$$I_{\widetilde{P}}(U;Y) = H(Q_U) - H_{\widetilde{P}}(U|Y) = H(Q_U). \tag{92}$$

Therefore, the right-hand side of (66) is lower bounded by $H(Q_U)$. Using (64), it follows that we can achieve the rate

$$\begin{aligned} H(Q_U) + Q_U(1) I_1^{\mathrm{LM}}(\hat{Q}_1) + Q_U(2) I_2^{\mathrm{LM}}(\hat{Q}_2) \\ = \log\left(e^{I_1^{\mathrm{LM}}(\hat{Q}_1)} + e^{I_2^{\mathrm{LM}}(\hat{Q}_2)}\right) \end{aligned} \tag{93}$$

where I_ν^{LM} is the LM rate for subchannel ν, and the equality follows by optimizing Q_U in the same way as [30, Sec. 16], yielding $Q_U(1) = \frac{e^{I_1^{\mathrm{LM}}(\hat{Q}_1)}}{e^{I_1^{\mathrm{LM}}(\hat{Q}_1)} + e^{I_2^{\mathrm{LM}}(\hat{Q}_2)}}$. Using similar arguments to [6], it can be shown that the LM rate with an optimized input distribution can be strictly less than (93) even for simple examples (e.g. binary symmetric subchannels).

2) Zero Undetected Error Capacity: It was shown by Csiszár and Narayan [3] that two special cases of the mismatched capacity are the zero-undetected erasures capacity [31] and the zero-error capacity [32]. Here we consider the zero-undetected erasures capacity, defined to be the highest achievable rate in the case that the decoder is required to know with certainty whether or not an error has occurred. For any DMC, the zero-undetected erasures capacity is equal to the mismatched capacity under the decoding metric $q(x,y) = \mathbb{1}\{W(y|x) > 0\}$ [3].

We consider an example from [33], where $\mathcal{X} = \mathcal{Y} = \{0,1,2\}$, and the channel is described by the entries of

$$\boldsymbol{W} = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.75 & 0.25 \\ 0.25 & 0 & 0.75 \end{bmatrix} \tag{94}$$

where x indexes the rows and y indexes the columns.

Using an exhaustive search to three decimal places, we found the optimized LM rate to be $R^*_{\mathrm{LM}} = 0.599$ bits/use, using the input distribution $Q = (0.449, 0.551, 0)$. It was stated in [33] that the rate obtained by considering the second-order product of the channel and metric (see [3]) is equal to $R^*_{\mathrm{LM2}} = 0.616$ bits/use. Using local optimization techniques, we verified that this rate is achieved with $Q = (0, 0.250, 0, 0.319, 0, 0, 0, 0.181, 0.250)$, where the order of the inputs is $(0,0), (0,1), (0,2), (1,0), \dots, (2,2)$.

The global optimization of (52)–(53) over $\mathcal{U}$ and Q_{UX} appears to be difficult. Setting $|\mathcal{U}| = 2$ and applying local optimization techniques using a number of starting points, we obtained an achievable rate of $R^*_{\mathrm{sc}} = 0.695$ bits/use, with $Q_U = (0.645, 0.355)$, $Q_{X|U}(\cdot|1) = (0.3, 0.7, 0)$ and $Q_{X|U}(\cdot|2) = (0,0,1)$. Thus, superposition coding not only yields an improvement over the single-letter LM rate, but also over the two-letter version. Note that since the decoding metric is the erasures-only metric, applying the LM rate to the k-th order product channel achieves the mismatched capacity in the limit as $k \to \infty$ [3]; however, in this example, a significant gap remains for $k = 2$.

3) A Case where Refined Superposition Coding Outperforms Standard Superposition Coding: Here we consider the channel and decoding metric described by the entries of

$$\boldsymbol{W} = \begin{bmatrix} 0.99 & 0.01 & 0 & 0 \\ 0.01 & 0.99 & 0 & 0 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{bmatrix} \tag{95}$$

$$\boldsymbol{q} = \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0.05 & 0.15 & 1 & 0.05 \\ 0.15 & 0.05 & 0.5 & 1 \end{bmatrix}. \tag{96}$$

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

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

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

Table I
Achievable rates (bits) for the mismatched channel (95)–(96).

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

The achievable rates for this example are summarized in Table I, where $Q^{(\mathrm{LM})}_{UX}$ denotes the distribution in which U is deterministic and the X-marginal maximizes the LM rate. While the achievable rate of Theorem 7 coincides with that of Theorem 5 under $Q^{(2)}_{UX}$, the former is significantly higher under $Q^{(1)}_{UX}$. 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 refined superposition coding yields a strict improvement over standard superposition coding (and hence over Lapidoth's rate [6]) after optimizing $\mathcal{U}$ and Q_{UX}. However, improvements for a fixed set of random-coding parameters are still of interest due to the fact that global optimizations are prohibitively complex in general.

V. Conclusion

We have provided techniques for studying multiuser random-coding ensembles for channel coding problems with mismatched decoding. The key initial step in each case is the application of a refined bound on the probability of a multiply-indexed union (*cf.* Appendix A), from which one can apply constant-composition coding and the method of types to obtain primal expressions and prove ensemble tightness, or cost-constrained random coding to obtain dual expressions and continuous-alphabet generalizations. We have demonstrated our techniques on both the mismatched MAC and the single-user channel with refined superposition coding, with the latter providing a new achievable rate at least as good as all previous rates in the literature.

After the initial preparation of this work, the superposition coding rate from Theorems 5–6 was used to find an example for which the LM rate is strictly smaller than the mismatched capacity for a binary-input DMC [34], thus providing a counter-example to the converse reported in [35]. Another work building on this paper is [36], which considers the matched relay channel, and shows that the utility of our refined union bounds is not restricted to mismatched decoders.

Appendix A
Upper and Lower Bounds on the Probability of a Multiply-Indexed Union

Bounds on the random-coding error probability in channel coding problems are often obtained using the truncated union bound, which states that for any set of events $\{A_i\}_{i=1}^N$,

$$\mathbb{P}\Big[\bigcup_i A_i\Big] \le \min\Big\{1, \sum_i \mathbb{P}[A_i]\Big\}. \tag{A.1}$$

In this paper, we are also interested in lower bounds on the probability of a union, which are used to prove ensemble tightness results. In particular, we make use of de Caen's lower bound [37], which states that

$$\mathbb{P}\Big[\bigcup_i A_i\Big] \ge \sum_i \frac{\mathbb{P}[A_i]^2}{\sum_{i'} \mathbb{P}[A_i \cap A_{i'}]}. \tag{A.2}$$

In the case that the events are pairwise independent and identically distributed, (A.2) proves the tightness of (A.1) to within a factor of $\frac{1}{2}$; see the proof of [38, Thm. 1].

In this section, we provide a number of upper and lower bounds on the probability of a multiply-indexed union. In several cases of interest, the upper and lower bounds coincide to within a constant factor, and generalize the above-mentioned tightness result of [38] to certain settings where pairwise independence need not hold.

Lemma 1. *Let $\{Z_1(i)\}_{i=1}^{N_1}$ and $\{Z_2(j)\}_{j=1}^{N_2}$ be independent sequences of identically distributed random variables on the alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively, with $Z_1(i) \sim P_{Z_1}$ and $Z_2(j) \sim P_{Z_2}$. For any set $\mathcal{A} \subseteq \mathcal{Z}_1 \times \mathcal{Z}_2$, we have:*

1) A general upper bound is given by

$$\mathbb{P}\bigg[\bigcup_{i,j}\Big\{\big(Z_1(i), Z_2(j)\big) \in \mathcal{A}\Big\}\bigg] \le \min\bigg\{1, N_1\mathbb{E}\bigg[\min\Big\{1, N_2\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A} \,\big|\, Z_1\big]\Big\}\bigg], N_2\mathbb{E}\bigg[\min\Big\{1, N_1\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A} \,\big|\, Z_2\big]\Big\}\bigg]\bigg\} \tag{A.3}$$

where $(Z_1, Z_2) \sim P_{Z_1} \times P_{Z_2}$.

2) If $\{Z_1(i)\}_{i=1}^{N_1}$ and $\{Z_2(j)\}_{j=1}^{N_2}$ are pairwise independent, then we have the lower bound

$$\mathbb{P}\bigg[\bigcup_{i,j}\Big\{\big(Z_1(i), Z_2(j)\big) \in \mathcal{A}\Big\}\bigg] \ge \frac{1}{4}\min\bigg\{1, N_1\frac{\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A}\big]^2}{\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A} \cap (Z_1,Z_2')\in\mathcal{A}\big]}, \tag{A.4}$$

$$N_2\frac{\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A}\big]^2}{\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A} \cap (Z_1',Z_2)\in\mathcal{A}\big]}, N_1N_2\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A}\big]\bigg\}, \tag{A.5}$$

where $(Z_1, Z_1', Z_2, Z_2') \sim P_{Z_1}(z_1)P_{Z_1}(z_1')P_{Z_2}(z_2)P_{Z_2}(z_2')$.

Proof: We first prove (A.3). Applying the union bound to the union over i gives

$$\mathbb{P}\bigg[\bigcup_{i,j}\Big\{\big(Z_1(i), Z_2(j)\big) \in \mathcal{A}\Big\}\bigg] \le N_1\mathbb{P}\bigg[\bigcup_{j}\Big\{\big(Z_1, Z_2(j)\big) \in \mathcal{A}\Big\}\bigg] \tag{A.6}$$

$$= N_1\mathbb{E}\bigg[\mathbb{P}\bigg[\bigcup_{j}\Big\{\big(Z_1, Z_2(j)\big) \in \mathcal{A}\Big\}\,\bigg|\, Z_1\bigg]\bigg]. \tag{A.7}$$

Applying the truncated union bound to the union over j, we recover the second term in the outer minimization in (A.3). The third term is obtained similarly by applying the union bounds in the opposite order, and the first term is trivial.

