entropy

MDPI

Article

Multi-Class Cost-Constrained Random Coding for Correlated Sources over the Multiple-Access Channel

Arezou Rezazadeh [1,*,†], Josep Font-Segura [2], Alfonso Martinez [2] and Albert Guillén i Fàbregas [2,3,4]

[1] Department of Electrical Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden
[2] Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain; josep.font@upf.edu (J.F.-S.); alfonso.martinez@upf.edu (A.M.); albert.guillen@upf.edu (A.G.i.F.)
[3] Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain
[4] Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK
* Correspondence: arezour@chalmers.se
† This work was carried out when the first author was with the Department of Information and Communication Technologies at Universitat Pompeu Fabra.

Abstract: This paper studies a generalized version of multi-class cost-constrained random-coding ensemble with multiple auxiliary costs for the transmission of N correlated sources over an N-user multiple-access channel. For each user, the set of messages is partitioned into classes and codebooks are generated according to a distribution depending on the class index of the source message and under the constraint that the codewords satisfy a set of cost functions. Proper choices of the cost functions recover different coding schemes including message-dependent and message-independent versions of independent and identically distributed, independent conditionally distributed, constant-composition and conditional constant composition ensembles. The transmissibility region of the scheme is related to the Cover-El Gamal-Salehi region. A related family of correlated-source Gallager source exponent functions is also studied. The achievable exponents are compared for correlated and independent sources, both numerically and analytically.

Keywords: multiple access channel; correlated sources; random coding; error exponents

Citation: Rezazadeh, A.; Font-Segura, J.; Martinez, A.; Guillén i Fàbregas, A. Multi-Class Cost-Constrained Random Coding for Correlated Sources over the Multiple-Access Channel. *Entropy* **2021**, *23*, 569. https://doi.org/10.3390/e23050569

Academic Editor: Victoria Kostina

Received: 12 April 2021
Accepted: 1 May 2021
Published: 3 May 2021

1. Introduction

In information theory, the fundamental problem of communication over a channel is studied from two complementary perspectives. First, one characterizes the transmissibility conditions, namely the circumstances under which the error probability asymptically vanishes as the blocklength goes to infinity. Second, one describes by means of error exponents the speed at which this error probability vanishes; the larger the exponent, the faster the error probability tends to zero. Since finding an exact expression for error probability is very difficult, a large body of work has investigated upper and lower bounds on the average error probability, or equivalently lower and upper bounds for the error exponent. In point-to-point, that is, single-user communication, using separate source-channel random coding [1,2], possibly with expurgation [1] (Eq. 5.7.10), yields lower bounds on the error exponent. In contrast, finding an upper bound to the error exponent satisfied by every code is more challenging. Generally, the hypothesis-testing method [3] is employed to derive upper bounds for the error exponent. Two well-known upper bounds to the error exponent are the sphere-packing exponent [4] and the minimum-distance exponent [5]. In fact, for rates greater than critical rate [1] (Sec. 5.6), the random-coding and sphere-packing bounds coincide with each other, while the expurgated and minimum-distance bounds coincide at rate zero.

For point-to-point communication, it was shown in ref. [1] (Prob. 5.16) that joint source-channel coding leads in general to a larger exponent than separate source-channel coding.

Additionally,using codewords with a composition dependent on the source message leads to a larger exponent than the case where codewords are drawn according to a fixed product distribution [6,7]. Moreover, a scheme where source messages are assigned to disjoint classes and encoded by codes that depend on the class index, attains the sphere-packing exponent in those cases where it is tight [8].

Many works have been devoted to studying the transmissibility and the error exponent for a two-user multiple-access channel (MAC) [9–11]. Separate source-channel coding for the MAC with independent sources was studied in refs. [9,12]. In ref. [13], a universal exponent for the MAC was derived by considering separate source-channel coding. In ref. [14], a transmissible region is derived for the MAC under mismatched decoding, where the decoding rule is fixed and possibly suboptimal. In ref. [15], it was shown that using structure coding can improve the error exponent of the MAC. The maximum-error-probability criterion and the impact of feedback for the MAC were studied in ref. [16]. By considering separate source-channel coding, lower and upper bounds for the error exponent of the MAC were respectively obtained in refs. [17,18]. For the MAC with independent sources, the idea of exploting the dependency between messages and codewords was studied in ref. [19]. In ref. [20], an achievable exponent for the MAC with independent sources was given in the dual domain, that is, as a lower dimensional problem over parameters in terms of Gallager functions. For the MAC with correlated sources, it was shown in ref. [11] that considering statistical dependency between messages and codewords for the MAC with correlated sources leads to a larger transmissible region. However, an example presented in ref. [21] shows that one can reliably transmit information through the MAC without satisfying the reliable transmission obtained in ref. [11]. In another line of work, superposition coding with Gacs Körner Witsenhausen (GKW) common part is used in ref. [22] to to describe the sufficient conditions lossless recoverability.

In contrast to single-user communication, the problem of reliable transmission of two correlated sources has not been solved yet and just the sufficient conditions of a reliable transmission has been derived. In ref. [23], by applying coding techniques, a new set of sufficient conditions were proposed. Moreover, in ref. [24] new sufficient conditions for the three-user MAC with correlated sources were studied. In ref. [25], an achievable exponent derived was presented in the primal domain, that is, as a multi-dimensional optimization problem over distributions that is generally difficult to analyze.

In this paper, we examine how statistical dependency between the messages and codewords improves the exponent, as well as its impact on the transmissibility region. In view of refs. [1] (Ch. 7) and [26], we study a generalized message-dependent cost-constrained random-coding ensemble with multiple cost functions. By choosing the proper cost functions, the multi-class cost-constrained ensemble subsumes multiple ensembles previously considered in the literature and recovers the transmissibility region in ref. [11].

The paper is organized as follows—in Section 2, we present the problem of transmission of N correlated sources over an N-input discrete memoryless multiple-access channel and provide the key definitions of error probability, transmissibility, random-coding ensemble, and achievable exponent. In Section 3, we review the existing random-coding ensembles, define a novel generalized multi-class cost-constraint ensemble and characterize its achievable exponent. In the discussion Section 4, we characterize the transmissibility region for our error exponent, relate the exponent to standard Gallager source and channel functions, and provide numerical results and formulas that allow us to rank the exponents attained by the various standard random-coding ensembles.

2. Problem Formulation

We study the simultaneous transmission of N correlated, discrete, memoryless sources over a channel; users are indexed by $\nu \in \mathcal{N} = \{1, 2, \ldots, N\}$. The source messages $\boldsymbol{u}_\nu$ of user ν have n symbols drawn from the alphabet $\mathcal{U}_\nu$. We denote by $\boldsymbol{u}_\sigma$ the ordered vector of source messages for all users in a set $\sigma \subset 2^{\mathcal{N}}$, i.e., a subset of the set of all user indices, and similarly by $\mathcal{U}_\sigma$ the Cartesian product of the source alphabets in the set σ. When

$\sigma = \mathcal{N}$, $\boldsymbol{u}_{\mathcal{N}}$ and $\mathcal{U}_{\mathcal{N}}$ denote the ordered vector of source messages for all users and the Cartesian product of the all source alphabets respectively. The sources are memoryless and are characterized by the joint probability distribution $\mathrm{P}_{\mathcal{N}}$

$$\mathrm{P}_{\mathcal{N}}(\boldsymbol{u}_{\mathcal{N}}) = \prod_{t=1}^{n} P_{\mathcal{N}}(u_{\mathcal{N},t}), \tag{1}$$

and by the symbol joint probability distribution $P_{\mathcal{N}}$. The source message and symbol marginal distributions of user $\nu \in \mathcal{N}$ are denoted by P_{ν} and P_{ν} respectively. Assuming that the sources are independent, the marginal distributions induce new joint (mismatched) probability distributions of sets of users $\sigma \subset 2^{\mathcal{N}}$. The induced independent-message and -symbol probabilities, denoted by $\mathrm{P}_{\sigma}^{\mathrm{ind}}$ and $P_{\sigma}^{\mathrm{ind}}$, are given by

$$\mathrm{P}_{\sigma}^{\mathrm{ind}}(\boldsymbol{u}_{\sigma}) = \prod_{\nu \in \sigma} \mathrm{P}_{\nu}(\boldsymbol{u}_{\nu}), \tag{2}$$

and similarly for $P_{\sigma}^{\mathrm{ind}}$.

Each user ν has an encoder that maps, without cooperation with the other users, the source message $\boldsymbol{u}_{\nu}$ onto a codeword $\boldsymbol{x}_{\nu}(\boldsymbol{u}_{\nu})$ also of length n and with symbols drawn from the alphabet $\mathcal{X}_{\nu}$. We denote the codebook of user ν by $\mathcal{C}_n^{\nu}$. We denote by $\boldsymbol{x}_{\sigma} \in \mathcal{X}_{\sigma}^{n}$ the vector of codewords for all users in a set $\sigma \subset 2^{\mathcal{N}}$. Both terminals simultaneously send these codewords over a discrete memoryless multiple access channel with output alphabet $\mathcal{Y}$. The symbolwise transition probability is denoted by W, and the channel is characterized by a conditional probability distribution

$$\mathrm{W}(\boldsymbol{y}|\boldsymbol{x}_{\mathcal{N}}) = \prod_{t=1}^{n} W(y_t|x_{\mathcal{N},t}), \tag{3}$$

where $\boldsymbol{y}$ is the received sequence of length n.

Based on $\boldsymbol{y}$, a joint decoder estimates all transmitted source messages $\boldsymbol{u}_{\mathcal{N}}$ according to the maximum a posteriori criterion:

$$\hat{\boldsymbol{u}}_{\mathcal{N}} = \underset{\boldsymbol{u}_{\mathcal{N}} \in \mathcal{U}_{\mathcal{N}}^{n}}{\arg\max}\, \mathrm{P}_{\mathcal{N}}(\boldsymbol{u}_{\mathcal{N}}) \mathrm{W}\big(\boldsymbol{y} \,|\, \boldsymbol{x}_{\mathcal{N}}(\boldsymbol{u}_{\mathcal{N}})\big), \tag{4}$$

where $\mathcal{U}_{\mathcal{N}}^{n}$ denotes the set of all possible source messages $\boldsymbol{u}_{\mathcal{N}}$. An error occurs if the decoded messages $\hat{\boldsymbol{u}}_{\mathcal{N}}$ differ from the transmitted $\boldsymbol{u}_{\mathcal{N}}$; we refer to $\hat{\boldsymbol{u}}_{\mathcal{N}} \neq \boldsymbol{u}_{\mathcal{N}}$ as an error event. The error probability for a given set of codebooks, $P_{\mathrm{e}}(\mathcal{C}_n^{\mathcal{N}})$, is thus given by

$$P_{\mathrm{e}}(\mathcal{C}_n^{\mathcal{N}}) \triangleq \Pr\{\hat{\boldsymbol{U}}_{\mathcal{N}} \neq \boldsymbol{U}_{\mathcal{N}}\}. \tag{5}$$

In our analysis, it will prove convenient to split the error event into $2^N - 1$ distinct types of error events indexed by the non-empty subsets in the power set of the user indices $2^{\mathcal{N}} \setminus \emptyset$, for example, $\tau \in \{\{1\}, \{2\}, \{1,2\}\}$ for $N = 2$. More precisely, the error event of type τ corresponds to the conditions $\hat{\boldsymbol{u}}_{\nu} \neq \boldsymbol{u}_{\nu}$ for all $\nu \in \tau$ and $\hat{\boldsymbol{u}}_{\nu} = \boldsymbol{u}_{\nu}$ for all $\nu \in \tau^c$, where τ^c is the complement of τ in the power set of the user indices.

We are interested in the asymptotics of the error probability for sufficiently large n, namely whether the error probability vanishes and how fast this probability tends to zero as it vanishes. The sources $U_{\mathcal{N}}$ are said to be transmissible over the channel if there exists a sequence of codebooks $\mathcal{C}_n^{\mathcal{N}}$ such that $\lim_{n\to\infty} P_{\mathrm{e}}(\mathcal{C}_n^{\mathcal{N}}) = 0$. To characterize the speed at which the error probability vanishes, we use the notion of exponent. An exponent E is said to be achievable if there exists a sequence of codebooks such that

$$\liminf_{n\to\infty} -\frac{1}{n} \log P_{\mathrm{e}}(\mathcal{C}_n^{\mathcal{N}}) \geq \mathrm{E}. \tag{6}$$

Source transmissibility and error-exponent achievability are typically studied by means of random coding. With random coding, one generates and studies sequences of ensembles of codebooks whose codewords are randomly drawn from a distribution $\mathrm{Q}_\nu(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu)$ independently for each user; as indicated by the notation, this distribution may possibly depend on the source message $\boldsymbol{u}_\nu$. The random-coding probability distribution for the channel input $\mathrm{Q}_\mathcal{N}(\boldsymbol{x}_\mathcal{N}|\boldsymbol{u}_\mathcal{N})$ combined for all users is given by

$$\mathrm{Q}_\mathcal{N}(\boldsymbol{x}_\mathcal{N}|\boldsymbol{u}_\mathcal{N}) = \prod_{\nu\in\mathcal{N}} \mathrm{Q}_\nu(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu). \tag{7}$$

The use of random coding allows us to study how the error probability averaged over the ensemble, denoted by $\bar{P}_\mathrm{e}$ vanishes as n grows. More importantly, it shows the existence of good codes in the ensemble such that their error probability vanishes. For the point-to-point and the multiple-access channels, a number of such random-coding ensembles have been studied in the literature, as reviewed in the following section, where we also present a multi-class cost-constrained ensemble subsuming all these ensembles and characterize the achievable exponent and transmissibility region of this ensemble.

2.1. Summary of Notation Used in the Paper

Sets are usually denoted by calligraphic upper case letter, e.g., $\mathcal{X}$, and the n-Cartesian product set of $\mathcal{X}$ is denoted by $\mathcal{X}^n$. The cardinality of a set such as $\mathcal{X}$ is denoted by $|\mathcal{X}|$. The indicator function representing an error event or that an element x belongs to a set $\mathcal{X}$ is denoted by $\mathbf{1}\{x \in \mathcal{X}\}$.

The number of users is denoted by N and user indices are typically represented by ν. The set of all users is denoted by $\mathcal{N}$. The power set of all subsets of $\mathcal{N}$ is denoted by $2^\mathcal{N}$ and the complement of a subset $\sigma \subset 2^\mathcal{N}$ is denoted by σ^c; sets in the power set of users are denoted that by Greek letters, for example, τ and σ. The number of source-message classes and of cost functions for user ν are respectively denoted by K_ν and L_ν; the sets of such classes are functions are respectively denoted by $\mathcal{K}_\nu$ and $\mathcal{L}_\nu$. Indices for source classes and cost functions are typically denoted by i_ν and ℓ_ν respectively.

Subscripts and superscripts in a quantity A may represent sets of user indices σ. Depending on the context, the quantity represents a list or a suitable product of variables for all elements in the set σ. For instance, for $\sigma = \{1,2\}$, $A^\sigma = (A^1, A^2)$ or $A_\sigma = (A_1, A_2)$. If the quantity is a probability distribution, its value for σ represents the probability distribution of the sequence, for example, $Q_\sigma^{i_\sigma}(x_\sigma) = \prod_{\nu\in\sigma} Q_\nu^{i_\nu}(x_\nu)$. If the quantity is a set, its value for σ is the Cartesian product, for example, $\mathcal{U}_\sigma = \mathcal{U}_1 \times \mathcal{U}_2$ for $\sigma = \{1,2\}$. If $\sigma = \emptyset$, then $A_\sigma = A^\sigma = 0$. If σ is a singleton, for example, $\sigma = \{2\}$, we simply write A_2 or A^2. We denote the operation that merges and sorts two lists A_{σ_1} and A_{σ_2} with $\sigma_1 \cap \sigma_2 = \emptyset$ into an ordered list containing all users in the union $\sigma_1 \cup \sigma_2$ by $[A_{\sigma_1}, A_{\sigma_2}]$. For sets of user indices, we denote such merging operation by $[\sigma_1, \sigma_2]$ and we have $[\sigma, \sigma^c] = \mathcal{N}$.

Scalar random variables are denoted by capital letters, for example, X and lowercase letters represent a particular realisation, for example, $x \in \mathcal{X}$. Capital bold letter denotes random vectors or sequences, for example, $\boldsymbol{X}$, while small bold letter $\boldsymbol{x} \in \mathcal{X}^n$ denote deterministic vectors or sequences. Probability distributions for vectors or sequences, typically of length n, (resp. for symbols) are represented by text-style letters, for example, P, Q, W (resp. math-style letters, for example, P, Q, W). Sequences symbols are usually affixed a subscript to indicate a user index; the t-th symbol in the sequence $\boldsymbol{x}_\nu$ is denoted by $x_{\nu,t}$.

The source-symbol distribution for user ν is denoted by $P_\nu(u_\nu)$. The joint distribution for users σ is denoted by $P_\sigma(u_\sigma)$; the joint distribution, computed as if the sources were independent, is denoted by $P_\sigma^{\mathrm{ind}}(u_\sigma)$. The conditional source distribution for users σ_1 given another set σ_2 is denoted by $P_{\sigma_1|\sigma_2}(u_{\sigma_1}|u_{\sigma_2})$. Vector or sequence distributions are defined analogously with P replaced by P. Channel input distributions are denoted by $Q_\nu(x_\nu)$, $Q_\nu^{i_\nu}(x_\nu)$, or $Q_{\nu,u_\nu}^{i_\nu}(x_\nu)$, where i_ν denotes the index of the class source message and $Q_{\nu,u_\nu}(x_\nu)$ is a shorthand for the conditional distribution $Q_\nu(x_\nu|u_\nu)$. Cost functions are similarly denoted by $a_\nu(x_\nu)$, $a_\nu^{i_\nu}(x_\nu)$, or $a_{\nu,u_\nu}^{i_\nu}(x_\nu)$. Vector or sequence distributions are defined

analogously with Q or a respectively replaced by Q or a. The conditional distribution for the channel output symbol (resp. sequence) is denoted by $W(y|x_{\mathcal{N}})$ (resp. $\mathrm{W}(y|x_{\mathcal{N}})$).

3. Multi-Class Cost-Constrained Ensemble with Statistical Dependency

3.1. Review of Random-Coding Ensembles

The simplest and oldest random-coding ensemble is the independent, identically distributed (*iid*) [1,12,17,27], where the symbols $x_{\nu,t}$ in all codewords $\boldsymbol{x}_\nu$ of a given user ν are generated independently according to the same input distributions $Q_\nu(x_{\nu,t})$ for all source messages $\boldsymbol{u}_\nu$. Throughout the paper, we shall identify ensembles by hyphenated acronyms, where the first part indicates the possible dependence of the codeword on the source message and the second part describes the generation of symbols in a codeword. This first ensemble is thus the message-independent iid (*mi-iid*) ensemble, since codewords have the same distribution for all source messages and symbols are independent of each other and independent of the source message symbols too. For the *mi-iid* ensemble, the random-coding distribution is given by

$$\mathrm{Q}_\nu^{\textit{mi-iid}}(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu) = \prod_{t=1}^{n} Q_\nu(x_{\nu,t}). \tag{8}$$

In the message-independent, independent-conditionally-distributed (*mi-icd*) ensemble, the codewords $\boldsymbol{x}_\nu$ of user ν are generated identically for all source messages $\boldsymbol{u}_\nu$, independently of the full message $\boldsymbol{u}_\nu$, and with symbols according to a set of $|\mathcal{U}_\nu|$ conditional probability distributions $Q_{\nu,u_\nu}(x_\nu) \triangleq Q_\nu(x_\nu|u_\nu)$. To this end, let $\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)$ denote the set of positions where the symbol $u_\nu \in \mathcal{U}$ appears in the sequence $\boldsymbol{u}_\nu$, namely

$$\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu) = \big\{t \in \{1,2,...,n\} : \boldsymbol{u}_{\nu,t} = u_\nu\big\}. \tag{9}$$

Within each subsequence of $\boldsymbol{u}_\nu$ where $\boldsymbol{u}_{\nu,t} = u_\nu$, represented by $\boldsymbol{u}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)$, symbols are drawn independently according to $Q_{\nu,u_\nu}(x_\nu)$. For this *mi-icd* ensemble, codewords are generated according to

$$\mathrm{Q}_\nu^{\textit{mi-icd}}(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu) = \prod_{u_\nu \in \mathcal{U}_\nu} \prod_{t \in \mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)} Q_{\nu,u_\nu}(x_{\nu,t}) \tag{10}$$

$$= \prod_{t=1}^{n} Q_{\nu,u_{\nu,t}}(x_{\nu,t}). \tag{11}$$

Compared to the *mi-iid* ensemble, the *mi-icd* ensemble can lead to a larger transmissible region for the multiple-access channel with correlated sources [11,21]. An example of generation of three codewords $\boldsymbol{x}_\nu^{(1)}$, $\boldsymbol{x}_\nu^{(2)}$ and $\boldsymbol{x}_\nu^{(3)}$ in the mi-icd ensemble is shown in Figure 1, for a given source sequence $\boldsymbol{u}_\nu = (\alpha,\beta,\beta,\gamma,\beta,\gamma,\gamma,\alpha,\beta,\alpha)$ with source alphabet $\mathcal{U} = \{\alpha,\beta,\gamma\}$. To generate each codeword $\boldsymbol{x}_\nu$ with alphabet $\mathcal{X} = \{a,c,e\}$, three subcodewords $\boldsymbol{x}_\nu(\mathcal{I}_\alpha(\boldsymbol{u}_\nu)$, $\boldsymbol{x}_\nu(\mathcal{I}_\beta(\boldsymbol{u}_\nu)$ and $\boldsymbol{x}_\nu(\mathcal{I}_\gamma(\boldsymbol{u}_\nu)$ are pairwise-independently generated with i. i. d. distributions $Q_{\nu,\alpha} = (1/3,1/3,1/3)$, $Q_{\nu,\beta} = (1/2,1/4,1/4)$ and $Q_{\nu,\gamma} = (1/3,2/3,0)$, respectively. Symbols generated according to $Q_{\nu,\alpha}$, $Q_{\nu,\beta}$ and $Q_{\nu,\gamma}$ are respectively represented as green circles, blue boxes and red diamonds in the figure. In the example, $\mathcal{I}_\alpha(\boldsymbol{u}_\nu) = \{1,8,10\}$, $\mathcal{I}_\beta(\boldsymbol{u}_\nu) = \{2,3,5,9\}$ and $\mathcal{I}_\gamma(\boldsymbol{u}_\nu) = \{4,6,7\}$. For instance, the subcodeword $\boldsymbol{x}_\nu^{(1)}(\mathcal{I}_\gamma(\boldsymbol{u}_\nu)$ has three symbols, each generated independently from $Q_{\nu,\gamma}$, leading to the red-diamond symbols $\boldsymbol{x}_\nu^{(1)}(\mathcal{I}_\gamma(\boldsymbol{u}_\nu) = (a,a,a)$.

