Lossless Source Coding in the Point-to-Point, Multiple Access, and Random Access Scenarios

Shuqing Chen, *Graduate Student Member, IEEE,* Michelle Effros, *Fellow, IEEE,*
and Victoria Kostina, *Member, IEEE**

Abstract—This work studies point-to-point, multiple access, and random access lossless source coding in the finite-blocklength regime. In each scenario, a random coding technique is developed and used to analyze third-order coding performance. Asymptotic results include a third-order characterization of the Slepian-Wolf rate region with an improved converse that relies on a connection to composite hypothesis testing. For dependent sources, the result implies that the independent encoders used by Slepian-Wolf codes can achieve the same third-order-optimal performance as a single joint encoder. The concept of random access source coding is introduced to generalize multiple access (Slepian-Wolf) source coding to the case where encoders decide independently whether or not to participate and the set of participating encoders is unknown *a priori* to both the encoders and the decoder. The proposed random access source coding strategy employs rateless coding with scheduled feedback. A random coding argument proves the existence of a single deterministic code of this structure that simultaneously achieves the third-order-optimal Slepian-Wolf performance for each possible active encoder set.

Index Terms—Lossless source coding, Slepian-Wolf, random access, finite blocklength, random coding, non-asymptotic information theory, Gaussian approximation, hypothesis testing, meta-converse.

I. Introduction

WE study the fundamental limits of fixed-length, finite-blocklength lossless source coding in three scenarios:

1) *Point-to-point*: A single source is compressed by a single encoder and decompressed by a single decoder.
2) *Multiple access*: Each source in a fixed set of sources is compressed by an independent encoder; all sources are decompressed by a joint decoder.
3) *Random access*: Each active source from some set of possible sources is compressed by an independent encoder; all active sources are decompressed by a joint decoder.

The information-theoretic limit in any lossless source coding scenario is the set of code sizes or rates at which a desired level of reconstruction error is achievable. Shannon's theory [3] analyzes this fundamental limit by allowing an arbitrarily long encoding blocklength in order to obtain a vanishing error probability. Finite-blocklength limits [4]–[7], which are of particular interest in delay-sensitive and computationally-constrained coding environments, allow a non-vanishing error probability and study refined asymptotics of the rates achievable with encoding blocklength n. Due to their non-vanishing error probability, the resulting codes are sometimes called "almost-lossless" source codes. We here use the term "source coding" to refer to this almost-lossless coding paradigm.

In point-to-point source coding, non-asymptotic bounds and asymptotic expansions of the minimum achievable rate appear in [4], [6], [8]–[10]. In [6], Kontoyiannis and Verdú analyze the optimal code to give a *third-order* characterization of the minimum achievable rate $R^*(n, \epsilon)$ at blocklength n and error probability ϵ. For a finite-alphabet, stationary, memoryless source with single-letter distribution P_X, entropy $H(X)$, and varentropy $V(X) > 0$,

$$R^*(n, \epsilon) \approx H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n}, \qquad (1)$$

where $Q^{-1}(\cdot)$ is the inverse complementary Gaussian distribution function, and any higher-order term is bounded by $O\left(\frac{1}{n}\right)$.

For a multiple access source code (MASC), also known as a Slepian-Wolf (SW) source code [11], the fundamental limit is the set of achievable rate tuples known as the rate region. The *first-order* rate region for stationary, memoryless and general sources appears in [11] and [9], [12], respectively. *Second-order* asymptotic expansions of the MASC rate region for stationary, memoryless sources appear in [13], [14]. Tan and Kosut's characterization [13] is similar in form to the first two terms of (1), with varentropy $V(X)$ replaced by the entropy dispersion matrix and third-order term bounded by $O\left(\frac{\log n}{n}\right)$.

For point-to-point source coding, our contributions include non-asymptotic characterizations of the performance of randomly designed codes using threshold and maximum-likelihood decoders. The former analysis demonstrates that combining random coding with the best possible threshold decoder cannot achieve $-\frac{\log n}{2n}$ in the third-order term in (1), and thus it is strictly sub-optimal. The latter shows that combining random coding with maximum likelihood decoding achieves the first three terms in (1). We derive both bounds by deriving and analyzing a source coding analog to the random coding union (RCU) bound from channel coding [5, Th. 16]. Our asymptotic expansion is achieved by a random code rather than the optimal code from [6]. Thus, there is no loss (up to

the third-order term) due to random code design, which in turn shows that many codes have near-optimal performance; further, since our RCU bound holds when restricted to linear compressors, there are many good linear codes. The RCU bound is also important because it generalizes to the MASC and other scenarios where the optimal code is not known.

Our MASC RCU bound yields a new MASC achievability bound (Theorem 18). Establishing a link to composite hypothesis testing (HT) yields a new MASC HT converse (Theorem 19), which extends the meta-converse for channel coding in [5] to source coding with multiple encoders. This converse recovers and improves the previous converse due to Han [9, Lemma 7.2.2] and is equivalent to the LP-based converse of Jose and Kulkarni [15], which is the current best MASC converse. Our analysis of composite HT, including both non-asymptotic and asymptotic characterizations, develops tools with potential application in other multiple-terminal communication scenarios and beyond. The MASC RCU bound and HT converse together yield the *third-order* MASC rate region for stationary, memoryless sources (Theorem 20), revealing a $-\frac{\log n}{2n}$ third-order term that is independent of the number of encoders. This tightens the $O\!\left(\frac{\log n}{n}\right)$ third-order bound from [13], which grows linearly with the source alphabet size and exponentially with the number of encoders. For dependent sources, the MASC's third-order-optimal sum rate equals the third-order-optimal rate achievable through joint encoding.

While a MASC assumes a fixed, known collection of encoders, the set of transmitters communicating with a given access point in applications like sensor networks, the internet of things, and random access communication may be unknown or time-varying. The information theory literature treats the resulting *channel coding* challenges in papers such as [16]–[18]. We introduce the notion of a *random access source code (RASC)* and tackle the resulting source coding challenges. The RASC extends the MASC to scenarios where some encoders are inactive, and the decoder seeks to reliably reconstruct the sources associated with the active encoders assuming that the set of active encoders is unknown *a priori*.

We propose and analyze a robust RASC with rateless encoders that transmit codewords symbol by symbol until the receiver tells them to stop. Unlike typical rateless codes, which allow arbitrary decoding times [19]–[22], our code employs a small set of decoding times. Single-bit feedback from the decoder to all encoders at each potential decoding time tells the encoders whether or not to continue transmitting.

We demonstrate (Theorem 24) that there exists a single deterministic RASC that *simultaneously* achieves, for every possible set of active encoders, the third-order-optimal MASC performance for the active source set. Since traditional random coding arguments do not guarantee the existence of a single deterministic code that meets multiple independent constraints, prior code designs for multiple-constraint scenarios (e.g., [21]) employ a family of codes indexed using common randomness shared by all communicators. We develop an alternative approach, deriving a refined random coding argument (Lemma 25) that demonstrates the existence of a single deterministic code that meets all our constraints simultaneously; this technique may eliminate the need for common randomness

in other communication scenarios. For stationary, memoryless, permutation-invariant sources, employing identical encoders at all transmitters reduces RASC design complexity.

Except where noted, all presented source coding results apply to both finite and countably infinite source alphabets.

The organization of this paper is as follows. Section II defines notation. Section III treats (point-to-point) source coding. Section IV studies composite HT, developing general tools for multiple-encoder communication scenarios. Section V treats the MASC. Section VI introduces and studies the RASC. Each of Sections III, V, and VI follows a similar flow:

1) For the (point-to-point) source code: Section III-A defines the problem. Section III-B provides historical background. Section III-C presents our new random coding achievability bounds and their asymptotic expansions.
2) For the MASC: Section V-A gives definitions. Section V-B provides historical background. Section V-C presents new non-asymptotic bounds. Section V-D presents the third-order MASC characterization, comparing MASC and point-to-point source coding performance. Section V-E bounds the impact of limited feedback (and cooperation) on the third-order-optimal MASC region.
3) For the RASC: Section VI-A defines the problem and describes our proposed code. Section VI-B highlights related work. Section VI-C derives converse and achievability characterizations for our proposed code's finite-blocklength performance. Section VI-D treats the simplified code for permutation-invariant sources.

Section VII contains concluding remarks. Proofs of auxiliary results appear in the appendices.

II. NOTATION

For any positive integer i, let $[i] \triangleq \{1, \ldots, i\}$. We use uppercase letters (e.g., X) for random variables, lowercase letters (e.g., x) for scalar values, calligraphic uppercase letters (e.g., $\mathcal{E}$) for subsets of a sample space (events) or index sets, and script uppercase letters (e.g., $\mathscr{Q}$) for subsets of a Euclidean space. We use both bold face and superscripts for vectors (e.g., $\mathbf{x} = x^n$, $\mathbf{1} = (1, \ldots, 1)$, and $\mathbf{0} = (0, \ldots, 0)$). Given a sequence $(x_1, x_2, \ldots)$ with element x_i in set $\mathcal{X}_i$ for each i and given an ordered index set $\mathcal{T} \subseteq \mathbb{N}$, we define vector $\mathbf{x}_{\mathcal{T}} \triangleq (x_i, \ i \in \mathcal{T})$ and set $\mathcal{X}_{\mathcal{T}} \triangleq \prod_{i \in \mathcal{T}} \mathcal{X}_i$. Given a set $\mathcal{X}$, $\mathcal{X}^n$ is the n-fold Cartesian product of $\mathcal{X}$. We denote matrices by sans serif uppercase letters (e.g., V) and the (i,j)-th element of matrix V by $[\mathsf{V}]_{i,j}$. Inequalities between two vectors of the same dimension indicate elementwise inequalities. Given vector $\mathbf{u} \in \mathbb{R}^d$ and set $\mathscr{Q} \subset \mathbb{R}^d$, $\mathbf{u} + \mathscr{Q}$ denotes the Minkowski sum of $\{\mathbf{u}\}$ and $\mathscr{Q}$, giving $\mathbf{u} + \mathscr{Q} \triangleq \{\mathbf{u} + \mathbf{q} : \mathbf{q} \in \mathscr{Q}\}$. For two functions $u(n)$ and $f(n)$, $u(n) = O(f(n))$ if there exist $c, n_0 \in \mathbb{R}_+$ such that $0 \leq u(n) \leq cf(n)$ for all $n > n_0$. For a d-dimensional function $\mathbf{u} : \mathbb{N} \to \mathbb{R}^d$, $\mathbf{u}(n) = O(f(n))\mathbf{1}$ if $u_i(n) = O(f(n))$ for all $i \in [d]$. For any finite set $\mathcal{A}$, $\mathcal{P}(\mathcal{A})$ represents the power set of $\mathcal{A}$ excluding the empty set, giving $\mathcal{P}(\mathcal{A}) \triangleq \{\mathcal{T} : \mathcal{T} \subseteq \mathcal{A}\} \backslash \emptyset$. We use $|\cdot|_+ \triangleq \max\{0, \cdot\}$. All uses of 'log' and 'exp', if not specified, employ an arbitrary common base, which determines the information unit.

Denote the standard and complementary Gaussian cumulative distribution functions (cdf) by $\Phi(z)$ and $Q(z)$, giving

$$\Phi(z) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} \, du \tag{2}$$

$$Q(z) \triangleq 1 - \Phi(z). \tag{3}$$

Function $Q^{-1}(\cdot)$ denotes the inverse of $Q(\cdot)$. The standard Gaussian probability density function is

$$\phi(z) \triangleq \Phi'(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \tag{4}$$

The d-dimensional generalization of the Gaussian cdf is

$$\Phi(\mathsf{V}; \mathbf{z}) \triangleq \Phi(\mathsf{V}; z_1, \dots, z_d) \tag{5}$$

$$\triangleq \frac{1}{\sqrt{(2\pi)^d |\mathsf{V}|}} \int_{-\infty}^{z_1} \dots \int_{-\infty}^{z_d} e^{-\frac{1}{2} \sum\limits_{i,j=1}^{d} u_i u_j [\mathsf{V}^{-1}]_{i,j}} \, du_d \dots du_1.$$

Given an ordered index set $\mathcal{T} \subset \mathbb{N}$, let $P_{\mathbf{X}_\mathcal{T}}$ be a distribution defined on countable alphabet $\mathcal{X}_\mathcal{T}$. For any $\mathcal{A}, \mathcal{B} \subseteq \mathcal{T}$ with $\mathcal{A} \cap \mathcal{B} = \emptyset$ and any $(\mathbf{x}_\mathcal{A}, \mathbf{x}_\mathcal{B}) \in \mathcal{X}_\mathcal{A} \times \mathcal{X}_\mathcal{B}$, the information and conditional information are defined as

$$\imath(\mathbf{x}_\mathcal{A}) \triangleq \log \frac{1}{P_{\mathbf{X}_\mathcal{A}}(\mathbf{x}_\mathcal{A})} \tag{6}$$

$$\imath(\mathbf{x}_\mathcal{A}|\mathbf{x}_\mathcal{B}) \triangleq \log \frac{1}{P_{\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B}}(\mathbf{x}_\mathcal{A}|\mathbf{x}_\mathcal{B})}. \tag{7}$$

The corresponding entropy, conditional entropy, varentropy, conditional varentropy, third centered moment of information, and third centered moment of conditional information are defined by, respectively,

$$H(\mathbf{X}_\mathcal{A}) \triangleq \mathbb{E}[\imath(\mathbf{X}_\mathcal{A})] \tag{8}$$

$$H(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B}) \triangleq \mathbb{E}[\imath(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B})] \tag{9}$$

$$V(\mathbf{X}_\mathcal{A}) \triangleq \mathrm{Var}[\imath(\mathbf{X}_\mathcal{A})] \tag{10}$$

$$V(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B}) \triangleq \mathrm{Var}[\imath(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B})] \tag{11}$$

$$T(\mathbf{X}_\mathcal{A}) \triangleq \mathbb{E}\big[|\imath(\mathbf{X}_\mathcal{A}) - H(\mathbf{X}_\mathcal{A})|^3\big] \tag{12}$$

$$T(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B}) \triangleq \mathbb{E}\big[|\imath(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B}) - H(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B})|^3\big]. \tag{13}$$

We also define random variables

$$V_c(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B}) \triangleq \mathbb{E}\Big[\big(\imath(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B}) - \mathbb{E}[\imath(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B})|\mathbf{X}_\mathcal{B}]\big)^2 \big| \mathbf{X}_\mathcal{B}\Big] \tag{14}$$

$$T_c(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B}) \triangleq \mathbb{E}\Big[|\imath(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B}) - \mathbb{E}[\imath(\mathbf{X}_\mathcal{A}|\mathbf{X}_\mathcal{B})|\mathbf{X}_\mathcal{B}]|^3 \big| \mathbf{X}_\mathcal{B}\Big]. \tag{15}$$

III. POINT-TO-POINT SOURCE CODING

A. Definitions

In point-to-point source coding, the encoder maps a discrete random variable X defined on finite or countably infinite alphabet $\mathcal{X}$ into a message from codebook $[M]$. The decoder reconstructs X from the compressed description. Formal definitions of codes and their information-theoretic limits follow. For prior definitions, see, for example, [9, Chapter 1].

Definition 1 (Point-to-point source code). *An (M, ϵ) code for a random variable X with discrete alphabet $\mathcal{X}$ comprises* an encoding function $\mathsf{f} \colon \mathcal{X} \to [M]$ and a decoding function $\mathsf{g} \colon [M] \to \mathcal{X}$ with error probability $\mathbb{P}[\mathsf{g}(\mathsf{f}(X)) \neq X] \leq \epsilon$.

Definition 2 (Block point-to-point source code). *An (n, M, ϵ) code is an (M, ϵ) code defined for a random vector X^n with discrete vector alphabet $\mathcal{X}^n$.*

Definition 3 (Minimum achievable rate). *The minimum code size $M^*(n, \epsilon)$ and rate $R^*(n, \epsilon)$ achievable at blocklength n and error probability ϵ are defined as*

$$M^*(n, \epsilon) \triangleq \min\{M : \exists\, (n, M, \epsilon) \text{ code}\}$$

$$R^*(n, \epsilon) \triangleq \frac{1}{n} \log M^*(n, \epsilon).$$

A discrete information source is a sequence of discrete random variables, $X_1, X_2, \dots$, specified by the transition probability kernels $P_{X_i|X^{i-1}}$, $i = 1, 2, \dots$ While Definition 2 applies to many classes of sources, including sources with memory and non-stationary sources, our asymptotic analysis focuses on stationary, memoryless sources, where $P_{X_i|X^{i-1}} = P_X$ for all $i = 1, 2, \dots$ (i.e., $X_1, X_2, \dots$ are i.i.d.).