To prove (A.5), we make use of de Caen's bound in (A.2). Noting by symmetry that each term in the outer summation is equal, and splitting the inner summation according to which of the (i,j) indices coincide with (i',j'), we obtain

$$\begin{aligned}\mathbb{P}\bigg[\bigcup_{i,j}\Big\{\big(Z_1(i), Z_2(j)\big) \in \mathcal{A}\Big\}\bigg] &\ge N_1N_2\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A}\big]^2 \\ &\quad\times\bigg((N_1-1)(N_2-1)\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A}\big]^2 \\ &\quad + (N_2-1)\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A} \cap (Z_1,Z_2')\in\mathcal{A}\big] \\ &\quad + (N_1-1)\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A} \cap (Z_1',Z_2)\in\mathcal{A}\big] \\ &\quad + \mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A}\big]\bigg)^{-1}. \end{aligned} \tag{A.8}$$

The lemma follows by upper bounding $N_\nu - 1$ by N_ν for $\nu = 1,2$, and upper bounding the four terms in the $(\cdot)^{-1}$ by four times the maximum of those terms. ■

The following lemma gives conditions under which a weakened version of (A.3) matches (A.5) to within a factor of four. Recall that ν^c denotes the item in $\{1,2\}$ differing from ν

Lemma 2. *Let $\{Z_1(i)\}_{i=1}^{N_1}$ and $\{Z_2(j)\}_{j=1}^{N_2}$ be independent sequences of identically distributed random variables on the alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively, with $Z_1(i) \sim P_{Z_1}$ and $Z_2(j) \sim P_{Z_2}$. Fix a set $\mathcal{A} \subseteq \mathcal{Z}_1 \times \mathcal{Z}_2$, and define*

$$\mathcal{A}_\nu \triangleq \Big\{z_\nu \in \mathcal{Z}_\nu \,:\, (z_1,z_2) \in \mathcal{A} \text{ for some } z_{\nu^c} \in \mathcal{Z}_{\nu^c}\Big\} \tag{A.9}$$

for $\nu = 1,2$.

1) A general upper bound is given by

$$\mathbb{P}\bigg[\bigcup_{i,j}\Big\{\big(Z_1(i), Z_2(j)\big) \in \mathcal{A}\Big\}\bigg] \le \min\Big\{1, N_1\mathbb{P}\big[Z_1\in\mathcal{A}_1\big], N_2\mathbb{P}\big[Z_2\in\mathcal{A}_2\big], N_1N_2\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A}\big]\Big\}, \tag{A.10}$$

where $(Z_1, Z_2) \sim P_{Z_1} \times P_{Z_2}$.

2) If (i) $\{Z_1(i)\}_{i=1}^{N_1}$ are pairwise independent, (ii) $\{Z_2(j)\}_{j=1}^{N_2}$ are pairwise independent, (iii) $\mathbb{P}\big[(z_1,Z_2)\in\mathcal{A}\big]$ is the same for all $z_1 \in \mathcal{A}_1$, and (iv) $\mathbb{P}\big[(Z_1,z_2)\in\mathcal{A}\big]$ is the same for all $z_2 \in \mathcal{A}_2$, then

$$\mathbb{P}\bigg[\bigcup_{i,j}\Big\{\big(Z_1(i), Z_2(j)\big) \in \mathcal{A}\Big\}\bigg] \ge \frac{1}{4}\min\Big\{1, N_1\mathbb{P}\big[Z_1\in\mathcal{A}_1\big], N_2\mathbb{P}\big[Z_2\in\mathcal{A}_2\big], N_1N_2\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A}\big]\Big\}. \tag{A.11}$$

Proof: We obtain (A.10) by weakening (A.3) in multiple ways. The second term in (A.10) follows since the inner probability in the second term of (A.3) is zero whenever $\mathbb{P}[Z_1 \notin \mathcal{A}]$, and since $\min\{1,\zeta\} \le 1$. The third term in (A.10) is obtained similarly, and the fourth term follows from the fact that $\min\{1,\zeta\} \le \zeta$.

The lower bound in (A.11) follows from (A.5), and since the additional assumptions in the second part of the lemma statement imply

$$\begin{aligned}
&\frac{\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A}\big]^2}{\mathbb{P}\big[(Z_1,Z_2)\in\mathcal{A}\cap(Z_1,Z_2')\in\mathcal{A}\big]}\\
&\quad=\frac{\mathbb{P}\big[Z_1\in\mathcal{A}_1\big]^2\mathbb{P}\big[(z_1,Z_2)\in\mathcal{A}\big]^2}{\mathbb{P}\big[Z_1\in\mathcal{A}_1\big]\mathbb{P}\big[(z_1,Z_2)\in\mathcal{A}\big]^2}, &\text{(A.12)}\\
&\quad=\mathbb{P}\big[Z_1\in\mathcal{A}_1\big] &\text{(A.13)}
\end{aligned}$$

where z_1 is an arbitrary element of $\mathcal{A}_1$. The third term in the minimization in (A.5) can be handled similarly. ■

A generalization of Lemma 2 to the probability of a union indexed by K values can be found in [13, Appendix D].

Appendix B
Equivalent Forms of Convex Optimization Problems

The achievable rates and error exponents derived in this paper are presented in both primal and dual forms, analogously to the LM rate in (5)–(6). The corresponding proofs of equivalence are more involved than that of the LM rate (see [4]). Here we provide two lemmas that are useful in proving the equivalences. The following lemma generalizes the result that (5) and (6) are equivalent, and is proved using Lagrange duality [39, Ch. 5].

Lemma 3. *Fix the finite alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$, the non-negative functions $f(z_1,z_2)$ and $g(z_1,z_2)$, the distributions $P_{Z_1}\in\mathcal{P}(\mathcal{Z}_1)$ and $P_{Z_2}\in\mathcal{P}(\mathcal{Z}_2)$, and a constant β. Then*

$$\min_{\substack{\widetilde{P}_{Z_1Z_2}\,:\,\widetilde{P}_{Z_1}=P_{Z_1},\widetilde{P}_{Z_2}=P_{Z_2},\\ \mathbb{E}_{\widetilde{P}}[\log f(Z_1,Z_2)]\ge\beta}} I_{\widetilde{P}}(Z_1;Z_2)-\mathbb{E}_{\widetilde{P}}[\log g(Z_1,Z_2)] \tag{B.1}$$

is equal to

$$\begin{aligned}
&\sup_{\lambda\ge0,\mu_1(\cdot)}\sum_{z_1}P_{Z_1}(z_1)\mu_1(z_1)+\lambda\beta\\
&-\sum_{z_2}P_{Z_2}(z_2)\log\sum_{\overline{z}_1}P_{Z_1}(\overline{z}_1)f(\overline{z}_1,z_2)^{\lambda}g(\overline{z}_1,z_2)e^{\mu_1(\overline{z}_1)},
\end{aligned}\tag{B.2}$$

where the supremum over $\mu_1(\cdot)$ is taken over all real-valued functions on $\mathcal{Z}_1$.

Proof: The Lagrangian [39, Sec. 5.1.1] of the optimization problem in (B.1) is given by

$$\begin{aligned}
L=\sum_{z_1,z_2}\widetilde{P}_{Z_1Z_2}(z_1,z_2)\bigg(\log\frac{\widetilde{P}(z_1,z_2)}{P_{Z_1}(z_1)P_{Z_2}(z_2)}-\log g(z_1,z_2)\\
-\lambda\log f(z_1,z_2)\bigg)+\sum_{z_1}\mu_1(z_1)\big(P_{Z_1}(z_1)-\widetilde{P}_{Z_1}(z_1)\big)\\
+\sum_{z_2}\mu_2(z_2)\big(P_{Z_2}(z_2)-\widetilde{P}_{Z_2}(z_2)\big)+\lambda\beta,
\end{aligned}\tag{B.3}$$

where $\lambda\ge0$, $\mu_1(\cdot)$ and $\mu_2(\cdot)$ are Lagrange multipliers. Since the objective in (B.1) is convex and the constraints are affine, the optimal value is equal to L for some choice of $\widetilde{P}_{Z_1Z_2}$ and the Lagrange multipliers satisfying the Karush-Kuhn-Tucker (KKT) conditions [39, Sec. 5.5.3].

We proceed to simplify (B.3) using the KKT conditions. Setting $\frac{\partial L}{\partial\widetilde{P}(z_1,z_2)}=0$ yields

$$\begin{aligned}
1+\log\frac{\widetilde{P}_{Z_1Z_2}(z_1,z_2)}{P_{Z_1}(z_1)P_{Z_2}(z_2)f(z_1,z_2)^{\lambda}g(z_1,z_2)}\\
-\mu_1(z_1)-\mu_2(z_2)=0.
\end{aligned}\tag{B.4}$$

Solving for $\widetilde{P}_{Z_1Z_2}(z_1,z_2)$ applying the constraint $\widetilde{P}_{Z_2}=P_{Z_2}$, and then solving for $\mu_2(z_2)$, we obtain

$$\mu_2(z_2)=1-\log\sum_{\overline{z}_1}P_{Z_1}(\overline{z}_1)f(\overline{z}_1,z_2)^{\lambda}g(\overline{z}_1,z_2)e^{\mu_1(\overline{z}_1)}. \tag{B.5}$$

Substituting (B.4) into (B.3) yields

$$L=-1+\sum_{z_1}\mu_1(z_1)P_{Z_1}(z_1)+\sum_{z_2}\mu_2(z_2)P_{Z_2}(z_2)+\lambda\beta, \tag{B.6}$$

and applying (B.5) yields (B.2) with the supremum omitted. It follows that (B.2) is an upper bound to (B.1).