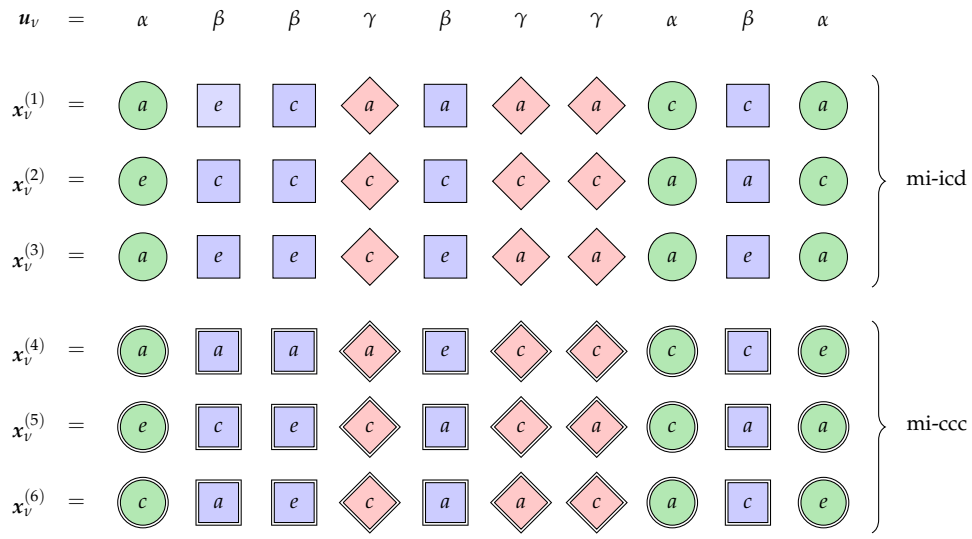

Figure 1. Example of codewords $\boldsymbol{x}_\nu^{(1)}$, $\boldsymbol{x}_\nu^{(2)}$ and $\boldsymbol{x}_\nu^{(3)}$ in the *mi-icd* ensemble and $\boldsymbol{x}_\nu^{(4)}$, $\boldsymbol{x}_\nu^{(5)}$ and $\boldsymbol{x}_\nu^{(6)}$ in the *mi-ccc* ensemble, for a given source sequence $\boldsymbol{u}_\nu$.

Next, we have the message-dependent iid (*md-iid*) ensemble [6,8,19,25,28], where codewords for each user are generated with i. i. d. symbols according to different distributions $Q_\nu^{i_\nu}(x_\nu)$ that depend on the full source message through the class index i_ν of the class the source message belongs to. More precisely, for user $\nu = 1, 2$ with source marginal distribution P_ν, the i_ν-th class $\mathcal{A}_\nu^{i_\nu}$, where $i_\nu \in \mathcal{K}_\nu = \{1, \ldots, K_\nu\}$, is defined as the set of all source messages whose probability $\mathrm{P}_\nu(\boldsymbol{u}_\nu)$ is within a given interval, that is,

$$\mathcal{A}_\nu^{i_\nu} = \left\{\boldsymbol{u}_\nu \in \mathcal{U}_\nu^n \,:\, \gamma_{\nu,i_\nu}^n < \mathrm{P}_\nu(\boldsymbol{u}_\nu) \leq \gamma_{\nu,i_\nu-1}^n\right\}, \tag{12}$$

where the thresholds $\gamma_{\nu,j}$ are $K_\nu + 1$ non-negative numbers, ordered from higher to lower, such that $0 = \gamma_{\nu,K_\nu} \leq \gamma_{\nu,K_\nu-1} \leq \ldots \leq \gamma_{\nu,1} < \gamma_{\nu,0} = 1$, and $\min_{u_\nu} P_\nu(u_\nu) < \gamma_{\nu,K_\nu-1}$ and $\gamma_{\nu,1} \leq \max_{u_\nu} P_\nu(u_\nu)$. The md-iid random-coding distribution is given by

$$\mathrm{Q}_\nu^{\textit{md-iid}}(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu) = \prod_{t=1}^{n} Q_\nu^{i_\nu(\boldsymbol{u}_\nu)}(x_{\nu,t}). \tag{13}$$

The exponent of this md-iid ensemble can be larger than that of the *mi-iid* ensemble for joint source-channel coding [8,20,28].

In the message-dependent, independent conditional symbol distributions (*md-icd*) ensemble, messages in the class i_ν for user ν are encoded with codewords whose symbols are generated independently according to the conditional input distribution $Q_{\nu,u_\nu}^{i_\nu}(x_\nu)$. The random-coding distribution of the md-icd ensemble is thus given by

$$\mathrm{Q}_\nu^{\textit{md-icd}}(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu) = \prod_{u_\nu \in \mathcal{U}_\nu} \prod_{t \in \mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)} Q_{\nu,u_\nu}^{i_\nu(\boldsymbol{u}_\nu)}(x_{\nu,t}). \tag{14}$$

In the message-independent, constant-composition (*mi-cc*) ensemble [29,30], codewords $\boldsymbol{x}_\nu$ are drawn independently with an empirical distribution $\hat{Q}_\nu(x_\nu)$ close to a given $Q_\nu(x_\nu)$, independently of the source message $\boldsymbol{u}_\nu$. For each user, codewords $\boldsymbol{x}_\nu$ are randomly picked from $\mathcal{T}_\nu^n(Q_\nu)$, the set of all sequences whose empirical distribution has a variational distance to Q_ν of at most $1/n$, that is

$$\mathcal{T}_\nu^n(Q_\nu) = \left\{\boldsymbol{x}_\nu \in \mathcal{X}_\nu^n : \max_{x_\nu} \left|\hat{Q}_\nu(x_\nu) - Q_\nu(x_\nu)\right| < \frac{1}{n}\right\}. \tag{15}$$

For this *mi-cc* ensemble, the random-coding distribution is given by

$$\mathrm{Q}_{\nu}^{\text{mi-cc}}(\boldsymbol{x}_{\nu}|\boldsymbol{u}_{\nu}) = \frac{1}{|\mathcal{T}_{\nu}^{n}(Q_{\nu})|}\mathbf{1}\{\boldsymbol{x}_{\nu} \in \mathcal{T}_{\nu}^{n}(Q_{\nu})\}. \tag{16}$$

While the *mi-cc* and *mi-iid* ensembles lead to identical transmissibility conditions, the former may achieve strictly larger exponents for suboptimal input distributions already in single-user settings [29].

The message-independent, conditional constant-composition (*mi-ccc*) ensemble combines features of the *mi-icd* and *mi-cc* ensembles. For each subsequence $\boldsymbol{u}_{\nu}(\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu}))$, the corresponding subcodewords $\boldsymbol{x}_{\nu}(\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu}))$ are drawn independently from the set $\mathcal{T}_{\nu}^{|\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu})|}(Q_{\nu,u_{\nu}})$ of subsequences with empirical distribution close to $Q_{\nu,u_{\nu}}(x_{\nu})$, namely

$$\mathcal{T}_{\nu}^{|\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu})|}(Q_{\nu,u_{\nu}}) = \left\{\boldsymbol{x}_{\nu} \in \mathcal{X}_{\nu}^{|\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu})|} : \max_{x_{\nu} \in \boldsymbol{x}_{\nu}} \left|\hat{Q}_{\nu}(x_{\nu}) - Q_{\nu,u_{\nu}}(x_{\nu})\right| < \frac{1}{|\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu})|}\right\}. \tag{17}$$

The random-coding distribution of the *mi-ccc* ensemble is given by

$$\mathrm{Q}_{\nu}^{\textit{mi-ccc}}(\boldsymbol{x}_{\nu}|\boldsymbol{u}_{\nu}) = \prod_{u_{\nu} \in \mathcal{U}_{\nu}} \frac{1}{\left|\mathcal{T}_{\nu}^{|\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu})|}(Q_{\nu,u_{\nu}})\right|}\mathbf{1}\left\{\boldsymbol{x}_{\nu}(\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu})) \in \mathcal{T}_{\nu}^{|\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu})|}(Q_{\nu,u_{\nu}})\right\}. \tag{18}$$

An example of the generation of three codewords $\boldsymbol{x}_{\nu}^{(4)}$, $\boldsymbol{x}_{\nu}^{(5)}$ and $\boldsymbol{x}_{\nu}^{(6)}$ in the *mi-ccc* ensemble is also shown in Figure 1 as a comparison to the md-iid ensemble, for the same source sequence $\boldsymbol{u}_{\nu}$, source alphabet $\mathcal{U} = \{\alpha, \beta, \gamma\}$ and input alphabet $\mathcal{X} = \{a, c, e\}$. Now, to generate each codeword $\boldsymbol{x}_{\nu}$, three subcodewords $\boldsymbol{x}_{\nu}(\mathcal{I}_{\alpha}(\boldsymbol{u}_{\nu}))$, $\boldsymbol{x}_{\nu}(\mathcal{I}_{\beta}(\boldsymbol{u}_{\nu}))$ and $\boldsymbol{x}_{\nu}(\mathcal{I}_{\gamma}(\boldsymbol{u}_{\nu}))$ are pairwise-independently, uniformly drawn in the type classes with empirical distributions $\hat{Q}_{\nu,\alpha}$, $\hat{Q}_{\nu,\beta}$ and $\hat{Q}_{\nu,\gamma}$ that are closest to $Q_{\nu,\alpha}$, $Q_{\nu,\beta}$ and $Q_{\nu,\gamma}$, respectively. Since in the example $|\mathcal{I}_{\alpha}(\boldsymbol{u}_{\nu})| = 3$, $|\mathcal{I}_{\beta}(\boldsymbol{u}_{\nu})| = 4$ and $|\mathcal{I}_{\gamma}(\boldsymbol{u}_{\nu})| = 3$, it follows that $\hat{Q}_{\nu,\alpha} = (1/3, 1/3, 1/3)$, $\hat{Q}_{\nu,\beta} = (1/2, 1/4, 1/4)$ and $\hat{Q}_{\nu,\gamma} = (1/3, 2/3, 0)$. Symbols generated according to $\hat{Q}_{\nu,\alpha}$, $\hat{Q}_{\nu,\beta}$ and $\hat{Q}_{\nu,\gamma}$ are respectively represented as green doubled circles, blue doubled boxes and red doubled diamonds in the figure. For instance, all subcodewords $\boldsymbol{x}_{\nu}^{(j)}(\mathcal{I}_{\gamma}(\boldsymbol{u}_{\nu}))$, for $j = 4, 5, 6$, have three symbols jointly generated from the constant-composition type $\hat{Q}_{\nu,\gamma}$, that is, exactly one *a* and two *c*s.

The message-dependent, constant-composition (*md-cc*) ensemble combines the features of having different distributions for different messages with constant-composition random coding. For messages in the class $i_{\nu} \in \{1, \ldots, K_{\nu}\}$ for user ν, codewords are drawn from the set of sequences with empirical distribution close to $Q_{\nu}^{i_{\nu}}(x_{\nu})$. For this ensemble, the random-coding distribution is given by

$$\mathrm{Q}_{\nu}^{\textit{md-cc}}(\boldsymbol{x}_{\nu}|\boldsymbol{u}_{\nu}) = \frac{1}{\left|\mathcal{T}_{\nu}^{n}(Q_{\nu}^{i_{\nu}(\boldsymbol{u}_{\nu})})\right|}\mathbf{1}\left\{\boldsymbol{x}_{\nu} \in \mathcal{T}_{\nu}^{n}(Q_{\nu}^{i_{\nu}(\boldsymbol{u}_{\nu})})\right\}. \tag{19}$$

Finally, the message-dependent, conditional constant-composition (*md-ccc*) ensemble combines several of the ensembles listed above. For a given message $\boldsymbol{u}_{\nu} = (u_{\nu,1}, ..., u_{\nu,n})$ in the i_{ν}-th class, that is, $\boldsymbol{u}_{\nu} \in \mathcal{A}_{\nu}^{i_{\nu}}$, the subsequence of $\boldsymbol{u}_{\nu}$ having the same symbol u_{ν}, that is, $\boldsymbol{u}_{\nu}(\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu}))$, is encoded with pairwise-independent codewords generated from the set of codewords with empirical distribution very close to $Q_{\nu,u_{\nu}}^{i_{\nu}}(x_{\nu})$. The random-coding distribution of the md-ccc ensemble is thus given by

$$\mathrm{Q}_{\nu}^{\textit{md-ccc}}(\boldsymbol{x}_{\nu}|\boldsymbol{u}_{\nu}) = \prod_{u_{\nu} \in \mathcal{U}_{\nu}} \frac{1}{\left|\mathcal{T}_{\nu}^{|\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu})|}(Q_{\nu,u_{\nu}}^{i_{\nu}(\boldsymbol{u}_{\nu})})\right|}\mathbf{1}\left\{\boldsymbol{x}_{\nu}(\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu})) \in \mathcal{T}_{\nu}^{|\mathcal{I}_{u_{\nu}}(\boldsymbol{u}_{\nu})|}(Q_{\nu,u_{\nu}}^{i_{\nu}(\boldsymbol{u}_{\nu})})\right\}. \tag{20}$$

3.2. Generalized Multi-Class Cost-Constrained Ensemble

Motivated by the ensembles listed in the previous section, and inspired by refs. [1] (Ch. 7) and [26] (Sec. II), we study a generalized message-dependent multi-class cost-constrained random-coding ensemble with multiple auxiliary costs.

For each user, we partition the set of source messages into K_ν disjoint classes with thresholds on the message probabilities as in Equation (12). Let the source message be in the i_ν-th class, that is, $i_\nu(\boldsymbol{u}_\nu) = i_\nu$. Given the source message $\boldsymbol{u}_\nu$ and the source symbol u_ν, we consider the subsequence $\boldsymbol{u}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)$, where $\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)$ is defined in Equation (9), and we denote the corresponding source subsequence and subcodeword by $\boldsymbol{u}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)$ and $\boldsymbol{x}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)$ respectively. For each user ν, class index i_ν, and source message symbol u_ν, the subcodeword $\boldsymbol{x}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)$ is drawn according to a symbolwise i. i. d. distribution $Q^{i_\nu}_{\nu,u_\nu}(x_\nu)$ conditioned on a set of cost constraints being satisfied. We consider L_ν additive cost functions $a^{i_\nu,\ell_\nu}_{\nu,u_\nu}(x_\nu)$, $\ell_\nu \in \mathcal{L}_\nu = \{1,\dots,L_\nu\}$. The total cost $\mathrm{a}^{i_\nu,\ell_\nu}_{\nu,u_\nu}\Big(\boldsymbol{x}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)\Big)$ of the subcodeword $\boldsymbol{x}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)$ is given by the sum of the symbol costs $a^{i_\nu,\ell_\nu}_{\nu,u_\nu}$, namely

$$\mathrm{a}^{i_\nu,\ell_\nu}_{\nu,u_\nu}\Big(\boldsymbol{x}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)\Big) = \sum_{j\in\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)} a^{i_\nu,\ell_\nu}_{\nu,u_\nu}(x_{\nu,j}). \tag{21}$$

We assume that the average cost $\phi^{i_\nu,\ell_\nu}_{\nu,u_\nu}$ under the conditional distribution $Q^{i_\nu}_{\nu,u_\nu}$ is zero:

$$\phi^{i_\nu,\ell_\nu}_{\nu,u_\nu} = \sum_{x_\nu\in\mathcal{X}_\nu} Q^{i_\nu}_{\nu,u_\nu}(x_\nu)\, a^{i_\nu,\ell_\nu}_{\nu,u_\nu}(x_\nu) = 0. \tag{22}$$

Finally, fix some parameters $\delta_\nu > 0$ and let $\mathcal{D}^{i_\nu}_\nu$ be the set of codewords for which the average empirical cost of its constituent subcodewords $\frac{1}{|\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)|}\mathrm{a}^{i_\nu,\ell_\nu}_{\nu,u_\nu}\Big(\boldsymbol{x}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)\Big)$ is close to the statistical mean $\phi^{i_\nu,\ell_\nu}_{\nu,u_\nu} = 0$ for all cost functions and source symbols, i.e.,

$$\mathcal{D}^{i_\nu}_{\nu,\boldsymbol{u}_\nu} \triangleq \left\{\boldsymbol{x}_\nu \in \mathcal{X}^n_\nu : \left|\frac{1}{|\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)|}\mathrm{a}^{i_\nu,\ell_\nu}_{\nu,u_\nu}\Big(\boldsymbol{x}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)\Big)\right| \le \frac{\delta_\nu}{|\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)|},\ u_\nu\in\mathcal{U}_\nu,\ \ell_\nu\in\mathcal{L}_\nu\right\}. \tag{23}$$

Codewords $\boldsymbol{x}_\nu$ are the combination of subcodewords $\boldsymbol{x}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)$ with respective positions in $\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)$. For this multi-class cost-constrained ensemble, the random-coding distribution is thus given by

$$\mathrm{Q}^{cost}_\nu(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu) = \frac{1}{\Xi_\nu}\prod_{u_\nu\in\mathcal{U}_\nu}\prod_{t\in\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)} Q^{i_\nu}_{\nu,u_\nu}(x_{\nu,t})\mathbf{1}\{\boldsymbol{x}_\nu\in\mathcal{D}^{i_\nu}_{\nu,\boldsymbol{u}_\nu}\} \tag{24}$$

$$= \frac{1}{\Xi_\nu}\prod_{t=1}^{n} Q^{i_\nu}_{\nu,u_{\nu,t}}(x_{\nu,t})\mathbf{1}\{\boldsymbol{x}_\nu\in\mathcal{D}^{i_\nu}_{\nu,\boldsymbol{u}_\nu}\}, \tag{25}$$

where Ξ_ν is a normalizing constant and the class index is determined by the source message, $i_\nu = i_\nu(\boldsymbol{u}_\nu)$.

The multi-class cost-constrained ensemble subsumes all the ensembles described in Section 3.1. First of all, the iid and icd ensembles are recovered by setting $L_\nu = 0$ and choosing the appropriate number of classes K_ν and random-coding distributions Q_ν, Q_{ν,u_ν}, $Q^{i_\nu}_\nu$ and $Q^{i_\nu}_{\nu,u_\nu}$. For all these cases, the set $\mathcal{D}^{i_\nu}_{\nu,\boldsymbol{u}_\nu}$ includes all generated codewords and the normalizing constant is $\Xi_\nu = 1$.

To recover the constant-composition ensembles, for which constraints force the subcodewords to belong to some set $\mathcal{T}^n_\nu(Q_\nu)$ or $\mathcal{T}^{|\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)|}_\nu(Q^{i_\nu}_{\nu,u_\nu})$, for each of the K_ν classes for user ν we set $\delta_\nu < 1$, $L_\nu = |\mathcal{X}_\nu|$ and bijectively map the channel input symbols to cost function indices $\ell_\nu(x_\nu)$ so that

$$a^{i_\nu,\ell_\nu}_{\nu,u_\nu}(x_\nu) = \mathbf{1}\{x_\nu = \ell_\nu\} - Q^{i_\nu}_{\nu,u_\nu}(\ell_\nu). \tag{26}$$

In case the ensemble does not depend on either i_ν or u_ν, these symbols are dropped from Equation (26). For example, for the md-cc ensemble, we have $a^{i_\nu}_{\nu,\ell_\nu}(\ell_\nu) = \mathbf{1}\{x_\nu = \ell_\nu\} - Q^{i_\nu}_\nu(x_\nu)$. In addition, the codeword set $\mathcal{D}^{i_\nu}_{\nu,\boldsymbol{u}_\nu}$ in Equation (23) is simplified as

$$\mathcal{D}^{i_\nu}_{\nu,\boldsymbol{u}_\nu} = \left\{ \boldsymbol{x}_\nu \in \mathcal{X}^n_\nu : \left| \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\{x_{\nu,t} = x\} - Q^{i_\nu}_\nu(x) \right| \le \frac{1}{n},\ x \in \mathcal{X}_\nu \right\}, \tag{27}$$

which is the same as $\mathcal{T}^n(Q^{i_\nu}_\nu)$ given a version of Equation (15) where Q_ν may depend on i_ν.

Again, choosing the right number of classes K_ν and random-coding distributions Q_ν, Q_{ν,u_ν}, $Q^{i_\nu}_\nu$, and $Q^{i_\nu}_{\nu,u_\nu}$ recovers the various constant-composition ensembles. By construction, the set $\mathcal{D}^{i_\nu}_{\nu,\boldsymbol{u}_\nu}$ includes only the (sub)codewords with empirical distribution close to respectively Q_ν, Q_{ν,u_ν}, $Q^{i_\nu}_\nu$, and $Q^{i_\nu}_{\nu,u_\nu}$, and the normalizing constant Ξ_ν is the probability of the corresponding type set (or product thereof). As an example, for the md-ccc ensemble, choosing the cost functions in Equation (26) as follows

$$\mathrm{a}^{i_\nu,\ell_\nu}_{\nu,u_\nu}\Big(\boldsymbol{x}_\nu\big(\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)\big)\Big) = \sum_{j\in\mathcal{I}_{u_\nu}(\boldsymbol{u}_\nu)} \mathbf{1}\{x_{\nu,j} = \ell_\nu\} - Q^{i_\nu}_{\nu,u_\nu}(\ell_\nu) \tag{28}$$

yields the following cost-constraint set, which is equivalent to Equation (17),

$$\mathcal{D}^{i_\nu}_{\nu,\boldsymbol{u}_\nu} = \left\{ \boldsymbol{x}_\nu \in \mathcal{X}^n_\nu : \left| \frac{\sum_{j\in\mathcal{I}_u(\boldsymbol{u}_\nu)} \mathbf{1}\{x_{\nu,j} = x\}}{\big|\mathcal{I}_u(\boldsymbol{u}_\nu)\big|} - Q^{i_\nu}_{\nu,u_\nu}(x) \right| \le \frac{1}{\big|\mathcal{I}_u(\boldsymbol{u}_\nu)\big|},\ u \in \mathcal{U}_\nu,\ x \in \mathcal{X}_\nu \right\}. \tag{29}$$

3.3. Exponent for the Generalized Multi-Class Cost-Constrained Ensemble

Theorem 1. *For the transmission of N correlated memorlyess sources with joint distribution $P_\mathcal{N}$, where $\mathcal{N} = \{1, 2 \ldots, N\}$, over a channel with input $x_\mathcal{N}$ over a memoryless channel with transition probabilitiy $W(y|x_\mathcal{N})$, consider a random-coding multi-class cost-constrained ensemble where source messages for each user $\nu \in \mathcal{N}$ are allocated, depending on their probabilities, into K_ν classes with thresholds $\{\gamma_{\nu,0}, \gamma_{\nu,1}, \ldots, \gamma_{\nu,K_\nu}\}$, as in Equation (12), and encoded onto codewords randomly generated with a distribution $\mathrm{Q}^{i_\nu}_\nu(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu)$ that depends on the source message according to Equation (24) through symbol distributions $Q^{i_\nu}_{\nu,u_\nu}$ that possibly depend on the source-message class index i_ν and source symbol $u_\nu$and L_ν cost functions $a^{i_\nu,\ell_\nu}_{\nu,u_\nu}$, $\ell_\nu \in \{1, 2, \ldots, L_\nu\}$. This random-coding ensemble attains the following exponent E^{cost}*

$$\mathrm{E}^{cost} = \min_{\tau\in 2^\mathcal{N}\setminus\emptyset,\, i_\mathcal{N}\in\mathcal{K}_\mathcal{N}} \ \max_{0\le\rho\le 1} \ \max_{\lambda^{L,U}_\mathcal{N}\ge 0,\, r^{\ell_\mathcal{N}}_{\mathcal{N}u_\mathcal{N}}\in\mathbf{R}} E^{i_\mathcal{N}}_\tau\big(\rho, \lambda^{L,U}_\mathcal{N}, r^{\ell_\mathcal{N}}_{\mathcal{N}u_\mathcal{N}}\big), \tag{30}$$

where the Gallager function $E^{i_\mathcal{N}}_\tau\big(\rho, \lambda^{L,U}_\mathcal{N}, r^{\ell_\mathcal{N}}_{\mathcal{N}u_\mathcal{N}}\big)$ is given by

$$E^{i_\mathcal{N}}_\tau\big(\rho, \lambda^{L,U}_\mathcal{N}, r^{\ell_\mathcal{N}}_{\mathcal{N}u_\mathcal{N}}\big) = \\ -\log \sum_{u_{\tau^c}, x_{\tau^c}, y} \left(\sum_{u_\tau, x_\tau} P_\mathcal{N}(u_\mathcal{N})^{\frac{1}{1+\rho}} \Lambda^{i_\mathcal{N}}_\mathcal{N}(u_\mathcal{N}) Q^{i_\tau}_{\tau,u_\tau}(x_\tau) R^{i_\mathcal{N}}_{\mathcal{N},u_\mathcal{N}}(x_\mathcal{N}) \big(Q^{i_{\tau^c}}_{\tau^c,u_{\tau^c}}(x_{\tau^c}) W(y|x_\mathcal{N})\big)^{\frac{1}{1+\rho}} \right)^{1+\rho}, \tag{31}$$

and the functions $\Lambda^{i_\sigma}_\sigma(u_\sigma)$ and $R^{i_\sigma}_{\sigma,u_\sigma}(x_\sigma)$ are respectively given by

$$\Lambda^{i_\sigma}_\sigma(u_\sigma) = \prod_{\nu\in\sigma} \left(\frac{\mathrm{P}_\nu(u_\nu)}{\gamma_{\nu,i_\nu}} \right)^{\lambda^L_\nu} \left(\frac{\gamma_{\nu,i_\nu - 1}}{\mathrm{P}_\nu(u_\nu)} \right)^{\lambda^U_\nu}, \tag{32}$$

$$R^{i_\sigma}_{\sigma,u_\sigma}(x_\sigma) = \prod_{\nu\in\sigma} \prod_{\ell_\nu\in\mathcal{L}_\nu} e^{r^{\ell_\nu}_{\nu u_\nu} a^{i_\nu,\ell_\nu}_{\nu,u_\nu}(x_\nu)}, \tag{33}$$

and implicitly depende on the set of optimization parameters $\big(\lambda^{L,U}_\mathcal{N}, r^{\ell_\mathcal{N}}_{\mathcal{N}u_\mathcal{N}}\big)$.