B. Background

Shannon's source coding theorem [3] describes the fundamental limit on the asymptotic performance for lossless source coding on a stationary, memoryless source, giving

$$\lim_{n \to \infty} R^*(n, \epsilon) = H(X), \quad \forall \epsilon \in (0, 1). \tag{16}$$

In the finite-blocklength regime, Kontoyiannis and Verdú [6] characterize $R^*(n, \epsilon)$ using upper and lower bounds that match in their first three terms and show an $O\big(\frac{1}{n}\big)$ fourth-order gap.

Theorem 1 (Kontoyiannis and Verdú [6]). *Consider a stationary, memoryless source with finite alphabet $\mathcal{X}$, single-letter distribution P_X, and varentropy $V(X) > 0$. Then[1] (achievability) for all $0 < \epsilon \leq \frac{1}{2}$ and all[2] $n > \left(\frac{T(X)}{V(X)^{3/2}\epsilon}\right)^2$,*

$$R^*(n, \epsilon) \leq H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) - \frac{\log_2 n}{2n} \tag{17}$$

$$+ \frac{1}{n} \log_2 \left(\frac{\log_2 e}{\sqrt{2\pi V(X)}} + \frac{T(X)}{V(X)^{3/2}}\right)$$

$$+ \frac{1}{n} \frac{T(X)}{V(X) \phi\Big(\Phi^{-1}\big(\Phi(Q^{-1}(\epsilon)) + \frac{T(X)}{V(X)^{3/2}\sqrt{n}}\big)\Big)};$$

(converse) for all $0 < \epsilon \leq \frac{1}{2}$ and all

$$n > \frac{1}{4}\left(1 + \frac{T(X)}{2V(X)^{3/2}}\right)^2 \frac{1}{(\phi(Q^{-1}(\epsilon))Q^{-1}(\epsilon))^2}, \tag{18}$$

$$R^*(n, \epsilon) \geq H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) - \frac{\log_2 n}{2n}$$

$$- \frac{1}{n} \frac{T(X) + 2V(X)^{3/2}}{2V(X)\phi(Q^{-1}(\epsilon))}. \tag{19}$$

---

[1]These bounds, which are stated in a base-2 logarithmic scale in [6], hold for any base. The base of the logarithm determines the information unit.

[2]According to [6], the achievability bound holds for any $n \geq 1$. Notice, however, that it only becomes meaningful when $n > \left(\frac{T(X)}{V(X)^{3/2}\epsilon}\right)^2$.

Remark 1. Although [6, Theorem 1] restricts attention to $0 < \epsilon \leq \frac{1}{2}$ and $\mathcal{X}$ finite, the proof in [6] applies for all $0 < \epsilon < 1$ and any countable source alphabet, achieving the same first three terms in (17) and (19) and fourth-order term $\pm O\left(\frac{1}{n}\right)$ (which varies with ϵ) provided that the third centered moment $T(X)$ of information random variable X is finite.

Remark 2. When $V(X) = 0$, the source is uniformly distributed over a finite alphabet (i.e., non-redundant), and $H(X) = \log|\mathcal{X}|$. The optimal code maps any $1-\epsilon$ fraction of possible source outcomes to unique codewords, giving

$$1-\epsilon \leq \frac{M^*(n,\epsilon)}{|\mathcal{X}|^n} \leq 1-\epsilon + \frac{1}{|\mathcal{X}|^n}.$$

As a result, when P_X is uniform,

$$H(X) - \frac{1}{n}\log\frac{1}{1-\epsilon} \leq R^*(n,\epsilon)$$
$$\leq H(X) - \frac{1}{n}\log\frac{1}{1-\epsilon} + \frac{\log e}{n(1-\epsilon)}\exp(-nH(X)), \quad (20)$$

which matches (17) up to the second order (since $V(X) = 0$) but omits the $-\frac{\log n}{2n}$ third-order term.

Remark 3. While it is not captured by our notation, $R^*(n,\epsilon)$ is a function of P_X. Since the $-\frac{\log n}{2n}$ third-order term appears in (17) and (19) but not in (20), the *bound* on $R^*(n,\epsilon)$, when viewed as a function of P_X, is discontinuous at the point where P_X equals the uniform distribution on $\mathcal{X}$. In contrast, $R^*(n,\epsilon)$, which is known and calculable, is continuous. The problem arises because Berry-Esseen type bounds are loose for small $V(X)$. Thus for any finite n, the achievability bound in (17) blows up as $V(X) \to 0$. See Figure 1. Theorem 1 states that for any $V(X) > 0$ there exists some $n_0 = n_0(P_X, \epsilon)$ such that for all $n > n_0$, $R^*(n,\epsilon)$ behaves like $-\frac{\log n}{2n}$ in the third-order term; the smaller the value of $V(X)$, the larger n_0 must be.

Achievability results based on Shannon's random coding argument [3] are important because they do not require knowledge of the optimal code, which is available only in a few special communication scenarios (e.g., [6], [7]). The following *random coding* achievability bound[3] is obtained by assigning source realizations to codewords independently and uniformly at random. The threshold decoder decodes to $x \in \mathcal{X}$ if and only if x is a unique source realization that (i) is compatible with the observed codeword under the given (random) code design, and (ii) has information $\imath(x)$ below $\log M - \gamma$.

Theorem 2 (e.g. [24], [25, Th. 9.4]). *There exists an (M,ϵ) code for discrete random variable X such that*

$$\epsilon \leq \mathbb{P}[\imath(X) > \log M - \gamma] + \exp(-\gamma), \, \forall \gamma > 0. \quad (21)$$

Particularizing (21) to a stationary, memoryless source with single-letter distribution P_X satisfying $V(X) > 0$ and $T(X) < \infty$, choosing $\log M$ and γ optimally, and applying the Berry-Esseen inequality (see Theorem 6 below) gives

$$R^*(n,\epsilon) \leq H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}(\epsilon) + \frac{\log n}{2n} + O\left(\frac{1}{n}\right). \quad (22)$$

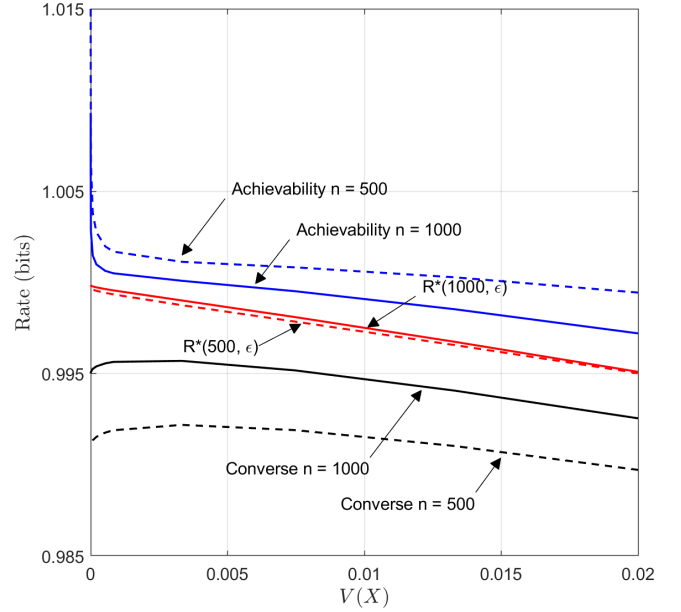[3]Tighter bounds based on the optimal code appear in [9, Lemma 1.3.1] and [23, Remark 5].

Fig. 1: Evaluations of the achievability bound in (17), the converse bound in (19), and the optimum $R^*(n,\epsilon)$, all shown as a function of $V(X) = p(1-p)(\log\frac{1-p}{p})^2$ for a Bernoulli-p source at $\epsilon = 0.1$.

Since the optimal application of Theorem 2 yields (22), which exceeds the bounds in Theorem 1 by $+\frac{\log n}{n}$ in the third-order term, we are left to wonder whether random code design, threshold decoding, or both yield third-order performance penalties. In [6, Th. 8], Kontoyiannis and Verdú precisely characterize the performance of a code designed with i.i.d. uniform random codeword generation and an optimal (maximum likelihood) decoder. Unfortunately, that result is difficult to use in the asymptotic analysis. In Section III-C Theorem 4, below, we derive a new random coding bound using a maximum likelihood decoder; this result demonstrates that random coding suffices to achieve the third-order optimal performance for a stationary, memoryless source.

C. New Achievability Bounds Based on Random Coding

We next use random code design to derive two new non-asymptotic achievability bounds for point-to-point source coding. We call these results the dependence testing (DT) bound and the random coding union (RCU) bound since they are the source coding analogues of the DT [5, Th. 17] and RCU [5, Th. 16] bounds in channel coding. The DT bound tightens Theorem 2, which is also based on threshold decoding.

Theorem 3 (DT bound). *Given a discrete random variable X, there exists an (M,ϵ) code with a threshold decoder for which*

$$\epsilon \leq \mathbb{E}[\exp\{-|\log M - \imath(X)|_+\}]. \quad (23)$$

Proof. Appendix A. ∎

The proof of Theorem 3 bounds the random coding performance of a threshold decoder with threshold $\log\gamma$ as

$$\epsilon \leq \mathbb{P}[\imath(X) > \log\gamma] + \frac{1}{M}\mathbb{U}[\imath(X) \leq \log\gamma], \quad (24)$$

where $\mathbb{U}[\cdot]$ denotes a mass with respect to the counting measure U_X on $\mathcal{X}$, which assigns unit weight to each $x \in \mathcal{X}$. As in a channel coding argument from [5], we apply the Neyman-Pearson lemma and find that the right-hand side of (24) equals $\frac{M+1}{M}$ times the minimum measure of the error event in a Bayesian binary hypothesis test between P_X with a priori probability $\frac{M}{M+1}$ and U_X with a priori probability $\frac{1}{M+1}$. (The Neyman-Pearson lemma generalizes to σ-finite measures like U_X [23, Remark 5].) This error measure is minimized by the test that compares the log likelihood ratio $\log \frac{U_X(X)}{P_X(X)}$ to the log ratio of a priori probabilities $\log \frac{M/(M+1)}{1/(M+1)}$, giving

$$H_0 : P_X, \text{ selected if } \imath(X) \leq \log M$$
$$H_1 : U_X, \text{ selected if } \imath(X) > \log M.$$

Taking $\gamma = M$ minimizes the right-hand side of (24), which implies that Theorem 3 is the tightest possible bound for random coding with threshold decoding.

Particularizing Theorem 3 to a stationary, memoryless source with a single-letter distribution P_X satisfying $V(X) > 0$ and $T(X) < \infty$ and invoking the Berry-Esseen inequality (see Theorem 6 below), we obtain the asymptotic expansion

$$R^*(n, \epsilon) \leq H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right). \quad (25)$$

Unfortunately, (25) is sub-optimal in its third-order term. Thus, random code design with threshold-based decoding fails to achieve the optimal third-order performance.

Next, we present the RCU bound, which employs random code design and maximum likelihood decoding.

Theorem 4 (RCU bound). *Given a discrete random variable X, there exists an (M, ϵ) code with a maximum likelihood decoder for which*

$$\epsilon \leq \mathbb{E}\left[\min\left\{1, \frac{1}{M}\mathbb{E}\left[\exp(\imath(\bar{X}))\mathbb{1}\{\imath(\bar{X}) \leq \imath(X)\}|X\right]\right\}\right], (26)$$

where $P_{X\bar{X}}(a, b) = P_X(a)P_X(b)$ for all $a, b \in \mathcal{X}$.

Proof. Our random code design randomly and independently draws encoder output $\mathsf{F}(x)$ for each $x \in \mathcal{X}$ from the uniform distribution on $[M]$. We use the maximum likelihood decoder

$$\mathsf{g}(c) = \arg\max_{x \in \mathcal{X}: \mathsf{F}(x)=c} P_X(x) = \arg\min_{x \in \mathcal{X}: \mathsf{F}(x)=c} \imath(x). \quad (27)$$

If multiple source symbols have the maximal probability mass, the decoder design chooses among them uniformly at random.

Under this random code construction, the expected error probability is bounded by the probability $\mathbb{P}[\mathcal{E}]$ of event

$$\mathcal{E} \triangleq \{\exists \bar{x} \in \mathcal{X}\backslash\{X\} \text{ s.t. } \imath(\bar{x}) \leq \imath(X), \mathsf{F}(\bar{x}) = \mathsf{F}(X)\}, \quad (28)$$

where probability measure $\mathbb{P}[\cdot]$ captures both the random source output X and the random encoding map F. The

resulting error bound is

$$\mathbb{E}[\mathbb{P}(\{g(F(X)) \neq X\}|F(\cdot))]$$
$$\leq \mathbb{P}[\mathcal{E}]$$
$$= \mathbb{E}\left[\mathbb{P}\left[\bigcup_{\bar{x} \in \mathcal{X}\backslash\{X\}} \{\imath(\bar{x}) \leq \imath(X), \mathsf{F}(\bar{x}) = \mathsf{F}(X)\}|X\right]\right] \quad (29)$$
$$\leq \mathbb{E}\left[\min\left\{1, \sum_{\substack{\bar{x} \in \mathcal{X}: \\ \bar{x} \neq X}} \mathbb{P}[\{\imath(\bar{x}) \leq \imath(X), \mathsf{F}(\bar{x}) = \mathsf{F}(X)\}|X]\right\}\right] \quad (30)$$
$$\leq \mathbb{E}\left[\min\left\{1, \frac{1}{M}\sum_{\bar{x} \in \mathcal{X}} \mathbb{1}\{\imath(\bar{x}) \leq \imath(X)\}\right\}\right] \quad (31)$$
$$= \mathbb{E}\left[\min\left\{1, \frac{1}{M}\mathbb{E}\left[\frac{1}{P_X(\bar{X})}\mathbb{1}\{\imath(\bar{X}) \leq \imath(X)\}|X\right]\right\}\right], \quad (32)$$

where (29) applies the law of iterated expectation, (30) bounds the probability by the minimum of the union bound and 1, (31) holds because the encoder outputs are drawn i.i.d. uniformly at random and independently of X, and (32) rewrites (31) in terms of the distribution $P_{X\bar{X}} = P_X P_X$.

The existence of the desired (M, ϵ) code follows since (32) equals the right-hand side of (26). ∎

Remark 4. By the argument employed in the proof of [25, Th. 9.5], we obtain the same RCU bound if we randomize only over linear encoding maps. Thus, there is no loss in performance when restricting to linear compressors.

We next show that the RCU bound recovers the first three terms of the achievability result in Theorem 1. Thus, the sub-optimal third-order terms in (22) and (25) result from the sub-optimal decoder rather than the random encoder design. This is important since optimal codes are not available for scenarios like the MASC studied in Section V, below.

Theorem 5 focuses on a stationary, memoryless source with single-letter distribution P_X satisfying

$$V(X) > 0 \quad (33)$$
$$T(X) < \infty. \quad (34)$$

Define constants

$$B \triangleq C_0 \frac{T(X)}{V(X)^{3/2}} \quad (35)$$
$$C \triangleq 2\left(\frac{\log 2}{\sqrt{2\pi V(X)}} + 2B(X)\right), \quad (36)$$

where C_0 is the absolute constant in the Berry-Esseen inequality for i.i.d. random variables. (See Theorem 6, below.)

Theorem 5 (Third-order-optimal achievability via random coding). *Consider a stationary, memoryless source satisfying the conditions in (33) and (34). For all $0 < \epsilon < 1$,*

$$R^*(n, \epsilon) \leq H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} + \xi(n), \quad (37)$$

where $\xi(n) = O\left(\frac{1}{n}\right)$ is bounded more precisely as follows.

1) For all $0 < \epsilon \leq \frac{1}{2}$ and $n > \left(\frac{B+C}{\epsilon}\right)^2$,

$$\xi(n) \leq \frac{1}{n}\log C + \frac{1}{n}\frac{B+C}{\phi\left(\Phi^{-1}\left(\Phi(Q^{-1}(\epsilon)) + \frac{B+C}{\sqrt{n}}\right)\right)}. \quad (38)$$

2) For all $\frac{1}{2} < \epsilon < 1$ and $n > \left(\frac{B+C}{\epsilon - \frac{1}{2}}\right)^2$,

$$\xi(n) \leq \frac{1}{n}\log C + \frac{1}{n}\frac{B+C}{\phi(Q^{-1}(\epsilon))}. \quad (39)$$

Before we show our proof of the asymptotic expansion in Theorem 5, we state two auxiliary results used in our analysis. The first is the classical Berry-Esseen inequality (e.g., [26, Chapter XVI.5]), stated here with the best known absolute constant C_0 from [27].