To obtain a matching lower bound, we make use of the log-sum inequality [40, Thm. 2.7.1] similarly to [4, Appendix A]. For any $\widetilde{P}_{Z_1Z_2}$ satisfying the constraints in (B.1), we can lower bound the objective in (B.1) as follows:

$$\begin{aligned}
&\sum_{z_1,z_2}\widetilde{P}_{Z_1Z_2}(z_1,z_2)\log\frac{\widetilde{P}(z_1,z_2)}{P_{Z_1}(z_1)P_{Z_2}(z_2)g(z_1,z_2)} &\text{(B.7)}\\
&\ge\sum_{z_1,z_2}\widetilde{P}_{Z_1Z_2}(z_1,z_2)\\
&\qquad\times\log\frac{\widetilde{P}(z_1,z_2)}{P_{Z_1}(z_1)P_{Z_2}(z_2)f(z_1,z_2)^{\lambda}g(z_1,z_2)}+\lambda\beta &\text{(B.8)}\\
&=\sum_{z_1,z_2}\widetilde{P}_{Z_1Z_2}(z_1,z_2)\\
&\qquad\times\log\frac{\widetilde{P}(z_1,z_2)}{P_{Z_1}(z_1)P_{Z_2}(z_2)f(z_1,z_2)^{\lambda}g(z_1,z_2)e^{\mu_1(\overline{z}_1)}}\\
&\qquad+\sum_{z_1}P_{Z_1}(z_1)\mu_1(z_1)+\lambda\beta, &\text{(B.9)}
\end{aligned}$$

where (B.8) holds for any $\lambda\ge0$ due to the constraint $\mathbb{E}_{\widetilde{P}}[\log f(Z_1,Z_2)]\ge\beta$, and (B.9) holds for any $\mu_1(\cdot)$ by an expansion of the logarithm. Applying the log-sum inequality, we can lower bound (B.9) by the objective in (B.2). Since $\lambda\ge0$ and $\mu_1(\cdot)$ are arbitrary, the proof is complete. ■

When using Lemma 3, we will typically be interested the case that either $g(\cdot,\cdot)=1$, or $f(\cdot,\cdot)=1$ and $\beta=0$.

The following lemma will allow certain convex optimization problems to be expressed in a form where, after some manipulations, Lemma 3 can be applied.

Lemma 4. *Fix a positive integer d and let $\mathcal{D}$ be a convex subset of $\mathbb{R}^d$. Let $f(\boldsymbol{z})$, $g(\boldsymbol{z})$, $g_1(\boldsymbol{z})$ and $g_2(\boldsymbol{z})$ be convex functions mapping $\mathbb{R}^d$ to $\mathbb{R}$ such that*

$$g_1(\boldsymbol{z})+g_2(\boldsymbol{z})\le g(\boldsymbol{z}) \tag{B.10}$$

for all $\boldsymbol{z}\in\mathcal{D}$. Then

$$\min_{\boldsymbol{z}\in\mathcal{D}}f(\boldsymbol{z})+\Big[\max\big\{g_1(\boldsymbol{z}),g_2(\boldsymbol{z}),g(\boldsymbol{z})\big\}\Big]^+ \tag{B.11}$$

is equal to

$$\max\left\{\min_{\boldsymbol{z}\in\mathcal{D}} f(\boldsymbol{z}) + \Big[\max\big\{g_1(\boldsymbol{z}), g(\boldsymbol{z})\big\}\Big]^+, \min_{\boldsymbol{z}\in\mathcal{D}} f(\boldsymbol{z}) + \Big[\max\big\{g_2(\boldsymbol{z}), g(\boldsymbol{z})\big\}\Big]^+\right\}. \quad \text{(B.12)}$$

Proof: We define the following functions ($\nu = 1, 2$):

$$\Phi_0(\boldsymbol{z}) \triangleq f(\boldsymbol{z}) + \big[g(\boldsymbol{z})\big]^+ \quad \text{(B.13)}$$

$$\Phi_\nu(\boldsymbol{z}) \triangleq f(\boldsymbol{z}) + \big[\max\big\{g_\nu(\boldsymbol{z}), g(\boldsymbol{z})\big\}\big]^+. \quad \text{(B.14)}$$

Since $f(\cdot)$, $g(\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$ are convex by assumption, it follows from the composition rules in [39, Sec. 3.2.4] that $\Phi_0(\cdot)$, $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$ are also convex.

We wish to show that

$$\min_{\boldsymbol{z}\in\mathcal{D}} \max\big\{\Phi_1(\boldsymbol{z}), \Phi_2(\boldsymbol{z})\big\} = \max\left\{\min_{\boldsymbol{z}\in\mathcal{D}} \Phi_1(\boldsymbol{z}), \min_{\boldsymbol{z}\in\mathcal{D}} \Phi_2(\boldsymbol{z})\right\}. \quad \text{(B.15)}$$

We define the following regions for $\nu = 1, 2$:

$$\mathcal{R}_\nu = \big\{\boldsymbol{z} : \Phi_\nu(\boldsymbol{z}) > \Phi_0(\boldsymbol{z})\big\}. \quad \text{(B.16)}$$

The key observation is that $\mathcal{R}_1$ and $\mathcal{R}_2$ are disjoint. To see this, we observe from (B.13)–(B.14) that any $\boldsymbol{z} \in \mathcal{R}_1 \cap \mathcal{R}_2$ satisfies $g_1(\boldsymbol{z}) > g(\boldsymbol{z})$ and $g_2(\boldsymbol{z}) > g(\boldsymbol{z})$. Combined with (B.10), these imply $g_1(\boldsymbol{z}) < 0$ and $g_2(\boldsymbol{z}) < 0$, and it follows from (B.13)–(B.14) that $\Phi_0(\boldsymbol{z}) = \Phi_1(\boldsymbol{z}) = \Phi_2(\boldsymbol{z})$, in contradiction with the assumption that $\boldsymbol{z} \in \mathcal{R}_1 \cap \mathcal{R}_2$. Thus, $\mathcal{R}_1 \cap \mathcal{R}_2$ is empty, which implies that $g_1(\boldsymbol{z})$ and $g_2(\boldsymbol{z})$ cannot simultaneously be the unique maximizers in (B.14) for both $\nu = 1$ and $\nu = 2$. Combining this with (B.13), we obtain

$$\Phi_0(\boldsymbol{z}) = \min\big\{\Phi_1(\boldsymbol{z}), \Phi_2(\boldsymbol{z})\big\}. \quad \text{(B.17)}$$

To prove (B.15), we use a proof by contradiction. Let the left-hand side and right-hand side be denoted by f^* and $\tilde{f}^*$ respectively. The inequality $f^* \ge \tilde{f}^*$ holds by definition, so we assume that $f^* > \tilde{f}^*$. Let $\boldsymbol{z}_1^*$ and $\boldsymbol{z}_2^*$ minimize Φ_1 and Φ_2 respectively on the right-hand side of (B.15), so that

$$\tilde{f}^* = \max\big\{\Phi_1(\boldsymbol{z}_1^*), \Phi_2(\boldsymbol{z}_2^*)\big\}. \quad \text{(B.18)}$$

The assumption $f^* > \tilde{f}^*$ implies that

$$\Phi_2(\boldsymbol{z}_1^*) > \Phi_1(\boldsymbol{z}_1^*) \quad \text{(B.19)}$$

$$\Phi_1(\boldsymbol{z}_2^*) > \Phi_2(\boldsymbol{z}_2^*). \quad \text{(B.20)}$$

Next, we define

$$\hat{\Phi}_\nu(\lambda) \triangleq \Phi_\nu\big(\lambda\boldsymbol{z}_1^* + (1-\lambda)\boldsymbol{z}_2^*\big) \quad \text{(B.21)}$$

for $\lambda \in [0,1]$ and $\nu = 0, 1, 2$. Since any convex function is also convex when restricted to a straight line [39, Section 3.1.1], it follows that $\hat{\Phi}_0$, $\hat{\Phi}_1$ and $\hat{\Phi}_2$ are convex in λ. From (B.19)–(B.20), we have

$$\hat{\Phi}_2(1) > \hat{\Phi}_1(1) \quad \text{(B.22)}$$

$$\hat{\Phi}_1(0) > \hat{\Phi}_2(0). \quad \text{(B.23)}$$

Since $\hat{\Phi}_1$ and $\hat{\Phi}_2$ are convex, they are also continuous (at least in the region that they are finite), and it follows that the two must intersect somewhere in $(0,1)$, say at λ^*. Therefore,

$$\hat{\Phi}_0(\lambda^*) = \min\big\{\hat{\Phi}_1(\lambda^*), \hat{\Phi}_2(\lambda^*)\big\} \quad \text{(B.24)}$$

$$= \max\big\{\hat{\Phi}_1(\lambda^*), \hat{\Phi}_2(\lambda^*)\big\} \quad \text{(B.25)}$$

$$\ge \min_{\boldsymbol{z}\in\mathcal{D}} \max\big\{\Phi_1(\boldsymbol{z}), \Phi_2(\boldsymbol{z})\big\} \quad \text{(B.26)}$$

$$= f^*, \quad \text{(B.27)}$$

where (B.24) follows from (B.17). Finally, we have the following contradiction: (i) Combining (B.27) with the assumption that $f^* > \tilde{f}^*$, we have

$$\hat{\Phi}_0(\lambda^*) > \tilde{f}^* = \max\{\hat{\Phi}_1(1), \hat{\Phi}_2(0)\}, \quad \text{(B.28)}$$

where the equality follows from (B.18); (ii) From (B.17), we have $\hat{\Phi}_0(\lambda) = \min\{\hat{\Phi}_1(\lambda), \hat{\Phi}_2(\lambda)\}$, and it follows from (B.22)–(B.23) that $\hat{\Phi}_0(1) = \hat{\Phi}_1(1)$ and $\hat{\Phi}_0(0) = \hat{\Phi}_2(0)$. Using the convexity of $\hat{\Phi}_0$ and Jensen's inequality, we have

$$\hat{\Phi}_0(\lambda^*) \le \lambda^*\hat{\Phi}_1(1) + (1-\lambda^*)\hat{\Phi}_2(0) \quad \text{(B.29)}$$

$$\le \max\{\hat{\Phi}_1(1), \hat{\Phi}_2(0)\}. \quad \text{(B.30)}$$

■

Appendix C
Multiple-Access Channel Proofs

A. Preliminary Lemma for Proving Theorem 3

The following lemma expresses (29) in a form that is more amenable to Lagrange duality techniques.