Proof. This result is proved in Appendix A. □

The random-coding exponent in Equation (30) depends on the partitioning of the source-message set into classes, the channel input distributions, and the codeword cost-constraint functions. The best possible generalized cost-constraint exponent is obtained by optimizing over the multi-class partitioning, the cost constraints and the input distributions. We briefly discuss the optimization w. r. t. the thresholds of the source messages partitioning in Appendix B. In the next section, we provide some numerical examples where we compute the optimal exponents for either independent or correlated sources, and find that the optimal number of classes is two. In ref. [31] (Sec. 3.2.1.1), we provide some indications of why this optimality of only two classes is harder to establish in multi-user scenarios, compared to the single-user case. In the next section, we use Equations (31) and (30) to respectively obtain the source and channel Gallager functions of the various ensembles in Section 3.1 and rank their achievable exponents and transmissibility regions.

4. Discussion

4.1. Gallager Functions for Correlated Sources

In this section, we evaluate the generalized Gallager function $E_\tau^{i_\mathcal{N}}(\rho, \lambda_\mathcal{N}^{\mathrm{L,U}}, r_{\mathcal{N}u_\mathcal{N}}^{\ell_\mathcal{N}})$ of the multi-class cost-constrained ensemble in Equation (31) for the various ensembles described in Section 3.1. In the cases where it is possible, we relate this Gallager function to the well-known [1] correlated-source and channel Gallager functions, respectively given by:

$$E_{\mathrm{s},\sigma}(\rho, P_\mathcal{N}) = \log \sum_{u_{\sigma^c}} \left(\sum_{u_\sigma} P_\mathcal{N}(u_\mathcal{N})^{\frac{1}{1+\rho}} \right)^{1+\rho}, \tag{34}$$

$$E_0(\rho, Q, W) = -\log \sum_{y} \left(\sum_{x} Q(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho}, \tag{35}$$

where $\sigma \in 2^\mathcal{N}$. Using that $[u_\sigma, u_{\sigma^c}] = u_\mathcal{N}$, the standard Gallager source function is given by $E_\mathrm{s}(\rho, P_\mathcal{N}) = E_{\mathrm{s},\mathcal{N}}(\rho, P_\mathcal{N})$, with $\mathcal{N} = \{1, \ldots, N\}$ the set of user indices.

For the simple *mi-iid* ensemble, with only one source class and no cost constraints, $K_\nu = 1$ and $L_\nu = 0$ for all $\nu \in \mathcal{N}$, and $\Lambda_\sigma^{i_\sigma}(u_\sigma) = R_{\sigma,u_\sigma}^{i_\sigma}(x_\sigma) = 1$ for all $\sigma \in 2^\mathcal{N}$. With no statistical dependency between messages and codewords, $Q_{\nu,u_\nu}(x_\nu) = Q_\nu(x_\nu)$. Setting $i_\mathcal{N} = 1$ and $\lambda_\mathcal{N}^{\mathrm{L,U}} = r_{\mathcal{N}u_\mathcal{N}}^{\ell_\mathcal{N}} = 0$ in Equation (31) gives the Gallager function $E_\tau^{mi\text{-}iid}(\rho, P_\mathcal{N}, Q_\mathcal{N}, W)$,

$$E_\tau^{mi-iid}(\rho, P_\mathcal{N}, Q_\mathcal{N}, W) = -\log \sum_{u_{\tau^c}, x_{\tau^c}, y} \left(\sum_{u_\tau, x_\tau} P_\mathcal{N}(u_\mathcal{N})^{\frac{1}{1+\rho}} Q_\tau(x_\tau) \left(Q_{\tau^c}(x_{\tau^c}) W(y|x_\mathcal{N})\right)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{36}$$

Isolating the summations over u_{τ^c} and u_τ, we can split the Gallager function as

$$E_\tau^{mi\text{-}iid}(\rho, P_\mathcal{N}, Q_\mathcal{N}, W) = E_0(\rho, Q_\tau, Q_{\tau^c} W) - E_{\mathrm{s},\tau}(\rho, P_\mathcal{N}), \tag{37}$$

where $Q_{\tau^c} W$ is a shorthand for $Q_{\tau^c}(x_{\tau^c}) W(y|x_\mathcal{N})$, the transition probability of a channel with input x_τ and output (x_{τ^c}, y).

For the *mi-icd* ensemble, we have a similar set-up as for the *mi-iid* ensemble, where $Q_{\nu,u_\nu}(x_\nu)$ now may depend on u_ν. In this case, the Gallager function $E_\tau^{mi\text{-}icd}(\cdot)$ is given by Equation (36) with $Q_\sigma(x_\sigma)$ replaced by $Q_{\sigma,u_\sigma}(x_\sigma)$, for $\sigma \in \{\tau, \tau^c\}$:

$$E_\tau^{mi\text{-}icd}(\rho, P_\mathcal{N}, Q_{\mathcal{N}\mathcal{U}}, W) = -\log \sum_{u_{\tau^c}, x_{\tau^c}, y} \left(\sum_{u_\tau, x_\tau} P_\mathcal{N}(u_\mathcal{N})^{\frac{1}{1+\rho}} Q_{\tau,u_\tau}(x_\tau) \left(Q_{\tau^c,u_{\tau^c}}(x_{\tau^c}) W(y|x_\mathcal{N})\right)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{38}$$

As the summations over u_{τ^c} and u_τ are not independent from the rest, the Gallager function does not split into source and channel functions unless the sources are independent, in

which case one can find an *mi-iid* ensemble with a tilted unconditional input distribution and identical exponent. To this end, and for a given conditional input distribution $Q_{v,u_v}(x_v)$, let us define a tilted distribution $Q_v^\rho(x_v)$ as

$$Q_v^\rho(x_v) = \sum_{u_v} \frac{P_v(u_v)^{\frac{1}{1+\rho}}}{\sum_{\bar{u}_v} P_v(\bar{u}_v)^{\frac{1}{1+\rho}}} Q_{v,u_v}(x_v). \tag{39}$$

From this equation, we have the following equality:

$$Q_v^\rho(x_v) \sum_{\bar{u}_v} P_v(\bar{u}_v)^{\frac{1}{1+\rho}} = \sum_{u_v} P_v(u_v)^{\frac{1}{1+\rho}} Q_{v,u_v}(x_v). \tag{40}$$

Substituting this identity together with $P_{\mathcal{N}}(u_{\mathcal{N}}) = P_\tau(u_\tau)P_{\tau^c}(u_{\tau^c})$ in Equation (38) and rearranging the result, we obtain the following Gallager function for independent sources:

$$E_\tau^{mi\text{-}icd}(\rho, P_{\mathcal{N}}, Q_{\mathcal{N}\mathcal{U}}, W) = E_0(\rho, Q_\tau^\rho, WQ_{\tau^c}) - E_s(\rho, P_\tau) \tag{41}$$

$$= E_\tau^{mi\text{-}iid}(\rho, P_{\mathcal{N}}, [Q_\tau^\rho, Q_{\tau^c}], W). \tag{42}$$

For the md-iid and md-icd ensembles, there are K_v source classes per user and no cost constraints, i.e., $L_v = 0$ and $R^{i_\sigma}_{\sigma,u_\sigma}(x_\sigma) = 1$ for $v \in \mathcal{N}$ and $\sigma \in 2^{\mathcal{N}}$. Settting $r^{\ell_{\mathcal{N}}}_{\mathcal{N}u_{\mathcal{N}}} = 0$ in Equation (31) gives the Gallager function $E^{md\text{-}icd}_{\tau,i_{\mathcal{N}}}(\cdot)$ for generic $i_{\mathcal{N}}$ [31] (Eq. (4.36)),

$$E^{md\text{-}icd}_{\tau,i_{\mathcal{N}}}(\rho, P_{\mathcal{N}}, Q^{i_{\mathcal{N}}}_{\mathcal{N}\mathcal{U}}, W)$$
$$= -\log \sum_{u_{\tau^c}, x_{\tau^c}, y} \left(\sum_{u_\tau, x_\tau} P_{\mathcal{N}}(u_{\mathcal{N}})^{\frac{1}{1+\rho}} \Lambda^{i_{\mathcal{N}}}_{\mathcal{N}}(u_{\mathcal{N}}) Q^{i_\tau}_{\tau,u_\tau}(x_\tau) \left(Q^{i_{\tau^c}}_{\tau^c,u_{\tau^c}}(x_{\tau^c}) W(y|x_{\mathcal{N}}) \right)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{43}$$

The Gallager function $E^{md\text{-}iid}_{\tau,i_{\mathcal{N}}}(\cdot)$ for the *md-iid* ensemble is obtained by setting $Q_{\sigma,u_\sigma}(x_\sigma) = Q_\sigma(x_\sigma)$, independent of u_v, for $\sigma \in \{\tau, \tau^c\}$ in Equation (43). As the summations over u_{τ^c} and u_τ are now independent from the rest, the Gallager function splits as

$$E^{md\text{-}iid}_{\tau,i_{\mathcal{N}}}(\rho, P_{\mathcal{N}}, Q^{i_{\mathcal{N}}}_{\mathcal{N}}, W) = E_0(\rho, Q^{i_\tau}_\tau, Q^{i_{\tau^c}}_{\tau^c} W) - E^{i_{\mathcal{N}}}_{s,\tau}(\rho, P_{\mathcal{N}}), \tag{44}$$

where we defined $E^{i_{\mathcal{N}}}_{s,\tau}(\rho, P_{\mathcal{N}})$, a modified Gallager E_s-function, as

$$E^{i_{\mathcal{N}}}_{s,\tau}(\rho, P_{\mathcal{N}}) = \log \sum_{u_{\tau^c}} \left(\sum_{u_\tau} P_{\mathcal{N}}(u_{\mathcal{N}})^{\frac{1}{1+\rho}} \Lambda^{i_{\mathcal{N}}}_{\mathcal{N}}(u_{\mathcal{N}}) \right)^{1+\rho}. \tag{45}$$

The maximization w. r. t. $\lambda^{\mathrm{L,U}}_{\mathcal{N}}$ in Equation (30) only affects the second term in the r. h. s. of Equation (44), since the function $\Lambda^{i_{\mathcal{N}}}_{\mathcal{N}}$ only appears in the source part of the exponent. In Appendix C, we discuss the properties of Equation (45) after the maximization w. r. t. $\lambda^{\mathrm{L,U}}_{\mathcal{N}}$ as a function of ρ, and establish some connections to the Gallager source function (34) and to the source functions for the single-user md-iid ensemble in ref. [8].

The Gallager functions for the constant-composition ensembles differ from the ones considered so far in the presence of $L_v = |\mathcal{X}_v|$ cost functions $a^{i_v,\ell_v}_{v,u_v}(x_v)$, given in Equation (26), for each input distribution $Q^{i_v}_{v,u_v}(x_v)$. These cost functions appear in the Gallager functions through the factors $R^{i_\sigma}_{\sigma,u_\sigma}(x_\sigma)$, for $\sigma \in \{\tau, \tau^c\}$ that multiply each appearance of $Q^{i_\sigma}_{\sigma,u_\sigma}(x_\sigma)$ in the function, and through their associated optimization parameters $r^{\ell_{\mathcal{N}}}_{\mathcal{N}u_{\mathcal{N}}}$. The expressions of the Gallager functions for these constant-composition ensembles can be easily inferred from this obversation, so we focus on the factor $R^{i_\sigma}_{\sigma,u_\sigma}(x_\sigma)$ itself.

For the *mi-cc* and md-cc ensembles, the cost functions $a_{\nu,u_\nu}^{i_\nu,\ell_\nu}(x_\nu)$, factor $R_{\sigma,u_\sigma}^{i_\sigma}(x_\sigma)$, and associated optimization parameter $r_{\nu u_\nu}^{\ell_\nu}$ are independent of u_ν, we thus write $a_\nu^{i_\nu,\ell_\nu}(x_\nu)$, $R_\sigma^{i_\sigma}(x_\sigma)$, and $r_\nu^{\ell_\nu}$. The expressions in Equations (26) and (33) for $\mathcal{L}_\nu = \mathcal{X}_\nu$ give

$$R_\nu^{i_\nu}(x_\nu) = e^{\sum_{\ell_\nu \in \mathcal{X}_\nu} r_\nu^{\ell_\nu} \left(\mathbf{1}\{x_\nu = \ell_\nu\} - Q_\nu^{i_\nu}(\ell_\nu)\right)}. \tag{46}$$

The exponent in Equation (46) can be evaluated as

$$\sum_{\ell_\nu \in \mathcal{X}_\nu} r_\nu^{\ell_\nu} \left(\mathbf{1}\{x_\nu = \ell_\nu\} - Q_\nu^{i_\nu}(\ell_\nu)\right) = r_\nu^{x_\nu} - \sum_{\ell_\nu \in \mathcal{X}_\nu} r_\nu^{\ell_\nu} Q_\nu^{i_\nu}(\ell_\nu) \tag{47}$$

$$= \alpha_{\tau,\nu}^{i_\nu}(x_\nu), \tag{48}$$

where we have defined a function $\alpha_{\tau,\nu}^{i_\nu}(x_\nu)$ that depends on τ and i_ν through the optimization parameters $r_\nu^{\ell_\nu}$. We can be easily verify that $\alpha_{\tau,\nu}^{i_\nu}$ has zero mean, in other words, $\sum_{x_\nu} \alpha_{\tau,\nu}^{i_\nu}(x_\nu) Q_\nu^{i_\nu}(x_\nu) = 0$. At this point, the parameters $r_\nu^{\ell_\nu}$ may be replaced by the equivalent real-valued functions $\alpha_{\tau,\nu}^{i_\nu}(x_\nu)$. We obtain the *mi-cc* Gallager function $E_\tau^{mi\text{-}cc}(\cdot)$ by setting $i_\mathcal{N} = 1$ and $\lambda_\mathcal{N}^{\mathrm{L,U}} = 0$ in Equation (31),

$$E_\tau^{mi\text{-}cc}(\rho, \alpha_{\tau,\mathcal{N}}, P_\mathcal{N}, Q_\mathcal{N}, W)$$

$$= -\log \sum_{u_{\tau^c}, x_{\tau^c}, y} \left(\sum_{u_\tau, x_\tau} P_\mathcal{N}(u_\mathcal{N})^{\frac{1}{1+\rho}} Q_\tau(x_\tau) e^{\alpha_{\tau,\mathcal{N}}(x_\mathcal{N})} \left(Q_{\tau^c}(x_{\tau^c}) W(y|x_\mathcal{N})\right)^{\frac{1}{1+\rho}} \right)^{1+\rho} \tag{49}$$

$$= -\log \sum_{x_{\tau^c}, y} \left(\sum_{x_\tau} Q_\tau(x_\tau) e^{\alpha_{\tau,\mathcal{N}}(x_\mathcal{N})} \left(Q_{\tau^c}(x_{\tau^c}) W(y|x_\mathcal{N})\right)^{\frac{1}{1+\rho}} \right)^{1+\rho} - E_{\mathrm{s},\tau}(\rho, P_\mathcal{N}), \tag{50}$$

where we split the Gallager function into channel and source terms in analogy to Equation (37).

In ref. [31] (Eq. (4.49)), the *md-cc* ensemble was studied for $N = 2$ users in both the primal and dual domains. The *md-cc* Gallager function $E_\tau^{md\text{-}cc}(\cdot)$ for N users is obtained by combining the derivation of Equation (50) with that of Equation (44) to yield

$$E_\tau^{md\text{-}cc}(\rho, \alpha_{\tau,\mathcal{N}}^{i_\mathcal{N}}, P_\mathcal{N}, Q_\mathcal{N}^{i_\mathcal{N}}, W) = -\log \sum_{x_{\tau^c}, y} \left(\sum_{x_\tau} Q_\tau(x_\tau) e^{\alpha_{\tau,\mathcal{N}}^{i_\mathcal{N}}(x_\mathcal{N})} \left(Q_{\tau^c}(x_{\tau^c}) W(y|x_\mathcal{N})\right)^{\frac{1}{1+\rho}} \right)^{1+\rho}$$

$$- \log \sum_{u_{\tau^c}} \left(\sum_{u_\tau} P_\mathcal{N}(u_\mathcal{N})^{\frac{1}{1+\rho}} \Lambda_\mathcal{N}^{i_\mathcal{N}}(u_\mathcal{N}) \right)^{1+\rho}. \tag{51}$$

As in previous cases, the exponent is obtained after maximization over $\alpha_{\tau,\mathcal{N}}^{i_\mathcal{N}}$.

Concluding our list, the cost functions $a_{\nu,u_\nu}^{i_\nu,\ell_\nu}(x_\nu)$, factors $R_{\sigma,u_\sigma}^{i_\sigma}(x_\sigma)$, and parameters $r_{\nu u_\nu}^{\ell_\nu}$ for the *mi-ccc* and *md-ccc* ensembles do depend on u_ν. In analogy to Equation (48), we define a zero-mean function $\beta_{\tau,\nu,u_\nu}^{i_\nu}(x_\nu)$ as

$$\beta_{\tau,\nu,u_\nu}^{i_\nu}(x_\nu) = r_{\nu u_\nu}^{x_\nu} - \sum_{\ell_\nu \in \mathcal{X}_\nu} r_{\nu u_\nu}^{\ell_\nu} Q_{\nu,u_\nu}^{i_\nu}(\ell_\nu), \tag{52}$$

and similarly for $\beta_{\tau,\nu,u_\nu}(x_\nu)$ for the *mi-ccc* ensemble. The Gallager function for the *mi-ccc* ensemble $E_\tau^{mi\text{-}ccc}(\cdot)$ is obtained by combining the derivations of Equation (50) and of Equation (38),

$$E_\tau^{mi\text{-}ccc}(\rho, \beta_{\tau,\mathcal{N},u_\mathcal{N}}, P_\mathcal{N}, Q_{\mathcal{N},\mathcal{U}}, W)$$

$$= -\log \sum_{u_{\tau^c}, x_{\tau^c}, y} \left(\sum_{u_\tau, x_\tau} P_\mathcal{N}(u_\mathcal{N})^{\frac{1}{1+\rho}} Q_{\tau,u_\tau}(x_\tau) e^{\beta_{\tau,\mathcal{N},u_\mathcal{N}}(x_\mathcal{N})} \left(Q_{\tau^c,u_{\tau^c}}(x_{\tau^c}) W(y|x_\mathcal{N})\right)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{53}$$

Similarly, for the *md-ccc* ensemble, and in agreeement with the 2-user case studied in ref. [31] (Eq. (4.45)), combining the derivations of Equations (50) and (43), yields

$$E_{\tau,i_{\mathcal{N}}}^{md\text{-}ccc}(\rho,\beta_{\tau,\mathcal{N},u_{\mathcal{N}}}^{i_{\mathcal{N}}},P_{\mathcal{N}},Q_{\mathcal{N},\mathcal{U}}^{i_{\mathcal{N}}},W)$$
$$=-\log\sum_{u_{\tau^c},x_{\tau^c},y}\left(\sum_{u_\tau,x_\tau}P_{\mathcal{N}}(u_{\mathcal{N}})^{\frac{1}{1+\rho}}\Lambda_{\mathcal{N}}^{i_{\mathcal{N}}}(u_{\mathcal{N}})Q_{\tau,u_\tau}^{i_\tau}(x_\tau)e^{\beta_{\tau,\mathcal{N},u_{\mathcal{N}}}^{i_{\mathcal{N}}}(x_{\mathcal{N}})}\left(Q_{\tau^c,u_{\tau^c}}^{i_{\tau^c}}(x_{\tau^c})W(y|x_{\mathcal{N}})\right)^{\frac{1}{1+\rho}}\right)^{1+\rho}. \tag{54}$$

4.2. Transmissibility

We may obtain the transmissibility conditions from the achievable exponents derived in Section 4.1, following the random-coding method described in ref. [1] (Th. 5.6.4). The analysis extends the transmissibility condition for joint source-channel coding in ref. [1] (Prob. 5.16), to account for statistical dependency of the codeword on the source message in the multiuser set-up. As mentioned above, the source $U_{\mathcal{N}}$ is transmissible over the channel W if there exists a sequence of codes with vanishing error probability, or equivalently, with strictly positive achievable error exponent E^{cost} in Equation (30). As an example, we present the derivation for the *mi-icd* ensemble where the class and cost functions in Equations (32) and (33) are inactive, namely $\Lambda_\sigma^{i_\sigma}(u_\sigma)=R_{\sigma,u_\sigma}^{i_\sigma}(x_\sigma)=1$ for all $\sigma\in 2^{\mathcal{N}}$, and leave the general case of $K_\nu>1$ classes and cost-constrained codewords as an open problem.