Theorem 6 (Berry-Esseen inequality)**.** *Let $Z_1, \ldots, Z_n$ be independent random variables such that $V \triangleq \frac{1}{n}\sum_{i=1}^n \mathrm{Var}[Z_i] > 0$ and $T \triangleq \frac{1}{n}\sum_{i=1}^n \mathbb{E}[|Z_i - \mathbb{E}[Z_i]|^3] < \infty$. Then for any real t and $n \geq 1$,*

$$\left|\mathbb{P}\left[\frac{1}{\sqrt{nV}}\sum_{i=1}^n(Z_i - \mathbb{E}[Z_i]) \geq t\right] - Q(t)\right| \leq \frac{C_0 T}{V^{3/2}\sqrt{n}}, \quad (40)$$

where $0.4097 \leq C_0 \leq 0.5583$ ($0.4097 \leq C_0 < 0.4690$ for identically distributed Z_i) [27].

We refer to $C_0 \cdot T/V^{3/2}$ as the Berry-Esseen constant.

The second result is from Polyanskiy et al. [5, Lemma 47].

Lemma 7 ([5, Lemma 47])**.** *In the setting of Theorem 6, it holds for any A and $n \geq 1$ that*

$$\mathbb{E}\left[\exp\left\{-\sum_{i=1}^n Z_i\right\}\mathbb{1}\left\{\sum_{i=1}^n Z_i \geq A\right\}\right]$$
$$\leq 2\left(\frac{\log 2}{\sqrt{2\pi V}} + 2C_0\frac{T}{V^{3/2}}\right)\frac{1}{\sqrt{n}}\exp(-A). \quad (41)$$

Proof of Theorem 5. We analyze the RCU bound of Theorem 4 for random variable X^n. For notational brevity, define

$$I_n \triangleq \imath(X^n) = \sum_{i=1}^n \imath(X_i), \quad \bar{I}_n \triangleq \imath(\bar{X}^n) = \sum_{i=1}^n \imath(\bar{X}_i). \quad (42)$$

By Theorem 4, there exists an (n, M, ϵ') code such that

$$\epsilon' \leq \mathbb{E}\left[\min\left\{1, \frac{1}{M}\mathbb{E}\left[\exp(\bar{I}_n)\mathbb{1}\{\bar{I}_n \leq I_n\}|X^n\right]\right\}\right], \quad (43)$$

where $P_{X^n \bar{X}^n} = P_X^n P_{\bar{X}}^n$, and each of I_n and $\bar{I}_n$ is a sum of i.i.d. random variables. Applying Lemma 7 with $Z_i = -\imath(\bar{X}_i)$ and $A = -I_n$ gives

$$\mathbb{E}\left[\exp(\bar{I}_n)\mathbb{1}\{\bar{I}_n \leq I_n\}|X^n\right] \leq \frac{C}{\sqrt{n}}\exp(I_n). \quad (44)$$

Plugging (44) in (43), we find

$$\epsilon' \leq \mathbb{E}\left[\min\left\{1, \frac{C}{M\sqrt{n}}\exp(I_n)\right\}\right] \quad (45)$$

$$= \mathbb{P}\left[I_n > \log\frac{M\sqrt{n}}{C}\right]$$

$$\quad + \frac{C}{M\sqrt{n}}\mathbb{E}\left[\exp(I_n)\mathbb{1}\left\{I_n \leq \log\frac{M\sqrt{n}}{C}\right\}\right] \quad (46)$$

$$\leq \mathbb{P}\left[I_n > \log M + \frac{1}{2}\log n - \log C\right] + \frac{C}{\sqrt{n}}, \quad (47)$$

where (45) plugs (44) into (43), (46) separates the cases $I_n > \log(M\sqrt{n}/C)$ and $I_n \leq \log(M\sqrt{n}/C)$, and (47) applies Lemma 7 to the second term in (46).

Denote for brevity

$$\delta_n \triangleq \frac{B+C}{\sqrt{n}}. \quad (48)$$

We now choose

$$\log M = nH(X) + \sqrt{nV(X)}Q^{-1}(\epsilon - \delta_n) - \frac{1}{2}\log n$$
$$+ \log C \quad (49)$$

and apply the Berry-Esseen inequality (Theorem 6) to (47), giving $\epsilon' \leq \epsilon$ and proving achievability bound

$$R^*(n, \epsilon) \leq \frac{\log M}{n}. \quad (50)$$

To obtain (37) from (49), we note that as long as $\delta_n < \epsilon$,

$$Q^{-1}(\epsilon - \delta_n) = \Phi^{-1}\left(\Phi(Q^{-1}(\epsilon)) + \delta_n\right) \quad (51)$$

$$= Q^{-1}(\epsilon) + \delta_n(\Phi^{-1})'(\xi_n) \quad (52)$$

$$= Q^{-1}(\epsilon) + \frac{\delta_n}{\phi(\Phi^{-1}(\xi_n))}, \quad (53)$$

where (51) applies the definition of the Gaussian cumulative distribution function $\Phi(\cdot)$ and its complement $Q(\cdot)$ from (2) and (3), (52) holds by a first-order Taylor bound for some $\xi_n \in \left[\Phi(Q^{-1}(\epsilon)), \Phi(Q^{-1}(\epsilon)) + \delta_n\right]$, and (53) holds by the inverse function theorem.

1) For $\epsilon \leq \frac{1}{2}$ and $\delta_n < \epsilon$, $\xi_n \geq \frac{1}{2}$ and $\phi(\Phi^{-1}(\xi_n))$ is decreasing in ξ_n. We can further bound the right-hand side of (53) and conclude that

$$Q^{-1}(\epsilon - \delta_n) \leq Q^{-1}(\epsilon) + \frac{\delta_n}{\phi(\Phi^{-1}(\Phi(Q^{-1}(\epsilon)) + \delta_n))}. \quad (54)$$

2) For $\epsilon > \frac{1}{2}$ and $\delta_n < \epsilon - \frac{1}{2}$, we have $\xi_n \leq \frac{1}{2}$ and $\phi(\Phi^{-1}(\xi_n))$ is increasing in ξ_n. We conclude that

$$Q^{-1}(\epsilon - \delta_n) \leq Q^{-1}(\epsilon) + \frac{\delta_n}{\phi(Q^{-1}(\epsilon))}. \quad (55)$$

Plugging (54) and (55) into (49) gives (38) and (39). $\blacksquare$

IV. COMPOSITE HYPOTHESIS TESTING

The meta-converse for channel coding [5, Th. 26][4] and its generalizations to lossy source coding [23] and joint source-channel coding [30], [31] apply binary hypothesis testing to derive converses in point-to-point communication problems. To extend this approach to multi-terminal coding (see, e.g., Section V Theorem 19, below), we develop a corresponding method using composite hypothesis testing. We first develop non-asymptotic tools and then analyze the asymptotics.

A composite hypothesis test $P_{Z|X} : \mathcal{X} \to \{0, 1\}$ tests a simple hypothesis against a composite hypothesis:

$$H_0 : X \sim P, \qquad\qquad \text{selected if } Z = 1$$
$$H_1 : X \sim Q_j \text{ for some } j \in [k], \text{ selected if } Z = 0,$$

---

[4]The quantum information theory literature contains an earlier approach to channel coding converses using binary hypothesis testing [28], [29, Ch. 4.6].

where X is the observation, P is the distribution under the simple hypothesis, and $\{Q_j\}_{j=1}^k$ is the collection of possible distributions under the composite hypothesis. The following definition generalizes the optimal β-function from binary to composite hypothesis testing. (See, for example, [32, Def. 1].)

Definition 4. *The set of achievable false-positive errors for power-α tests between distribution P and collection of distributions $\{Q_j\}_{j=1}^k$ is the subset of $[0,1]^k$ defined as*

$$\beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big) \triangleq \{\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k) :$$
$$\exists \, test \; s.t. \; \mathbb{P}[Z=1] \geq \alpha, \; \mathbb{Q}_j[Z=1] \leq \beta_j, \; \forall j \in [k]\}, \quad (56)$$

where $\mathbb{P}[\cdot]$ denotes a probability with respect to P, and for each $j \in [k]$, $\mathbb{Q}_j[\cdot]$ denotes a probability with respect to Q_j.

Like binary hypothesis tests (see [23, Remark 5]), composite hypothesis tests can be generalized to allow P and $\{Q_j\}_{j=1}^k$ to be σ-finite measures; in such cases, $\beta_\alpha(P, \{Q_j\}_{j=1}^k)$ may not be a subset of $[0,1]^k$. We apply this generalization in Section V-C2 to derive our new MASC converse.

In [32], Huang and Moulin study the asymptotics of the set $\beta_\alpha(P, \{Q_j\}_{j=1}^k)$, giving a third-order-optimal characterization [32, Th. 1]. As noted in [33, Appendix D], there is a gap in their converse proof (see also Remark 6, below). We here present a comprehensive analysis of composite hypothesis testing, starting with non-asymptotic characterizations and then particularizing them to give a new proof of [32, Th. 1].

A. Non-Asymptotic Bounds

The analysis of $\beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big)$ in [32] uses the test that achieves the minimal (boundary) points of that set. For each minimal point $\boldsymbol{\beta}$, there exists a vector $\mathbf{a} = (a_1, \ldots, a_k) \geq \mathbf{0}$, $\mathbf{a} \neq \mathbf{0}$, such that the generalized Neyman-Pearson test

$$P_{Z|X}(1|x) = \begin{cases} 1, & \text{for } x \text{ s.t. } P(x) > \sum_{j=1}^k a_j Q_j(x) \\ 0, & \text{for } x \text{ s.t. } P(x) < \sum_{j=1}^k a_j Q_j(x) \\ \lambda, & \text{for } x \text{ s.t. } P(x) = \sum_{j=1}^k a_j Q_j(x), \end{cases} \quad (57)$$

achieves $\boldsymbol{\beta}$; here $\lambda \in [0,1]$ is chosen so that $\mathbb{P}[Z=1] = \alpha$. While the above test is optimal, the achievability and converse bounds that follow simplify the asymptotic analysis.

Lemma 8 (Achievability). *For any $\gamma_j \geq 0$, $j \in [k]$, there exists a composite hypothesis test $P_{Z|X}$ for which*

$$\mathbb{P}[Z=1] = \mathbb{P}\left[\bigcap_{j\in[k]} \left\{\frac{P(X)}{Q_j(X)} \geq \gamma_j\right\}\right] \quad (58)$$

$$\mathbb{Q}_j[Z=1] \leq \mathbb{E}_P\left[\frac{Q_j(X)}{P(X)} 1\left\{\frac{P(X)}{Q_j(X)} \geq \gamma_j\right\}\right], \; \forall j \in [k]. \quad (59)$$

Proof. Fix any $\gamma_j \geq 0$, $j \in [k]$. Consider the (sub-optimal) likelihood-ratio threshold test[5]:

$$P_{Z|X}(1|x) = \begin{cases} 1, & \text{if } \frac{P(X)}{Q_j(X)} \geq \gamma_j, \; \forall j \in [k] \\ 0, & \text{otherwise.} \end{cases} \quad (60)$$

Under this test, (58) follows immediately, and (59) holds by

$$\mathbb{Q}_j[Z=1] = \mathbb{Q}_j\left[\bigcap_{j\in[k]} \left\{\frac{P(X)}{Q_j(X)} \geq \gamma_j\right\}\right] \quad (61)$$

$$\leq \mathbb{Q}_j\left[\frac{P(X)}{Q_j(X)} \geq \gamma_j\right] \quad (62)$$

$$= \sum_{x\in\mathcal{X}} P(x) \cdot \frac{Q_j(x)}{P(x)} 1\left\{\frac{P(x)}{Q_j(x)} \geq \gamma_j\right\} \quad (63)$$

$$= \mathbb{E}_P\left[\frac{Q_j(X)}{P(X)} 1\left\{\frac{P(X)}{Q_j(X)} \geq \gamma_j\right\}\right]. \quad (64)$$

$\blacksquare$

The following converse bound extends [5, Eq. (102)] from binary hypothesis testing to composite hypothesis testing.

Lemma 9 (Converse). *For any α, if $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k) \in \beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big)$, then*

$$\alpha - \sum_{j=1}^k \gamma_j \beta_j \leq \mathbb{P}\left[\bigcap_{j\in[k]} \left\{\frac{P(X)}{Q_j(X)} \geq \gamma_j\right\}\right], \quad (65)$$

where $\gamma_j \geq 0$, $j \in [k]$ are arbitrary constants.

Proof. Appendix B. $\blacksquare$

Lemma 10 extends the argument of [34, Lemma 1] from binary to composite hypothesis testing.

Lemma 10 (Variational lemma). *For any α, if $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k) \in \beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big)$, then*

$$\alpha - \sum_{j=1}^k \gamma_j \beta_j \leq 1 - \sum_{x\in\mathcal{X}} \min\left\{P(x), \sum_{j=1}^k \gamma_j Q_j(x)\right\}, \quad (66)$$

where $\gamma_j \geq 0$, $j \in [k]$, are arbitrary constants and equality is achieved by a generalized Neyman-Pearson test.

Proof. Appendix C. $\blacksquare$

Given any $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$, define

$$\epsilon^*(\boldsymbol{\beta}) \triangleq \inf\{\epsilon \in [0,1] : \exists \, test \; s.t.$$
$$\mathbb{P}[Z=1] \geq 1 - \epsilon, \; \mathbb{Q}_j[Z=1] \leq \beta_j, \; \forall j \in [k]\}. \quad (67)$$

Then Lemma 10 gives

$$\epsilon^*(\boldsymbol{\beta}) = \quad (68)$$
$$\sup_{\gamma_1, \ldots, \gamma_k \geq 0} \left\{\sum_{x\in\mathcal{X}} \min\left\{P(x), \sum_{j=1}^k \gamma_j Q_j(x)\right\} - \sum_{j=1}^k \gamma_j \beta_j\right\}.$$

---

[5]In [32], Huang and Moulin also use this sub-optimal likelihood-ratio threshold in their asymptotic achievability analysis.

Remark 5. We can derive Lemma 9 from the variational characterization in Lemma 10 by noting that

$$1 - \sum_{x \in \mathcal{X}} \min\left\{ P(x), \sum_{j=1}^{k} \gamma_j Q_j(x) \right\}$$

$$\leq 1 - \sum_{x \in \mathcal{X}} P(x) 1\left\{ P(x) < \sum_{j=1}^{k} \gamma_j Q_j(x) \right\} \quad (69)$$

$$= \mathbb{P}\left[P(X) \geq \sum_{j=1}^{k} \gamma_j Q_j(X) \right] \quad (70)$$

$$\leq \mathbb{P}\left[\bigcap_{j \in [k]} \left\{ \frac{P(X)}{Q_j(X)} \geq \gamma_j \right\} \right]. \quad (71)$$

Lemmas 9 and 10 are useful beyond the asymptotic analysis of composite hypothesis testing. They also make it possible to recover previous converse bounds from our new MASC meta-converse, as presented in Section V-C2 below.

B. Asymptotics for I.I.D. Distributions

We here characterize the asymptotics of $\beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big)$ when each of P and $\{Q_j\}_{j=1}^k$ is a product of n identical single-shot distributions, i.e., $P(X^n) = \prod_{i=1}^{n} P(X_i)$ and $Q_j(X^n) = \prod_{i=1}^{n} Q_j(X_i)$, $j \in [k]$.

We begin with notation. For each $j \in [k]$, define

$$D_j \triangleq \mathbb{E}_P\left[\log \frac{P(X)}{Q_j(X)} \right] \quad (72)$$

$$V_j \triangleq \mathrm{Var}_P\left[\log \frac{P(X)}{Q_j(X)} \right] \quad (73)$$

$$T_j \triangleq \mathbb{E}_P\left[\left| \log \frac{P(X)}{Q_j(X)} - D_j \right|^3 \right]. \quad (74)$$

Define vector $\mathbf{D}$ and matrix V as

$$\mathbf{D} \triangleq (D_j, j \in [k]) \quad (75)$$

$$\mathsf{V} \triangleq \mathbf{Cov}_P\left[\left(\log \frac{P(X)}{Q_j(X)}, j \in [k] \right) \right]. \quad (76)$$

Let $\mathbf{Z} \in \mathbb{R}^d$ be a Gaussian random vector with mean zero and covariance matrix V. Define the multidimensional counterpart of the function $Q^{-1}(\cdot)$ as

$$\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) \triangleq \left\{ \mathbf{z} \in \mathbb{R}^d : \mathbb{P}[\mathbf{Z} \leq \mathbf{z}] \geq 1 - \epsilon \right\}. \quad (77)$$

The set $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ appears in characterizations such as [13], [32]. When V is non-singular, the boundary of $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ approaches $z_i = \sqrt{[\mathsf{V}]_{i,i}} Q^{-1}(\epsilon)$ in each dimension $i \in [d]$, as illustrated in Figure 2(a). For $\epsilon \leq 1/2$, $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ lies in the positive orthant of $\mathbb{R}^d$; for $\epsilon > 1/2$, $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ extends outside of the positive orthant. If $\epsilon' < \epsilon$, then $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon') \subset \mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$. See Figure 2(b) for plots of the boundaries of $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ in $\mathbb{R}^2$. If V is singular with rank $r < d$, then $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ lies in an r-dimensional subspace of $\mathbb{R}^d$.