Lemma 5. *The achievable rate condition in* (29) *holds if the following holds for at least one of* $\nu = 1, 2$*:*

$$R_1 + R_2 \le \min_{\substack{\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_{12}(Q_1\times Q_2\times W) \\ I_{\widetilde{P}}(X_\nu;Y) \le R_\nu}} D(\widetilde{P}_{X_1X_2Y} \| Q_1 \times Q_2 \times P_Y) \quad \text{(C.1)}$$

Proof: We first write the condition in (29) as

$$0 \le \min_{\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_{12}(Q_1\times Q_2\times W)} \max\big\{D(\widetilde{P}_{X_1X_2Y} \| Q_1\times Q_2\times P_Y) - (R_1+R_2), I_{\widetilde{P}}(X_1;Y) - R_1,\, I_{\widetilde{P}}(X_2;Y) - R_2\big\}, \quad \text{(C.2)}$$

where the equivalence is seen by noting that this condition is always satisfied when the minimizer satisfies $I_{\widetilde{P}}(X_1;Y) > R_1$ or $I_{\widetilde{P}}(X_2;Y) > R_2$. Next, we claim that this condition is equivalent to the following holding for at least one of $\nu = 1, 2$:

$$0 \le \min_{\widetilde{P}_{X_1X_2Y} \in \mathcal{T}_{12}(Q_1\times Q_2\times W)} \max\big\{ D(\widetilde{P}_{X_1X_2Y} \| Q_1\times Q_2\times P_Y) - (R_1+R_2), I_{\widetilde{P}}(X_\nu;Y) - R_\nu\big\}. \quad \text{(C.3)}$$

This is seen by applying Lemma 4 with the following identifications ($\nu = 1, 2$):

$$f(\boldsymbol{z}) = 0 \quad \text{(C.4)}$$

$$g(\boldsymbol{z}) = D\big(\widetilde{P}_{X_1X_2Y} \| Q_1 \times Q_2 \times P_Y\big) - R_1 - R_2 \quad \text{(C.5)}$$

$$g_\nu(\boldsymbol{z}) = I_{\widetilde{P}}(X_\nu;Y) - R_\nu. \quad \text{(C.6)}$$

From the last two lines and the identity

$$D\big(\widetilde{P}_{X_1X_2Y}\|Q_1\times Q_2\times P_Y\big)=I_{\widetilde{P}}(X_1;Y)+I_{\widetilde{P}}(X_2;Y)+I_{\widetilde{P}}(X_1;X_2|Y),\quad\text{(C.7)}$$

which holds under the constraints present in the definition of $\mathcal{T}_{12}$ in (14), we see that the condition in (B.10) is satisfied.

Finally, the lemma follows from (C.3) by reversing the step used to obtain (C.2). ∎

B. Proof of First Part of Theorem 3

Each expression in the theorem statement is derived similarly, so we focus on (35). We claim that (C.1) holds if and only if

$$R_1\le\max_{\rho_2\in[0,1]}\min_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}(P_{X_1X_2Y})}I_{\widetilde{P}}(X_1;Y)+\rho_2I_{\widetilde{P}}(X_2;X_1,Y)-\rho_2R_2,\quad\text{(C.8)}$$

where here and in the remainder of the proof we write $P_{X_1X_2Y}\triangleq Q_1\times Q_2\times W$. To see this, we first note that by the identity

$$D\big(P_{X_1X_2Y}\|Q_1\times Q_2\times P_Y\big)=I_P(X_1;Y)+I_P(X_2;X_1,Y),\quad\text{(C.9)}$$

(C.3) (with $\nu=1$) is equivalent to

$$R_1\le\min_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}(Q_1\times Q_2\times W)}I_{\widetilde{P}}(X_1;Y)+\big[I_{\widetilde{P}}(X_2;X_1,Y)-R_2\big]^+.\quad\text{(C.10)}$$

Next, we apply the identity $[\alpha]^+=\max_{0\le\rho_1\le1}\rho_1\alpha$. The resulting objective is linear in ρ_1 and jointly convex in $(P_{X_1X_2Y},\widetilde{P}_{X_1X_2Y})$, so we can apply Fan's minimax theorem [41] to interchange the maximization and minimizations, thus yielding (C.8).

We define the sets

$$\mathcal{T}_{12}'(P_{X_1X_2Y},\hat{P}_{X_1Y})\triangleq\Big\{\widetilde{P}_{X_1X_2Y}\in\mathcal{P}(\mathcal{X}_1\times\mathcal{X}_2\times\mathcal{Y})\,:\,\widetilde{P}_{X_2}=P_{X_2},\widetilde{P}_{X_1Y}=\hat{P}_{X_1Y},\ \mathbb{E}_{\widetilde{P}}[\log q(X_1,X_2,Y)]\ge\mathbb{E}_P[\log q(X_1,X_2,Y)]\Big\}\quad\text{(C.11)}$$

$$\mathcal{T}_{12}''(P_{X_1X_2Y})\triangleq\Big\{\hat{P}_{X_1Y}\in\mathcal{P}(\mathcal{X}_1\times\mathcal{Y})\,:\,\hat{P}_{X_1}=P_{X_1},\hat{P}_Y=P_Y\Big\}.\quad\text{(C.12)}$$

It follows that $\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}(P_{X_1X_2Y})$ (see (14)) if and only if $\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}'(P_{X_1X_2Y},\hat{P}_{X_1Y})$ for some $\hat{P}_{X_1Y}\in\mathcal{T}_{12}''(P_{X_1X_2Y})$. We can therefore replace the minimization over $\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}(P_{X_1X_2Y})$ in (C.8) with minimizations over $\hat{P}_{X_1Y}\in\mathcal{T}_{12}''(P_{X_1X_2Y})$ and $\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}'(P_{X_1X_2Y},\hat{P}_{X_1Y})$.

We prove the theorem by performing the minimization in several steps, and performing multiple applications of Lemma 3. Each such application will yield an overall optimization of the form $\sup\min\sup\{\cdot\}$, and we will implicitly use Fan's minimax theorem [41] to obtain an equivalent expression of the form $\sup\sup\min\{\cdot\}$. Thus, we will leave the optimization of the dual variables (i.e. the suprema) until the final step.

Step 1: We first consider the minimization of the term $I_{\widetilde{P}}(X_1;X_2,Y)$ over $\widetilde{P}_{X_1X_2Y}$ when $P_{X_1X_2Y}\in\mathcal{S}(\boldsymbol{Q})$ and $\hat{P}_{X_1Y}\in\mathcal{T}_{12}''(P_{X_1X_2Y})$ are fixed, and thus all of the terms in the objective in (C.8) other than $I_{\widetilde{P}}(X_1;X_2,Y)$ are fixed. The minimization is given by

$$\mathsf{F}_1\triangleq\min_{\widetilde{P}_{X_1X_2Y}\in\mathcal{T}_{12}'(P_{X_1X_2Y},\hat{P}_{X_1Y})}I_{\widetilde{P}}(X_1;X_2,Y).\quad\text{(C.13)}$$

Applying Lemma 3 with $P_{Z_1}=P_{X_2}$, $P_{Z_2}=\hat{P}_{X_1Y}$ and $\mu_1(\cdot)=a_2(\cdot)$, we obtain the dual expression

$$\mathsf{F}_1=-\sum_{x_1,y}\hat{P}_{X_1Y}(x_1,y)\log\sum_{\overline{x}_2}P_{X_2}(\overline{x}_2)q(x_1,\overline{x}_2,y)^se^{a_2(\overline{x}_2)}+s\sum_{x_1,x_2,y}P_{X_1X_2Y}(x_1,x_2,y)\log q(x_1,x_2,y)+\sum_{x_2}P_{X_2}(x_2)a_2(x_2).\quad\text{(C.14)}$$

Step 2: After Step 1, the overall objective (see (C.8)) is given by

$$I_{\hat{P}}(X_1;Y)+\rho_2\mathsf{F}_1-\rho_2R_2,\quad\text{(C.15)}$$

where we have replaced $I_{\widetilde{P}}(X_1;Y)$ by $I_{\hat{P}}(X_1;Y)$ due to the constraint $\widetilde{P}_{X_1Y}=\hat{P}_{X_1Y}$ in (C.11). Since the only terms involving $\hat{P}_{X_1Y}$ are $I_{\hat{P}}(X_1;Y)$ and the first term in (C.14), we consider the minimization

$$\mathsf{F}_2\triangleq\min_{\hat{P}_{X_1Y}\in\mathcal{T}_{12}''(P_{X_1X_2Y})}I_{\hat{P}}(X_1;Y)-\rho_2\sum_{x_1,y}\hat{P}_{X_1Y}(x_1,y)\times\log\sum_{\overline{x}_2}P_{X_2}(\overline{x}_2)q(x_1,\overline{x}_2,y)^se^{a_2(\overline{x}_2)}.\quad\text{(C.16)}$$

Applying Lemma 3 with $P_{Z_1}=P_{X_1}$, $P_{Z_2}=P_Y$ and $\mu_1(\cdot)=a_1(\cdot)$, we obtain

$$\mathsf{F}_2=\sum_{x_1}P_{X_1}(x_1)a_1(x_1)-\sum_yP_Y(y)\log\sum_{\overline{x}_1}P_{X_1}(\overline{x}_1)\times\bigg(\sum_{\overline{x}_2}P_{X_2}(\overline{x}_2)q(\overline{x}_1,\overline{x}_2,y)^se^{a_2(\overline{x}_2)}\bigg)^{\rho_2}e^{a_1(\overline{x}_1)}.\quad\text{(C.17)}$$

Step 3: From (C.14), (C.15) and (C.17), the overall objective is now given by

$$\mathsf{F}_3\triangleq\mathsf{F}_2-\rho_2R_2+\rho_2\sum_{x_1,x_2,y}P_{X_1X_2Y}(x_1,x_2,y)\log q(x_1,x_2,y)^se^{a_2(x_2)}.\quad\text{(C.18)}$$

Substituting (C.17) and performing some rearrangements, we obtain the objective in (35), and conclude the proof by taking the supremum over ρ_2, s, $a_1(\cdot)$ and $a_2(\cdot)$.