For the *mi-icd* case, and similarly to Gallager's E_0-function [1] (Th. 5.6.3), the Gallager function $E_\tau^{mi\text{-}icd}(\cdot)$ in Equation (38) is concave ($\cap$) with respect to ρ and satisfies $E_\tau^{mi\text{-}icd}(\rho=0,\cdot)=0$. For every $\tau\subset 2^{\mathcal{N}}\setminus\emptyset$, let $\hat{\rho}_\tau$ be the optimizer given by

$$\hat{\rho}_\tau=\arg\max_{0\le\rho\le 1}E_\tau^{mi\text{-}icd}(\rho,P_{\mathcal{N}},Q_{\mathcal{N},\mathcal{U}},W). \tag{55}$$

Therefore, the achievable exponent is strictly positive, namely $E_\tau^{mi\text{-}icd}(\hat{\rho}_\tau,\cdot)>0$, as far as the slope of the $E_\tau^{mi\text{-}icd}(\rho,\cdot)$ function is strictly positive at $\rho=0$, that is

$$\frac{\partial}{\partial\rho}E_\tau^{mi\text{-}icd}(\rho,P_{\mathcal{N}},Q_{\mathcal{N},\mathcal{U}},W)\Big|_{\rho=0}>0. \tag{56}$$

Taking the derivative with respect to ρ at both sides of Equation (38), after some algebraic manipulations, we find that (56) is equivalent to

$$\sum_{u_{\tau^c}}P_{\tau^c}(u_{\tau^c})\sum_{x_{\tau^c},y}\sum_{u_\tau,x_\tau}P_{\tau|\tau^c}(u_\tau|u_{\tau^c})Q_{\tau,u_\tau}(x_\tau)Q_{\tau^c,u_{\tau^c}}(x_{\tau^c})W(y|x_{\mathcal{N}})\times$$
$$\times\log\frac{P_{\tau|\tau^c}(u_\tau|u_{\tau^c})Q_{\tau^c,u_{\tau^c}}(x_{\tau^c})W(y|x_{\mathcal{N}})}{\sum_{\bar{u}_\tau,\bar{x}_\tau}P_{\tau|\tau^c}(\bar{u}_\tau|u_{\tau^c})Q_{\tau,\bar{u}_\tau}(\bar{x}_\tau)Q_{\tau^c,u_{\tau^c}}(x_{\tau^c})W(y|[\bar{x}_\tau,x_{\tau^c}])}>0. \tag{57}$$

We next write the expression in the left hand-side of the inequality (57) in terms of entropy and mutual information. We denote as $H(P)$ the entropy of a source with distribution P [32] (Eq. (2.1)) and by $I(Q,W)$ the mutual information of a channel W with input distribution Q [32] (Eq. (2.28)). For $\sigma\subset 2^{\mathcal{N}}$, we define a channel input distribution $Q_{\tau|\sigma}$, that is conditioned to the source messages u_σ, as

$$Q_{\tau|\sigma}(x_\tau|u_\sigma)=\sum_{u_\tau\in\mathcal{U}_\ll}P_{\tau|\sigma}(u_\tau|u_\sigma)Q_{\tau,u_\tau}(x_\tau). \tag{58}$$

Therefore, the transmissibility condition (57) can be compactly expressed as

$$H\left(P_{\tau|\tau^c}\right)<I\left(Q_{\tau|\tau^c},W|P_{\tau^c}Q_{\tau^c|\tau^c}\right),\quad \tau\subset 2^{\mathcal{N}}\setminus\emptyset. \tag{59}$$

As it is, $Q_{\tau^c|\tau^c}$ is "transparent", as it cancels inside the fraction, and the channel law may also be written as $Q_{\tau^c|\tau^c}W$, removing the conditioning in the mutual information. With $\mathcal{N}=\{1,2\}$ in Equation (59), we recover the achievable Cover-El Gamal-Salehi region [11] (Eq. (3)).

4.3. Numerical Examples

In this section, we present two simple examples showing that the exponent of the md-iid ensemble can be larger than that of the *mi-iid* ensemble with only two classes (and associated input distributions) for each user. First, we consider two correlated discrete memoryless sources, $N = 2$ and $\mathcal{N} = \{1, 2\}$, with alphabet $\mathcal{U}_\nu = \{0, 1\}$ for both users $\nu \in \mathcal{N}$, and probability distribution $P_{\mathcal{N}}(u_1, u_2)$ given in matrix form as

$$P_{\mathcal{N}} = \begin{pmatrix} 0.0005 & 0.0095 \\ 0.0005 & 0.9895 \end{pmatrix}. \tag{60}$$

The sources are sent over a discrete memoryless multiple-access channel with input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \{1, 2, 3, 4, 5, 6\}$ and output alphabet $\mathcal{Y} = \{1, 2, 3, 4\}$. The channel transition probabilites are given by a 36×4 matrix W, such that $W(y|x_1, x_2)$ is the row $x_1 + 6(x_2 - 1)$. The transition matrix W is given by

$$W = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \\ W_5 \\ W_6 \end{pmatrix}, \tag{61}$$

where the 6×4 submatrices W_ℓ, $\ell = 1, \ldots, 6$ are given as follows. First, the submatrix W_1 corresponds to the point-to-point channel discussed in ref. [8] (Sec. IV.C), given by

$$W_1 = \begin{pmatrix} 1 - 3k_1 & k_1 & k_1 & k_1 \\ k_1 & 1 - 3k_1 & k_1 & k_1 \\ k_1 & k_1 & 1 - 3k_1 & k_1 \\ k_1 & k_1 & k_1 & 1 - 3k_1 \\ 0.5 - k_2 & 0.5 - k_2 & k_2 & k_2 \\ k_2 & k_2 & 0.5 - k_2 & 0.5 - k_2 \end{pmatrix}, \tag{62}$$

for $k_1 = 0.045$ and $k_2 = 0.01$. Let the m-th row of matrix W_1 is denoted by $W_1(m)$. The matrix W_2 (resp. W_3) is a 6×4 matrix whose rows are all $W_1(5)$ (resp. $W_1(6)$). The matrices W_4, W_5 and W_6 are respectively given by

$$W_4 = \begin{pmatrix} W_1(2) \\ W_1(3) \\ W_1(4) \\ W_1(1) \\ W_1(6) \\ W_1(5) \end{pmatrix}, \quad W_5 = \begin{pmatrix} W_1(3) \\ W_1(4) \\ W_1(1) \\ W_1(2) \\ W_1(5) \\ W_1(6) \end{pmatrix}, \quad W_6 = \begin{pmatrix} W_1(4) \\ W_1(1) \\ W_1(2) \\ W_1(3) \\ W_1(6) \\ W_1(5) \end{pmatrix}. \tag{63}$$

The optimal achievable exponent [8] (Sec. IV.C) for the single-user channel W_1 in Equation (62) is related to two different distributions $Q^\star$ and $Q^\dagger$, given in vector form by

$$Q^\star = (0, 0, 0, 0, 1/2, 1/2), \tag{64}$$

$$Q^\dagger = (1/4, 1/4, 1/4, 1/4, 0, 0). \tag{65}$$

We let each user employ these distributions in the md-iid ensemble with input distribution in Equation (13) according to the source message partitioning in Equation (12) with $K_\nu = 2$ classes per user and thresholds $\gamma_{\mathcal{N}} = (\gamma_1, \gamma_2)$. Since we consider two input distributions for each user, the channel Gallager function $\max_{\rho \in [0,1]} E_0(\rho, Q_\tau^{i_\tau}, WQ_{\tau^c}^{i_{\tau^c}})$ is not concave in ρ [8]. To find the md-iid exponent $E^{md\text{-}iid}$, we optimize over the class thresholds following the method in Appendix B with the Gallager function in Equation (44), exploit the properties of the source function in Equation (45) in Appendix C, and also find the optimal input

distribution assignment of $Q_\nu^{i_\nu}$ for each $\nu \in \{1,2\}$. In our setting, we have four possible assignments, namely

$$\Omega_1 : Q_1^1 = Q_2^1 = Q^\star,\ Q_1^2 = Q_2^2 = Q^\dagger, \tag{66}$$
$$\Omega_2 : Q_1^1 = Q_2^2 = Q^\star,\ Q_1^2 = Q_2^1 = Q^\dagger, \tag{67}$$
$$\Omega_3 : Q_1^2 = Q_2^1 = Q^\star,\ Q_1^1 = Q_2^2 = Q^\dagger, \tag{68}$$
$$\Omega_4 : Q_1^2 = Q_2^2 = Q^\star,\ Q_1^1 = Q_2^1 = Q^\dagger. \tag{69}$$

We start our numerical discussion by assessing which of the possible four assignments in Equations (66)–(69) leads to a higher error exponent. For each possible pair of thresholds (γ_1, γ_2), we numerically calculate the optimal assignment $\Omega^\star(\gamma_\mathcal{N})$ given by

$$\Omega^\star(\gamma_\mathcal{N}) = \arg\max_{\Omega_j} \min_{i_\mathcal{N}} \min_{\tau} E_\tau^{i_\mathcal{N}}(\gamma_\mathcal{N}), \tag{70}$$

and the corresponding achievable error exponent $E^{md\text{-}iid}(\gamma_\mathcal{N})$ as

$$E^{cost}(\gamma_\mathcal{N}) = \max_{\Omega_j} \min_{i_\mathcal{N}} \min_{\tau} E_\tau^{i_\mathcal{N}}(\gamma_\mathcal{N}), \tag{71}$$

where the exponent function $E_\tau^{i_\mathcal{N}}(\gamma_\mathcal{N})$ is given in Equation (A55). Figures 2 and 3 respectively show $\Omega^\star(\gamma_\mathcal{N})$ and $E^{cost}(\gamma_\mathcal{N})$ for the valid range of $\gamma_\mathcal{N}$. For most pair of thresholds (γ_1, γ_2), assignments Ω_1 and Ω_3 lead to the highest exponent among the possible assignments, while assignments Ω_2 and Ω_4 are optimal only for a marginal region. Using this information, and combined with the values of the achievable exponents in Figure 3, we determine the message-dependent exponent

$$E^{md\text{-}iid} = \max_{\gamma_\mathcal{N}} E^{cost}(\gamma_\mathcal{N}). \tag{72}$$

In this example, we obtained the achievable exponent $E^{md\text{-}iid} = 0.2611$, corresponding to the input distribution assignment Ω_1 in Equation (66) and optimal source message partitioning $\gamma_1^\star = 0.8469$ and $\gamma_2^\star = 0.6581$. The optimal point $\gamma_\mathcal{N}^\star$ is shown by a white (black) bullet in Figure 2 (Figure 3).

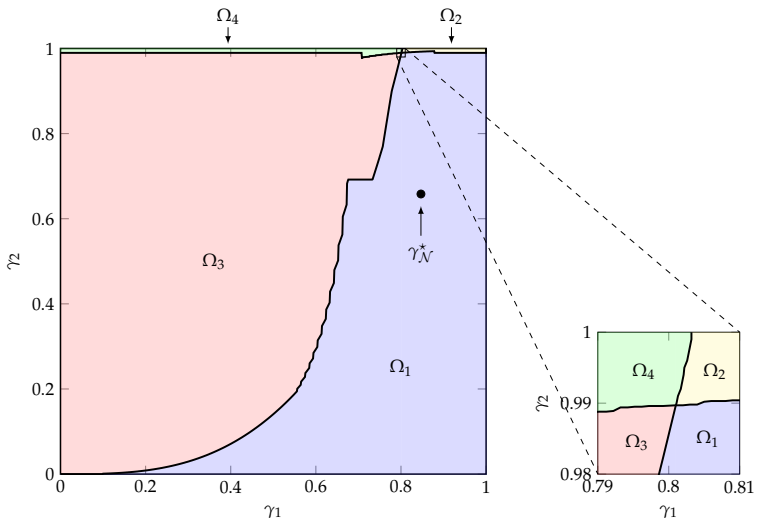

Figure 2. Correlated-sources optimal assignment $\Omega^\star(\gamma_\mathcal{N})$ in Equation (70) for all pairs of thresholds (γ_1, γ_2).

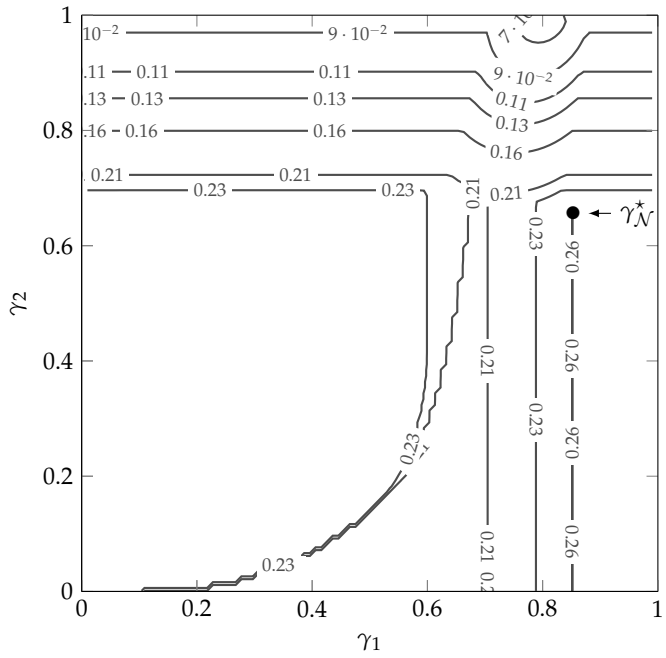

Figure 3. Correlated-sources error exponent $E^{cost}(\gamma_{\mathcal{N}})$ in Equation (71) for all pairs of thresholds (γ_1, γ_2).

Alternatively, we may first optimize over $\gamma_{\mathcal{N}}$ and then over the assignments Ω_j. To do so, we solve the system of Equation (A58) in Appendix B to numerically determine the optimal thresholds $\gamma^\star_{\mathcal{N}}$, and compute the exponent $E^{cost}(\Omega_j)$ as

$$E^{cost}(\Omega_j) = \min_{i_{\mathcal{N}}} \min_{\tau} E^{i_{\mathcal{N}}}_{\tau}(\gamma^\star_{\mathcal{N}}), \tag{73}$$

where the exponent function $E^{i_{\mathcal{N}}}_{\tau}(\gamma_{\mathcal{N}})$ is given in Equation (A55). We provide in Table 1 the values of the optimal thresholds $\gamma^\star_{\mathcal{N}}$ and exponents $E^{i_{\mathcal{N}}}_{\tau}(\gamma^\star_{\mathcal{N}})$ under the different assignment Ω_j, for the three types of error τ and the four possible user classes $i_{\mathcal{N}}$. For each assignment, the minimum over $i_{\mathcal{N}}$ and τ as in Equation (73) is highlighted in gray, leading to the exponent $E^{cost}(\Omega_j)$. The message-dependent exponent is then

$$E^{md\text{-}iid} = \max_{j} E^{cost}(\Omega_j), \tag{74}$$

recovering the error exponent $E^{md\text{-}iid} = 0.2611$ for input distribution assignment Ω_1 obtained using the previous method in Equation (71).

In the second example, we consider the transmission of two independent discrete memoryless sources with identical source alphabets $\mathcal{U}_\nu = \{0, 1\}$ with distributions induced by the marginals of Equation (60), given by $P_1(0) = 0.01$ and $P_2(0) = 0.001$. These sources are transmitted over the multiple-access channel with transition probability given by Equation (61), and are encoded using the md-iid ensemble with the input distribution assignments Ω_j in Equations (66)–(69). Following the footsteps of the correlated sources case, in Table 2 we calculate optimal thresholds $\gamma^\star_{\mathcal{N}}$ and exponents $E^{i_{\mathcal{N}}}_{\tau}(\gamma^\star_{\mathcal{N}})$ for the possible input distribution assignments and determine the exponent of the md-iid ensemble using Equations (73) and (74). In this case, the optimal assignment is again Ω_1, with optimal source message partitioning specified by the thresholds $\gamma^\star_1 = 0.8779$ and $\gamma^\star_2 = 0.6933$, achieving an exponent of $E^{md\text{-}iid} = 0.2458$, slightly smaller than that of correlated sources.

Table 1. Correlated-sources optimal thresholds $\gamma^\star_{\mathcal{N}}$ and exponents $E_\tau^{i_{\mathcal{N}}}(\gamma^\star_{\mathcal{N}})$ in Equation (73) for assignments Ω_j in Equations (66)–(69). For each assignment, the minimum over $i_{\mathcal{N}}$ and τ is highlighted in gray.

	Assignment Ω_1				**Assignment Ω_2**			
	$\gamma_1^\star = 0.8469$ $\gamma_2^\star = 0.6581$				$\gamma_1^\star = 1$ $\gamma_2^\star = 1$			
(i_1, i_2)	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$
$\tau = \{1\}$	0.3131	0.2735	0.3120	0.2611	0.0642	0.3268	0.1005	0.3604
$\tau = \{2\}$	0.3986	0.4369	0.2611	0.4119	0.3959	0.3986	0.4323	0.3110
$\tau = \{1,2\}$	0.2611	0.2972	0.2630	0.2883	0.2108	0.2108	0.2360	0.2637
	Assignment Ω_3				**Assignment Ω_4**			
	$\gamma_1^\star = 0.5605$ $\gamma_2^\star = 0.6709$				$\gamma_1^\star = 0.6985$ $\gamma_2^\star = 0.9033$			
(i_1, i_2)	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$
$\tau = \{1\}$	0.3120	0.2503	0.2763	0.2897	0.0879	0.3605	0.0879	0.3112
$\tau = \{2\}$	0.2503	0.3898	0.5675	0.5731	0.3664	0.2503	0.4720	0.4684
$\tau = \{1,2\}$	0.2630	0.2816	0.2503	0.3012	0.2360	0.2632	0.2097	0.2097

Table 2. Independent-sources *md-iid* optimal thresholds $\gamma^\star_{\mathcal{N}}$ and exponents $E_\tau^{i_{\mathcal{N}}}(\gamma^\star_{\mathcal{N}})$ in Equation (73) for assignments Ω_j in Equations (66)–(69). For each assignment, the minimum over $i_{\mathcal{N}}$ and τ is highlighted in gray.

	Assignment Ω_1				**Assignment Ω_2**			
	$\gamma_1^\star = 0.8779$ $\gamma_2^\star = 0.6933$				$\gamma_1^\star = 0.8776$ $\gamma_2^\star = 1$			
(i_1, i_2)	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$
$\tau = \{1\}$	0.3343	0.2458	0.3089	0.2458	0.0913	0.3341	0.0913	0.3089
$\tau = \{2\}$	0.3850	0.3987	0.2458	0.3788	0.4555	0.3850	0.4357	0.2459
$\tau = \{1,2\}$	0.2730	0.2870	0.2685	0.2863	0.3430	0.2728	0.2956	0.2685
	Assignment Ω_3				**Assignment Ω_4**			
	$\gamma_1^\star = 0.61$ $\gamma_2^\star = 0.7043$				$\gamma_1^\star = 0.7092$ $\gamma_2^\star = 1$			
(i_1, i_2)	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$
$\tau = \{1\}$	0.3089	0.2367	0.2681	0.3078	0.0913	0.3117	0.0913	0.2648
$\tau = \{2\}$	0.2367	0.3672	0.5425	0.5538	0.4269	0.2367	0.5393	0.4683
$\tau = \{1,2\}$	0.2685	0.2811	0.2367	0.3133	0.3006	0.2685	0.2740	0.2164

For the sake of completeness and purpose of comparison, we also calculate the exponent for the *mi-iid* ensemble described in Equation (8). In the absence of message dependence, for a given assignment Ω_j, the mi-iid exponent is given by

$$E^{no\text{-}cost}(\Omega_j) = \min_\tau E_\tau, \tag{75}$$

where the exponent function E_τ is given by $E_\tau = \max_\rho E_\tau^{mi\text{-}iid}(\rho, P_{\mathcal{N}}, Q_{\mathcal{N}}, W)$ and $E_\tau^{mi\text{-}iid}$ is the Gallager function in Equation (37), described in the previous subsection. For both the correlated and independent sources described above, Table 3 presents the achievable exponents E_τ for each type of error τ and input distribution assignment (Q_1, Q_2), where Q_1 and Q_2 are either of $Q^\star$ and $Q^\dagger$ in Equations (64) and (65). In our numerical example for correlated sources, the assignment with highest exponent is $(Q_1, Q_2) = (Q^\dagger, Q^\star)$, giving an exponent of $E^{mi\text{-}iid} = 0.2503$, slightly smaller than that of the md-iid ensemble. In contrast, the *mi-iid* exponent for independent sources, according to the second part of Table 3 is found to be $E^{mi\text{-}iid} = 0.2367$ with input distribution $(Q_1, Q_2) = (Q^\star, Q^\dagger)$. In this case, the md-iid exponent $E^{md\text{-}iid}$ is around 4% larger that the *mi-iid*; this situation is in contrast with to-point communication, where the gain in exponent achieved by an ensemble with two distributions is typically smaller, for example, 1% in ref. [8]. Hence, message-dependent

random coding with two class distributions, compared to iid random coding, may lead to a higher error exponent gain in the MAC than in point-to-point communication.

Table 3. Mi-iid exponents E_τ in Equation (75) for two correlated and two independent sources vs several input distribution assigments (Q_1, Q_2). For each assignment, the minimum over τ is highlighted in gray.

		Correlated Sources		
(Q_1, Q_2)	$(Q^\star, Q^\star)$	$(Q^\star, Q^\dagger)$	$(Q^\dagger, Q^\star)$	$(Q^\dagger, Q^\dagger)$
$\tau = \{1\}$	0.2682	0.0642	0.3120	0.0879
$\tau = \{2\}$	0.3986	0.3986	0.2503	0.3696
$\tau = \{1,2\}$	0.2097	0.2097	0.2630	0.2360
		Independent Sources		
(Q_1, Q_2)	$(Q^\star, Q^\star)$	$(Q^\star, Q^\dagger)$	$(Q^\dagger, Q^\star)$	$(Q^\dagger, Q^\dagger)$
$\tau = \{1\}$	0.2648	0.3089	0.0627	0.0865
$\tau = \{2\}$	0.3850	0.2367	0.3850	0.3559
$\tau = \{1,2\}$	0.2164	0.2685	0.2164	0.2421

4.4. Comparison of the Random-Coding Achievable Error Exponents

From the numerical results presented in Section 4.3, as well as from refs. [8,20,28,31], the message-dependent ensembles attain in general a larger exponent than their message-independent counterparts. We now compare the random-coding exponents for the ensembles presented in Section 3.1, whose Gallager functions were obtained in Section 4.1.

For independent sources, we found in Equation (42) that for a given conditional input distribution $Q_{v,u_v}(x_v)$ and ρ, there exists an iiid distribution $Q_{v,\rho}$ given by Equation (39) with identical Gallager function. Thus, the *mi-iid* and *mi-icd* ensembles attains the same exponent, after maximization over the input distributions. Similarly, we conclude that md-iid and md-icd-ensembles attain the same exponent.