Theorem 11 derives a third-order-optimal characterization of $\beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big)$ under assumptions

$$V_j > 0, \, \forall j \in [k] \quad (78)$$

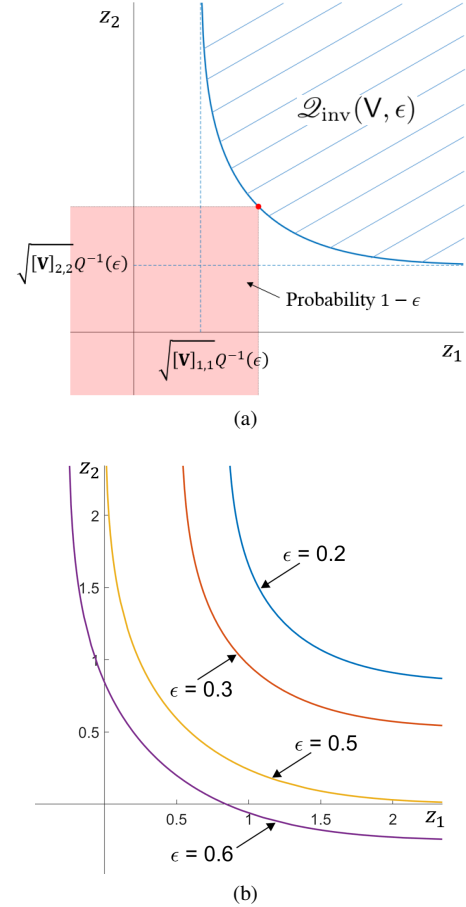$$T_j < \infty, \, \forall j \in [k]. \quad (79)$$

Fig. 2: Illustrations of $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) \subset \mathbb{R}^2$. (a) A schematic drawing of $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$. (b) A graph plotting the boundaries of $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) \subset \mathbb{R}^2$ with various values of ϵ when V is the identity matrix.

Define the inner and outer bounding sets

$$\mathscr{B}_{\mathrm{in}}^*(n, \epsilon) \triangleq \exp\left\{ -n\mathbf{D} + \sqrt{n}\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) - \frac{\log n}{2}\mathbf{1} + O(1)\mathbf{1} \right\}$$

$$\mathscr{B}_{\mathrm{out}}^*(n, \epsilon) \triangleq \exp\left\{ -n\mathbf{D} + \sqrt{n}\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) - \frac{\log n}{2}\mathbf{1} - O(1)\mathbf{1} \right\},$$

where vector $\mathbf{D}$ and matrix V are defined in (75) and (76).

Theorem 11 (Third-order-optimal asymptotics)**.** *Assume that* P *and* $\{Q_j\}_{j=1}^k$ *are product distributions composed of* n *identical single-shot distributions that satisfy* (78) *and* (79)*. For any* $\alpha \in (0, 1)$*, the set* $\beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big)$ *satisfies*

$$\mathscr{B}_{\mathrm{in}}^*(n, \epsilon) \subseteq \beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big) \subseteq \mathscr{B}_{\mathrm{out}}^*(n, \epsilon), \quad (80)$$

where $\epsilon = 1 - \alpha$*.*

Remark 6. In [32, Th. 1], Huang and Moulin claim the third-order-optimal result in Theorem 11 when V is non-singular. Unfortunately, there is a gap in their converse proof. Applying [32, Lemma 2] to get [32, Eq. (13)] requires that vector $\mathbf{b}$ is independent of n. However, they consider any $\mathbf{b} \in \mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$, which may grow with n because set $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ is unbounded. Thus, [32, Eq. (13)] does not always hold.

We resolve this issue with a new proof of Theorem 11 that leverages Lemmas 8 and 9. We first show two auxiliary results.

The multidimensional Berry-Esseen theorem bounds the probability of a sum of i.i.d. random vectors. Bentkus' theorem [35, Th. 1.1] for the case with mean zero and identity covariance achieves the best known dependence on dimension. Tan and Kosut extend [35, Th. 1.1] to non-singular covariance matrices [13, Cor. 8]. We here extend [13, Cor. 8] to covariance matrices with non-zero rank.

Lemma 12. *Let $\mathbf{U}_1, \ldots, \mathbf{U}_n$ be i.i.d. random vectors in $\mathbb{R}^d$ with mean zero and covariance matrix V. Let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{V})$ be a Gaussian vector in $\mathbb{R}^d$. Define $r \triangleq \mathrm{rank}(\mathsf{V})$. Let T be a $d \times r$ matrix whose columns are the r normalized eigenvectors of V with non-zero eigenvalues. Define i.i.d. random vectors $\mathbf{W}_1, \ldots, \mathbf{W}_n \in \mathbb{R}^r$ such that $\mathbf{U}_i = \mathsf{T}\mathbf{W}_i$ for $i \in [n]$. Let $\mathsf{V}_r \triangleq \mathrm{Cov}[\mathbf{W}_1]$ and $\beta_r \triangleq \mathbb{E}[\|\mathbf{W}_1\|_2^3]$. If $r \geq 1$, then for all n,*

$$\sup_{\mathbf{z} \in \mathbb{R}^d} \left| \mathbb{P}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{U}_i \leq \mathbf{z}\right] - \mathbb{P}[\mathbf{Z} \leq \mathbf{z}] \right| \leq \frac{400 d^{1/4} \beta_r}{\lambda_{\min}(\mathsf{V}_r)^{3/2} \sqrt{n}}, \tag{81}$$

where $\lambda_{\min}(\mathsf{V}_r) > 0$ is the smallest eigenvalue of matrix V_r.

Proof. Appendix D. ∎

If $r = d$, then $\mathsf{V}_r = \mathsf{V}$ and Lemma 12 recovers [13, Cor. 8]. The following lemma is useful for our asymptotic analysis.

Lemma 13. *Fix an arbitrary $d \times d$ positive-semidefinite matrix V and $0 < \epsilon < 1$. Then, the following results hold.*

1) There exist constants D_1 and $\delta_1 > 0$ such that for all $0 \leq \delta < \delta_1$,

$$\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) \subseteq \mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon - \delta) - D_1 \delta \mathbf{1}. \tag{82}$$

2) There exist constants D_2 and $\delta_2 > 0$ such that for all $0 \leq \delta < \delta_2$,

$$\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) \supseteq \mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon + \delta) + D_2 \delta \mathbf{1}. \tag{83}$$

Proof. Appendix E. ∎

Proof of Theorem 11. Define random variables

$$I_j \triangleq \sum_{i=1}^n \log \frac{P(X_i)}{Q_j(X_i)}, \ j \in [k] \tag{84}$$

and random vector

$$\mathbf{I} \triangleq (I_j, \ j \in [k]). \tag{85}$$

For brevity, denote

$$\boldsymbol{\gamma} \triangleq (\gamma_j, \ j \in [k]). \tag{86}$$

To prove the achievability part of Theorem 11, we particularize Lemma 8 to product distributions $P^{\otimes n}$ and $\{Q_j^{\otimes n}\}_{j=1}^k$ to obtain that for any $\boldsymbol{\gamma} \geq \mathbf{0}$, there exists a test $P_{Z|X^n}$ for which

$$\mathbb{P}[Z = 1] = \mathbb{P}[\mathbf{I} \geq \log \boldsymbol{\gamma}] \tag{87}$$

$$\mathbb{Q}_j[Z = 1] \leq \mathbb{E}_P[\exp(-I_j)\mathbf{1}\{I_j \geq \log \gamma_j\}], \ \forall j \in [k]. \tag{88}$$

Take any $\boldsymbol{\gamma}$ such that

$$\log \boldsymbol{\gamma} \in n\mathbf{D} - \sqrt{n}\mathscr{Q}_{\mathrm{inv}}\left(\mathsf{V}, \epsilon - \frac{B}{\sqrt{n}}\right), \tag{89}$$

where B is the constant on the right side of (81) for $\mathbf{I}_n$, which is finite under assumptions (78) and (79). Applying Lemma 12 to (87) gives

$$\mathbb{P}[Z = 1] = \mathbb{P}\left[\frac{1}{\sqrt{n}}(-\mathbf{I} + n\mathbf{D}) \leq \frac{1}{\sqrt{n}}(-\log \boldsymbol{\gamma} + n\mathbf{D})\right] \tag{90}$$

$$\geq \mathbb{P}\left[\mathbf{Z} \leq \frac{1}{\sqrt{n}}(-\log \boldsymbol{\gamma} + n\mathbf{D})\right] - \frac{B}{\sqrt{n}} \tag{91}$$

$$\geq 1 - \epsilon, \tag{92}$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{V})$ and matrix V is defined in (76). Applying Lemma 7 to (88) gives

$$\mathbb{Q}_j[Z = 1] \leq \frac{K_j}{\sqrt{n}} \exp(-\log \gamma_j), \tag{93}$$

where

$$K_j \triangleq 2\left(\frac{\log 2}{\sqrt{2\pi V_j}} + 2C_0 \frac{T_j}{V_j^{3/2}}\right) \tag{94}$$

is a finite positive constant by the assumptions in (78) and (79). Plugging (89) into (93) and noting (92) gives

$$\beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big) \tag{95}$$
$$\supseteq \exp\left\{-n\mathbf{D} + \sqrt{n}\mathscr{Q}_{\mathrm{inv}}\left(\mathsf{V}, \epsilon - \frac{C}{\sqrt{n}}\right) - \frac{\log n}{2}\mathbf{1} + O(1)\mathbf{1}\right\}$$
$$\supseteq \exp\left\{-n\mathbf{D} + \sqrt{n}\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) - \frac{\log n}{2}\mathbf{1} + O(1)\mathbf{1}\right\}, \tag{96}$$

where (96) follows from Lemma 13-1.

For the converse, recall from Lemma 9 that if $\epsilon = 1 - \alpha$, then any $\boldsymbol{\beta} \in \beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big)$ must satisfy

$$\epsilon \geq 1 - \mathbb{P}\left[\bigcap_{j=1}^k \left\{-I_j \leq \log \frac{1}{\gamma_j}\right\}\right] - \sum_{j=1}^k \gamma_j \beta_j \tag{97}$$

for all $\gamma_j \geq 0$, $j \in [k]$. Take

$$\gamma_j = \frac{1}{\beta_j \sqrt{n}}, \ j \in [k]. \tag{98}$$

Then, (97) becomes

$$\epsilon \geq 1 - \mathbb{P}\left[-\mathbf{I} \leq \log \boldsymbol{\beta} + \frac{\log n}{2}\mathbf{1}\right] - \frac{k}{\sqrt{n}} \tag{99}$$

$$\geq 1 - \mathbb{P}\left[\mathbf{Z} \leq \frac{1}{\sqrt{n}}\left(\log \boldsymbol{\beta} + n\mathbf{D} + \frac{\log n}{2}\mathbf{1}\right)\right] - \frac{B+k}{\sqrt{n}}, \tag{100}$$

where (100) applies Lemma 12 and B is the constant in the right side of (81). By the definition of $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ in (77), (100) implies that

$$\boldsymbol{\beta} \in \exp\left\{-n\mathbf{D} + \sqrt{n}\mathscr{Q}_{\mathrm{inv}}\left(\mathsf{V}, \epsilon + \frac{B+k}{\sqrt{n}}\right) - \frac{\log n}{2}\mathbf{1}\right\}. \tag{101}$$

Applying Lemma 13-2, we conclude from (101) that

$$\beta_\alpha\big(P, \{Q_j\}_{j=1}^k\big)$$
$$\subseteq \exp\left\{-n\mathbf{D} + \sqrt{n}\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) - \frac{\log n}{2}\mathbf{1} - O(1)\mathbf{1}\right\}. \tag{102}$$
∎

V. MULTIPLE ACCESS SOURCE CODING

To simplify notation, we focus on MASCs with two encoders. Our definitions and results generalize to more than two encoders, as briefly noted in Remark 12 below.

A. Definitions

In a MASC [11], also known as a Slepian-Wolf source code, independent encoders compress a pair of random variables (X_1, X_2) with discrete alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$. Encoder i, $i \in [2]$, observes only X_i, which it maps to a codeword in $[M_i]$; a single decoder jointly decodes the pair of codewords to reconstruct (X_1, X_2). We first define codes for abstract random objects and then particularize to random objects that live in an alphabet endowed with a Cartesian product structure.

Definition 5 (MASC)**.** *An (M_1, M_2, ϵ) MASC for random variables (X_1, X_2) with discrete alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ comprises two encoding functions $\mathsf{f}_1 \colon \mathcal{X}_1 \to [M_1]$ and $\mathsf{f}_2 \colon \mathcal{X}_2 \to [M_2]$ and a decoding function, $\mathsf{g} \colon [M_1] \times [M_2] \to \mathcal{X}_1 \times \mathcal{X}_2$ with error probability $\mathbb{P}[\mathsf{g}(\mathsf{f}_1(X_1), \mathsf{f}_2(X_2)) \neq (X_1, X_2)] \leq \epsilon$.*

In block coding, encoders individually observe X_1^n and X_2^n drawn from distribution $P_{X_1^n X_2^n}$ on $\mathcal{X}_1^n \times \mathcal{X}_2^n$. Our block MASC definition is similar to those in [13] and [14].

Definition 6 (Block MASC)**.** *An (n, M_1, M_2, ϵ) MASC is an (M_1, M_2, ϵ) MASC for random vectors (X_1^n, X_2^n) on $\mathcal{X}_1^n \times \mathcal{X}_2^n$. The code rate $\mathbf{R} = (R_1, R_2)$ is given by*

$$R_1 \triangleq \frac{1}{n} \log M_1, \; R_2 \triangleq \frac{1}{n} \log M_2. \quad (103)$$

Definition 7 $((n, \epsilon)$-rate region)**.** *Rate $\mathbf{R} = (R_1, R_2)$ is (n, ϵ)-achievable if there exists an (n, M_1, M_2, ϵ) MASC with $R_1 \leq \frac{1}{n} \log M_1$ and $R_2 \leq \frac{1}{n} \log M_2$. The (n, ϵ)-rate region $\mathscr{R}^*(n, \epsilon)$ is the closure of the set of (n, ϵ)-achievable rate pairs.*

While definitions 6 and 7 apply to arbitrary discrete random variables (X_{1i}, X_{2i}), $i = 1, 2, \ldots$, with transition probability kernels $P_{(X_1 X_2)_i | (X_1 X_2)^{i-1}}$, our asymptotic analysis focuses on stationary, memoryless sources, where $P_{(X_1 X_2)_i | (X_1 X_2)^{i-1}} = P_{X_1 X_2}$ for all $i = 1, 2, \ldots$

For any rate $\mathbf{R} = (R_1, R_2)$ and distribution $P_{X_1 X_2}$, define

$$\overline{\mathbf{R}} \triangleq \begin{bmatrix} R_1 \\ R_2 \\ R_1 + R_2 \end{bmatrix}, \; \overline{\mathbf{H}} \triangleq \begin{bmatrix} H(X_1|X_2) \\ H(X_2|X_1) \\ H(X_1, X_2) \end{bmatrix}. \quad (104)$$

B. Background

In [11], Slepian and Wolf prove that if (X_1^n, X_2^n) are stationary and memoryless, then for every $\epsilon \in (0, 1)$,

$$\lim_{n \to \infty} \mathscr{R}^*(n, \epsilon) = \{(R_1, R_2) : R_1 \geq H(X_1|X_2)$$
$$R_2 \geq H(X_2|X_1)$$
$$R_1 + R_2 \geq H(X_1, X_2)\}, \quad (105)$$

(i.e., the strong converse holds). We call this region the asymptotic MASC rate region.

In [12], Miyake and Kanaya give achievability and converse bounds for finite-blocklength coding on finite-alphabet sources. In [9], Han gives corresponding results for sources with countable alphabets. While these results are stated in [9] for general sources whose alphabets adopt n-fold Cartesian product structures, we here describe them in an abstract form.

Theorem 14 (Achievability, Han [9, Lemma 7.2.1])**.** *Given discrete random variables (X_1, X_2), there exists an (M_1, M_2, ϵ) MASC satisfying*

$$\epsilon \leq \mathbb{P}[\{\imath(X_1|X_2) \geq \log M_1 - \gamma\}$$
$$\cup \{\imath(X_2|X_1) \geq \log M_2 - \gamma\}$$
$$\cup \{\imath(X_1, X_2) \geq \log M_1 M_2 - \gamma\}] + 3 \exp(-\gamma), \quad (106)$$

where $\gamma > 0$ is an arbitrary constant.