C. Proof of Theorem 4

We begin with the following proposition, which shows that the exponents $(E_{r,1}^{\mathrm{cc}},E_{r,2}^{\mathrm{cc}},E_{r,12}^{\mathrm{cc}'})$ (see (17) and (27)) under ML decoding coincide with those by Liu and Hughes in the absence of time-sharing [12].

Proposition 3. *Under ML decoding (i.e. $q(x_1,x_2,y) = W(y|x_1,x_2)$), $E^{\rm cc}_{r,\nu}$ and $E^{\rm cc'}_{r,12}$ can be expressed as*

$$E^{\rm cc}_{r,\nu}(\boldsymbol{Q},R_\nu) = \min_{P_{X_1X_2Y}\in\mathcal{S}(\boldsymbol{Q})} D(P_{X_1X_2Y}\|Q_1\times Q_2\times W) + \big[I_P(X_\nu;X_{\nu^c},Y)-R_\nu\big]^+ \quad \text{(C.19)}$$

$$E^{\rm cc'}_{r,12}(\boldsymbol{Q},R_1,R_2) = \min_{P_{X_1X_2Y}\in\mathcal{S}(\boldsymbol{Q})} D(P_{X_1X_2Y}\|Q_1\times Q_2\times W) + \big[D(P_{X_1X_2Y}\|Q_1\times Q_2\times P_Y)-(R_1+R_2)\big]^+. \quad \text{(C.20)}$$

Proof: The proof is similar to that of [15, Lemma 9], so we provide only an outline, and we focus on the type-12 exponent. Consider any pair $(P_{X_1X_2Y},\widetilde{P}_{X_1X_2Y})$ satisfying the constraints of (27). If $D(\widetilde{P}_{X_1X_2Y}\|Q_1\times Q_2\times P_Y)\ge D(P_{X_1X_2Y}\|Q_1\times Q_2\times P_Y)$, we can lower bound the objective of (27) by that of (C.20). In the remaining case, we may use the constraint $\mathbb{E}_{\widetilde{P}}[\log W]\ge\mathbb{E}_P[\log W]$ to lower bound the objective in (27) by that of (C.20) with $\widetilde{P}_{X_1X_2Y}$ in place of $P_{X_1X_2Y}$. This proves that (C.20) lower bounds (27), and the matching upper bound follows immediately from the fact that $\widetilde{P}_{X_1X_2Y}=P_{X_1X_2Y}$ satisfies the constraints of the minimization in (27). ■

We know that $E^{\rm cc}_{r,12}\ge E^{\rm cc'}_{r,12}$ always holds, and hence the left-hand side of (45) is greater than or equal to the right-hand side. It remains to prove the reverse inequality. From the definition of $\mathcal{T}_{12}(P_{X_1X_2Y})$, $\widetilde{P}_{X_1X_2Y}=P_{X_1X_2Y}$ always satisfies the constraints of (18), and hence

$$E^{\rm cc}_{r,12}(\boldsymbol{Q},R_1,R_2)\le F_{12}(\boldsymbol{Q},R_1,R_2), \quad \text{(C.21)}$$

where

$$F_{12}(\boldsymbol{Q},R_1,R_2)\triangleq \min_{P_{X_1X_2Y}\in\mathcal{S}(\boldsymbol{Q})} D(P_{X_1X_2Y}\|Q_1\times Q_2\times W) + \Big[\max\Big\{I_P(X_1;Y)-R_1, I_P(X_2;Y)-R_2, D\big(P_{X_1X_2Y}\|Q_1\times Q_2\times P_Y\big)-R_1-R_2\Big\}\Big]^+. \quad \text{(C.22)}$$

We will prove (45) by showing that

$$\min\big\{E^{\rm cc}_{r,1}(\boldsymbol{Q},R_1),E^{\rm cc}_{r,2}(\boldsymbol{Q},R_2),F_{12}(\boldsymbol{Q},R_1,R_2)\big\} \le \min\big\{E^{\rm cc}_{r,1}(\boldsymbol{Q},R_1),E^{\rm cc}_{r,2}(\boldsymbol{Q},R_2),E^{\rm cc'}_{r,12}(\boldsymbol{Q},R_1,R_2)\big\}. \quad \text{(C.23)}$$

It suffices to show that whenever F_{12} exceeds $E^{\rm cc'}_{r,12}$, F_{12} also greater than or equal to either $E^{\rm cc}_{r,1}$ or $E^{\rm cc}_{r,2}$. Comparing (C.20) and (C.22), the objective in (C.22) only exceeds that of (C.20) when the maximum in (C.22) is achieved by $I_P(X_1;Y)-R_1$ or $I_P(X_2;Y)-R_2$. We show that the former implies $F_{12}\ge E^{\rm cc}_{r,2}$; it can similarly be shown that the latter implies $F_{12}\ge E^{\rm cc}_{r,1}$. If $I_P(X_1;Y)-R_1$ achieves the maximum, we have

$$I_P(X_1;Y)-R_1\ge D\big(P_{X_1X_2Y}\|Q_1\times Q_2\times P_Y\big)-R_1-R_2. \quad \text{(C.24)}$$

Using the identity (C.9), we can write (C.24) as $I_P(X_2;X_1,Y)\le R_2$. For any $P_{X_1X_2Y}$ satisfying this property, the objective in (C.19) (with $\nu=2$) equals $D(P_{X_1X_2Y}\|Q_1\times Q_2\times W)$, and thus cannot exceed the objective in (C.22). It follows that $F_{12}\ge E^{\rm cc}_{r,2}$.

Appendix D
Refined Superposition Coding Proofs

A. A Preliminary Lemma

Similarly to Lemma 5 for the MAC, the following lemma gives an alternative expression for (82) that is more amenable to Lagrange duality techniques.

Lemma 6. *The condition in* (82) *holds if and only if the following holds for at least one of $u=1,2$:*

$$R_0\le \min_{\widetilde{P}_{UXY}\in\mathcal{T}_0(Q_{UX}\times W)} I_{\widetilde{P}}(U;Y)+\Big[\max\Big\{ Q_U(u)\big(I_{\widetilde{P}}(X;Y|U=u)-R_{1u}\big), I_{\widetilde{P}}(X;Y|U)-R_1\Big\}\Big]^+ \quad \text{(D.1)}$$

Proof: This is a special case of Lemma 4 in Appendix B with the following identifications ($u=1,2$):

$$f(\boldsymbol{z})=I_{\widetilde{P}}(U;Y) \quad \text{(D.2)}$$

$$g(\boldsymbol{z})=I_{\widetilde{P}}(X;Y|U)-R_1 \quad \text{(D.3)}$$

$$g_u(\boldsymbol{z})=Q_U(u)\big(I_{\widetilde{P}}(X;Y|U=u)-R_{1u}\big), \quad \text{(D.4)}$$

where we recall that $R_1=\sum_u Q_U(u)R_{1u}$. In this case, the condition in (B.10) holds with equality. ■

B. Proof of First Part of Theorem 8

We show the equivalence of (D.1) ($\nu=1$) and (87) ($u=1$); identical arguments apply for $\nu=u=2$. The primal expression is written in terms of a minimization over $\widetilde{P}_{UXY}$. It is convenient to "split" this distribution into three distributions: $\widetilde{P}_{UY}$, $\hat{P}_{XY}\triangleq\widetilde{P}_{XY|U}(\cdot,\cdot|1)$ and $\hat{\hat{P}}_{XY}\triangleq\widetilde{P}_{XY|U}(\cdot,\cdot|2)$. Using a similar argument to the start of Section C-B, we can write the right-hand side of (D.1) as

$$\sup_{\rho_1\in[0,1],\rho_2\in[0,1]}\ \min_{\widetilde{P}_{UY},\hat{P}_{XY},\hat{\hat{P}}_{XY}} I_{\widetilde{P}}(U;Y)+\rho_1Q_U(1)I_{\hat{P}}(X;Y) + \rho_1\rho_2Q_U(2)I_{\hat{\hat{P}}}(X;Y)-\rho_1R_{11}-\rho_1\rho_2R_{12}. \quad \text{(D.5)}$$

Defining $P_{UXY}\triangleq Q_{UX}\times W$, the minimization is subject to the constraints (i) $\widetilde{P}_U=Q_U$, (ii) $\hat{P}_X=Q_{X|U}(\cdot|1)$, (iii) $\hat{\hat{P}}_X=Q_{X|U}(\cdot|2)$, (iv) $\widetilde{P}_Y=P_Y$, (v) $\hat{P}_Y=\widetilde{P}_{Y|U}(\cdot|1)$, (vi) $\hat{\hat{P}}_Y=\widetilde{P}_{Y|U}(\cdot|2)$, (vii) $Q_U(1)\mathbb{E}_{\hat{P}}[\log q(X,Y)]+Q_U(2)\mathbb{E}_{\hat{\hat{P}}}[\log q(X,Y)]\ge\mathbb{E}_P[\log q(X,Y)]$.