In ref. [31] (Prop. 2.9), it was proved that for point-to-point communication, the exponent of the *mi-ccc* ensemble may be lower than that of the *mi-cc* ensemble. The same steps actually prove the same result for the MAC with independent sources. Thus, for the MAC with independent sources we have

$$\mathrm{E}^{\textit{mi-ccc}} \leq \mathrm{E}^{\textit{mi-cc}} \leq \mathrm{E}^{\textit{md-cc}}, \qquad \mathrm{E}^{\textit{md-ccc}} \leq \mathrm{E}^{\textit{md-cc}}, \tag{76}$$

$$\mathrm{E}^{\textit{mi-iid}} \leq \mathrm{E}^{\textit{mi-cc}} \leq \mathrm{E}^{\textit{md-cc}}, \qquad \mathrm{E}^{\textit{md-iid}} \leq \mathrm{E}^{\textit{md-cc}}, \tag{77}$$

and $\mathrm{E}^{\textit{md-cc}}$ is thus largest among the ensembles in Section 3.1 for an arbitrary input distribution. As discussed in ref. [29] (Th. 4), for optimal input distributions both $\mathrm{E}^{\textit{md-cc}}$ and $\mathrm{E}^{\textit{md-iid}}$ may coincide.

Concerning the optimal partitioning into message classes, for point-to-point communication it is known that partitioning the source-message set into two classes is sufficient to attain the optimal error exponent [8,31] (Prop. 2.7). However, the proof of ref. [31] (Prop. 2.7) cannot be easily generalized to the MAC with independent sources. At the same time, we could not find an example showing that assigning more than two input distributions leads to a larger exponent. Hence, finding the sufficient number of input distributions is for the message-dependent exponent is an open problem.

The comparisons in Equations (76) and (77) for correlated sources require, in general, a more sophisticated machinery and we consider here two simple cases. For the message-dependent md-icd and md-ccc ensembles, we observe that compared to $\mathrm{E}^{\textit{md-icd}}_{\tau,i_\mathcal{N}}$ in Equation (43) the $\mathrm{E}^{\textit{md-ccc}}_{\tau,i_\mathcal{N}}$ exponent in Equation (54) contains an additional term $\beta^{i_\mathcal{N}}_{\tau,\mathcal{N},u_\mathcal{N}}(x_\mathcal{N})$ to guarantee the constant-composition distribution as in Equation (52). This allows to recover $\mathrm{E}^{\textit{md-ccc}}_{\tau,i_\mathcal{N}}$ by setting $\beta^{i_\mathcal{N}}_{\tau,\mathcal{N},u_\mathcal{N}}(x_\mathcal{N}) = 0$ in $\mathrm{E}^{\textit{md-icd}}_{\tau,i_\mathcal{N}}$ and to prove that $\mathrm{E}^{\textit{md-icd}} \leq \mathrm{E}^{\textit{md-ccc}}$ after maximizing w. r. t. $\beta^{i_\mathcal{N}}_{\tau,\mathcal{N},u_\mathcal{N}}(x_\mathcal{N})$. Similarly for the ensembles with statistical independence between messages and codewords, we observe that the constant-composition exponent

$\mathrm{E}^{md\text{-}cc}_{\tau,i_{\mathcal{N}}}$ in Equation (51) also contains the additional term $\alpha^{i_{\mathcal{N}}}_{\tau,\mathcal{N}}(x_{\mathcal{N}})$ compared to its iid counterpart $\mathrm{E}^{md\text{-}iid}_{\tau,i_{\mathcal{N}}}$ in Equation (44), yielding $\mathrm{E}^{md\text{-}iid} \leq \mathrm{E}^{md\text{-}cc}$. Put together, for correlated sources it holds that

$$\mathrm{E}^{md\text{-}icd} \leq \mathrm{E}^{md\text{-}ccc}, \qquad \mathrm{E}^{md\text{-}iid} \leq \mathrm{E}^{md\text{-}cc}, \tag{78}$$

suggesting that, as in the case of single-user communication, the use of constant-composition input distributions may lead to higher exponents than the symbol-wise independent distributions when transmitting correlated sources over the MAC.

Summarizing, proper choices of the cost functions recover the different coding schemes considered in Section 3.1, including message-dependent and message-independent versions of iid, independent conditionally distributed, constant-composition, and conditional constant composition ensembles. Thanks to the flexibility of the generalized cost-constraint random-coding ensemble, the achievable exponents of the various ensembles can be compared and ranked, both numerically and analytically.

Author Contributions: Conceptualization, A.M. and A.G.i.F.; methodology J.F.-S., software, A.R.; formal analysis, A.R.; witting A.R., J.F.-S. and A.M. All authors have equally contributed to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been funded in part by the European Research Council under grant 725411, and by the Spanish Ministry of Economy and Competitiveness under grant TEC2016-78434-C3-1-R.

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Theorem 1

We start by bounding the average error probability over the generalized cost-constraint ensemble, $\bar{P}_{\mathrm{e}}$. Counting ties as errors, the random coding union bound [2] (Th. 16) for joint source-channel is

$$\bar{P}_{\mathrm{e}} \leq \sum_{\boldsymbol{u}_{\mathcal{N}},\boldsymbol{x}_{\mathcal{N}},\boldsymbol{y}} \mathrm{P}_{\mathcal{N}}(\boldsymbol{u}_{\mathcal{N}})\mathrm{Q}_{\mathcal{N}}(\boldsymbol{x}_{\mathcal{N}}|\boldsymbol{u}_{\mathcal{N}})\mathrm{W}(\boldsymbol{y}|\boldsymbol{x}_{\mathcal{N}}) \min\left\{1, \sum_{\hat{\boldsymbol{u}}_{\mathcal{N}}\neq\boldsymbol{u}_{\mathcal{N}}} \Pr\left\{\frac{\mathrm{P}_{\mathcal{N}}(\hat{\boldsymbol{u}}_{\mathcal{N}})\mathrm{W}(\boldsymbol{y}|\hat{\boldsymbol{X}}_{\mathcal{N}})}{\mathrm{P}_{\mathcal{N}}(\boldsymbol{u}_{\mathcal{N}})\mathrm{W}(\boldsymbol{y}|\boldsymbol{x}_{\mathcal{N}})} \geq 1\right\}\right\}, \tag{A1}$$

where $\mathrm{Q}_{\mathcal{N}}(\boldsymbol{x}_{\mathcal{N}}|\boldsymbol{u}_{\mathcal{N}})$ is given by Equation (7), with every user using the generalized cost-constrained input distribution $\mathrm{Q}_{\nu}(\boldsymbol{x}_{\nu}|\boldsymbol{u}_{\nu})$ as in Equation (24), and $\hat{\boldsymbol{x}}_{\mathcal{N}}$ has the same distribution as $\boldsymbol{x}_{\mathcal{N}}$ but conditioned on $\hat{\boldsymbol{u}}_{\mathcal{N}}$ rather than $\boldsymbol{u}_{\mathcal{N}}$, i.e., $\mathrm{Q}_{\mathcal{N}}(\boldsymbol{x}_{\mathcal{N}}|\hat{\boldsymbol{u}}_{\mathcal{N}})$. The summation over $\hat{\boldsymbol{u}}_{\mathcal{N}} \neq \boldsymbol{u}_{\mathcal{N}}$ can be split into $2^N - 1$ distinct types of error events indexed by the non-empty subsets in the power set of the user indices $2^{\mathcal{N}} \setminus \emptyset$, e.g., $\tau \in \{\{1\},\{2\},\{1,2\}\}$ for $N = 2$, such that $\hat{\boldsymbol{u}}_{\tau^c} = \boldsymbol{u}_{\tau^c}$ and $\hat{\boldsymbol{u}}_{\nu} \neq \boldsymbol{u}_{\nu}$ for all $\nu \in \tau$.

Since $\min\{1, a+b\} \leq \min\{1,a\} + \min\{1,b\}$, we bound $\bar{P}_{\mathrm{e}}$ as

$$\bar{P}_{\mathrm{e}} \leq \sum_{\tau\in 2^{\mathcal{N}}\setminus\emptyset} \bar{P}^{\tau}_{\mathrm{e}}, \tag{A2}$$

where $\bar{P}^{\tau}_{\mathrm{e}}$ is in turn given by

$$\bar{P}^{\tau}_{\mathrm{e}} = \sum_{\boldsymbol{u}_{\mathcal{N}},\boldsymbol{x}_{\mathcal{N}},\boldsymbol{y}} \mathrm{P}_{\mathcal{N}}(\boldsymbol{u}_{\mathcal{N}})\mathrm{Q}_{\mathcal{N}}(\boldsymbol{x}_{\mathcal{N}}|\boldsymbol{u}_{\mathcal{N}})\mathrm{W}(\boldsymbol{y}|\boldsymbol{x}_{\mathcal{N}}) \min\left\{1, \sum_{\substack{\hat{\boldsymbol{u}}_{\mathcal{N}}:\hat{\boldsymbol{u}}_{\tau^c}=\boldsymbol{u}_{\tau^c},\\ \hat{\boldsymbol{u}}_{\nu}\neq\boldsymbol{u}_{\nu},\nu\in\tau}} \Pr\left\{\frac{\mathrm{P}_{\mathcal{N}}(\hat{\boldsymbol{u}}_{\mathcal{N}})\mathrm{W}(\boldsymbol{y}|[\boldsymbol{x}_{\tau^c},\hat{\boldsymbol{X}}_{\tau}])}{\mathrm{P}_{\mathcal{N}}(\boldsymbol{u}_{\mathcal{N}})\mathrm{W}(\boldsymbol{y}|\boldsymbol{x}_{\mathcal{N}})} \geq 1\right\}\right\}, \tag{A3}$$

where the inner probability is computed according to the distribution $\mathrm{Q}_{\tau}(\boldsymbol{x}_{\tau}|\boldsymbol{u}_{\tau})$, including only users $\boldsymbol{u}_{\tau}$ in the set τ as $\hat{\boldsymbol{x}}_{\tau^c} = \boldsymbol{x}_{\tau^c}$. We recall that $[\boldsymbol{x}_{\tau^c}, \hat{\boldsymbol{X}}_{\tau}]$ is the sorted merger of the channel inputs for users in the sets τ^c and τ, in this case $\boldsymbol{x}_{\tau^c}$ and $\hat{\boldsymbol{X}}_{\tau}$ respectively.

Next, we split the summation over $\boldsymbol{u}_{\mathcal{N}}$ in Equation (A3) into classes $i_{\mathcal{N}} \in \mathcal{K}_{\mathcal{N}}$ defined by Equation (12), summing then over the messages belonging to the Cartesian product of the sets $\mathcal{A}_{\mathcal{N}}^{i_{\mathcal{N}}}$. We note that codewords are generated according to distributions that depend on the class index of the sources. Let $\mathcal{D}_{\mathcal{N},\boldsymbol{u}_{\mathcal{N}}}^{i_{\mathcal{N}}}$ be the Cartesian product of the sets of codewords $\mathcal{D}_{\nu,\boldsymbol{u}_\nu}^{i_\nu}$ in Equation (23) for $\nu = 1, 2, \ldots, N$, and define

$$\mathrm{Q}_{\mathcal{N}}^{i_{\mathcal{N}}}(\boldsymbol{x}_{\mathcal{N}}|\boldsymbol{u}_{\mathcal{N}}) = \prod_{\nu\in\mathcal{N}} \mathrm{Q}_{\nu}^{i_\nu}(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu), \tag{A4}$$

where $\mathrm{Q}_{\nu}^{i_\nu}(\boldsymbol{x}_\nu|\boldsymbol{u}_\nu)$ is given by either Equation (24) or Equation (25). Then, the double outer summation of Equation (A3) over $\boldsymbol{u}_{\mathcal{N}}$ and $\boldsymbol{x}_{\mathcal{N}}$ can be written as

$$\sum_{\boldsymbol{u}_{\mathcal{N}},\boldsymbol{x}_{\mathcal{N}}} \mathrm{P}_{\mathcal{N}}(\boldsymbol{u}_{\mathcal{N}})\mathrm{Q}_{\mathcal{N}}(\boldsymbol{x}_{\mathcal{N}}|\boldsymbol{u}_{\mathcal{N}}) = \sum_{i_{\mathcal{N}}\in\mathcal{K}_{\mathcal{N}}} \sum_{\substack{\boldsymbol{u}_{\mathcal{N}}\in\mathcal{A}_{\mathcal{N}}^{i_{\mathcal{N}}} \\ \boldsymbol{x}_{\mathcal{N}}\in\mathcal{D}_{\mathcal{N},\boldsymbol{u}_{\mathcal{N}}}^{i_{\mathcal{N}}}}} \mathrm{P}_{\mathcal{N}}(\boldsymbol{u}_{\mathcal{N}})\mathrm{Q}_{\mathcal{N}}^{i_{\mathcal{N}}}(\boldsymbol{x}_{\mathcal{N}}|\boldsymbol{u}_{\mathcal{N}}) \tag{A5}$$

$$= \sum_{i_{\tau^c}\in\mathcal{K}_{\tau^c}} \sum_{\substack{\boldsymbol{u}_{\tau^c}\in\mathcal{A}_{\tau^c}^{i_{\tau^c}} \\ \boldsymbol{x}_{\tau^c}\in\mathcal{D}_{\tau^c,\boldsymbol{u}_{\tau^c}}^{i_{\tau^c}}}} \mathrm{P}_{\tau^c}(\boldsymbol{u}_{\tau^c})\mathrm{Q}_{\tau^c}^{i_{\tau^c}}(\boldsymbol{x}_{\tau^c}|\boldsymbol{u}_{\tau^c}) \sum_{i_\tau\in\mathcal{K}_\tau} \sum_{\substack{\boldsymbol{u}_\tau\in\mathcal{A}_\tau^{i_\tau} \\ \boldsymbol{x}_\tau\in\mathcal{D}_{\tau,\boldsymbol{u}_\tau}^{i_\tau}}} \mathrm{P}_{\tau|\tau^c}(\boldsymbol{u}_\tau|\boldsymbol{u}_{\tau^c})\mathrm{Q}_{\tau}^{i_\tau}(\boldsymbol{x}_\tau|\boldsymbol{u}_\tau), \tag{A6}$$

where we split the summations over $\boldsymbol{u}_{\mathcal{N}}$ and $\boldsymbol{x}_{\mathcal{N}}$ into separate summations over $\boldsymbol{u}_{\tau^c}$ and $\boldsymbol{u}_\tau$, similarly with $\boldsymbol{x}_{\tau^c}$ and $\boldsymbol{x}_\tau$ with the corresponding rearrangements in the probabilities, and written the term $\mathrm{Q}_{\tau^c}^{i_{\tau^c}}(\boldsymbol{x}_{\tau^c}|\boldsymbol{u}_{\tau^c})$ in a similar way to Equation (A4). The inner summation of Equation (A3) can be split in an analogous manner based on the classes to which $\hat{\boldsymbol{u}}_\tau$ belongs, now indexed by the variable $j_\tau \in \mathcal{K}_\tau$. Applying this fact together with Markov's inequality

$$\Pr\{A \geq 1\} \leq \min_{s\geq 0} \mathrm{E}[A^s] \tag{A7}$$

to upper bound the probability with a parameter $s \geq 0$ that implicitly depends on the error-event type τ and indices i_{τ^c}, i_τ, and j_τ. We bound the inner summation of Equation (A3) as

$$\sum_{\substack{\hat{\boldsymbol{u}}_{\mathcal{N}}:\hat{\boldsymbol{u}}_{\tau^c}=\boldsymbol{u}_{\tau^c}, \\ \hat{\boldsymbol{u}}_\nu\neq\boldsymbol{u}_\nu,\nu\in\tau}} \Pr\left\{\frac{\mathrm{P}_{\mathcal{N}}(\hat{\boldsymbol{u}}_{\mathcal{N}})\mathrm{W}(\boldsymbol{y}|[\boldsymbol{x}_{\tau^c},\hat{\boldsymbol{X}}_\tau])}{\mathrm{P}_{\mathcal{N}}(\boldsymbol{u}_{\mathcal{N}})\mathrm{W}(\boldsymbol{y}|\boldsymbol{x}_{\mathcal{N}})} \geq 1\right\} \leq$$

$$\leq \sum_{j_\tau\in\mathcal{K}_\tau} \min_{s\geq 0} \left\{ \sum_{\substack{\hat{\boldsymbol{u}}_\tau\in\mathcal{A}_\tau^{j_\tau} \\ \hat{\boldsymbol{x}}_\tau\in\mathcal{D}_{\tau,\hat{\boldsymbol{u}}_\tau}^{j_\tau}}} \mathrm{Q}_{\tau}^{j_\tau}(\hat{\boldsymbol{x}}_\tau|\hat{\boldsymbol{u}}_\tau) \left(\frac{\mathrm{P}_{\tau|\tau^c}(\hat{\boldsymbol{u}}_\tau|\boldsymbol{u}_{\tau^c})\mathrm{W}(\boldsymbol{y}|[\boldsymbol{x}_{\tau^c},\hat{\boldsymbol{x}}_\tau])}{\mathrm{P}_{\tau|\tau^c}(\boldsymbol{u}_\tau|\boldsymbol{u}_{\tau^c})\mathrm{W}(\boldsymbol{y}|[\boldsymbol{x}_{\tau^c},\boldsymbol{x}_\tau])}\right)^s \right\}, \tag{A8}$$

where we also used that $\mathrm{P}_{\mathcal{N}}(\hat{\boldsymbol{u}}_{\mathcal{N}}) = \mathrm{P}_{\tau^c}(\hat{\boldsymbol{u}}_{\tau^c})\mathrm{P}_{\tau|\tau^c}(\hat{\boldsymbol{u}}_\tau|\hat{\boldsymbol{u}}_{\tau^c}) = \mathrm{P}_{\tau^c}(\boldsymbol{u}_{\tau^c})\mathrm{P}_{\tau|\tau^c}(\hat{\boldsymbol{u}}_\tau|\boldsymbol{u}_{\tau^c})$ to rewrite the message probabilities in Equation (A8) and we expressed the codeword $\boldsymbol{x}_{\mathcal{N}}$ as $[\boldsymbol{x}_{\tau^c},\boldsymbol{x}_\tau]$.

Inserting Equations (A6) and (A8) into Equation (A3) and using the following inequality for $A \geq 0$,

$$\min\{1, A\} \leq \min_{\rho\in[0,1]} A^\rho, \tag{A9}$$

where $\rho \in [0, 1]$, we further bound $\bar{P}_{\mathrm{e}}^{\tau}$ as

$$\bar{P}_{\mathrm{e}}^{\tau} \leq \sum_{i_{\tau^c}\in\mathcal{K}_{\tau^c}} \sum_{i_\tau\in\mathcal{K}_\tau} \sum_{j_\tau\in\mathcal{K}_\tau} \min_{s\geq 0} \min_{\rho\in[0,1]} \bar{P}_{\mathrm{e},i_{\tau^c}i_\tau}^{\tau,j_\tau}, \tag{A10}$$

where after some minor rearrangements $\bar{P}_{\mathrm{e},i_{\tau^c}i_\tau}^{\tau,j_\tau}$ is in turn given by

$$\bar{P}_{\mathrm{e},i_{\tau^c}i_\tau}^{\tau,j_\tau} = \sum_{\substack{\boldsymbol{u}_{\tau^c}\in\mathcal{A}_{\tau^c}^{i_{\tau^c}}\\ \boldsymbol{x}_{\tau^c}\in\mathcal{D}_{\tau^c,\boldsymbol{u}_{\tau^c}}^{i_{\tau^c}}}} \mathrm{P}_{\tau^c}(\boldsymbol{u}_{\tau^c})\mathrm{Q}_{\tau^c}^{i_{\tau^c}}(\boldsymbol{x}_{\tau^c}|\boldsymbol{u}_{\tau^c}) \sum_{\boldsymbol{y}\in\mathcal{Y}^n} \sum_{\substack{\boldsymbol{u}_\tau\in\mathcal{A}_\tau^{i_\tau}\\ \boldsymbol{x}_\tau\in\mathcal{D}_{\tau,\boldsymbol{u}_\tau}^{i_\tau}}} \mathrm{P}_{\tau|\tau^c}(\boldsymbol{u}_\tau|\boldsymbol{u}_{\tau^c})\mathrm{Q}_\tau^{i_\tau}(\boldsymbol{x}_\tau|\boldsymbol{u}_\tau)\mathrm{W}\big(\boldsymbol{y}|[\boldsymbol{x}_{\tau^c},\boldsymbol{x}_\tau]\big)$$
$$\left(\sum_{\substack{\hat{\boldsymbol{u}}_\tau\in\mathcal{A}_\tau^{j_\tau}\\ \hat{\boldsymbol{x}}_\tau\in\mathcal{D}_{\tau,\hat{\boldsymbol{u}}_\tau}^{j_\tau}}} \mathrm{Q}_\tau^{j_\tau}(\hat{\boldsymbol{x}}_\tau|\hat{\boldsymbol{u}}_\tau)\left(\frac{\mathrm{P}_{\tau|\tau^c}(\hat{\boldsymbol{u}}_\tau|\boldsymbol{u}_{\tau^c})\mathrm{W}\big(\boldsymbol{y}|[\boldsymbol{x}_{\tau^c},\hat{\boldsymbol{x}}_\tau]\big)}{\mathrm{P}_{\tau|\tau^c}(\boldsymbol{u}_\tau|\boldsymbol{u}_{\tau^c})\mathrm{W}\big(\boldsymbol{y}|[\boldsymbol{x}_{\tau^c},\boldsymbol{x}_\tau]\big)}\right)^s\right)^\rho. \tag{A11}$$

Note that, for some conveniently chosen variables z_0 and z_{i_τ}, sets $\mathcal{Z}_0$ and $\mathcal{Z}_{i_\tau}$, as well as functions $f_0(z_0)$ and $f_{i_\tau}^s(z_0,z_{i_\tau})$, with $i_\tau\in\mathcal{K}_\tau$, we can express $\bar{P}_{\mathrm{e},i_{\tau^c}i_\tau}^{\tau,j_\tau}$ as

$$\bar{P}_{\mathrm{e},i_{\tau^c}i_\tau}^{\tau,j_\tau} = \sum_{z_0\in\mathcal{Z}_0} f_0(z_0)\left(\sum_{z_{i_\tau}\in\mathcal{Z}_{i_\tau}} f_{i_\tau}^{1-s_{i_\tau,j_\tau}\rho_{i_\tau,j_\tau}}(z_0,z_{i_\tau})\right)\left(\sum_{z_{j_\tau}\in\mathcal{Z}_{j_\tau}} f_{j_\tau}^{s_{i_\tau,j_\tau}}(z_0,z_{j_\tau})\right)^{\rho_{i_\tau,j_\tau}}. \tag{A12}$$

In Equation (A12), the variable z_0 stands for the triplet $(\boldsymbol{u}_{\tau^c},\boldsymbol{x}_{\tau^c},\boldsymbol{y})$, the alphabet $\mathcal{Z}_0$ for the Cartesian product $\mathcal{A}_{\tau^c}^{i_{\tau^c}}\times\mathcal{D}_{\tau^c,\boldsymbol{u}_{\tau^c}}^{i_{\tau^c}}\times\mathcal{Y}^n$ and the function $f_0(z_0)$ is given by $\mathrm{P}_{\tau^c}(\boldsymbol{u}_{\tau^c})\mathrm{Q}_{\tau^c}^{i_{\tau^c}}(\boldsymbol{x}_{\tau^c}|\boldsymbol{u}_{\tau^c})$. The variable z_{i_τ} stands for the pair $(\boldsymbol{u}_\tau,\boldsymbol{x}_\tau)$, the alphabet $\mathcal{Z}_{i_\tau}$ for the Cartesian product $\mathcal{A}_\tau^{i_\tau}\times\mathcal{D}_{\tau,\boldsymbol{u}_\tau}^{i_\tau}$ and the function $f_{i_\tau}^s(z_0,z_{i_\tau})$ is given by $\mathrm{P}_{\tau|\tau^c}(\boldsymbol{u}_\tau|\boldsymbol{u}_{\tau^c})^s\mathrm{Q}_\tau^{i_\tau}(\boldsymbol{x}_\tau|\boldsymbol{u}_\tau)\mathrm{W}\big(\boldsymbol{y}|[\boldsymbol{x}_{\tau^c},\boldsymbol{x}_\tau]\big)^s$.