Theorem 15 (Converse, Han [9, Lemma 7.2.2])**.** *Any (M_1, M_2, ϵ) MASC on discrete random variables (X_1, X_2) satisfies*

$$\epsilon \geq \mathbb{P}[\{\imath(X_1|X_2) \geq \log M_1 + \gamma\}$$
$$\cup \{\imath(X_2|X_1) \geq \log M_2 + \gamma\}$$
$$\cup \{\imath(X_1, X_2) \geq \log M_1 M_2 + \gamma\}] - 3 \exp(-\gamma), \quad (107)$$

where $\gamma > 0$ is an arbitrary constant.

In [15], Jose and Kulkarni derive a new linear programming (LP) finite-blocklength converse, tightening the bound in Theorem 15 with an extra non-negative term (see [15, Cor. 13]).

Theorem 16 (LP-based converse, [15, Th. 12])**.** *Any (M_1, M_2, ϵ) MASC on discrete random variables (X_1, X_2) satisfies*

$$\epsilon \geq \sup_{\phi_1, \phi_2, \phi_3} \left\{ \sum_{\substack{x_1 \in \mathcal{X}_1 \\ x_2 \in \mathcal{X}_2}} \min \left\{ P_{X_1 X_2}(x_1, x_2), \sum_{j=1}^{3} \phi_j(x_1, x_2) \right\} \right.$$
$$- M_1 \sum_{x_2 \in \mathcal{X}_2} \max_{\hat{x}_1 \in \mathcal{X}_1} \phi_1(\hat{x}_1, x_2) - M_2 \sum_{x_1 \in \mathcal{X}_1} \max_{\hat{x}_2 \in \mathcal{X}_2} \phi_2(x_1, \hat{x}_2)$$
$$\left. - M_1 M_2 \max_{\hat{x}_1 \in \mathcal{X}_1, \hat{x}_2 \in \mathcal{X}_2} \phi_3(\hat{x}_1, \hat{x}_2) \right\}, \quad (108)$$

where the supremum is over $\phi_1, \phi_2, \phi_3 \colon \mathcal{X}_1 \times \mathcal{X}_2 \to [0, 1]$ such that $0 \leq \phi_1(x_1, x_2), \phi_2(x_1, x_2), \phi_3(x_1, x_2) \leq P_{X_1 X_2}(x_1, x_2)$ for all $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$.

The best prior asymptotic expansion of the MASC rate region is the second-order characterization developed independently in [13], [14]. In [13], Tan and Kosut introduce an *entropy dispersion matrix*, which serves a role similar to the scalar dispersion in the point-to-point case [5], [6], [23].

Definition 8 (Tan and Kosut [13, Def. 7])**.** *The entropy dispersion matrix V for random variables (X_1, X_2) is the covariance matrix $\mathsf{V} \triangleq \mathrm{Cov}[\bar{\imath}(X_1, X_2)]$ of random vector*

$$\bar{\imath}(X_1, X_2) \triangleq \begin{bmatrix} \imath(X_1|X_2) \\ \imath(X_2|X_1) \\ \imath(X_1, X_2) \end{bmatrix}. \quad (109)$$

Note that V is a 3×3 positive-semidefinite matrix with $V(X_1|X_2)$, $V(X_2|X_1)$, and $V(X_1, X_2)$ on the diagonal.

Tan and Kosut [13] give a second-order characterization of the MASC rate region for finite-alphabet stationary, memoryless sources in terms of the asymptotic rate region and the entropy dispersion matrix. Their result, reproduced below, exhibits an $O\left(\frac{\log n}{n}\right)$ gap in the third-order term.

Define

$$\mathscr{R}_{\text{in}}(n,\epsilon) \triangleq \left\{ \mathbf{R} \in \mathbb{R}^2 : \overline{\mathbf{R}} \in \overline{\mathbf{H}} + \frac{\mathscr{Q}_{\text{inv}}(\mathsf{V},\epsilon)}{\sqrt{n}} + \frac{\nu \log n}{n}\mathbf{1} \right\}$$
(110)

$$\mathscr{R}_{\text{out}}(n,\epsilon) \triangleq \left\{ \mathbf{R} \in \mathbb{R}^2 : \overline{\mathbf{R}} \in \overline{\mathbf{H}} + \frac{\mathscr{Q}_{\text{inv}}(\mathsf{V},\epsilon)}{\sqrt{n}} - \frac{\log n}{n}\mathbf{1} \right\},$$
(111)

where $\overline{\mathbf{R}}$ and $\overline{\mathbf{H}}$ are defined in (104), V is the entropy dispersion matrix for (X_1, X_2) (Definition 8), $\nu \triangleq |\mathcal{X}_1||\mathcal{X}_2| + \kappa + \frac{3}{2}$, and κ is the absolute finite positive constant from [13, Def. 6].

Theorem 17 (Tan and Kosut [13, Th. 1]). *Consider finite-alphabet, stationary, memoryless sources* (X_1, X_2) *with* $P_{X_1 X_2}(x_1, x_2) > 0$ *for every* $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. *For any* $0 < \epsilon < 1$ *and all* n *sufficiently large,*

$$\mathscr{R}_{\text{in}}(n,\epsilon) \subseteq \mathscr{R}^*(n,\epsilon) \subseteq \mathscr{R}_{\text{out}}(n,\epsilon).$$
(112)

Remark 7. The inner boundary defined in (110) is achievable by a universal coding scheme [13, Sec. VI]. The outer bounding region in (111) is based on [9, Lemma 7.2.2].

In [14], Nomura and Han use [9, Lemma 7.2.1] and [9, Lemma 7.2.2] to derive a second-order MASC coding theorem for stationary, memoryless, dependent sources. Their result is equivalent to Theorem 17 up to the second-order term and applies also for countable alphabets. Neither [13] nor [14] finds the precise third-order term. In Sections V-C and V-D, below, we give new non-asymptotic MASC bounds and then apply them to precisely characterize the third-order asymptotics.

C. New Non-Asymptotic Bounds

1) Achievability: We present a MASC RCU bound, extending Theorem 4 to the multiple-encoder case.

Theorem 18 (MASC RCU bound). *Given discrete random variables* (X_1, X_2), *there exists an* (M_1, M_2, ϵ) *MASC with*

$$\epsilon \leq \mathbb{E}[\min\{1, A_1 + A_2 + A_{12}\}]$$
(113)

where

$$A_1 \triangleq \frac{1}{M_1}\mathbb{E}\left[\exp\left(\imath(\bar{X}_1'|X_2)\right) \\ \mathbf{1}\{\imath(\bar{X}_1'|X_2) \leq \imath(X_1|X_2)\}|X_1, X_2\right]$$
(114)

$$A_2 \triangleq \frac{1}{M_2}\mathbb{E}\left[\exp\left(\imath(\bar{X}_2'|X_1)\right) \\ \mathbf{1}\{\imath(\bar{X}_2'|X_1) \leq \imath(X_2|X_1)\}|X_1, X_2\right]$$
(115)

$$A_{12} \triangleq \frac{1}{M_1 M_2}\mathbb{E}\left[\exp\left(\imath(\bar{X}_1, \bar{X}_2)\right) \\ \mathbf{1}\{\imath(\bar{X}_1, \bar{X}_2) \leq \imath(X_1, X_2)\}|X_1, X_2\right]$$
(116)

$$P_{X_1 X_2 \bar{X}_1 \bar{X}_2 \bar{X}_1' \bar{X}_2'}(a, b, \bar{a}, \bar{b}, \bar{a}', \bar{b}') \\ = P_{X_1 X_2}(a, b) P_{X_1 X_2}(\bar{a}, \bar{b}) P_{X_1|X_2}(\bar{a}'|b) P_{X_2|X_1}(\bar{b}'|a).$$
(117)

Proof. For every $x_i \in \mathcal{X}_i$, $i \in [2]$, draw $\mathsf{F}_i(x_i)$ i.i.d. uniformly at random from $[M_i]$. The maximum likelihood decoder is defined for each $(c_1, c_2) \in [M_1] \times [M_2]$ by

$$\mathsf{g}(c_1, c_2) = \arg \min_{\substack{(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2: \\ \mathsf{F}_1(x_1) = c_1, \mathsf{F}_2(x_2) = c_2}} \imath(x_1, x_2),$$
(118)

where ties are broken equiprobably at random in the code design. This decoder is optimal for the given encoder.

We bound the random code's expected error probability by the probability of the union of events

$$\mathcal{E}_1 \triangleq \{\exists \bar{x}_1 \in \mathcal{X}_1 \backslash \{X_1\} : \\ \imath(\bar{x}_1|X_2) \leq \imath(X_1|X_2), \mathsf{F}_1(\bar{x}_1) = \mathsf{F}_1(X_1)\}$$
(119)

$$\mathcal{E}_2 \triangleq \{\exists \bar{x}_2 \in \mathcal{X}_2 \backslash \{X_2\} : \\ \imath(\bar{x}_2|X_1) \leq \imath(X_2|X_1), \mathsf{F}_2(\bar{x}_2) = \mathsf{F}_2(X_2)\}$$
(120)

$$\mathcal{E}_{12} \triangleq \{\exists \bar{x}_1 \in \mathcal{X}_1 \backslash \{X_1\}, \bar{x}_2 \in \mathcal{X}_2 \backslash \{X_2\} : \\ \imath(\bar{x}_1, \bar{x}_2) \leq \imath(X_1, X_2), \\ \mathsf{F}_1(\bar{x}_1) = \mathsf{F}_1(X_1), \mathsf{F}_2(\bar{x}_2) = \mathsf{F}_2(X_2)\}.$$
(121)

By a derivation similar to that in the proof of Theorem 4,

$$\mathbb{E}[\mathbb{P}[\{g(\mathsf{F}_1(X_1), \mathsf{F}_2(X_2)) \neq (X_1, X_2)\}]$$
$$\leq \mathbb{P}[\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_{12}]$$
(122)
$$\leq \mathbb{E}\Bigg[\min \Bigg\{ 1, \\ \sum_{\bar{x}_1 \in \mathcal{X}_1 \backslash \{X_1\}} \mathbb{P}[\imath(\bar{x}_1|X_2) \leq \imath(X_1|X_2), \mathsf{F}_1(\bar{x}_1) = \mathsf{F}_1(X_1)|X_1, X_2] \\ + \sum_{\bar{x}_2 \in \mathcal{X}_2 \backslash \{X_2\}} \mathbb{P}[\imath(\bar{x}_2|X_1) \leq \imath(X_2|X_1), \mathsf{F}_2(\bar{x}_2) = \mathsf{F}_2(X_2)|X_1, X_2] \\ + \sum_{\substack{\bar{x}_1 \in \mathcal{X}_1 \backslash \{X_1\} \\ \bar{x}_2 \in \mathcal{X}_2 \backslash \{X_2\}}} \mathbb{P}[\imath(\bar{x}_1, \bar{x}_2) \leq \imath(X_1, X_2), \\ \mathsf{F}_1(\bar{x}_1) = \mathsf{F}_1(X_1), \mathsf{F}_2(\bar{x}_2) = \mathsf{F}_2(X_2)|X_1, X_2] \Bigg\}\Bigg]$$
(123)
$$\leq \mathbb{E}\Bigg[\min \Bigg\{ 1, \frac{1}{M_1} \sum_{\bar{x}_1 \in \mathcal{X}_1} \mathbf{1}\{\imath(\bar{x}_1|X_2) \leq \imath(X_1|X_2)\} \\ + \frac{1}{M_2} \sum_{\bar{x}_2 \in \mathcal{X}_2} \mathbf{1}\{\imath(\bar{x}_2|X_1) \leq \imath(X_2|X_1)\} \\ + \frac{1}{M_1 M_2} \sum_{\bar{x}_1 \in \mathcal{X}_1, \bar{x}_2 \in \mathcal{X}_2} \mathbf{1}\{\imath(\bar{x}_1, \bar{x}_2) \leq \imath(X_1, X_2)\} \Bigg\}\Bigg],$$
(124)

and (124) is equal to the right side of (113) as desired. ∎

Figure 4 in Section V-D1 plots the point-to-point (Theorem 4) and MASC (Theorem 18) RCU bounds.

2) Converse: The MASC composite hypothesis testing converse employs the set $\beta_\alpha\left(P, \{Q_j\}_{j=1}^k\right)$ (see Definition 4) and its generalization to σ-finite measures.

Theorem 19 (Hypothesis testing (HT) converse). *Let* $P_{X_1 X_2}$ *be the source distribution defined on* $\mathcal{X}_1 \times \mathcal{X}_2$. *Let* $Q_{X_1 X_2}^{(1)}$,

$Q_{X_1X_2}^{(2)}$, and $Q_{X_1X_2}^{(3)}$ be any σ-finite measures defined on $\mathcal{X}_1 \times \mathcal{X}_2$. Any (M_1, M_2, ϵ) MASC satisfies

$$(\beta_1^*, \beta_2^*, \beta_3^*) \in \beta_{1-\epsilon}\left(P_{X_1X_2}, \left\{Q_{X_1X_2}^{(1)}, Q_{X_1X_2}^{(2)}, Q_{X_1X_2}^{(3)}\right\}\right), \tag{125}$$

where

$$\beta_1^* \triangleq M_1 \sum_{x_2 \in \mathcal{X}_2} \max_{\hat{x}_1 \in \mathcal{X}_1} Q_{X_1X_2}^{(1)}(\hat{x}_1, x_2), \tag{126}$$

$$\beta_2^* \triangleq M_2 \sum_{x_1 \in \mathcal{X}_1} \max_{\hat{x}_2 \in \mathcal{X}_2} Q_{X_1X_2}^{(2)}(x_1, \hat{x}_2), \tag{127}$$

$$\beta_3^* \triangleq M_1 M_2 \max_{\hat{x}_1 \in \mathcal{X}_1, \hat{x}_2 \in \mathcal{X}_2} Q_{X_1X_2}^{(3)}(\hat{x}_1, \hat{x}_2). \tag{128}$$

Proof. Consider an (M_1, M_2, ϵ) MASC with stochastic encoders $P_{F_1|X_1}$ and $P_{F_2|X_2}$ and stochastic decoder $P_{\hat{X}_1\hat{X}_2|F_1F_2}$, where F_1 and F_2 are the encoder outputs, and $(\hat{X}_1, \hat{X}_2)$ is the decoder output. Fix distributions $\{Q_{X_1X_2}^{(j)}\}_{j=1}^3$ on $\mathcal{X}_1 \times \mathcal{X}_2$. Then $Z = 1\{(\hat{X}_1, \hat{X}_2) = (X_1, X_2)\}$ defines a (sub-optimal) composite HT for testing $P_{X_1X_2}$ against $\{Q_{X_1X_2}^{(j)}\}_{j=1}^3$, for which $\mathbb{P}[Z = 1] \geq 1 - \epsilon$ and

$$\mathbb{Q}^{(1)}[Z = 1]$$
$$= \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} Q_{X_1X_2}^{(1)}(x_1, x_2) \cdot \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} P_{F_1|X_1}(m_1|x_1)$$
$$\cdot P_{F_2|X_2}(m_2|x_2) \cdot P_{\hat{X}_1\hat{X}_2|F_1F_2}(x_1, x_2|m_1, m_2) \tag{129}$$
$$\leq \sum_{x_2 \in \mathcal{X}_2} \max_{\hat{x}_1 \in \mathcal{X}_1} Q_{X_1X_2}^{(1)}(\hat{x}_1, x_2) \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} P_{F_2|X_2}(m_2|x_2)$$
$$\cdot \sum_{x_1 \in \mathcal{X}_1} P_{F_1|X_1}(m_1|x_1) P_{\hat{X}_1\hat{X}_2|F_1F_2}(x_1, x_2|m_1, m_2) \tag{130}$$
$$\leq M_1 \sum_{x_2 \in \mathcal{X}_2} \max_{\hat{x}_1 \in \mathcal{X}_1} Q_{X_1X_2}^{(1)}(\hat{x}_1, x_2), \tag{131}$$

where (130) follows since $\max_{\hat{x}_1 \in \mathcal{X}_1} Q_{X_1X_2}^{(1)}(\hat{x}_1, x_2)$ is independent of x_1, and (131) follows by bounding the probability in the sum over $x_1 \in \mathcal{X}_1$ by 1. Similarly,

$$\mathbb{Q}^{(2)}[Z = 1] \leq M_2 \sum_{x_1 \in \mathcal{X}_1} \max_{\hat{x}_2 \in \mathcal{X}_2} Q_{X_1X_2}^{(2)}(x_1, \hat{x}_2), \tag{132}$$

$$\mathbb{Q}^{(3)}[Z = 1]$$
$$= \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} Q_{X_1X_2}^{(3)}(x_1, x_2) \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} P_{F_1|X_1}(m_1|x_1) \cdot$$
$$P_{F_2|X_2}(m_2|x_2) P_{\hat{X}_1\hat{X}_2|F_1F_2}(x_1, x_2|m_1, m_2) \tag{133}$$
$$\leq \max_{\hat{x}_1 \in \mathcal{X}_1, \hat{x}_2 \in \mathcal{X}_2} Q_{X_1X_2}^{(3)}(\hat{x}_1, \hat{x}_2) \cdot$$
$$\sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} P_{\hat{X}_1\hat{X}_2|F_1F_2}(x_1, x_2|m_1, m_2) \tag{134}$$
$$= M_1 M_2 \max_{\hat{x}_1 \in \mathcal{X}_1, \hat{x}_2 \in \mathcal{X}_2} Q_{X_1X_2}^{(3)}(\hat{x}_1, \hat{x}_2). \tag{135}$$

Thus (125) holds by the definition of $\beta_{1-\epsilon}(P, \{Q^{(j)}\}_{j=1}^k)$. $\blacksquare$

To recover Han's converse (Theorem 15) from Theorem 19, let P_{X_1} and P_{X_2} be the marginals of $P_{X_1X_2}$ and let U_{X_1},

U_{X_2}, and $U_{X_1X_2}$ be the counting measures over $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{X}_1 \times \mathcal{X}_2$. By Theorem 19, any (M_1, M_2, ϵ) MASC satisfies

$$(M_1, M_2, M_1 M_2)$$
$$\in \beta_{1-\epsilon}(P_{X_1X_2}, \{U_{X_1}P_{X_2}, P_{X_1}U_{X_2}, U_{X_1X_2}\}). \tag{136}$$

Applying Lemma 9 to (136) with $k = 3$ gives

$$\epsilon \geq \mathbb{P}\left[\left\{\imath(X_1|X_2) \geq \log\frac{1}{\gamma_1}\right\} \cup \left\{\imath(X_2|X_1) \geq \log\frac{1}{\gamma_2}\right\}\right.$$
$$\left. \cup \left\{\imath(X_1, X_2) \geq \log\frac{1}{\gamma_3}\right\}\right] - \gamma_1 M_1 - \gamma_2 M_2 - \gamma_3 M_1 M_2.$$

Setting $\gamma_1 = \frac{\exp(-\gamma)}{M_1}$, $\gamma_2 = \frac{\exp(-\gamma)}{M_2}$, and $\gamma_3 = \frac{\exp(-\gamma)}{M_1 M_2}$ for an arbitrary $\gamma > 0$ recovers Theorem 15.