Similarly to Section C-B, we apply the minimization in several steps, making repeated use of Lemma 3. We implicitly apply Fan's minimax theorem [41] after each step, so that the supremum over the dual variables can be left until the end. We provide less detail than the amount given in Section C-B, since the general steps are similar.

Step 1: For given joint distributions $\widetilde{P}_{UY}$ and $\hat{P}_{XY}$, the minimization $\min_{\hat{\hat{P}}_{XY}} I_{\hat{\hat{P}}}(X;Y)$ subject to the constraints (iii), (vi) and (vii) has a dual expression given by

$$\mathsf{F}_1\triangleq-\mathsf{F}_{1,1}+\mathsf{F}_{1,2}+\mathsf{F}_{1,3}-sQ_U(1)\mathsf{F}_{1,4}, \quad \text{(D.6)}$$

where

$$\mathsf{F}_{1,1} \triangleq \sum_y \widetilde{P}_{Y|U}(y|2) \log \sum_{\overline{x}} Q_{X|U}(\overline{x}|2) q(\overline{x},y)^{sQ_U(2)} e^{a_2(\overline{x})} \tag{D.7}$$

$$\mathsf{F}_{1,2} \triangleq \sum_{x_2} Q_{X|U}(x|2) a_2(x) \tag{D.8}$$

$$\mathsf{F}_{1,3} \triangleq s \sum_{x,y} P_{XY}(x,y) \log q(x,y) \tag{D.9}$$

$$\mathsf{F}_{1,4} \triangleq \sum_{x,y} \hat{P}_{XY}(x,y) \log q(x,y), \tag{D.10}$$

and where $s \geq 0$ and $a_2(\cdot)$ are dual variables.

Step 2: For a given joint distribution $\widetilde{P}_{UY}$, the minimization $\min_{\hat{P}_{XY}} I_{\hat{P}}(X;Y) - s\rho_2 Q_U(2)\mathsf{F}_{1,4}(\hat{P}_{XY})$ subject to (ii) and (v) has a dual expression given by

$$\mathsf{F}_2 \triangleq -\mathsf{F}_{2,1} + \mathsf{F}_{2,2}, \tag{D.11}$$

where

$$\mathsf{F}_{2,1} \triangleq \sum_y \widetilde{P}_{Y|U}(y|1) \log \sum_{\overline{x}} Q_{X|U}(\overline{x}|1) q(\overline{x},y)^{s\rho_2 Q_U(2)} e^{a_1(\overline{x})} \tag{D.12}$$

$$\mathsf{F}_{2,2} \triangleq \sum_x Q_{X|U}(x|1) a_1(x), \tag{D.13}$$

and where $a_1(\cdot)$ is a dual variable.

Step 3: Next, we consider the minimization $\min_{\widetilde{P}_{UY}} I_{\widetilde{P}}(U;Y) - \rho_1 Q_U(1)\mathsf{F}_{2,1} - \rho_1\rho_2 Q_U(2)\mathsf{F}_{1,1}$ subject to (i) and (iv). The objective can equivalently be expressed as

$$\mathsf{F}_3 \triangleq I_{\widetilde{P}}(U;Y) - \sum_u \rho_1(u) \sum_y \widetilde{P}_{UY}(u,y) \times \log \sum_{\overline{x}} Q_{X|U}(\overline{x}|u) q(\overline{x},y)^{s_1(u)} e^{a(u,\overline{x})} \tag{D.14}$$

using the definitions in (88) along with $a(u,x) \triangleq a_u(x)$. The dual expression is given by

$$\mathsf{F}_3 = \sum_u Q_U(u) b(u) - \sum_y P_Y(y) \log \sum_{\overline{u}} Q_U(\overline{u}) \times \bigg(\sum_{\overline{x}} Q_{X|U}(\overline{x}|\overline{u}) q(\overline{x},y)^{s_1(\overline{u})} e^{a(\overline{u},\overline{x})} \bigg)^{\rho_1(u)} e^{b(\overline{u})}, \tag{D.15}$$

where $b(\cdot)$ is a dual variable.

Step 4: The final objective is given by

$$\mathsf{F}_4 \triangleq \mathsf{F}_3 + \rho_1 Q_U(1)\mathsf{F}_{2,1} + \rho_1\rho_2 Q_U(2)\big(\mathsf{F}_{1,2} + \mathsf{F}_{1,3}\big) - \sum_{u=1,2} \rho_1(u) Q_U(u) R_{1u}. \tag{D.16}$$

After applying some algebraic manipulations, we obtain the dual expression

$$-\sum_{u=1,2} \rho_1(u) Q_U(u) R_{1u} + \sum_{u,x,y} P_{UXY}(u,x,y) \times \log \frac{\Big(q(x,y)^{s_1(u)} e^{a(u,x)}\Big)^{\rho_1(u)} e^{b(u)}}{\sum_{\overline{u}} Q_U(\overline{u}) \Big(\sum_{\overline{x}} Q_{X|U}(\overline{x}|\overline{u}) q(\overline{x},y)^{s_1(\overline{u})} e^{a(\overline{u},\overline{x})} \Big)^{\rho_1(\overline{u})} e^{b(\overline{u})}}. \tag{D.17}$$

To conclude the proof, we show that the variable $b(u)$ can be removed from the numerator and denominator in (D.17) without affecting the dual optimization. For $\rho_1 > 0$ and $\rho_2 > 0$, this follows by factoring $b(u)$ into $a(u,x)$. Using the identity $\mathbb{E}[e^{b(U)}] \geq e^{\mathbb{E}[b(U)]}$ (by Jensen's inequality), we find that the optimal value of the objective is zero when $\rho_1 = 0$, regardless of whether $b(u)$ is present. For the remaining case, namely $\rho_1 > 0$ and $\rho_2 = 0$, the objective depends on $a(u,x)$ only for $u = 1$. Moreover, since (D.17) depends on $b(\cdot)$ only through the difference $b(2) - b(1)$, we may set $b(2) = 0$ without loss of generality. The remaining parameter $b(1)$ can be factored into $a(1,x)$.

C. Proof of Second Part of Theorem 8

We focus on the derivation of (87) with $u = 1$, since the case $u = 2$ is handled similarly. The ideas used in the derivation are similar to those for the MAC (see the proof of Theorem 3), but the details are more involved.

Applying Lemma 1 to the union in (68), with $Z_1(i) = \boldsymbol{X}_1^{(1,i)}$ and $Z_2(j) = \boldsymbol{X}_2^{(1,j)}$, we obtain

$$\overline{p}_{e,0} \leq \mathbb{E}\Bigg[\min\Bigg\{1, (M_0 - 1)\mathbb{E}\Bigg[\min\Bigg\{1, M_{11}\mathbb{E}\Bigg[\min\Bigg\{1, M_{12}\mathbb{P}\bigg[\frac{q^n(\overline{\boldsymbol{X}},\boldsymbol{Y})}{q^n(\boldsymbol{X},\boldsymbol{Y})} \geq 1 \,\bigg|\, \overline{\boldsymbol{X}}_1 \bigg] \Bigg\} \,\Bigg|\, \overline{\boldsymbol{U}} \Bigg] \Bigg\} \,\Bigg|\, \boldsymbol{U},\boldsymbol{X},\boldsymbol{Y} \Bigg] \Bigg\} \Bigg]. \tag{D.18}$$

Using (61), Markov's inequality, and $\min\{1,\zeta\} \leq \zeta^\rho$ ($\rho \in [0,1]$), we obtain[2]

$$\overline{p}_{e,0} \leq (M_0 M_{11}^{\rho_1} M_{12}^{\rho_1\rho_2})^{\rho_0} \sum_{\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2} P_{\boldsymbol{U}}(\boldsymbol{u}) P_{\boldsymbol{X}_1}(\boldsymbol{x}_1) P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) \times \sum_{\boldsymbol{y}} W^n(\boldsymbol{y}|\Xi(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2)) \Bigg(\sum_{\overline{\boldsymbol{u}}} P_{\boldsymbol{U}}(\overline{\boldsymbol{u}}) \times \Bigg(\sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) \bigg(\frac{q^{n_1}(\overline{\boldsymbol{x}}_1, \boldsymbol{y}_1(\overline{\boldsymbol{u}}))}{q^{n_1}(\boldsymbol{x}_1, \boldsymbol{y}_1(\boldsymbol{u}))} \bigg)^{\rho_2 s} \Bigg)^{\rho_1} \times \Bigg(\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2) \bigg(\frac{q^{n_2}(\overline{\boldsymbol{x}}_2, \boldsymbol{y}_2(\overline{\boldsymbol{u}}))}{q^{n_2}(\boldsymbol{x}_2, \boldsymbol{y}_2(\boldsymbol{u}))} \bigg)^{s} \Bigg)^{\rho_1\rho_2} \Bigg)^{\rho_0}, \tag{D.19}$$

where $s \geq 0$ and $\rho_1, \rho_2 \in [0,1]$ are arbitrary. Using the definition of the ensemble in (84)–(86), we obtain