The optimization parameters s and ρ in Equation (A10) implictly depend on the error-event type τ and the indices i_{τ^c}, i_τ, and j_τ. For new parameters $\bar{\rho}_{\ell_\tau}\in[0,1]$, $\ell_\tau\in\mathcal{K}_\tau$, setting

$$s_{i_\tau,j_\tau} = \frac{1}{1+\bar{\rho}_{j_\tau}}, \tag{A13}$$
$$\rho_{i_\tau,j_\tau} = \frac{\bar{\rho}_{i_\tau}(1+\bar{\rho}_{j_\tau})}{1+\bar{\rho}_{i_\tau}}, \tag{A14}$$

In Equation (A12), we obtain the following partial upper bound in Equation (A10) as

$$\min_{s\geq 0}\min_{\rho\in[0,1]} \bar{P}_{\mathrm{e},i_{\tau^c}i_\tau}^{\tau,j_\tau} \leq \min_{\bar{\rho}_{\ell_\tau}\in[0,1]^{K_\tau}} \sum_{z_0\in\mathcal{Z}_0} f_0(z_0) \sum_{z_{i_\tau}\in\mathcal{Z}_{i_\tau}} f_{i_\tau}^{\frac{1}{1+\bar{\rho}_{i_\tau}}}(z_0,z_{i_\tau})\left(\sum_{z_{j_\tau}\in\mathcal{Z}_{j_\tau}} f_{j_\tau}^{\frac{1}{1+\bar{\rho}_{j_\tau}}}(z_0,z_{j_\tau})\right)^{\frac{\bar{\rho}_{i_\tau}(1+\bar{\rho}_{j_\tau})}{1+\bar{\rho}_{i_\tau}}}. \tag{A15}$$

Here, we have kept implicit the dependence on τ and i_{τ^c} of the optimization parameter $\bar{\rho}_{\ell_\tau}$. Now, applying Hölder's inequality [33] (Th. 13) in the form

$$\sum_{i\in\mathcal{K}}\alpha_i a_i b_i \leq \left(\sum_{i\in\mathcal{K}}\alpha_i a_i^p\right)^{\frac{1}{p}}\left(\sum_{i\in\mathcal{K}}\alpha_i b_i^{\frac{p}{p-1}}\right)^{\frac{p-1}{p}}, \quad \text{for } p\in[1,\infty), \tag{A16}$$

to the expression in Equation (A15) with $p_{i_\tau,j_\tau} = 1+\bar{\rho}_{i_\tau}\geq 1$, we obtain

$$\min_{s\geq 0}\min_{\rho\in[0,1]} \bar{P}_{\mathrm{e},i_{\tau^c}i_\tau}^{\tau,j_\tau} \leq \min_{\bar{\rho}_{\ell_\tau}\in[0,1]^{K_\tau}} \left(\sum_{z_0\in\mathcal{Z}_0} f_0(z_0)\left(\sum_{z_{i_\tau}\in\mathcal{Z}_{i_\tau}} f_{i_\tau}^{\frac{1}{1+\bar{\rho}_{i_\tau}}}(z_0,z_{i_\tau})\right)^{1+\bar{\rho}_{i_\tau}}\right)^{\frac{1}{1+\bar{\rho}_{i_\tau}}}$$
$$\left(\sum_{z_0\in\mathcal{Z}_0} f_0(z_0)\left(\sum_{z_{j_\tau}\in\mathcal{Z}_{j_\tau}} f_{j_\tau}^{\frac{1}{1+\bar{\rho}_{j_\tau}}}(z_0,z_{j_\tau})\right)^{1+\bar{\rho}_{j_\tau}}\right)^{\frac{\bar{\rho}_{i_\tau}}{1+\bar{\rho}_{i_\tau}}}. \tag{A17}$$

Next, putting Equation (A17) back in Equation (A10) and using the following inequality, proved in Appendix A.1, for $A_i \geq 0$ and $0 \leq s_i \leq 1$

$$\sum_{i,j \in \mathcal{K}} A_i^{s_i} A_j^{1-s_i} \leq 2|\mathcal{K}| \sum_{i \in \mathcal{K}} A_i \tag{A18}$$

in the double summation over i_τ and j_τ in Equation (A10), the following upper bound holds

$$\sum_{i_\tau \in \mathcal{K}_\tau} \sum_{j_\tau \in \mathcal{K}_\tau} \min_{s \geq 0} \min_{\rho \in [0,1]} \bar{P}^{\tau, j_\tau}_{\mathrm{e}, i_{\tau^c} i_\tau} \leq 2K_\tau \sum_{i_\tau \in \mathcal{K}_\tau} \min_{\rho \in [0,1]} \bar{P}^{\tau}_{\mathrm{e}, i_{\tau^c} i_\tau}, \tag{A19}$$

where we have moved the optimization over $\bar{\rho}_{\ell_\tau}$ inside the summation over i_τ and renamed $\bar{\rho}_{\ell_\tau}$ as ρ, with the dependence on the index i_τ kept implicit. Moreover, the expression for $\bar{P}^{\tau}_{\mathrm{e}, i_{\tau^c} i_\tau}$ is in fact given by $\bar{P}^{\tau, j_\tau}_{\mathrm{e}, i_{\tau^c} i_\tau}$ in Equation (A11) after setting $i_\tau = j_\tau$, $s = \frac{1}{1+\rho}$ and rearranging terms, that is,

$$\bar{P}^{\tau}_{\mathrm{e}, i_{\tau^c} i_\tau} = \sum_{\substack{\boldsymbol{u}_{\tau^c} \in \mathcal{A}_{\tau^c}^{i_{\tau^c}} \\ \boldsymbol{x}_{\tau^c} \in \mathcal{D}_{\tau^c, \boldsymbol{u}_{\tau^c}}^{i_{\tau^c}}}} \mathrm{P}_{\tau^c}(\boldsymbol{u}_{\tau^c}) \mathrm{Q}_{\tau^c}^{i_{\tau^c}}(\boldsymbol{x}_{\tau^c} | \boldsymbol{u}_{\tau^c}) \sum_{\boldsymbol{y} \in \mathcal{Y}^n} \left(\sum_{\substack{\boldsymbol{u}_\tau \in \mathcal{A}_\tau^{i_\tau} \\ \boldsymbol{x}_\tau \in \mathcal{D}_{\tau, \boldsymbol{u}_\tau}^{i_\tau}}} \mathrm{P}_{\tau | \tau^c}(\boldsymbol{u}_\tau | \boldsymbol{u}_{\tau^c})^{\frac{1}{1+\rho}} \mathrm{Q}_\tau^{i_\tau}(\boldsymbol{x}_\tau | \boldsymbol{u}_\tau) \mathrm{W}(\boldsymbol{y} | \boldsymbol{x}_{\mathcal{N}})^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{A20}$$

It remains to factorize Equation (A20) into a product of symbol distributions in order to obtain a single-letter expression for the exponent. We start by upper bounding the summations over the input messages $\boldsymbol{u}_{\tau^c}$ and $\boldsymbol{u}_\tau$. For a list of users σ with corresponding messages $\boldsymbol{u}_\sigma$, list of class indices i_σ and some function $p_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma)$, we have that

$$\sum_{\boldsymbol{u}_\sigma \in \mathcal{A}_\sigma^{i_\sigma}} p_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma) = \sum_{\boldsymbol{u}_\sigma \in \mathcal{U}_\sigma^n} p_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma) \mathbf{1}\{\boldsymbol{u}_\sigma \in \mathcal{A}_\sigma^{i_\sigma}\}, \tag{A21}$$

where we used the definition of the message sets $\mathcal{A}_\sigma^{i_\sigma}$ in Equation (12) and the identity

$$\sum_{i \in \mathcal{K}} f_i = \sum_{i \in \mathbf{N}} f_i \mathbf{1}\{i \in \mathcal{K}\}. \tag{A22}$$

Using the upper bound

$$\mathbf{1}\{a < b \leq c\} \leq \min_{\lambda^{\mathrm{L}}, \lambda^{\mathrm{U}} \geq 0} \left(\frac{b}{a}\right)^{\lambda^{\mathrm{L}}} \left(\frac{c}{b}\right)^{\lambda^{\mathrm{U}}} \tag{A23}$$

for $a, b, c > 0$ with $\lambda^{\mathrm{L}}, \lambda^{\mathrm{U}} \geq 0$, together with the fact that the source-message classes are defined separately for each user to express the source message probabilities in terms of $\mathrm{P}_\sigma^{\mathrm{ind}}(\boldsymbol{u}_\sigma) = \prod_{\nu \in \sigma} \mathrm{P}_\nu(\boldsymbol{u}_\nu)$ similarly to Equation (2), we upper bound the r. h. s. of Equation (A21) as

$$\sum_{\boldsymbol{u}_\sigma \in \mathcal{A}_\sigma^{i_\sigma}} p_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma) \leq \min_{\lambda_\sigma^{\mathrm{L,U}} \geq 0} \sum_{\boldsymbol{u}_\sigma \in \mathcal{U}_\sigma^n} p_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma) \left(\frac{\mathrm{P}_\sigma^{\mathrm{ind}}(\boldsymbol{u}_\sigma)}{\gamma_{\sigma, i_\sigma}^n}\right)^{\lambda_\sigma^{\mathrm{L}}} \left(\frac{\gamma_{\sigma, i_\sigma - 1}^n}{\mathrm{P}_\sigma^{\mathrm{ind}}(\boldsymbol{u}_\sigma)}\right)^{\lambda_\sigma^{\mathrm{U}}}, \tag{A24}$$

where we jointly wrote $\lambda_\sigma^{\mathrm{L}}$ and $\lambda_\sigma^{\mathrm{U}}$ as $\lambda_\sigma^{\mathrm{L,U}}$. Defininining

$$\Lambda_\sigma^{i_\sigma}(u_\sigma) = \left(\frac{\mathrm{P}_\sigma^{\mathrm{ind}}(u_\sigma)}{\gamma_{\sigma, i_\sigma}}\right)^{\lambda_\sigma^{\mathrm{L}}} \left(\frac{\gamma_{\sigma, i_\sigma - 1}}{\mathrm{P}_\sigma^{\mathrm{ind}}(u_\sigma)}\right)^{\lambda_\sigma^{\mathrm{U}}}. \tag{A25}$$

and taking into account that the sources are memoryless, we obtain that the summations w. r. t. the source messages $\boldsymbol{u}_\tau$ and $\boldsymbol{u}_{\tau^c}$ in Equation (A20) are upper bounded as

$$\sum_{\boldsymbol{u}_\sigma \in \mathcal{A}_\sigma^{i_\sigma}} p_\sigma^{i_\sigma i_\sigma}(\boldsymbol{u}_\sigma) \leq \min_{\lambda_\sigma^{\mathrm{L,U}}} \sum_{\boldsymbol{u}_\sigma \in \mathcal{U}_\sigma^n} p_\sigma^{i_{\sigma^c} i_\sigma}(\boldsymbol{u}_\sigma) \prod_{t=1}^{n} \Lambda_\sigma^{i_\sigma}(u_{\sigma,t}). \tag{A26}$$

respecively for $\sigma = \tau$ and $\sigma = \tau^c$.

We proceed in a similar manner for the summations w. r. t. the codewords $\boldsymbol{x}_\tau$ and $\boldsymbol{x}_{\tau^c}$ in Equation (A20). For a list of users σ and some function $q_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma, \boldsymbol{x}_\sigma)$ implicitly defined, the summation over channel codewords $\boldsymbol{x}_\sigma \in \mathcal{D}_{\sigma,\boldsymbol{u}_\sigma}^{i_\sigma}$ can be upper bounded as:

$$\sum_{\boldsymbol{x}_\sigma \in \mathcal{D}_{\sigma,\boldsymbol{u}_\sigma}^{i_\sigma}} q_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma, \boldsymbol{x}_\sigma) = \sum_{\boldsymbol{x}_\sigma \in \mathcal{X}_\sigma^n} q_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma, \boldsymbol{x}_\sigma) \mathbf{1}\{\boldsymbol{x}_\sigma \in \mathcal{D}_{\sigma,\boldsymbol{u}_\sigma}^{i_\sigma}\} \tag{A27}$$

$$= \sum_{\boldsymbol{x}_\sigma \in \mathcal{X}_\sigma^n} q_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma, \boldsymbol{x}_\sigma) \prod_{v \in \sigma} \prod_{v_v \in \mathcal{U}_v} \prod_{\ell_v \in \mathcal{L}_v} \mathbf{1}\left\{ \left| \mathrm{a}_{v,v_v}^{i_v,\ell_v}(\boldsymbol{x}_v(\mathcal{I}_{v_v}(\boldsymbol{u}_v))) \right| \leq \delta_v \right\} \tag{A28}$$

$$\leq \min_{r_{\sigma v_\sigma}^{\ell_\sigma}, \bar{r}_{\sigma v_\sigma}^{\ell_\sigma}} \sum_{\boldsymbol{x}_\sigma \in \mathcal{X}_\sigma^n} q_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma, \boldsymbol{x}_\sigma) \prod_{v_\sigma \in \mathcal{U}_\sigma} \prod_{\ell_\sigma \in \mathcal{L}_\sigma} e^{r_{\sigma v_\sigma}^{\ell_\sigma} \mathrm{a}_{\sigma,v_\sigma}^{i_\sigma,\ell_\sigma}(\boldsymbol{x}_\sigma(\mathcal{I}_{v_\sigma}(\boldsymbol{u}_\sigma))) + \bar{r}_{\sigma v_\sigma}^{\ell_\sigma} \delta_\sigma}, \tag{A29}$$

where we used Equation (A22) in Equation (A27), the fact that the codeword ensembles are defined separately for each user together with the definition of the ensemble cost constraints in Equation (23) and subcodewords $\boldsymbol{x}_v(\mathcal{I}_{u_v}(\boldsymbol{u}_v))$ in Equation (A28), and a variant of Equation (A23) proved in Appendix A.2,

$$\mathbf{1}\{|a| \leq \delta\} \leq \min_{r,\bar{r}} e^{ra + \bar{r}\delta} \tag{A30}$$

for $r \in \mathbf{R}$ and $\bar{r} \geq 0$, in Equation (A29) for each indicator function of Equation (A28) and combined the product of exponentials over σ as a single exponential using the list notation. We continue by rewriting the double product over v_σ and ℓ_σ in Equation (A29) as follows

$$\prod_{v_\sigma \in \mathcal{U}_\sigma} \prod_{\ell_\sigma \in \mathcal{L}_\sigma} e^{r_{\sigma v_\sigma}^{\ell_\sigma} \mathrm{a}_{\sigma,v_\sigma}^{i_\sigma,\ell_\sigma}(\boldsymbol{x}_\sigma(\mathcal{I}_{v_\sigma}(\boldsymbol{u}_\sigma))) + \bar{r}_{\sigma v_\sigma}^{\ell_\sigma} \delta_\sigma} = \prod_{v_\sigma \in \mathcal{U}_\sigma} \prod_{\ell_\sigma \in \mathcal{L}_\sigma} e^{\bar{r}_{\sigma v_\sigma}^{\ell_\sigma} \delta_\sigma} \prod_{t \in \mathcal{I}_{v_\sigma}(\boldsymbol{u}_\sigma)} e^{r_{\sigma v_\sigma}^{\ell_\sigma} a_{\sigma,v_\sigma}^{i_\sigma,\ell_\sigma}(x_{\sigma,t}(\mathcal{I}_{v_\sigma}(\boldsymbol{u}_\sigma)))} \tag{A31}$$

$$= \beta_\sigma \prod_{t=1}^{n} R_{\sigma,u_{\tau,t}}^{i_\tau}(x_{\tau,t}), \tag{A32}$$

where in Equation (A31) we wrote the cost function in terms of the symbol costs and in Equation (A32) we rearranged terms and introduced a factor β_σ that depends on the list $\{\bar{r}_{\sigma v_\sigma}^{\ell_\sigma}\}$ and a function $R_{\sigma,u_\sigma}^{i_\sigma}(x_\sigma)$ that depends on the list $\{r_{\sigma v_\sigma}^{\ell_\sigma}\}$ and are respectively given by

$$\beta_\sigma = \prod_{v_\sigma \in \mathcal{U}_\sigma} \prod_{\ell_\sigma \in \mathcal{L}_\sigma} e^{\bar{r}_{\sigma v_\sigma}^{\ell_\sigma} \delta_\sigma}, \tag{A33}$$

$$R_{\sigma,u_\sigma}^{i_\sigma}(x_\sigma) = \prod_{\ell_\sigma \in \mathcal{L}_\sigma} e^{r_{\sigma u_\sigma}^{\ell_\sigma} a_{\sigma,u_\sigma}^{i_\sigma,\ell_\sigma}(x_\sigma)}. \tag{A34}$$

Replacing Equation (A32) back into Equation (A29), we obtain that the summations over the codewords are upper bounded as

$$\sum_{\boldsymbol{x}_\sigma \in \mathcal{D}_{\sigma,\boldsymbol{u}_\sigma}^{i_\sigma}} q_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma, \boldsymbol{x}_\sigma) \leq \min_{r_{\sigma u_\sigma}^{\ell_\sigma}, \bar{r}_{\sigma u_\sigma}^{\ell_\sigma}} \sum_{\boldsymbol{x}_\sigma \in \mathcal{X}_\sigma^n} q_\sigma^{i_\sigma}(\boldsymbol{u}_\sigma, \boldsymbol{x}_\sigma) \beta_\sigma \prod_{t=1}^{n} R_{\sigma,u_{\sigma,t}}^{i_\sigma}(x_{\tau,t}). \tag{A35}$$

for both user lists $\sigma = \tau$ and $\sigma = \tau^c$.

We now combine Equations (A26) and (A35) for $\sigma = \tau$ to bound the summation inside the parenthesis in Equation (A20) as

$$\sum_{\substack{\boldsymbol{u}_\tau \in \mathcal{A}_\tau^{i_\tau} \\ \boldsymbol{x}_\tau \in \mathcal{D}_{\tau,\boldsymbol{u}_\tau}^{i_\tau}}} \mathrm{P}_{\tau|\tau^c}(\boldsymbol{u}_\tau|\boldsymbol{u}_{\tau^c})^{\frac{1}{1+\rho}} \mathrm{Q}_\tau^{i_\tau}(\boldsymbol{x}_\tau|\boldsymbol{u}_\tau) \mathrm{W}(\boldsymbol{y}|\boldsymbol{x}_\mathcal{N})^{\frac{1}{1+\rho}} \leq$$

$$\leq \sum_{\substack{\boldsymbol{u}_\tau \in \mathcal{U}_\tau^n \\ \boldsymbol{x}_\tau \in \mathcal{X}_\tau^n}} \mathrm{P}_{\tau|\tau^c}(\boldsymbol{u}_\tau|\boldsymbol{u}_{\tau^c})^{\frac{1}{1+\rho}} \min_{\substack{\lambda_\tau^{\mathrm{L,U}} \\ r_{\tau u_\tau}^{\ell_\tau}, \bar{r}_{\tau u_\tau}^{\ell_\tau}}} \frac{\beta_\tau}{\Xi_\tau} \prod_{t=1}^n \Lambda_\tau^{i_\tau}(u_{\tau,t}) Q_{\tau,u_{\tau,t}}^{i_\tau}(x_{\tau,t}) R_{\tau,u_{\tau,t}}^{i_\tau}(x_{\tau,t}) \mathrm{W}(\boldsymbol{y}|\boldsymbol{x}_\mathcal{N})^{\frac{1}{1+\rho}}, \tag{A36}$$

where we expressed the distribution $\mathrm{Q}_\tau^{i_\tau}(\boldsymbol{x}_\tau, \boldsymbol{u}_\tau)$ in terms of the symbol-wise iid distribution $Q_{\tau,u_\tau}^{i_\tau}(x)$ as in Equation (25). Since both source and channel are memoryless, we may now factorize and rearrange the expression in Equation (A36) into single-letter, symbolwise factors as

$$\sum_{\substack{\boldsymbol{u}_\tau \in \mathcal{A}_\tau^{i_\tau} \\ \boldsymbol{x}_\tau \in \mathcal{D}_{\tau,\boldsymbol{u}_\tau}^{i_\tau}}} \mathrm{P}_{\tau|\tau^c}(\boldsymbol{u}_\tau|\boldsymbol{u}_{\tau^c})^{\frac{1}{1+\rho}} \mathrm{Q}_\tau^{i_\tau}(\boldsymbol{x}_\tau|\boldsymbol{u}_\tau) \mathrm{W}(\boldsymbol{y}|\boldsymbol{x}_\mathcal{N})^{\frac{1}{1+\rho}} \leq \min_{\substack{\lambda_\tau^{\mathrm{L,U}} \\ r_{\tau u_\tau}^{\ell_\tau}, \bar{r}_{\tau u_\tau}^{\ell_\tau}}} \frac{\beta_\tau}{\Xi_\tau} \prod_{t=1}^n g_\tau^{i_\tau}(u_{\tau^c,t}, x_{\tau^c,t}, y_t), \tag{A37}$$

where, for a list of users σ, the function $g_\sigma^{i_\sigma}(u_{\sigma^c}, x_{\sigma^c}, y)$ is defined as

$$g_\sigma^{i_\sigma}(u_{\sigma^c}, x_{\sigma^c}, y) = \sum_{\substack{u_\sigma \in \mathcal{U}_\sigma \\ x_\sigma \in \mathcal{X}_\sigma}} P_{\sigma|\sigma^c}(u_\sigma|u_{\sigma^c})^{\frac{1}{1+\rho}} \Lambda_\sigma^{i_\sigma}(u_\sigma) Q_{\sigma,u_\sigma}^{i_\sigma}(x_\sigma) R_{\sigma,u_\sigma}^{i_\sigma}(x_\sigma) W(y|x_\mathcal{N})^{\frac{1}{1+\rho}}. \tag{A38}$$

Although not explicitly, the function $g_\tau^{i_\tau}(u_{\tau^c}, x_{\tau^c}, y)$ in Equation (A37) depends on several optimization parameters, namely $\rho, \lambda_\tau^{\mathrm{L,U}}, r_{\tau u_\tau}^{\ell_\tau}, \bar{r}_{\tau u_\tau}^{\ell_\tau}$, which depend in turn on the error-event type τ and class indices i_τ and i_{τ^c}.