To show that Theorem 19 is equivalent to the LP-based converse (Theorem 16), we apply (68) to Theorem 19, showing that any (M_1, M_2, ϵ) MASC satisfies

$$\epsilon \geq \sup_{Q_{X_1X_2}^{(1)}, Q_{X_1X_2}^{(2)}, Q_{X_1X_2}^{(3)}} \epsilon^*(\beta_1^*, \beta_2^*, \beta_3^*) \tag{137}$$

$$= \sup_{Q_{X_1X_2}^{(1)}, Q_{X_1X_2}^{(2)}, Q_{X_1X_2}^{(3)}} \sup_{\gamma_1, \gamma_2, \gamma_3 \geq 0} \left\{ \right.$$
$$\sum_{\substack{x_1 \in \mathcal{X}_1 \\ x_2 \in \mathcal{X}_2}} \min\left\{ P_{X_1X_2}(x_1, x_2), \sum_{j=1}^3 \gamma_j Q_{X_1X_2}^{(j)}(x_1, x_2) \right\}$$
$$- \gamma_1 M_1 \sum_{x_2 \in \mathcal{X}_2} \max_{\hat{x}_1 \in \mathcal{X}_1} Q_{X_1X_2}^{(1)}(\hat{x}_1, x_2)$$
$$- \gamma_2 M_2 \sum_{x_1 \in \mathcal{X}_1} \max_{\hat{x}_2 \in \mathcal{X}_2} Q_{X_1X_2}^{(2)}(x_1, \hat{x}_2)$$
$$\left. - \gamma_3 M_1 M_2 \max_{\hat{x}_1 \in \mathcal{X}_1, \hat{x}_2 \in \mathcal{X}_2} Q_{X_1X_2}^{(3)}(\hat{x}_1, \hat{x}_2) \right\}. \tag{138}$$

The outer supremum is over σ-finite measures $Q_{X_1X_2}^{(1)}$, $Q_{X_1X_2}^{(2)}$, and $Q_{X_1X_2}^{(3)}$. In Appendix F, we show that the bounds in (138) and (108) are equivalent, establishing the equivalence between the MASC HT (Theorem 19) and LP (Theorem 16) converses.

Remark 8. When one of the sources is deterministic, the MASC HT converse reduces to the point-to-point HT converse [23, Eq. (64)]. For example, if X_2 is deterministic, then (136) reduces to

$$(M_1, 1, M_1) \in \beta_{1-\epsilon}(P_{X_1X_2}, \{U_{X_1}P_{X_2}, P_{X_1}U_{X_2}, U_{X_1X_2}\}),$$

which further reduces to

$$M_1 \geq \beta_{1-\epsilon}(P_{X_1}, U_{X_1}),$$

where $\beta_\alpha(P, Q)$ is the optimal β-function for binary hypothesis testing between distributions P and Q.

D. Asymptotics: Third-Order MASC Rate Region

The following third-order asymptotic characterization of the MASC rate region for stationary, memoryless sources closes the $O\left(\frac{\log n}{n}\right)$ gap between (110) and (111).

Consider stationary, memoryless sources with single-letter joint distribution $P_{X_1 X_2}$ for which

$$V(X_1, X_2) > 0, \ \mathbb{E}[V_c(X_1|X_2)] > 0, \ \mathbb{E}[V_c(X_2|X_1)] > 0, \quad (139)$$

$$T(X_1, X_2) < \infty, \ T(X_1|X_2) < \infty, \ T(X_2|X_1) < \infty, \quad (140)$$

$$\mathbb{E}\left[T_c^2(X_1|X_2)\right] < \infty, \ \mathbb{E}\left[T_c^2(X_2|X_1)\right] < \infty. \quad (141)$$

When (139) holds[6], rank(V) $\geq$ 1. Technical assumptions (139), (140), and (141) are required to ensure applicability of the multidimensional Berry-Esseen theorem and Lemma 7 in our asymptotic analysis. Assumption (140) is satisfied automatically if the alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ are finite.

Define the set

$$\overline{\mathscr{R}}^*(n, \epsilon) \triangleq \left\{ \overline{\mathbf{R}} \in \mathbb{R}^3 : \overline{\mathbf{R}} = \overline{\mathbf{H}} + \frac{\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)}{\sqrt{n}} - \frac{\log n}{2n} \mathbf{1} \right\}, \quad (142)$$

where vector $\overline{\mathbf{H}}$ is defined in (104), V is the entropy dispersion matrix for (X_1, X_2), and $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ is defined in (77). Note that $\mathscr{R}^*(n, \epsilon) \subset \mathbb{R}^2$ (see Definition 7) but $\overline{\mathscr{R}}^*(n, \epsilon) \subset \mathbb{R}^3$. Define the inner and outer bounding sets

$$\mathscr{R}_{\mathrm{in}}^*(n, \epsilon) \triangleq \left\{ \mathbf{R} \in \mathbb{R}^2 : \overline{\mathbf{R}} \in \overline{\mathscr{R}}^*(n, \epsilon) + O\left(\frac{1}{n}\right)\mathbf{1} \right\} \quad (143)$$

$$\mathscr{R}_{\mathrm{out}}^*(n, \epsilon) \triangleq \left\{ \mathbf{R} \in \mathbb{R}^2 : \overline{\mathbf{R}} \in \overline{\mathscr{R}}^*(n, \epsilon) - O\left(\frac{1}{n}\right)\mathbf{1} \right\}. \quad (144)$$

Theorem 20 (Third-order MASC rate region)**.** *Consider a pair of stationary, memoryless sources with single-letter joint distribution $P_{X_1 X_2}$ satisfying (139)–(141). For any $0 < \epsilon < 1$, the (n, ϵ)-rate region $\mathscr{R}^*(n, \epsilon)$ satisfies*

$$\mathscr{R}_{\mathrm{in}}^*(n, \epsilon) \subseteq \mathscr{R}^*(n, \epsilon) \subseteq \mathscr{R}_{\mathrm{out}}^*(n, \epsilon). \quad (145)$$

Since the upper and lower bounds in Theorem 20 agree up to their third-order terms, we call $\overline{\mathscr{R}}^*(n, \epsilon)$ the *third-order MASC rate region*. Figure 3 plots the boundaries of $\overline{\mathscr{R}}^*(n, \epsilon)$ at different values of n for an example pair of sources.

Remark 9. As noted in Remark 2, for point-to-point source coding, zero varentropy means that the source is uniform; the $-\frac{\log n}{2n}$ third-order term is absent in that case. While condition (139) limits Theorem 20 to sources with positive varentropies, Appendix G considers the case where one or more varentropies are zero. Roughly, each zero varentropy yields a zero dispersion, and the absence of a $-\frac{\log n}{2n}$ third-order term, similar to the point-to-point case. Furthermore, if $V(X_1|X_2) > 0$ but $\mathbb{E}[V_c(X_1|X_2)] = 0$, the corresponding achievable third order term increases from $-\frac{\log n}{2n}$ to 0.[7] This means that the optimal third order term lies in $[-\frac{\log n}{2n}, 0]$ in that case.

---

[6]In fact, the weaker condition $V(X_1|X_2) > 0$, $V(X_2|X_1) > 0$, $V(X_1, X_2) > 0$ suffices.
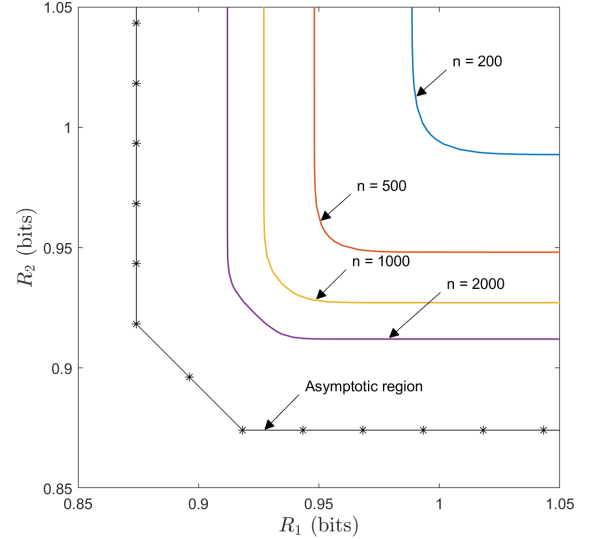[7]This is seen by modifying the reasoning in (172)–(187) in the proof of Theorem 20 below.

Fig. 3: Third-order MASC rate regions $\overline{\mathscr{R}}^*(n, \epsilon)$ at $\epsilon = 10^{-3}$ for stationary, memoryless sources (X_1, X_2) with $p_{X_1, X_2}(0,0) = 1/2$, $p_{X_1, X_2}(0,1) = p_{X_1, X_2}(1,1) = p_{X_1, X_2}(1,1) = 1/6$.

Proof of Theorem 20: achievability. We apply Theorem 18 to stationary, memoryless sources with $P_{X_1^n X_2^n} = P_{X_1 X_2}^n$ and then apply Lemmas 7 and 12 to analyze the bound. Let

$$I_1 \triangleq \imath(X_1^n | X_2^n) = \sum_{i=1}^n \imath(X_{1i} | X_{2i}) \quad (146)$$

$$I_2 \triangleq \imath(X_2^n | X_1^n) = \sum_{i=1}^n \imath(X_{2i} | X_{1i}) \quad (147)$$

$$I_{12} \triangleq \imath(X_1^n, X_2^n) = \sum_{i=1}^n \imath(X_{1i}, X_{2i}) \quad (148)$$

$$\bar{I}_1 \triangleq \imath(\bar{X}_1^{n\prime} | X_2^n) = \sum_{i=1}^n \imath(\bar{X}_{1i}' | X_{2i}) \quad (149)$$

$$\bar{I}_2 \triangleq \imath(\bar{X}_2^{n\prime} | X_1^n) = \sum_{i=1}^n \imath(\bar{X}_{2i}' | X_{1i}) \quad (150)$$

$$\bar{I}_{12} \triangleq \imath(\bar{X}_1^n, \bar{X}_2^n) = \sum_{i=1}^n \imath(\bar{X}_{1i}, \bar{X}_{2i}) \quad (151)$$

where $(X_{1i}, X_{2i}, \bar{X}_{1i}, \bar{X}_{2i}, \bar{X}_{1i}', \bar{X}_{2i}')$, $i = 1, \ldots, n$, are drawn i.i.d. according to the joint distribution defined in (117). With this notation, the random variables A_1, A_2, A_{12} defined in (114), (115), (116) particularize as

$$A_1 = \frac{1}{M_1} \mathbb{E}\left[\exp(\bar{I}_1) 1\{\bar{I}_1 \leq I_1\} | X_1^n, X_2^n\right] \quad (152)$$

$$A_2 = \frac{1}{M_2} \mathbb{E}\left[\exp(\bar{I}_2) 1\{\bar{I}_2 \leq I_2\} | X_1^n, X_2^n\right] \quad (153)$$

$$A_{12} = \frac{1}{M_1 M_2} \mathbb{E}\left[\exp(\bar{I}_{12}) 1\{\bar{I}_{12} \leq I_{12}\} | X_1^n, X_2^n\right] \quad (154)$$

By Theorem 18, there exists an (n, M_1, M_2, ϵ') MASC such that

$$\epsilon' \leq \mathbb{E}[\min\{1, A_1 + A_2 + A_{12}\}] \quad (155)$$
$$= \mathbb{E}[(A_1 + A_2 + A_{12}) 1\{A_1 + A_2 + A_{12} \leq 1\}]$$

$$+\mathbb{P}[A_1+A_2+A_{12}>1] \tag{156}$$

$$\leq \mathbb{E}[A_11\{A_1\leq 1\}]+\mathbb{E}[A_21\{A_2\leq 1\}] \tag{157}$$

$$+\mathbb{E}[A_{12}1\{A_{12}\leq 1\}]+\mathbb{P}[3A_1>1\cup 3A_2>1\cup 3A_{12}>1]$$

To bound each of the terms in (157), we first bound the random variables A_1, A_2, and A_{12} by random variables $\bar{A}_1$, $\bar{A}_2$, and $\bar{A}_{12}$ that are easier to work with.

Denote constants

$$K_1 \triangleq \frac{2\log 2}{\sqrt{2\pi V(X_1|X_2)}}+\frac{2C_0T(X_1|X_2)}{V(X_1|X_2)^{3/2}} \tag{158}$$

$$K_2 \triangleq \frac{2\log 2}{\sqrt{2\pi V(X_2|X_1)}}+\frac{2C_0T(X_2|X_1)}{V(X_2|X_1)^{3/2}} \tag{159}$$

$$K_{12} \triangleq \frac{2\log 2}{\sqrt{2\pi V(X_1,X_2)}}+\frac{2C_0T(X_1,X_2)}{V(X_1,X_2)^{3/2}} \tag{160}$$

that are finite by assumptions (139) and (140). Define

$$V_1(x_1^n) \triangleq \frac{1}{n}\sum_{i=1}^n \text{Var}[\iota(X_{2i}|X_{1i}=x_{1i})] \tag{161}$$

$$T_1(x_1^n) \triangleq \frac{1}{n}\sum_{i=1}^n \mathbb{E}\Big[\big|\iota(X_{2i}|X_{1i}=x_{1i})-\mathbb{E}[\iota(X_{2i}|X_{1i}=x_{1i})]\big|^3\Big] \tag{162}$$

for $x_1^n\in\mathcal{X}_1^n$. Define $V_2(x_2^n)$ and $T_2(x_2^n)$ for $x_2^n\in\mathcal{X}_2^n$ analogously.

Applying Lemma 7 to A_{12} yields

$$A_{12}\leq \bar{A}_{12}\triangleq \frac{K_{12}\exp(I_{12})}{M_1M_2\sqrt{n}}. \tag{163}$$

To bound A_1, we consider the cases $V_2(x_2^n)>0$ and $V_2(x_2^n)=0$ separately. If $V_2(x_2^n)>0$, then

$$K_1(x_2^n)\triangleq \frac{2\log 2}{\sqrt{2\pi V_2(x_2^n)}}+\frac{2C_0T_2(x_2^n)}{V_2(x_2^n)^{3/2}} \tag{164}$$

is finite by assumption (140), and Lemma 7 yields

$$A_1\leq \frac{K_1(X_2^n)\exp(I_1)}{M_1\sqrt{n}}, \quad \text{if } V_2(X_2^n)>0. \tag{165}$$

If $V_2(x_2^n)=0$, then $I_1=\bar{I}_1=H(X_1^n|X_2^n=x_2^n)$ irrespective of the realization of X_1^n, and

$$A_1=\frac{\exp(I_1)}{M_1}, \quad \text{if } V_2(X_2^n)=0. \tag{166}$$

Putting (165) and (166) together yields

$$A_1\leq \bar{A}_1\triangleq \begin{cases}\frac{K_1(X_2^n)\exp(I_1)}{M_1\sqrt{n}}, & V_2(X_2^n)>0\\ \frac{\exp(I_1)}{M_1}, & V_2(X_2^n)=0.\end{cases} \tag{167}$$

Similarly,

$$A_2\leq \bar{A}_2\triangleq \begin{cases}\frac{K_2(X_1^n)\exp(I_2)}{M_2\sqrt{n}}, & V_1(X_1^n)>0\\ \frac{\exp(I_2)}{M_2}, & V_1(X_1^n)=0,\end{cases} \tag{168}$$

where $K_2(x_1^n)$ is defined analogously to (164).