$$\overline{p}_{e,0} \,\dot{\leq}\, (M_0 M_{11}^{\rho_1} M_{12}^{\rho_1\rho_2})^{\rho_0} \times \sum_{\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2} P_{\boldsymbol{U}}(\boldsymbol{u}) P_{\boldsymbol{X}_1}(\boldsymbol{x}_1) P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) \sum_{\boldsymbol{y}} W^n(\boldsymbol{y}|\Xi(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2)) \times \Bigg(\sum_{\overline{\boldsymbol{u}}} P_{\boldsymbol{U}}(\overline{\boldsymbol{u}}) \Bigg(\sum_{\overline{\boldsymbol{x}}_1} P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1) \times \bigg(\frac{q^{n_1}(\overline{\boldsymbol{x}}_1, \boldsymbol{y}_1(\overline{\boldsymbol{u}}))}{q^{n_1}(\boldsymbol{x}_1, \boldsymbol{y}_1(\boldsymbol{u}))} \bigg)^{\rho_2 s} \frac{e^{a_1^{n_1}(\overline{\boldsymbol{x}}_1)}}{e^{a_1^{n_1}(\boldsymbol{x}_1)}} \Bigg)^{\rho_1} \times \Bigg(\sum_{\overline{\boldsymbol{x}}_2} P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2) \bigg(\frac{q^{n_2}(\overline{\boldsymbol{x}}_2, \boldsymbol{y}_2(\overline{\boldsymbol{u}}))}{q^{n_2}(\boldsymbol{x}_2, \boldsymbol{y}_2(\boldsymbol{u}))} \bigg)^{s} \frac{e^{a_2^{n_2}(\overline{\boldsymbol{x}}_2)}}{e^{a_2^{n_2}(\boldsymbol{x}_2)}} \Bigg)^{\rho_1\rho_2} \Bigg)^{\rho_0}, \tag{D.20}$$

[2]In the case of continuous alphabets, the summations should be replaced by integrals as necessary.

where for $u=1,2$, $a_u(\cdot)$ is one of the $L_u=2$ cost functions in (85), and $a_u^{n_u}(\boldsymbol{x}_u)\triangleq\sum_{i=1}^{n_u}a_u(x_{u,i})$. For each $(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{y})$, we write the argument to the summation over $\boldsymbol{y}$ in (D.20) as a product of two terms, namely

$$\begin{aligned}\mathsf{T}_1 \triangleq\, & W^n(\boldsymbol{y}|\Xi(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2)) \\ & \times\Big(q^{n_1}(\boldsymbol{x}_1,\boldsymbol{y}_1(\boldsymbol{u}))^{-\rho_1\rho_2 s}e^{-\rho_1 a_1^{n_1}(\boldsymbol{x}_1)} \\ & \times q^{n_2}(\boldsymbol{x}_2,\boldsymbol{y}_2(\boldsymbol{u}))^{-\rho_1\rho_2 s}e^{-\rho_1\rho_2 a_2^{n_2}(\boldsymbol{x}_2)}\Big)^{\rho_0} \quad \text{(D.21)}\end{aligned}$$

$$\begin{aligned}\mathsf{T}_2 \triangleq & \Bigg(\sum_{\overline{\boldsymbol{u}}}P_{\boldsymbol{U}}(\overline{\boldsymbol{u}}) \\ & \times\Bigg(\sum_{\overline{\boldsymbol{x}}_1}P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1)q^{n_1}\big(\overline{\boldsymbol{x}}_1,\boldsymbol{y}_1(\overline{\boldsymbol{u}})\big)^{\rho_2 s}e^{a_1^{n_1}(\overline{\boldsymbol{x}}_1)}\Bigg)^{\rho_1} \\ & \times\Bigg(\sum_{\overline{\boldsymbol{x}}_2}P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2)q^{n_2}\big(\overline{\boldsymbol{x}}_2,\boldsymbol{y}_2(\overline{\boldsymbol{u}})\big)^{s}e^{a_2^{n_2}(\overline{\boldsymbol{x}}_2)}\Bigg)^{\rho_1\rho_2}\Bigg)^{\rho_0}. \quad \text{(D.22)}\end{aligned}$$

Since $P_{\boldsymbol{X}_u}(\boldsymbol{x}_u)$ is upper bounded by a subexponential prefactor times $\prod_{i=1}^n Q_{X|U}(x_{u,i}|u)$ for $u=1,2$ (see Proposition 1), we have

$$\begin{aligned}&\sum_{\overline{\boldsymbol{x}}_1}P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1)q^{n_1}\big(\overline{\boldsymbol{x}}_1,\boldsymbol{y}_1(\overline{\boldsymbol{u}})\big)^{\rho_2 s}e^{a_1^{n_1}(\overline{\boldsymbol{x}}_1)} \\ &\qquad \dot{\le}\prod_{i=1}^{n_1}\sum_{\overline{x}_1}Q_{X|U}(\overline{x}_1|1)q(\overline{x}_1,y_{1,i}(\overline{\boldsymbol{u}}))^{\rho_2 s}e^{a_1(\overline{x}_1)} \quad \text{(D.23)} \\ &\sum_{\overline{\boldsymbol{x}}_2}P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2)q^{n_2}\big(\overline{\boldsymbol{x}}_2,\boldsymbol{y}_2(\overline{\boldsymbol{u}})\big)^{s}e^{a_2^{n_2}(\overline{\boldsymbol{x}}_2)} \\ &\qquad \dot{\le}\prod_{i=1}^{n_2}\sum_{\overline{x}_2}Q_{X|U}(\overline{x}_2|2)q(\overline{x}_2,y_{2,i}(\overline{\boldsymbol{u}}))^{s}e^{a_2(\overline{x}_2)}, \quad \text{(D.24)}\end{aligned}$$

where for $u=1,2$, $y_{u,i}(\overline{\boldsymbol{u}})$ is the i-th entry of $\boldsymbol{y}_u(\overline{\boldsymbol{u}})$. Using the definitions in (88) along with $a(u,x)\triangleq a_u(x)$, we therefore obtain

$$\begin{aligned}&\Bigg(\sum_{\overline{\boldsymbol{x}}_1}P_{\boldsymbol{X}_1}(\overline{\boldsymbol{x}}_1)q^{n_1}\big(\overline{\boldsymbol{x}}_1,\boldsymbol{y}_1(\overline{\boldsymbol{u}})\big)^{\rho_2 s}e^{a_1^{n_1}(\overline{\boldsymbol{x}}_1)}\Bigg)^{\rho_1} \\ &\quad\times\Bigg(\sum_{\overline{\boldsymbol{x}}_2}P_{\boldsymbol{X}_2}(\overline{\boldsymbol{x}}_2)q^{n_2}\big(\overline{\boldsymbol{x}}_2,\boldsymbol{y}_2(\overline{\boldsymbol{u}})\big)^{s}e^{a_2^{n_2}(\overline{\boldsymbol{x}}_2)}\Bigg)^{\rho_1\rho_2} \quad \text{(D.25)} \\ &\dot{\le}\Bigg(\prod_{i=1}^{n_1}\sum_{\overline{x}_1}Q_{X|U}(\overline{x}_1|1)q(\overline{x}_1,y_{1,i}(\overline{\boldsymbol{u}}))^{\rho_2 s}e^{a_1(\overline{x}_1)}\Bigg)^{\rho_1} \\ &\quad\times\Bigg(\prod_{i=1}^{n_2}\sum_{\overline{x}_2}Q_{X|U}(\overline{x}_2|2)q(\overline{x}_2,y_{2,i}(\overline{\boldsymbol{u}}))^{s}e^{a_2(\overline{x}_2)}\Bigg)^{\rho_1\rho_2} \quad \text{(D.26)} \\ &=\prod_{i=1}^{n}\Bigg(\sum_{\overline{x}}Q_{X|U}(\overline{x}|\overline{u}_i)q(\overline{x},y_i)^{s_1(\overline{u}_i)}e^{a(\overline{u}_i,\overline{x})}\Bigg)^{\rho_1(\overline{u}_i)}. \quad \text{(D.27)}\end{aligned}$$

Hence, and using the fact that $P_{\boldsymbol{U}}(\boldsymbol{u})\,\dot{\le}\,Q_U^n(\boldsymbol{u})$ (see [14, Ch. 2]), we obtain

$$\begin{aligned}\mathsf{T}_2\,\dot{\le}\,&\prod_{i=1}^{n}\Bigg(\sum_{\overline{u}}Q_U(\overline{u}) \\ &\times\Bigg(\sum_{\overline{x}}Q(\overline{x}|\overline{u})q(\overline{x},y_i)^{s_1(\overline{u})}e^{a(\overline{u},\overline{x})}\Bigg)^{\rho_1(\overline{u})}\Bigg)^{\rho_0}. \quad \text{(D.28)}\end{aligned}$$

A similar argument (without the need for the $\dot{\le}$ steps) gives

$$\mathsf{T}_1=\prod_{i=1}^{n}W(y_i|x_i)\Big(q(x_i,y_i)^{-\rho_1(u_i)s_1(u_i)}e^{-\rho_1(u_i)a(u_i,x_i)}\Big)^{\rho_0}, \quad \text{(D.29)}$$

where we have used the fact that $W^n(\boldsymbol{y}|\Xi(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2))=W^{n_1}(\boldsymbol{y}_1(\boldsymbol{u})|\boldsymbol{x}_1)W^{n_2}(\boldsymbol{y}_2(\boldsymbol{u})|\boldsymbol{x}_2)$. Substituting (D.28) and (D.29) into (D.20), we obtain

$$\begin{aligned}\overline{p}_{e,0}\,\dot{\le}\,&(M_0M_{11}^{\rho_1}M_{12}^{\rho_1\rho_2})^{\rho_0}\sum_{\boldsymbol{u},\boldsymbol{x}}P_{\boldsymbol{U}\boldsymbol{X}}(\boldsymbol{u},\boldsymbol{x})\prod_{i=1}^{n}\sum_{y}W(y|x_i) \\ &\times\Bigg(\sum_{\overline{u}}Q_U(\overline{u})\Bigg(\sum_{\overline{x}}Q_{X|U}(\overline{x}|\overline{u}_i) \\ &\times\bigg(\frac{q(\overline{x},y_i)}{q(x_i,y_i)}\bigg)^{s_1(\overline{u}_i)}\frac{e^{a(\overline{u}_i,\overline{x})}}{e^{a(u_i,x_i)}}\Bigg)^{\rho_1(\overline{u}_i)}\Bigg)^{\rho_0}, \quad \text{(D.30)}\end{aligned}$$

where

$$\begin{aligned}P_{\boldsymbol{U}\boldsymbol{X}}(\boldsymbol{u},\boldsymbol{x})\triangleq&\sum_{\boldsymbol{x}_1,\boldsymbol{x}_2}P_{\boldsymbol{U}}(\boldsymbol{u})P_{\boldsymbol{X}_1}(\boldsymbol{x}_1)P_{\boldsymbol{X}_2}(\boldsymbol{x}_2) \\ &\times\mathbb{1}\{\boldsymbol{x}=\Xi(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2)\}. \quad \text{(D.31)}\end{aligned}$$