Again, we use Equations (A26) and (A35) for $\sigma = \tau^c$ and the fact that the source is memoryless to upper bound the summation outside the parenthesis in Equation (A20) as

$$\sum_{\substack{\boldsymbol{u}_{\tau^c} \in \mathcal{A}_{\tau^c}^{i_{\tau^c}} \\ \boldsymbol{x}_{\tau^c} \in \mathcal{D}_{\tau^c,\boldsymbol{u}_{\tau^c}}^{i_{\tau^c}}}} \mathrm{P}_{\tau^c}(\boldsymbol{u}_{\tau^c}) \mathrm{Q}_{\tau^c}^{i_{\tau^c}}(\boldsymbol{x}_{\tau^c}|\boldsymbol{u}_{\tau^c}) \leq$$

$$\leq \min_{\substack{\lambda_{\tau^c}^{\mathrm{L,U}} \\ r_{\tau^c u_{\tau^c}}^{\ell_{\tau^c}}, \bar{r}_{\tau^c u_{\tau^c}}^{\ell_{\tau^c}}}} \sum_{\substack{\boldsymbol{u}_{\tau^c} \in \mathcal{U}_{\tau^c}^n \\ \boldsymbol{x}_{\tau^c} \in \mathcal{X}_{\tau^c}^n}} \frac{\beta_{\tau^c}}{\Xi_{\tau^c}} \prod_{t=1}^n P_{\tau^c}(u_{\tau^c,t}) \Lambda_{\tau^c}^{i_{\tau^c}}(u_{\tau^c,t}) Q_{\tau^c,u_{\tau^c,t}}^{i_{\tau^c}}(x_{\tau^c,t}) R_{\tau^c,u_{\tau^c,t}}^{i_{\tau^c}}(x_{\tau^c,t}). \tag{A39}$$

Substituting Equations (A37) and (A39) in Equation (A20), the resulting expression back into Equation (A19) and then into Equation (A10), we get

$$\bar{P}_{\mathrm{e}}^\tau \leq 2K_\tau \sum_{\substack{i_{\tau^c} \in \mathcal{K}_{\tau^c} \\ i_\tau \in \mathcal{K}_\tau}} \min_{\substack{\rho, \lambda_{\tau^c}^{\mathrm{L,U}} \\ r_{\tau^c u_{\tau^c}}^{\ell_{\tau^c}}, \bar{r}_{\tau^c u_{\tau^c}}^{\ell_{\tau^c}}}} \sum_{\substack{\boldsymbol{u}_{\tau^c} \in \mathcal{U}_{\tau^c}^n \\ \boldsymbol{x}_{\tau^c} \in \mathcal{X}_{\tau^c}^n}} \frac{\beta_{\tau^c}}{\Xi_{\tau^c}} \left(\prod_{t=1}^n P_{\tau^c}(u_{\tau^c,t}) \Lambda_{\tau^c}^{i_{\tau^c}}(u_{\tau^c,t}) Q_{\tau^c,u_{\tau^c,t}}^{i_{\tau^c}}(x_{\tau^c,t}) R_{\tau^c,u_{\tau^c,t}}^{i_{\tau^c}}(x_{\tau^c,t}) \right)$$

$$\left(\sum_{\boldsymbol{y} \in \mathcal{Y}^n} \left(\frac{\beta_\tau}{\Xi_\tau} \right)^{1+\rho} \min_{\substack{\lambda_\tau^{\mathrm{L,U}} \\ r_{\tau u_\tau}^{\ell_\tau}, \bar{r}_{\tau u_\tau}^{\ell_\tau}}} \prod_{t=1}^n g_\tau^{i_\tau}(u_{\tau^c,t}, x_{\tau^c,t}, y_t)^{1+\rho} \right). \tag{A40}$$

Let us define now the function $h_\sigma^{i_\sigma, i_{\sigma^c}}$ of the user set σ and the class indices i_σ and i_{σ^c} as

$$h_\sigma^{i_\sigma, i_{\sigma^c}} = \sum_{u_\sigma \in \mathcal{U}_\sigma, x_\sigma \in \mathcal{X}_\sigma, y \in \mathcal{Y}} P_\sigma(u_\sigma) \Lambda_\sigma^{i_\sigma}(u_\sigma) Q_{\sigma,u_\sigma}^{i_\sigma}(x_\sigma) R_{\sigma,u_\sigma}^{i_\sigma}(x_\sigma) \, g_{\sigma^c}^{i_{\sigma^c}}(u_\sigma, x_\sigma, y)^{1+\rho}. \tag{A41}$$

With this definition, we can rewrite Equation (A40) in a compact manner as

$$\bar{P}_{\mathrm{e}}^{\tau} \leq 2K_{\tau} \sum_{i_{\mathcal{N}} \in \mathcal{K}_{\mathcal{N}}} \min_{\rho, \lambda_{\mathcal{N}}^{\mathrm{L,U}}, r_{\mathcal{N}u_{\mathcal{N}}}^{\ell_{\mathcal{N}}}, \bar{r}_{\mathcal{N}u_{\mathcal{N}}}^{\ell_{\mathcal{N}}}} \frac{\beta_{\tau^c}}{\Xi_{\tau^c}} \left(\frac{\beta_{\tau}}{\Xi_{\tau}} \right)^{1+\rho} \prod_{t=1}^{n} h_{\tau^c}^{i_{\tau^c}, i_{\tau}}, \tag{A42}$$

where we have also combined the complementary sets i_{τ^c} and i_{τ} into $i_{\mathcal{N}}$, and similarly for $\lambda_{\mathcal{N}}^{\mathrm{L}}$, $\lambda_{\mathcal{N}}^{\mathrm{U}}$, $r_{\mathcal{N}u_{\mathcal{N}}}^{\ell_{\mathcal{N}}}$, and $\bar{r}_{\mathcal{N}u_{\mathcal{N}}}^{\ell_{\mathcal{N}}}$. Finally, substituting Equation (A42) into Equation (A10) and then back into Equation (A2), taking (minus) the logarithm of the bound on $\bar{P}_{\mathrm{e}}$, dividing the result by n, and the limit as $n \to \infty$, we obtain a lower bound $\mathrm{E}_{\mathcal{KL}}^{cost}$ to the exponent of the generalized cost-constrained ensemble $\mathrm{E}_{\mathcal{KL}}^{cost}$, namely

$$\mathrm{E}_{\mathcal{KL}}^{cost} = \min_{\tau, i_{\mathcal{N}}} \max_{\rho, \lambda_{\mathcal{N}}^{\mathrm{L,U}}, r_{\mathcal{N}u_{\mathcal{N}}}^{\ell_{\mathcal{N}}}} \left\{ -\log h_{\tau^c}^{i_{\tau^c}, i_{\tau}} \right\}, \tag{A43}$$

where we have used that as $n \to \infty$, the quantities $2K_{\tau}$, $\frac{\beta_{\tau^c}}{\Xi_{\tau^c}}$, and $\left(\frac{\beta_{\tau}}{\Xi_{\tau}}\right)^{1+\rho}$ are subexponential in the blocklength n and do not contribute to the exponent, accordingly removed $\bar{r}_{\tau^c u_{\tau^c}}^{\ell_{\tau^c}}$ and $\bar{r}_{\tau u_{\tau}}^{\ell_{\tau}}$ from the optimization parameter list, and finally used that the exponential decay of the error probability in Equation (A2) will be dominated by the worst error type τ and the worst classes assignment $i_{\mathcal{N}}$. It will prove convenient to the express the exponent in terms of a Gallager function $E_{\tau}^{i_{\mathcal{N}}}\left(\rho, \lambda_{\mathcal{N}}^{\mathrm{L,U}}, r_{\mathcal{N}u_{\mathcal{N}}}^{\ell_{\mathcal{N}}}\right)$, defined as

$$E_{\tau}^{i_{\mathcal{N}}}\left(\rho, \lambda_{\mathcal{N}}^{\mathrm{L,U}}, r_{\mathcal{N}u_{\mathcal{N}}}^{\ell_{\mathcal{N}}}\right) = -\log h_{\tau^c}^{i_{\tau^c}, i_{\tau}}. \tag{A44}$$

Substituted the expression for $h_{\tau^c}^{i_{\tau^c}, i_{\tau}}$ in Equation (A40), where $\Lambda_{\tau}^{i_{\tau}}(u_{\tau})$ and $R_{\nu, u_{\nu}}^{i_{\nu}}(x_{\nu})$ are respectively given by Equations (A25) and (A34), we may express $E_{\tau}^{i_{\mathcal{N}}}\left(\rho, \lambda_{\mathcal{N}}^{\mathrm{L,U}}, r_{\mathcal{N}u_{\mathcal{N}}}^{\ell_{\mathcal{N}}}\right)$ as

$$\begin{aligned} &E_{\tau}^{i_{\mathcal{N}}}\left(\rho, \lambda_{\mathcal{N}}^{\mathrm{L,U}}, r_{\mathcal{N}u_{\mathcal{N}}}^{\ell_{\mathcal{N}}}\right) = \\ &\quad -\log \Bigg(\sum_{u_{\tau^c}, x_{\tau^c}, y} P_{\tau^c}(u_{\tau^c}) \Lambda_{\tau^c}^{i_{\tau^c}}(u_{\tau^c}) Q_{\tau^c, u_{\tau^c}}^{i_{\tau^c}}(x_{\tau^c}) R_{\tau^c, u_{\tau^c}}^{i_{\tau^c}}(x_{\tau^c}) \\ &\qquad \left(\sum_{u_{\tau}, x_{\tau}} P_{\tau|\tau^c}(u_{\tau}|u_{\tau^c})^{\frac{1}{1+\rho}} \Lambda_{\tau}^{i_{\tau}}(u_{\tau}) Q_{\tau, u_{\tau}}^{i_{\tau}}(x_{\tau}) R_{\tau, u_{\tau}}^{i_{\tau}}(x_{\tau}) W(y|x_{\mathcal{N}})^{\frac{1}{1+\rho}} \right)^{1+\rho} \Bigg), \end{aligned} \tag{A45}$$

or equivalently in the alternative form

$$\begin{aligned} &E_{\tau}^{i_{\mathcal{N}}}\left(\rho, \lambda_{\mathcal{N}}^{\mathrm{L,U}}, r_{\mathcal{N}u_{\mathcal{N}}}^{\ell_{\mathcal{N}}}\right) = \\ &-\log \sum_{u_{\tau^c}, x_{\tau^c}, y} \left(\sum_{u_{\tau}, x_{\tau}} P_{\mathcal{N}}(u_{\mathcal{N}})^{\frac{1}{1+\rho}} \Lambda_{\mathcal{N}}^{i_{\mathcal{N}}}(u_{\mathcal{N}}) Q_{\tau, u_{\tau}}^{i_{\tau}}(x_{\tau}) R_{\mathcal{N}, u_{\mathcal{N}}}^{i_{\mathcal{N}}}(x_{\mathcal{N}}) \left(Q_{\tau^c, u_{\tau^c}}^{i_{\tau^c}}(x_{\tau^c}) W(y|x_{\mathcal{N}})^{\frac{1}{1+\rho}} \right)^{1+\rho}, \end{aligned} \tag{A46}$$

where in Equation (A46) we have moved the product

$$P_{\tau^c}(u_{\tau^c}) \Lambda_{\tau^c}^{i_{\tau^c}}(u_{\tau^c}) Q_{\tau^c, u_{\tau^c}}^{i_{\tau^c}}(x_{\tau^c}) R_{\tau^c, u_{\tau^c}}^{i_{\tau^c}}(x_{\tau^c})$$

inside the parenthesis and merged terms in τ and τ^c as done above, as well as redefined the optimization parameters $\frac{\lambda_{\tau^c}^{\mathrm{L}}}{1+\rho}$, $\frac{\lambda_{\tau^c}^{\mathrm{L}}}{1+\rho}$, and $\frac{r_{\tau^c u_{\tau^c}}^{\ell_{\tau^c}}}{1+\rho}$ as $\lambda_{\tau^c}^{\mathrm{L}}$, $\lambda_{\tau^c}^{\mathrm{U}}$, and $r_{\tau^c u_{\tau^c}}^{\ell_{\tau^c}}$ respectively.

Appendix A.1. Proof of Equation (A18)

A sketch of the proof of the inequality in Equation (A18) proceeds as follows:

$$\sum_{i,j\in\mathcal{K}} A_i^{s_i} A_j^{1-s_i} \le \sum_{i,j\in\mathcal{K}} \left(s_i A_i + (1-s_i) A_j\right) \tag{A47}$$

$$\le \sum_{i,j\in\mathcal{K}} (A_i + A_j) \tag{A48}$$

$$= 2|\mathcal{K}| \sum_{i\in\mathcal{K}} A_i \tag{A49}$$

where Equation (A47) follows from the inequality between arithmetic and geometric means and in Equation (A48) we used that $0 \le s \le 1$.

Appendix A.2. Proof of Equation (A30)

We have the following

$$\mathbf{1}\{|a| \le \delta\} = \mathbf{1}\{-\delta \le a\}\mathbf{1}\{a \le \delta\} \tag{A50}$$

$$= \mathbf{1}\{e^{-\delta} \le e^{a}\}\mathbf{1}\{e^{a} \le e^{+\delta}\} \tag{A51}$$

$$\le e^{r_-(a+\delta)} e^{r_+(\delta-a)} \tag{A52}$$

$$= e^{ra+\bar{r}\delta} \tag{A53}$$

for $r_-, r_+ \ge 0$ or equivalently $r = r_- - r_+ \in \mathbf{R}$ and $\bar{r} = r_- + r_+ \ge 0$. The bound in Equation (A53) can be optimized w. r. t. r and $\hat{r}$.

Appendix B. Computation of the Optimum Multi-Class Thresholds

In this section we find some conditions describing the optimum partitioning of the source-message set into classes for the optimization of the exponent in Equation (30). For simplicity, let each user $\nu \in \mathcal{N}$ have two classes, $K_\nu = 2$.

From the class definition in Equation (12) with $K_\nu = 2$, we have that $\gamma_{\nu,2} = 0$ and $\gamma_{\nu,0} = 1$, so we need find just one optimum $\gamma_{\nu,1}$ for each user, which redefine as γ_ν. Optimizing the exponent in Equation (30) over $\gamma_\mathcal{N}$ gives

$$\max_{0\le\gamma_\mathcal{N}\le 1} \mathrm{E}^{cost} = \max_{0\le\gamma_\mathcal{N}\le 1} \min_{i_\mathcal{N}} \min_{\tau} \max_{\rho,\lambda_\mathcal{N}^{\mathrm{L,U}},r_{\mathcal{N}u_\mathcal{N}}^{\ell_\mathcal{N}}} E_\tau^{i_\mathcal{N}}\left(\rho, \lambda_\mathcal{N}^{\mathrm{L,U}}, r_{\mathcal{N}u_\mathcal{N}}^{\ell_\mathcal{N}}\right), \tag{A54}$$

where one of the parameters $\lambda_\mathcal{N}^{\mathrm{L}}$ or $\lambda_\mathcal{N}^{\mathrm{U}}$ is zero for each $i_\mathcal{N}$, as the corresponding constraint is absent. For each $\gamma_\mathcal{N}$, we have a minimization over 2^N assignments $i_\mathcal{N}$. Following the same steps as in refs. [31] (Sec. 4.1.2) and [31] (Lemma 4.3), we find that $E_\tau^{i_\mathcal{N}}(\gamma_\mathcal{N})$ defined, with some abuse of notation, as

$$E_\tau^{i_\mathcal{N}}(\gamma_\mathcal{N}) = \max_{\rho,\lambda_\mathcal{N}^{\mathrm{L,U}},r_{\mathcal{N}u_\mathcal{N}}^{\ell_\mathcal{N}}} E_\tau^{i_\mathcal{N}}\left(\rho, \lambda_\mathcal{N}^{\mathrm{L,U}}, r_{\mathcal{N}u_\mathcal{N}}^{\ell_\mathcal{N}}\right), \tag{A55}$$

is a non-decreasing (resp. non-increasing) function with respect to γ_ν for $i_\mathcal{N} = [i_\nu, i_{\nu^c}]$ with $i_\nu = 1$ (resp. $i_\nu = 2$), irrespective of the values of i_{ν^c} and of ν. For the sake of completeness, we present an independent proof of this fact here. Let $i_\nu = 1$ and τ be arbitrary. Using Equation (31), the function $E_\tau^{i_\mathcal{N}}(\gamma_\mathcal{N}, \rho)$ has the form $-\log\left(\sum_z f_1(z)/\gamma_\nu^{\lambda_\nu^{\mathrm{L}}}\right)$ for some function $f_1(z)$, as all $\gamma_\mathcal{N}$ are independent from each other, regardless the value of i_{ν^c}. Since $\lambda_\nu^{\mathrm{L}} \ge 0$, the function $E_\tau^{i_\mathcal{N}}(\gamma_\mathcal{N}, \rho)$ in Equation (A55) is non-decreasing with respect to γ_ν. When $i_\nu = 2$, this function $E_\tau^{i_\mathcal{N}}(\gamma_\mathcal{N}, \rho)$ has the form $-\log\left(\sum_z f_2(z)\gamma_\nu^{\lambda_\nu^{\mathrm{U}}}\right)$ for some $f_2(z)$, and is therefore non-increasing. This behavior will not change after taking maximization over ρ. As the minimum of monotonic functions is monotonic, the function $E_\tau^{i_\mathcal{N}}(\gamma_\mathcal{N})$ is non-decreasing (non-increasing) with respect to γ_ν, when $i_\nu = 1$ ($i_\nu = 2$).

For any ν and fixed γ_{ν^c}, we may write the optimization problem in Equation (A54) as

$$\max_{\gamma_{\nu^c}} \max_{\gamma_\nu} \min_{i_\nu} \min_{i_{\nu^c}} \min_{\tau} E_\tau^{[i_\nu, i_{\nu^c}]}\big([\gamma_\nu, \gamma_{\nu^c}]\big). \tag{A56}$$

The optimization problem $\max_{\gamma_\nu} \min_{i_\nu} \min_{i_{\nu^c}} \min_\tau E_\tau^{[i_\nu, i_{\nu^c}]}\big([\gamma_\nu, \gamma_{\nu^c}]\big)$ satisfies the following lemma, proved in Appendix B.1, with $\gamma = \gamma_\nu$, $i = i_\nu$, and $k_i(\gamma) = \min_{i_{\nu^c}} \min_\tau E_\tau^{[i_\nu, i_{\nu^c}]}\big([\gamma_\nu, \gamma_{\nu^c}]\big)$.

Lemma A1. *Let $k_1(\gamma)$ and $k_2(\gamma)$ be respectively continuous non-decreasing and non-increasing functions with respect to $\gamma \in [0,1]$. The optimal $\gamma^\star$ maximizing $\min_{i=1,2} k_i(\gamma)$ satisfies the following equation*

$$k_1(\gamma^\star) = k_2(\gamma^\star). \tag{A57}$$

When Equation (A57) does not have any solution, we have $\gamma^\star = 0$ if $k_1(0) > k_2(0)$, and $\gamma^\star = 1$ otherwise.

Therefore, the optimal $\gamma_\nu^\star$ satisfies

$$\min_{i_{\nu^c}} \min_\tau E_\tau^{[1, i_{\nu^c}]}\big([\gamma_\nu^*, \gamma_{\nu^c}]\big) = \min_{i_{\nu^c}} \min_\tau E_\tau^{[2, i_{\nu^c}]}\big([\gamma_\nu^*, \gamma_{\nu^c}]\big), \tag{A58}$$

if Equation (A58) has a solution. If not, the inequality $\min_{i_{\nu^c}} \min_\tau E_\tau^{[1, i_{\nu^c}]}\big([0, \gamma_{\nu^c}]\big) > \min_{i_{\nu^c}} \min_\tau E_\tau^{[1, i_{\nu^c}]}\big([0, \gamma_{\nu^c}]\big)$ is satisfied, we have $\gamma_\nu^\star = 0$ or $\gamma_\nu^\star = 1$ otherwise. Since Equation (A58) holds for any ν, evaluating it for each ν gives a system of equations for the computation of the optimal thresholds.

In ref. [31] (Sec. 3.2.1.1), we give a graphical interpretation of the solutions to Equation (A58) and outline the relevant differences with the single-user case. We observe a strong coupling between the exponent and the thresholds that prevents to find the optimal number of classes, suggesting that, unlike the single-user case, two classes might not be sufficient.

Appendix B.1. Proof of Lemma A1

The relative behaviour of a non-decreasing function with a non-increasing function can be categorized in three cases.

1. If $k_1(0) < k_2(0)$ and $k_1(1) > k_2(1)$, there exists a $\gamma^\star$ such that $k_1(\gamma^\star) = k_2(\gamma^\star)$. In this case, the function $\min_i k_i(\gamma)$ is non-decreasing from $[0, \gamma^\star)$, and non-increasing from $(\gamma^\star, 1]$. Thus, the maximum over γ of $\min_i k_i(\gamma)$ occurs at $\gamma = \gamma^\star$.
2. If $k_1(0) < k_2(0)$ and $k_1(1) < k_2(1)$, $k_1(\gamma)$ and $k_2(\gamma)$ do not cross in $\gamma \in [0,1]$. Hence, we have $\min_i k_i(\gamma) = k_1(\gamma)$ and obviously since it is an non-decreasing function the maximum over γ occurs at $\gamma = \gamma^\star = 1$.
3. When $k_1(0) \geq k_2(0)$, we have $\min_i k_i(\gamma) = k_2(\gamma)$ and hence $\gamma^\star = 0$.