Next, we apply Lemma 7 again to further bound each of the first three terms in (157):

$$\mathbb{E}[A_11\{A_1\leq 1\}]\leq \mathbb{E}\big[\bar{A}_11\{\bar{A}_1\leq 1\}\big]\leq \frac{K_1}{\sqrt{n}} \tag{169}$$

$$\mathbb{E}[A_21\{A_2\leq 1\}]\leq \mathbb{E}\big[\bar{A}_21\{\bar{A}_2\leq 1\}\big]\leq \frac{K_2}{\sqrt{n}} \tag{170}$$

$$\mathbb{E}[A_{12}1\{A_{12}\leq 1\}]\leq \mathbb{E}\big[\bar{A}_{12}1\{\bar{A}_{12}\leq 1\}\big]\leq \frac{K_{12}}{\sqrt{n}}. \tag{171}$$

We proceed to bound the last term in (157). For fixed constants $s_1<\mathbb{E}[V_c(X_2|X_1)]$ and $s_2<\mathbb{E}[V_c(X_1|X_2)]$, define the events $\mathcal{S}_1$ and $\mathcal{S}_2$ that X_1^n and X_2^n are typical, respectively:

$$\mathcal{S}_1\triangleq \{V_1(X_1^n)\geq \mathbb{E}[V_c(X_2|X_1)]-s_1,$$
$$T_1(X_1^n)\leq \mathbb{E}[T_c(X_2|X_1)]+s_1\} \tag{172}$$

$$\mathcal{S}_2\triangleq \{V_2(X_2^n)\geq \mathbb{E}[V_c(X_1|X_2)]-s_2,$$
$$T_2(X_2^n)\leq \mathbb{E}[T_c(X_1|X_2)]+s_2\} \tag{173}$$

Note that

$$\bar{A}_1\leq \bar{\bar{A}}_1\triangleq \frac{\bar{K}_1\exp(I_1)}{M_1\sqrt{n}}, \quad \text{if } \mathcal{S}_2 \text{ occurs} \tag{174}$$

$$\bar{A}_2\leq \bar{\bar{A}}_2\triangleq \frac{\bar{K}_2\exp(I_2)}{M_2\sqrt{n}}, \quad \text{if } \mathcal{S}_1 \text{ occurs}, \tag{175}$$

where

$$\bar{K}_1\triangleq \frac{2\log 2}{\sqrt{2\pi(\mathbb{E}[V_c(X_1|X_2)]-s_2)}}+\frac{2C_0(\mathbb{E}[T_c(X_1|X_2)]+s_2)}{(\mathbb{E}[V_c(X_1|X_2)]-s_2)^{3/2}} \tag{176}$$

$$\bar{K}_2\triangleq \frac{2\log 2}{\sqrt{2\pi(\mathbb{E}[V_c(X_2|X_1)]-s_1)}}+\frac{2C_0(\mathbb{E}[T_c(X_2|X_1)]+s_1)}{(\mathbb{E}[V_c(X_2|X_1)]-s_1)^{3/2}} \tag{177}$$

are both finite by the assumptions in (139) and (140).

Applying the union bound to $\mathbb{P}[\mathcal{S}_k^c]$, $k\in\{1,2\}$, and Chebyshev's inequality

$$\mathbb{P}[|Z-\mathbb{E}[Z]|>\delta]\leq \frac{\text{Var}[Z]}{\delta^2} \tag{178}$$

to both terms, we observe that for each $k\in\{1,2\}$,

$$\mathbb{P}[\mathcal{S}_k^c]\leq \frac{S_k}{n}, \tag{179}$$

where

$$S_1\triangleq \frac{1}{s_1^2}\big(\mathbb{E}[V_c^2(X_2|X_1)]+\mathbb{E}[T_c^2(X_2|X_1)]\big) \tag{180}$$

$$S_2\triangleq \frac{1}{s_2^2}\big(\mathbb{E}[V_c^2(X_1|X_2)]+\mathbb{E}[T_c^2(X_1|X_2)]\big) \tag{181}$$

are finite by assumption (141).

We are now prepared to apply Lemma 12 to the last term in (157). Pick any pair of rates (R_1,R_2) satisfying

$$\overline{\mathbf{R}}\in\overline{\mathscr{R}}^*\left(n,\epsilon-\frac{B}{\sqrt{n}}-\frac{K_1+K_2+K_{12}}{\sqrt{n}}-\frac{S_1+S_2}{n}\right)+\frac{1}{n}\mathbf{C}, \tag{182}$$

where the set $\overline{\mathscr{R}}^*(n,\epsilon)$ is defined in (142), $\mathbf{C}\triangleq \big(\log(3\bar{K}_1),\log(3\bar{K}_2),\log(3K_{12})\big)^T$, and B is the Bentkus

constant in the right-side of (81) for zero-mean i.i.d. random vectors

$$\mathbf{I}_i \triangleq \begin{bmatrix} \imath(X_{1i}|X_{2i}) \\ \imath(X_{2i}|X_{1i}) \\ \imath(X_{1i},X_{2i}) \end{bmatrix} - \overline{\mathbf{H}}, \text{ for } i = 1, \ldots, n. \quad (183)$$

Note that $B < \infty$ by assumption (140). We have

$$\mathbb{P}[3A_1 \le 1 \cap 3A_2 \le 1 \cap 3A_{12} \le 1]$$

$$\ge \mathbb{P}\big[3\bar{A}_1 \le 1 \cap 3\bar{A}_2 \le 1 \cap 3\bar{A}_{12} \le 1 \cap \mathcal{S}_1 \cap \mathcal{S}_2\big] \quad (184)$$

$$\ge \mathbb{P}\big[3\bar{\bar{A}}_1 \le 1 \cap 3\bar{\bar{A}}_2 \le 1 \cap 3\bar{A}_{12} \le 1\big] - \mathbb{P}[\mathcal{S}_1^c \cup \mathcal{S}_2^c] \quad (185)$$

$$= \mathbb{P}\left[\sum_{i=1}^n \mathbf{I}_i \le n\left(\overline{\mathbf{R}} - \overline{\mathbf{H}} + \frac{\log n}{2n}\mathbf{1} - \frac{1}{n}\mathbf{C}\right)\right] - \mathbb{P}[\mathcal{S}_1^c \cup \mathcal{S}_2^c] \quad (186)$$

$$\ge 1 - \epsilon + \frac{K_1 + K_2 + K_{12}}{\sqrt{n}}, \quad (187)$$

where (185) applies (174) and $\mathbb{P}[\mathcal{A} \cap \mathcal{B}] \ge \mathbb{P}[\mathcal{A}] - \mathbb{P}[\mathcal{B}^c]$, and (187) applies (182), Lemma 12 and (179).

Substituting (169), (170), (171), and (187) into (157) yields $\epsilon' \le \epsilon$, and the proof is complete since the set of (R_1, R_2) satisfying (182) contains $\mathscr{R}^*_{\text{in}}(n, \epsilon)$ by Lemma 13-1. ∎

Proof of Theorem 20: converse. We invoke Theorem 19 with $P_{X_1 X_2} = P^n_{X_1 X_2}$, $Q^{(1)}_{X_1 X_2} = U^n_{X_1} P^n_{X_2}$, $Q^{(2)}_{X_1 X_2} = P^n_{X_1} U^n_{X_2}$, and $Q^{(3)}_{X_1 X_2} = U^n_{X_1 X_2}$, where P_{X_1} and P_{X_2} are the marginals of $P_{X_1 X_2}$, and U_{X_1}, U_{X_2}, and $U_{X_1 X_2}$ are the counting measures over $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{X}_1 \times \mathcal{X}_2$. Applying Theorem 11 to $\beta_{1-\epsilon}(P_{X_1 X_2}, \{U_{X_1} P_{X_2}, P_{X_1} U_{X_2}, U_{X_1 X_2}\})$ under the assumptions in (139) and (140), we conclude that in order to attain error probability ϵ, M_1 and M_2 must satisfy

$$(M_1, M_2, M_1 M_2)$$
$$\in \exp\left\{n\overline{\mathbf{H}} + \sqrt{n}\mathscr{Q}_{\text{inv}}(\mathsf{V}, \epsilon) - \frac{\log n}{2}\mathbf{1} - O(1)\mathbf{1}\right\}, \quad (188)$$

which is equivalent to $(R_1, R_2) \in \mathscr{R}^*_{\text{out}}(n, \epsilon)$ (144), as desired. ∎

Remark 10. The converse of Theorem 20 can also be proved using Han's converse (Theorem 15) with $\gamma = \frac{\log n}{2}$ and Lemmas 12 and 13 in a way similar to that in the achievability proof above, except that we would use Lemma 13 to bound $\mathscr{Q}_{\text{inv}}\left(\mathsf{V}, \epsilon + O\left(\frac{1}{\sqrt{n}}\right)\right) \subseteq \mathscr{Q}_{\text{inv}}\left(\mathsf{V}, \epsilon\right) - O\left(\frac{1}{\sqrt{n}}\right)\mathbf{1}$ instead of $\mathscr{Q}_{\text{inv}}\left(\mathsf{V}, \epsilon - O\left(\frac{1}{\sqrt{n}}\right)\right)$. Our HT converse (Theorem 19) is stronger than Han's converse, but the gap is in the fourth- or higher-order terms, as illustrated through computation in Figure 4. Han's achievability bound (Theorem 14) with the third-order optimal choice of $\gamma = \frac{\log n}{2}$ leads to the third order term of $+\frac{\log n}{2n}$ instead of $-\frac{\log n}{2n}$. Thus Han's achievability is weaker than our RCU bound (Theorem 18) in the third order term.

Remark 11. Tan and Kosut's converse (Theorem 17) is also based on Han's converse. Instead of deriving an outer bound on $\mathscr{Q}_{\text{inv}}\left(\mathsf{V}, \epsilon + O\left(\frac{1}{\sqrt{n}}\right)\right)$ as given in Lemma 13, they apply multivariate Taylor approximation to expand the probability, giving a bound that is loose in the third-order term.

Remark 12. Theorem 20 generalizes to any finite number of encoders. Let $\mathcal{T} \subset \mathbb{N}$ be a nonempty ordered set with a unique index for each encoder. For any vector $\mathbf{R}_\mathcal{T} \in \mathbb{R}^{|\mathcal{T}|}$, define the $(2^{|\mathcal{T}|}-1)$-dimensional vector of its partial sums as

$$\overline{\mathbf{R}}_\mathcal{T} \triangleq \left(\sum_{i \in \mathcal{A}} R_i, \hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})\right). \quad (189)$$

For any distribution $P_{\mathbf{X}_\mathcal{T}}$ defined on $\mathcal{X}_\mathcal{T}$ and any $\mathbf{x}_\mathcal{T} \in \mathcal{X}_\mathcal{T}$, define $(2^{|\mathcal{T}|}-1)$-dimensional vectors

$$\bar{\boldsymbol{\imath}}_\mathcal{T}(\mathbf{x}_\mathcal{T}) \triangleq \left(\imath(\mathbf{x}_{\hat{\mathcal{T}}}|\mathbf{x}_{\mathcal{T} \setminus \hat{\mathcal{T}}}), \hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})\right) \quad (190)$$

$$\overline{\mathbf{H}}_\mathcal{T} \triangleq \mathbb{E}[\bar{\boldsymbol{\imath}}_\mathcal{T}(\mathbf{X}_\mathcal{T})], \quad (191)$$

and $(2^{|\mathcal{T}|}-1) \times (2^{|\mathcal{T}|}-1)$ entropy dispersion matrix

$$\mathsf{V}_\mathcal{T} \triangleq \text{Cov}[\bar{\boldsymbol{\imath}}_\mathcal{T}(\mathbf{X}_\mathcal{T})] \quad (192)$$

for random vector $\mathbf{X}_\mathcal{T}$. Define set

$$\overline{\mathscr{R}}^*_\mathcal{T}(n, \epsilon) \triangleq \overline{\mathbf{H}}_\mathcal{T} + \frac{\mathscr{Q}_{\text{inv}}(\mathsf{V}_\mathcal{T}, \epsilon)}{\sqrt{n}} - \frac{\log n}{2n}\mathbf{1}. \quad (193)$$

Thus, $\mathscr{R}^*_\mathcal{T}(n, \epsilon) \subset \mathbb{R}^{|\mathcal{T}|}$ while $\overline{\mathscr{R}}^*_\mathcal{T}(n, \epsilon) \subset \mathbb{R}^{2^{|\mathcal{T}|}-1}$. Finally,

$$\mathscr{R}^*_{\text{in},\mathcal{T}}(n, \epsilon) \triangleq \left\{\mathbf{R}_\mathcal{T} \in \mathbb{R}^{|\mathcal{T}|} : \overline{\mathbf{R}}_\mathcal{T} \in \overline{\mathscr{R}}^*_\mathcal{T}(n, \epsilon) + O\left(\frac{1}{n}\right)\mathbf{1}\right\} \quad (194)$$

$$\mathscr{R}^*_{\text{out},\mathcal{T}}(n, \epsilon) \triangleq \left\{\mathbf{R}_\mathcal{T} \in \mathbb{R}^{|\mathcal{T}|} : \overline{\mathbf{R}}_\mathcal{T} \in \overline{\mathscr{R}}^*_\mathcal{T}(n, \epsilon) - O\left(\frac{1}{n}\right)\mathbf{1}\right\}. \quad (195)$$

If every element of $\bar{\boldsymbol{\imath}}_\mathcal{T}(\mathbf{X}_\mathcal{T})$ has a positive variance and a finite third centered moment, then for any $0 < \epsilon < 1$,

$$\mathscr{R}^*_{\text{in},\mathcal{T}}(n, \epsilon) \subseteq \mathscr{R}^*_\mathcal{T}(n, \epsilon) \subseteq \mathscr{R}^*_{\text{out},\mathcal{T}}(n, \epsilon). \quad (196)$$

1) Comparison with Point-to-Point Source Coding: Figure 4 compares joint (point-to-point) compression of (X_1^n, X_2^n) to the MASC sum-rate at the symmetrical rate point $(R_1 = R_2)$. The gap between the MASC and point-to-point HT converses captures a penalty due to separate encoding. For small n, the third-order Gaussian approximation (without the $O\left(\frac{1}{n}\right)$ term) is more accurate at $\epsilon = 10^{-1}$ than at $\epsilon = 10^{-3}$ since the $O\left(\frac{1}{n}\right)$ term blows up as ϵ approaches 0.

It is well-known that optimal MASCs incur no first-order penalty in achievable sum rate when compared to joint coding [9], [11], [12]. We next investigate the higher-order penalty of the MASC's independent encoders.

Tan and Kosut introduce a quantity known as the *local dispersion* [13, Def. 4] to characterize the second-order speed of convergence to any asymptotic MASC rate point from any direction. For any non-corner point on the diagonal boundary of the asymptotic MASC rate region, the sum rate's second-order coefficient is optimal when approached either vertically or horizontally. Approaching corner points incurs a positive second-order penalty relative to point-to-point coding.

Two corollaries of Theorem 20, below, bound the MASC penalty by considering the achievable sum rate $R_1 + R_2$ for different choices of R_1 and R_2. We treat the cases where X_1 and X_2 are dependent and X_1 and X_2 are independent separately, assuming throughout that (139) and (140) hold.