If $P_{\boldsymbol{U}\boldsymbol{X}}$ were i.i.d. on Q_{UX}, then (D.30) would yield an error exponent that is positive when (87) ($u=1$) holds with strict inequality, by taking $\rho_0\to 0$ similarly to Theorem 3. The same can be done in the present setting by upper bounding $P_{\boldsymbol{U}\boldsymbol{X}}$ by a subexponential prefactor times Q_{UX}^n, analogously to (D.23)–(D.24). More precisely, we have

$$\begin{aligned}&P_{\boldsymbol{U}\boldsymbol{X}}(\boldsymbol{u},\boldsymbol{x}) \\ &\dot{\le}\sum_{\boldsymbol{x}_1,\boldsymbol{x}_2}P_{\boldsymbol{U}}(\boldsymbol{u})\Bigg(\prod_{i=1}^{n}Q_{X|U}(x_{1,i}|1)\Bigg) \\ &\qquad\times\Bigg(\prod_{i=1}^{n}Q_{X|U}(x_{2,i}|2)\Bigg)\mathbb{1}\{\boldsymbol{x}=\Xi(\boldsymbol{u},\boldsymbol{x}_1,\boldsymbol{x}_2)\} \quad \text{(D.32)} \\ &=P_{\boldsymbol{U}}(\boldsymbol{u})Q_{X|U}^n(\boldsymbol{x}|\boldsymbol{u}) \quad \text{(D.33)} \\ &\dot{\le}\,Q_U^n(\boldsymbol{u})Q_{X|U}^n(\boldsymbol{x}|\boldsymbol{u})=Q_{UX}^n(\boldsymbol{u},\boldsymbol{x}). \quad \text{(D.34)}\end{aligned}$$

References

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

[2] 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.

[3] 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.

[4] 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.

[5] V. Balakirsky, "A converse coding theorem for mismatched decoding at the output of binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1889–1902, Nov. 1995.

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

[7] A. Ganti, A. Lapidoth, and E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

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

[9] A. Somekh-Baruch, "On achievable rates and error exponents for channels with mismatched decoding," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 727–740, Feb. 2015.

[10] 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.
[11] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.
[12] Y. Liu and B. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 376–386, March 1996.
[13] J. Scarlett, "Reliable communication under mismatched decoding," Ph.D. dissertation, University of Cambridge, 2014, [Online: http://itc.upf.edu/biblio/1061].
[14] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.
[15] R. Gallager, "Fixed composition arguments and lower bounds to error probability," http://web.mit.edu/gallager/www/notes/notes5.pdf.
[16] P. Elias, "Coding for two noisy channels," in *Third London Symp. Inf. Theory*, 1955.
[17] A. Nazari, A. Anastasopoulos, and S. Pradhan, "Error exponent for multiple-access channels: Lower bounds," *IEEE Trans. Inf*, vol. 60, no. 9, pp. 5095–5115, Sept. 2014.
[18] A. G. D'yachkov, "Bounds on the average error probability for a code ensemble with fixed composition," *Prob. Inf. Transm.*, vol. 16, no. 4, pp. 3–8, 1980.
[19] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
[20] R. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, March 1985.
[21] ——, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.
[22] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. Guillén i Fàbregas, "Expurgated random-coding ensembles: Exponents, refinements and connections," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4449–4462, Aug. 2014.
[23] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Cost-constrained random coding and applications," in *Inf. Theory and Apps. Workshop*, San Diego, CA, Feb. 2013.
[24] M. Bierbaum and H. Wallmeier, "A note on the capacity region of the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 4, pp. 484–484, July 1979.
[25] R. Gallager, "Capacity and coding for degraded broadcast channels," *Prob. Peredachi Inf.*, vol. 10, no. 3, pp. 3–14, 1974.
[26] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 60–64, Jan. 1977.
[27] 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.
[28] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Tel Aviv University, 2003.
[29] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, Taipei, 2004.
[30] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423, July and Oct. 1948.
[31] M. S. Pinsker and A. Sheverdjaev, "Zero error capacity with erasure," *Prob. Inf. Transm.*, vol. 6, no. 1, pp. 20–24, 1970.
[32] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, Sept. 1956.
[33] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list, and detection zero-error capacities for low noise and a relation to identification," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 55–62, Jan. 1996.
[34] J. Scarlett, A. Somekh-Baruch, A. Martinez, and A. Guillén i Fàbregas, "A counter-example to the mismatched decoding converse for binary-input discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5387–5395, Oct. 2015.
[35] V. Balakirsky, "Coding theorem for discrete memoryless channels with given decision rule," in *Algebraic Coding*. Springer Berlin / Heidelberg, 1992, vol. 573, pp. 142–150.
[36] V. Y. F. Tan, "On the reliability function of the discrete memoryless relay channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1550–1573, April 2015.
[37] D. de Caen, "A lower bound on the probability of a union," *Discrete Math.*, vol. 169, pp. 217–220, 1997.
[38] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Ensemble-tight error exponents for mismatched decoders," in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, Oct. 2012, pp. 1951–1958.
[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[40] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 2006.
[41] K. Fan, "Minimax theorems," *Proc. Nat. Acad. Sci.*, vol. 39, pp. 42–47, 1953.

Jonathan Scarlett (S'14 – M'15) was born in Melbourne, Australia, in 1988. In 2010, he received the B.Eng. degree in electrical engineering and the B.Sci. degree in computer science from the University of Melbourne, Australia. In 2011, he was a research assistant at the Department of Electrical & Electronic Engineering, University of Melbourne. From October 2011 to August 2014, he was a Ph.D. student in the Signal Processing and Communications Group at the University of Cambridge, United Kingdom. He is now a post-doctoral researcher with the Laboratory for Information and Inference Systems at the École Polytechnique Fédérale de Lausanne, Switzerland. His research interests are in the areas of information theory, signal processing, machine learning, and high-dimensional statistics. He received the Cambridge Australia Poynton International Scholarship, and the 'EPFL Fellows' postdoctoral fellowship co-funded by Marie Curie.

Alfonso Martinez (SM'11) was born in Zaragoza, Spain, in October 1973. He is currently a Ramón y Cajal Research Fellow at Universitat Pompeu Fabra, Barcelona, Spain. He obtained his Telecommunications Engineering degree from the University of Zaragoza in 1997. In 1998-2003 he was a Systems Engineer at the research centre of the European Space Agency (ESAESTEC) in Noordwijk, The Netherlands. His work on APSK modulation was instrumental in the definition of the physical layer of DVB-S2. From 2003 to 2007 he was a Research and Teaching Assistant at Technische Universiteit Eindhoven, The Netherlands, where he conducted research on digital signal processing for MIMO optical systems and on optical communication theory. Between 2008 and 2010 he was a post-doctoral fellow with the Information-Theoretic Learning Group at Centrum Wiskunde & Informatica (CWI), in Amsterdam, The Netherlands. In 2011 he was a Research Associate with the Signal Processing and Communications Lab at the Department of Engineering, University of Cambridge, Cambridge, U.K. His research interests lie in the fields of information theory and coding, with emphasis on digital modulation and the analysis of mismatched decoding; in this area he has coauthored a monograph on "Bit-Interleaved Coded Modulation." More generally, he is intrigued by the connections between information theory, optical communications, and physics, particularly by the links between classical and quantum information theory.

Albert Guillén i Fàbregas (S'01 – M'05 – SM'09) received the Telecommunication Engineering Degree and the Electronics Engineering Degree from Universitat Politècnica de Catalunya and Politecnico di Torino, respectively in 1999, and the Ph.D. in Communication Systems from École Polytechnique Fédérale de Lausanne (EPFL) in 2004. Since 2011 he has been an ICREA Research Professor at Universitat Pompeu Fabra. He is also an Adjunct Researcher at the University of Cambridge. He has held appointments at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurècom, University of South Australia, University of Cambridge, as well as visiting appointments at EPFL, École Nationale des Télécommunications (Paris), Universitat Pompeu Fabra, University of South Australia, Centrum Wiskunde & Informatica and Texas A&M University in Qatar. His research interests are in the areas of information theory, coding theory and communication theory. Dr. Guillén i Fàbregas is a Member of the Young Academy of Europe, and received the Starting Grant from the European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference, the 2004 Best Doctoral Thesis Award from the Spanish Institution of Telecommunications Engineers, and a Research Fellowship of the Spanish Government to join ESA. He is also an Associate Editor of the IEEE Transactions on Information Theory, an Editor of the Foundations and Trends in Communications and Information Theory, Now Publishers and was an Editor of the IEEE Transactions on Wireless Communications.