Appendix C. Properties of the Modified Gallager Source Function

In this appendix, we study the modified Gallager source function $E_{\mathrm{s},\tau}^{i_\mathcal{N}}$ in Equation (45) involved in the achievable exponent for the md-iid ensemble. For the sake of simplicity, we consider the rather illustrative case of $N = 2$ users, each having a $K_\nu = 2$-class partition of the source messages with thresholds $i_\mathcal{N}$ where $\mathcal{N} = \{1, 2\}$. From the definition of the sets $\mathcal{A}_\nu^{i_\nu}$ in Equation (12) with $\gamma_{\nu,0} = 1$, $\gamma_{\nu,1} = \gamma_\nu$ and $\gamma_{\nu,2} = 0$, the two message sets

$$\mathcal{A}_\nu^1(\gamma_\nu) = \big\{\boldsymbol{u}_\nu \in \mathcal{U}_\nu^n \,:\, \mathrm{P}_\nu(\boldsymbol{u}_\nu) \geq \gamma_\nu^n\big\}, \tag{A59}$$

$$\mathcal{A}_\nu^2(\gamma_\nu) = \big\{\boldsymbol{u}_\nu \in \mathcal{U}_\nu^n \,:\, \mathrm{P}_\nu(\boldsymbol{u}_\nu) < \gamma_\nu^n\big\}, \tag{A60}$$

are specified using a single threshold γ_ν for each user $\nu \in \{1,2\}$. With some abuse of notation, we include the optimization w. r. t. $\lambda_{\mathcal{N}}^{\mathrm{L,U}}$ and make explicit the dependence on the thresholds $\gamma_{\mathcal{N}}$ in the expression of the source function $E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}$ in Equation (45), namely

$$E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}(\rho, P_{\mathcal{N}}, \gamma_{\mathcal{N}}) = \min_{\lambda_{\mathcal{N}}^{\mathrm{L,U}} \geq 0} \log \sum_{u_{\tau^c}} \left(\sum_{u_\tau} P_{\mathcal{N}}(u_{\mathcal{N}})^{\frac{1}{1+\rho}} \Lambda_{\mathcal{N}}^{i_{\mathcal{N}}}(u_{\mathcal{N}}) \right)^{1+\rho}. \tag{A61}$$

For $i_\nu = 1$, the set $\mathcal{A}_\nu^1(\gamma_\nu)$ in Equation (A59) has no upper threshold, hence we find that the optimal parameter λ_ν^{U} in this case is $\hat{\lambda}_\nu^{\mathrm{U}} = 0$. Similarly for $i_\nu = 2$, we obtain that $\hat{\lambda}_\nu^{\mathrm{L}} = 0$. As a consequence and without any loss of generality, we define $\lambda_\nu = \lambda_\nu^{\mathrm{L}}$ for $i_\nu = 1$, and $\lambda_\nu = \lambda_\nu^{\mathrm{U}}$ for $i_\nu = 2$, and further simplify Equation (A61) to the following optimization problem

$$E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}(\rho, P_{\mathcal{N}}, \gamma_{\mathcal{N}}) = \min_{\lambda_{\mathcal{N}} \geq 0} \log \sum_{u_{\tau^c}} \left(\sum_{u_\tau} P_{\mathcal{N}}(u_{\mathcal{N}})^{\frac{1}{1+\rho}} \left(\frac{\gamma_1}{P_1(u_1)} \right)^{(-1)^{i_1}\lambda_1} \left(\frac{\gamma_2}{P_2(u_2)} \right)^{(-1)^{i_2}\lambda_2} \right)^{1+\rho}, \tag{A62}$$

where we also used the definition of the functions $\Lambda_\sigma^{i_\sigma}$ in Equation (32) with $\sigma = \{1,2\}$. We recall that P_1 and P_2 are the marginal distributions for users $\nu = 1$ and $\nu = 2$, respectively, and the indices $i_\nu \in \{1,2\}$ indicate that user ν transmits a source message selected from the class $\mathcal{A}_\nu^{i_\nu}(\gamma_\nu)$ in Equations (A59) and (A60). It can be shown that the objective function in the r. h. s. of Equation (A62) is convex w. r. t. both λ_1 and λ_2. Hence, the minimizers $\hat{\lambda}_1$ and $\hat{\lambda}_2$ in the source function $E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}(\rho, P_{\mathcal{N}}, \gamma_{\mathcal{N}})$ are respectively given by $\hat{\lambda}_1 = \max\{\lambda_1^\star, 0\}$ and $\hat{\lambda}_2 = \max\{\lambda_2^\star, 0\}$, where $\lambda_1^\star$ and $\lambda_2^\star$ are the unique solution after setting the partial derivatives of the r. h. s. of Equation (A62) to zero. Two special cases can be obtained from Equation (A62).

The first case is when $\gamma_\nu = 1$ for $\nu \in \{1,2\}$, implying that no message partition happens whatsoever. In such a case, we have that $\hat{\lambda}_1 = \hat{\lambda}_2 = 0$ and Equation (A62) reduces to the joint source-channel coding source function for correlated-sources in Equation (34), i.e.,

$$E_{\mathrm{s},\tau}(\rho, P_{\mathcal{N}}) = \log \sum_{u_{\tau^c}} \left(\sum_{u_\tau} P_{\mathcal{N}}(u_{\mathcal{N}})^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{A63}$$

The second one is the case of independent sources. Substituting $P_{\mathcal{N}} = P_1 P_2$ in Equation (A62), after some algebra, we obtain that $E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}(\rho, P_{\mathcal{N}}, \gamma_{\mathcal{N}})$ can be split into two terms as

$$E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}(\rho, P_{\mathcal{N}}, \gamma_{\mathcal{N}}) = E_{\mathrm{s}}^{i_\tau}(\rho, P_\tau, \gamma_\tau) + E_{\mathrm{s}}^{i_{\tau^c}}(0, P_{\tau^c}, \gamma_{\tau^c}), \tag{A64}$$

where we defined the function $E_{\mathrm{s}}^i(\rho, P, \gamma)$ as

$$E_{\mathrm{s}}^i(\rho, P, \gamma) = \min_{\lambda \geq 0} \log \left(\sum_u P(u)^{\frac{1}{1+\rho}} \left(\frac{\gamma}{P(u)} \right)^{(-1)^i \lambda} \right)^{1+\rho} \tag{A65}$$

for arbitrary class index $i \in \{1,2\}$, source distribution P and threshold γ. First, we find that the unique solution after setting the derivative of the r. h. s. of Equation (A65) to zero, denoted as $\lambda^\star$, is implicitly given by

$$\frac{\sum_u P(u)^{\frac{1}{1+\alpha^\star}} \log P(u)}{\sum_u P(u)^{\frac{1}{1+\alpha^\star}}} = \log(\gamma), \tag{A66}$$

where we made the convenient change of variable

$$\frac{1}{1+\alpha^\star} = \frac{1}{1+\rho} - (-1)^i \lambda^\star. \tag{A67}$$

Although not made explicit, $\lambda^\star$ depends on the triplet (i, P, ρ, γ). When $\lambda^\star < 0$, or equivalently when

$$(-1)^i \left(\frac{1}{1+\rho} - \frac{1}{1+\alpha^\star} \right) < 0 \tag{A68}$$

we have that $\hat{\lambda} = \max(0, \lambda^\star) = 0$, implying that Equation (A65) simplifies to

$$E_{\mathrm{s}}^i(\rho, P, \gamma) = E_{\mathrm{s}}(\rho, P), \tag{A69}$$

where $E_{\mathrm{s}}(\rho, P)$ is the Gallager source function

$$E_{\mathrm{s}}(\rho, P) = \log \left(\sum_u P(u)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \tag{A70}$$

Otherwise, when $\hat{\lambda} = \lambda^\star \geq 0$, a regime given by the following inequality

$$(-1)^i \left(\frac{1}{1+\rho} - \frac{1}{1+\alpha^\star} \right) \geq 0 \tag{A71}$$

we may substitute $\lambda = \lambda^\star$ in the objective function in Equation (A65) to obtain

$$E_{\mathrm{s}}^i(\rho, P, \gamma) = (1+\rho) \log \left(\sum_u P(u)^{\frac{1}{1+\alpha^\star}} \right) + \frac{\alpha^\star - \rho}{1+\alpha^\star} \log(\gamma), \tag{A72}$$

where we wrote the expression in terms of $\alpha^\star$. Using Equation (A66) into Equation (A72) to replace $\log(\gamma)$, we get

$$E_{\mathrm{s}}^i(\rho, P, \gamma) = (1+\rho) \log \left(\sum_u P(u)^{\frac{1}{1+\alpha^\star}} \right) + \frac{\alpha^\star - \rho}{1+\alpha^\star} \frac{\sum_u P(u)^{\frac{1}{1+\alpha^\star}} \log P(u)}{\sum_u P(u)^{\frac{1}{1+\alpha^\star}}}. \tag{A73}$$

After some algebra, we are able to express the former equation in terms of the derivative of the E_{s}-function in Equation (A70), given by

$$E_{\mathrm{s}}'(\rho, P) = \log \left(\sum_u P(u)^{\frac{1}{1+\rho}} \right) - \frac{1}{1+\rho} \frac{\sum_u P(u)^{\frac{1}{1+\rho}} \log P(u)}{\sum_u P(u)^{\frac{1}{1+\rho}}}, \tag{A74}$$

and the E_{s}-function itself, as

$$E_{\mathrm{s}}^i(\rho, P, \gamma) = E_{\mathrm{s}}(\alpha^\star, P) + (\rho - \alpha^\star) E_{\mathrm{s}}'(\alpha^\star, P). \tag{A75}$$

We may finally combine Equations (A69) and (A75), with the respective ranges in Equations (A68) and (A71) to write the $E_{\mathrm{s}}^i(\rho, P, \gamma)$ function in Equation (A65) piecewise as

$$E_{\mathrm{s}}^1(\rho, P, \gamma) = \begin{cases} E_{\mathrm{s}}(\rho, P) & \frac{1}{1+\rho} \geq \frac{1}{1+\alpha^\star}, \\ E_{\mathrm{s}}(\alpha^\star, P) + E_{\mathrm{s}}'(\alpha^\star)(\rho - \alpha^\star) & \frac{1}{1+\rho} < \frac{1}{1+\alpha^\star}, \end{cases} \tag{A76}$$

and

$$E_{\mathrm{s}}^{2}(\rho,P,\gamma)=\begin{cases}E_{\mathrm{s}}(\rho,P) & \frac{1}{1+\rho}<\frac{1}{1+\alpha^{\star}},\\ E_{\mathrm{s}}(\alpha^{\star},P)+E_{\mathrm{s}}'(\alpha^{\star})(\rho-\alpha^{\star}) & \frac{1}{1+\rho}\geq\frac{1}{1+\alpha^{\star}}.\end{cases}\tag{A77}$$

where $\alpha^{\star}$ is the solution to the implicit Equation (A66), hence recovering the source error exponent functions of the md-iid ensemble described in ref. [8] (Lemma 1). The source functions $E_{\mathrm{s}}^{1}(\rho,P,\gamma)$ and $E_{\mathrm{s}}^{2}(\rho,P,\gamma)$ follow the Gallager function in Equation (A70) for a certain interval of ρ, and are the straight-line tangent beyond that interval. The tangent point $\alpha^{\star}$ is a function of the distribution P and of the multi-class threshold γ.

Once $E_{\mathrm{s}}^{i}(\rho,P,\gamma)$ in Equation (A65) is fully characterized, we may now discuss the correlated-sources error function $E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}(\rho,P_{\mathcal{N}},\gamma_{\mathcal{N}})$ in Equation (A64) in terms of the error type τ. We start with the third error type $\tau=\{1,2\}$, for which since $\tau^{c}=\emptyset$, we have that

$$E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}(\rho,P_{\mathcal{N}},\gamma_{\mathcal{N}})=E_{\mathrm{s}}^{i_1}(\rho,P_1,\gamma_1)+E_{\mathrm{s}}^{i_2}(\rho,P_2,\gamma_2),\tag{A78}$$

namely the superposition of two E_{s}^{i} functions as the ones in Equations (A76) and (A77), one for each user. For the remaining of this appendix, we consider the more informative error types $\tau=\{1\}$ and $\tau=\{2\}$ for the four possible pairs of class indices i_1 and i_2 in Equations (A59) and (A60), since in this case $E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}$ in Equation (A64) is either directly an $E_{\mathrm{s}}(\rho,P_{\tau})$ function or the straight-line tangent to it, in both cases shifted by a constant term given by $E_{\mathrm{s}}^{i_{\tau^{c}}}(0,P_{\tau^{c}},\gamma_{\tau^{c}})$.

Figure A1 shows the family of $E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}$ source functions respectively for independent and correlated sources, as a function of ρ where $P_{\mathcal{N}}$ given by Equation (60) and $\tau=\{1\}$. For independent sources, we observe that the source functions $E_{\mathrm{s},\tau}^{1,1}$ and $E_{\mathrm{s},\tau}^{2,1}$ follow the solid blue line depicting $\mathrm{E}_{s}(\rho,P_{\tau})$ as in Equation (A70) for a certain interval of ρ, and then take the tangent line beyond. A similar behavior is observed for the sources functions $E_{\mathrm{s},\tau}^{1,2}$ and $E_{\mathrm{s},\tau}^{2,2}$, which in this case follow or are tangent to the solid black line, the solid blue Gallager's source function shifted by the constant function $E_{\mathrm{s}}^{i_{\tau^{c}}}(0,P_{\tau^{c}},\gamma_{\tau^{c}})$ as in Equation (A64).

For correlated sources, the source functions $E_{\mathrm{s},\tau}^{1,1}$ and $E_{\mathrm{s},\tau}^{2,1}$ follow the generalized Gallager's source function given by Equation (A63) for a certain interval, but unlike independent sources they are not straight lines but a *curve* tangent to $\mathrm{E}_{s,\tau}$ beyond that interval. Some intuition about this fact can be gained from the primal form of the source function $E_{\mathrm{s},\tau}^{i_{\mathcal{N}}}$. Consider, for instance, the source function $E_{\mathrm{s},\tau}^{2,1}$ in Figure A1 for correlated sources, for which $i_1=2$ and $i_2=1$. The primal form of this source function $E_{\mathrm{s},\tau}^{2,1}$ can be obtained as a constrained optimization problem w. r. t. some auxiliary joint distribution $\hat{P}_{\mathcal{N}}$. The interval in ρ where $E_{\mathrm{s},\tau}^{2,1}$ does not follow $E_{\mathrm{s},\tau}$ in the dual form (approximately for $\rho\leq 0.5$ in the figure) corresponds to the case where only one of the two constraints on the auxiliary distribution $\hat{P}_{\mathcal{N}}$ is actually active in the primal form, where the constraint is given by $\sum_{u_{\mathcal{N}}}\hat{P}_{\mathcal{N}}(u_{\mathcal{N}})\log P_{\nu}(u_{\nu})=\log(\gamma_{\nu})$. This implies that, unlike the case of independent sources where each source has its auxiliary distribution $\hat{P}_1$ and $\hat{P}_2$ constrained, for correlated sources the joint auxiliary distribution $\hat{P}_{\mathcal{N}}$ is not fully constrained but is the union of joint distributions with one constrained marginal distribution. This partial constraint manifests as a curve in ρ, rather than a straight line, in the dual form. A similar behavior is observed for $E_{\mathrm{s},\tau}^{1,2}$ and $E_{\mathrm{s},\tau}^{2,2}$, which instead of following the source function for joint source-channel coding in Equation (A63) for some intervals of ρ, they follow the curve

$$\min_{\lambda_{\nu}\geq 0}\log\sum_{u_{\tau^{c}}}\left(\sum_{u_{\tau}}P_{\mathcal{N}}(u_{\mathcal{N}})^{\frac{1}{1+\rho}}\left(\frac{P_{\nu}(u_{\nu})}{\gamma_{\nu}}\right)^{-\frac{(-1)^{i_{\nu}}\lambda_{\nu}}{1+\rho}}\right)^{1+\rho},\tag{A79}$$

corresponding to Equation (A62) when the constraint for one source is not active, i.e., $\hat{\lambda}_{\nu^{c}}=0$.

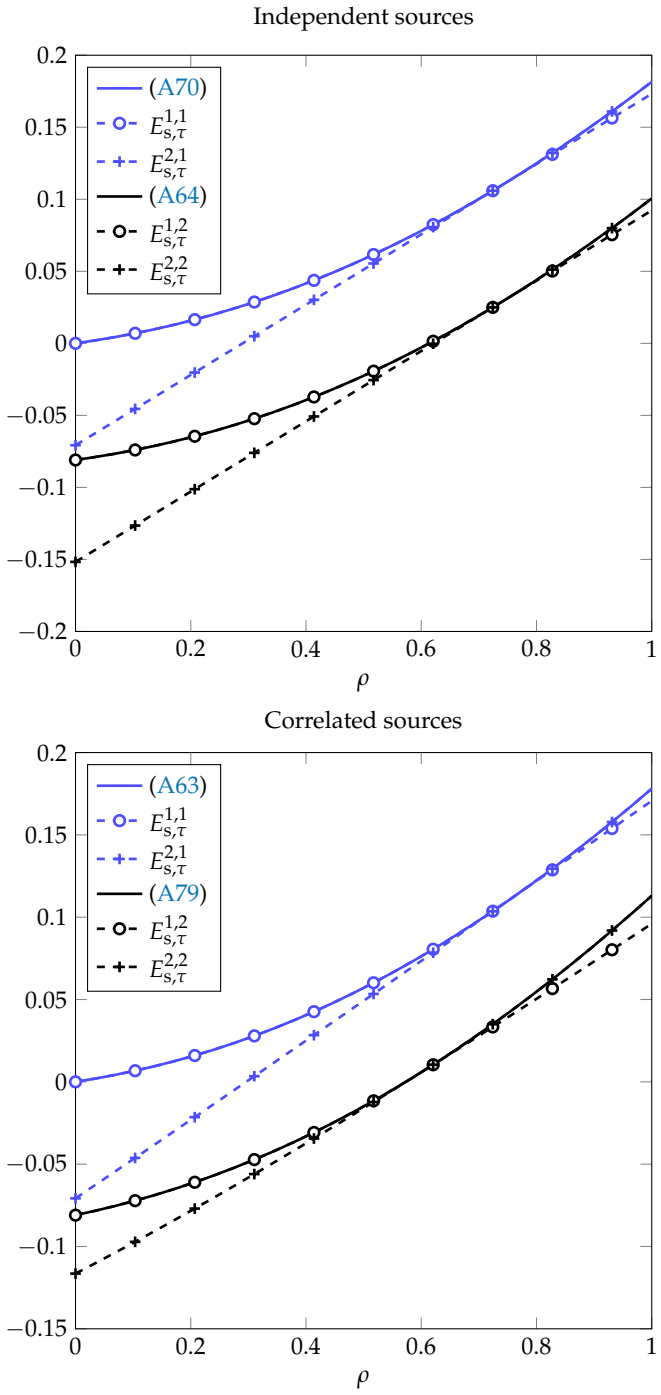

Figure A1. Example of the source functions $E_{s,\tau}^{i,N}$ in Equation (A62) for independent and correlated sources and error type $\tau \in \{\{1\},\{2\}\}$.

References

1. Gallager, R. *Information Theory and Reliable Communication*; John Wiley & Sons: Hoboken, NJ, USA, 1968.
2. Polyanskiy, Y.; Poor, H.V.; Verdú, S. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory* **2010**, *56*, 2307–2359. [CrossRef]
3. Jelinek, F. *Probabilistic Information Theory-Discrete and Memoryless Models*; McGraw-Hill Book Company: New York, NY, USA, 1968.
4. Shannon, C.; Gallager, R.; Berlekamp, E. Lower bounds to error probability for coding on discrete memoryless channels. I. *Inf. Control* **1967**, *10*, 65–103. [CrossRef]
5. Shannon, C.; Gallager, R.; Berlekamp, E. Lower bounds to error probability for coding on discrete memoryless channels. II. *Inf. and Control* **1967**, *10*, 522–552. [CrossRef]

6. Csiszár, I. Joint source-channel error exponent. *Probl. Control Inf. Theory* **1980**, *9*, 315–328.
7. Zhong, Y.; Alajaji, F.; Campbell, L.L. On the joint source-channel coding error exponent for discrete memoryless systems. *IEEE Trans. Inf. Theory* **2006**, *52*, 1450–1468. [CrossRef]
8. Tauste Campo, A.; Vazquez-Vilar, G.; Guillén i Fàbregas, A.; Martinez, A.; Koch, T. A derivation of the source-channel error exponent using nonidentical product distributions. *IEEE Trans. Inf. Theory* **2014**, *60*, 3209–3217. [CrossRef]
9. Ahlswede, R. Multi-way communication channels. In Proceedings of the 2nd International Symposium on Information Theory, Tsaghkadzor, Armenia, USSR, 2–8 September 1971.
10. Slepian, D.; Wolf, J.K. A coding theorem for multiple access channel with correlated sources. *Bell Syst. Tech. J.* **1973**, *52*, 1037–1076. [CrossRef]
11. Cover, T.M.; El Gamal, A.; Salehi, M. Multiple access channels with arbitrarily correlated sources. *IEEE Trans. Inf. Theory* **1980**, *26*, 648–657. [CrossRef]
12. Gallager, R. A perspective on multiple access channels. *IEEE Trans. Inf. Theory* **1985**, *31*, 124–142. [CrossRef]
13. Liu, Y.S.; Hughes, B. A new universal random coding bound for the multiple-access channel. *IEEE Trans. Inf. Theory* **1996**, *42*, 376–386.
14. Lapidoth, A. Mismatched decoding and the multiple-access channel. *IEEE Trans. Inf. Theory* **1996**, *42*, 1439–1452. [CrossRef]
15. Haim, E.; Kochman, Y.; Erez, U. Improving the MAC error exponent using distributed structure. In Proceedings of the 2011 IEEE International Symposium on Information Theory (ISIT), Saint Petersburg, Russia, 31 July–5 August 2011.
16. Cai, N. The maximum error probability criterion, random encoder, and feedback, in multiple input channels. *Entropy* **2014**, *16*, 1211–1242. [CrossRef]
17. Nazari, A.; Anastasopoulos, A.; Pradhan, S.S. Error exponent for multiple-access channels: Lower bounds. *IEEE Trans. Inf. Theory* **2014**, *60*, 5095–5115. [CrossRef]
18. Nazari, A.; Pradhan, S.S.; Anastasopoulos, A. Error exponent for multiple access channels: Upper bounds. *IEEE Trans. Inf. Theory* **2015**, *61*, 3605–3621. [CrossRef]
19. Farkas, L.; Kói, T. Random access and source-channel coding error exponents for multiple access channels. *IEEE Trans. Inf. Theory* **2015**, *61*, 3029–3040. [CrossRef]
20. Rezazadeh, A.; Font-Segura, J.; Martinez, A.; Guillén i Fàbregas, A. Multiple-Access Channel with Independent Sources: Error Exponent Analysis. In Proceedings of the 2018 IEEE Information Theory Workshop (ITW), Guangzhou, China, 25–29 November 2018.
21. Dueck, G. A note on the multiple access channel with correlated sources (Corresp.). *IEEE Trans. Inf. Theory* **1981**, *27*, 232–235. [CrossRef]
22. Padakandla, A. Communicating correlated sources over MAC and interference channels II: Joint source-channel coding. *IEEE Trans. Inf. Theory* **2021**. [CrossRef]
23. Padakandla, A. Communicating correlated sources over a MAC. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017.
24. Heidari, M.; Shirani, F.; Pradhan, S.S. New sufficient conditions for multiple-access channel with correlated sources. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016.
25. Rezazadeh, A.; Font-Segura, J.; Martinez, A.; Guillén i Fàbregas, A. An achievable error exponent for the multiple access channel with correlated sources. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017.
26. Scarlett, J.; Martinez, A.; Guillén i Fàbregas, A. Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations. *IEEE Trans. Inf. Theory* **2014**, *60*, 2647–2666. [CrossRef]
27. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. Journal* **1948**, *27*, 379–423. [CrossRef]
28. Rezazadeh, A.; Font-Segura, J.; Martinez, A.; Guillén i Fàbregas, A. Joint source-channel coding for the multiple-access channel with correlated sources. In Proceedings of the 2019 International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019.
29. Gallager, R. Fixed Composition Arguments and Lower Bounds to Error Probability. Available online: http://web.mit.edu/gallager/www/notes/notes5.pdf (accessed on 3 May 2021).
30. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.
31. Rezazadeh, A. Error Exponent Analysis for the Multiple-Access Channel with Correlated Sources. Ph.D. Thesis, Universitat Pompeu Fabra, Barcelona, Spain, 2019.
32. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006.
33. Hardy, G.H.; Littlewood, J.E.; Pólya, G. *Inequalities*, 2nd ed.; Cambridge University Press: Cambridge, UK, 1934.