$\epsilon = 10^{-1}$
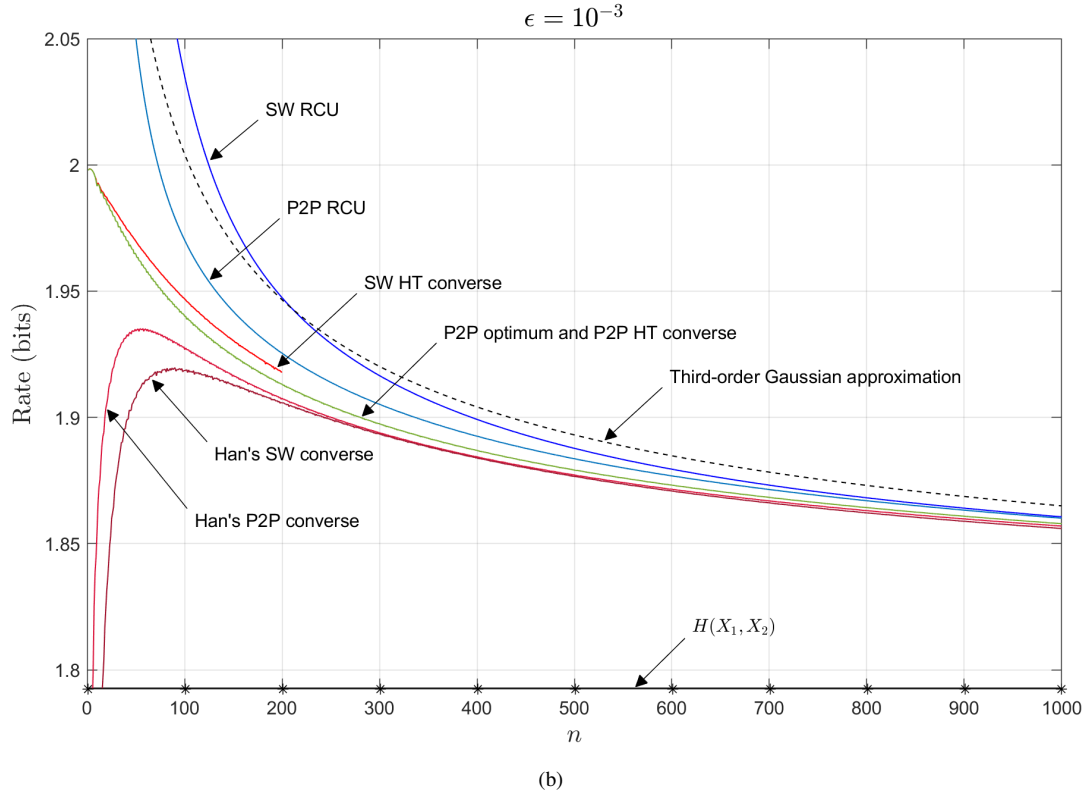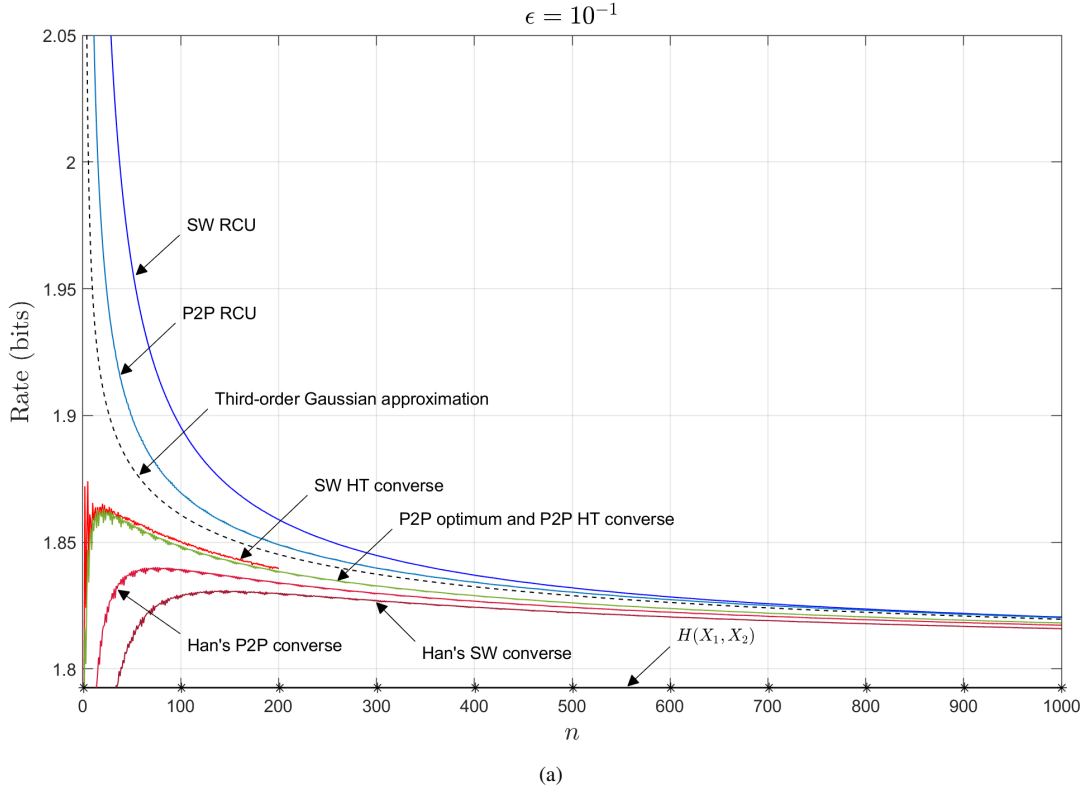
(a)

$\epsilon = 10^{-3}$

(b)

Fig. 4: Rate-blocklength trade-offs at (a) $\epsilon = 10^{-1}$ and (b) $\epsilon = 10^{-3}$ for a pair of binary, stationary, memoryless sources with joint distribution $p_{X_1 X_2}(0,0) = 1/2$, $p_{X_1 X_2}(0,1) = p_{X_1 X_2}(1,0) = p_{X_1 X_2}(1,1) = 1/6$. Due to computational limitations, we only plot the MASC HT converse for small blocklengths ($n \leq 200$). We evaluate the MASC HT converse with the sub-optimal choice of measures in (136), Han's point-to-point (P2P) converse is from [9, Lemma 1.3.2], Han's MASC converse is from Theorem 15 ([9, Lemma 7.2.2]), and the P2P HT converse is given in [23, Appendix A], which coincides with the optimum $R^*(n, \epsilon)$.

When X_1 and X_2 are dependent, $H(X_1) + H(X_2) > H(X_1, X_2) > H(X_1|X_2) + H(X_2|X_1)$, and the asymptotic sum-rate boundary contains non-corner and corner points. Corollary 21, below, shows that a MASC incurs no first-, second-, or third-order performance penalty relative to joint coding at non-corner points (i.e., when $R_1 < H(X_1)$ and $R_2 < H(X_2)$); in contrast, a MASC suffers a second-order performance penalty at corner points (i.e., when $R_1 = H(X_1)$ or $R_2 = H(X_2)$). See Figure 5(a) for an illustration.

Corollary 21. *Suppose that X_1 and X_2 are dependent.*

1) Fix constants $\delta_1, \delta_2, G > 0$ and $\epsilon \in (0,1)$. Then there exists some constant $n(\delta_1, \delta_2, G)$ such that if

$$R_1 \le H(X_1) - \delta_1 \qquad (197)$$
$$R_2 \le H(X_2) - \delta_2 \qquad (198)$$
$$R_1 + R_2 = H(X_1, X_2) + \sqrt{\frac{V(X_1, X_2)}{n}} Q^{-1}\left(\epsilon - \frac{G}{\sqrt{n}}\right)$$
$$- \frac{\log n}{2n} \qquad (199)$$

then $\mathbf{R} = (R_1, R_2) \in \overline{\mathscr{R}}^*(n, \epsilon)$ for all $n > n(\delta_1, \delta_2, G)$.

2) Fix $\epsilon \in (0,1)$. If

$$R_2 \ge H(X_2|X_1) + \frac{r^*}{\sqrt{n}} - \frac{\log n}{2n} + \frac{G}{n} \qquad (200)$$

for some $G > 0$, then $\mathbf{R} = (H(X_1), R_2) \in \overline{\mathscr{R}}^*(n, \epsilon)$. Conversely, if $\mathbf{R} = (H(X_1), R_2) \in \overline{\mathscr{R}}^*(n, \epsilon)$, then

$$R_2 \ge H(X_2|X_1) + \frac{r^*}{\sqrt{n}} - \frac{\log n}{2n}, \qquad (201)$$

where r^* is the solution to equation

$$\Phi(\mathsf{V}_2; r, r) = 1 - \epsilon, \qquad (202)$$

and V_2 is the covariance matrix for random vector $(\imath(X_2|X_1), \imath(X_1, X_2))$.

Proof. Appendix H. ∎

For independent sources, the asymptotic sum-rate boundary contains only the single (corner) point $(R_1, R_2) = (H(X_1), H(X_2))$, and the entropy dispersion matrix

$$\begin{bmatrix} V(X_1) & 0 \\ 0 & V(X_2) \end{bmatrix}.$$

is singular.

The next result concerns the third-order-optimal sum rate

$$\overline{R}^*_{\text{sum}}(n, \epsilon) \triangleq \min\left\{ R_1 + R_2 : \exists \mathbf{R} = (R_1, R_2) \right.$$
$$\left. \text{s.t. } \overline{\mathbf{R}} \in \overline{\mathscr{R}}^*(n, \epsilon) \right\}. \qquad (203)$$

According to Theorem 20, $\overline{R}^*_{\text{sum}}(n, \epsilon)$ characterizes the best achievable sum rate in SW source coding up to an $O(\frac{1}{n})$ gap.

Corollary 22. *For X_1, X_2 independent and $\epsilon \in (0,1)$,*

$$\overline{R}^*_{\text{sum}}(n, \epsilon) \qquad (204)$$
$$= H(X_1) + H(X_2) + \frac{\sqrt{V(X_1)} r_1^* + \sqrt{V(X_2)} r_2^*}{\sqrt{n}} - \frac{\log n}{2n},$$

$$a_1 = \sqrt{\frac{V(X_1, X_2)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} \pm O\left(\frac{1}{n}\right) \quad a_2 = \frac{r^*}{\sqrt{n}} - \frac{\log n}{2n} \pm O\left(\frac{1}{n}\right)$$

(a)

$$b_1 = \sqrt{\frac{V(X_1)}{n}} r_1^* - \frac{\log n}{2n} \pm O\left(\frac{1}{n}\right) \quad b_2 = \sqrt{\frac{V(X_2)}{n}} r_2^* - \frac{\log n}{2n} \pm O\left(\frac{1}{n}\right)$$
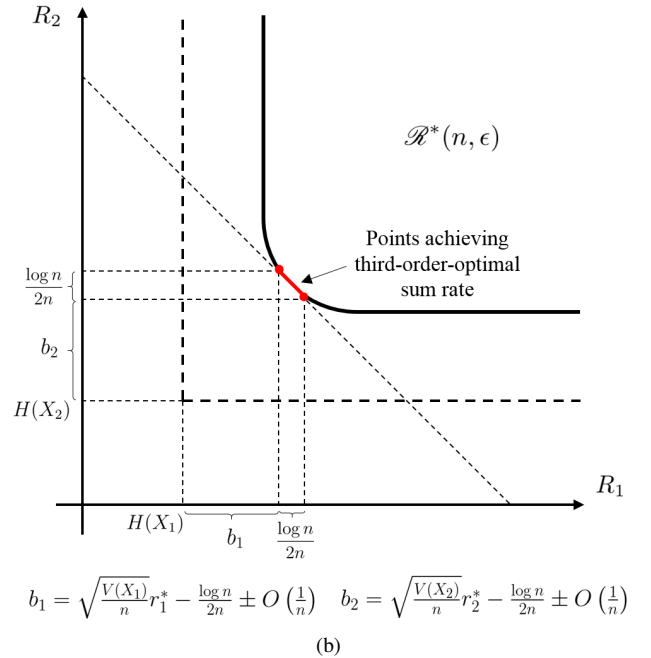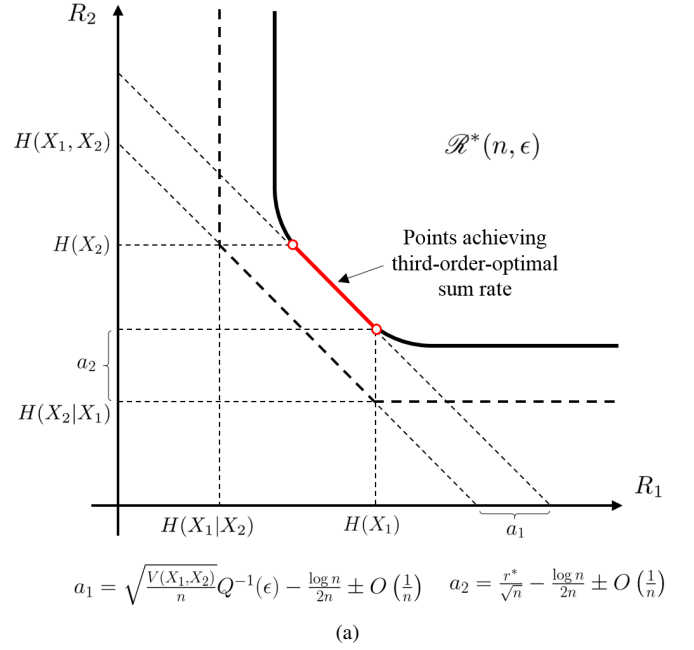
(b)

Fig. 5: Schematic plots of the (n, ϵ)-rate region and the third-order-optimal sum rate when (a) X_1, X_2 are dependent, (b) X_1, X_2 are independent. In (a), the boundary of $\mathscr{R}^*(n, \epsilon)$ between $H(X_1)$ and $H(X_2)$ (excluding the end points) contains rate points that achieve the optimal point-to-point rate up to the third order, while the end points do not achieve that optimal rate. The value of r^* in (a) is defined in (202); the values of r_1^*, r_2^* in (b) are defined in (207).

which is achieved by $\mathbf{R} = (R_1, R_2)$ with

$$R_1 = H(X_1) + \sqrt{\frac{V(X_1)}{n}} r_1^* - \lambda \frac{\log n}{2n} \qquad (205)$$

$$R_2 = H(X_2) + \sqrt{\frac{V(X_2)}{n}} r_2^* - (1 - \lambda) \frac{\log n}{2n}, \qquad (206)$$

for any $\lambda \in [0,1]$ *and*

$$(r_1^*, r_2^*) = \arg \min_{\substack{(r_1, r_2): \\ \Phi(r_1)\Phi(r_2) \geq 1-\epsilon}} \left(\sqrt{V(X_1)}r_1 + \sqrt{V(X_2)}r_2 \right).$$

$$(207)$$

Proof. Appendix I. ∎

By Corollary 22, for independent sources a unique (r_1^*, r_2^*) captures the best MASC second-order sum-rate; the third-order term is achieved at all points on a segment of the rate region boundary. See Figure 5(b). Under assumption (139),

$$\min_{\substack{(r_1, r_2): \\ \Phi(r_1)\Phi(r_2) \geq 1-\epsilon}} \left(\sqrt{V(X_1)}r_1 + \sqrt{V(X_2)}r_2 \right)$$
$$> \sqrt{V(X_1)+V(X_2)}Q^{-1}(\epsilon), \qquad (208)$$

where $V(X_1)+V(X_2) = V(X_1, X_2)$ for (X_1, X_2) independent. Here (208) follows since its left-hand side solves

$$\min_{(a_1, a_2)} (a_1 + a_2)$$

$$\text{s.t. } \Phi\left(\frac{a_1}{\sqrt{V(X_1)}} \right) \Phi\left(\frac{a_2}{\sqrt{V(X_2)}} \right) \geq 1-\epsilon, \qquad (209)$$

and the constraint in (209) requires $a_1 > \sqrt{V(X_1)}Q^{-1}(\epsilon)$ and $a_2 > \sqrt{V(X_2)}Q^{-1}(\epsilon)$, which gives the bound since

$$\sqrt{V(X_1)} + \sqrt{V(X_2)} > \sqrt{V(X_1)+V(X_2)}. \qquad (210)$$

Therefore, when X_1 and X_2 are independent, a MASC incurs a positive second-order sum-rate penalty relative to joint coding. Closed-form expressions for this penalty are available in special cases. When $V(X_1) = V(X_2)$, $r_1^* = r_2^* = Q^{-1}(1-\sqrt{1-\epsilon})$, and the penalty is

$$2\sqrt{\frac{V(X_1)}{n}}Q^{-1}(1-\sqrt{1-\epsilon}) - \sqrt{\frac{2V(X_1)}{n}}Q^{-1}(\epsilon). \qquad (211)$$

When X_1 and X_2 are i.i.d., the penalty equals the penalty for coding a vector X^{2n} of $2n$ i.i.d. outputs from P_X by applying an independent (n, ϵ) (point-to-point) code with error probability $1-\sqrt{1-\epsilon}$ to each of $(X_1, \ldots, X_n)$ and $(X_{n+1}, \ldots, X_{2n})$ instead of a single $(2n, \epsilon)$ code to vector X^{2n}.

E. Limited Feedback and Cooperation

The RASC proposed in Section VI employs limited feedback. We here analyze the impact of feedback on the underlying MASC. In our feedback model, the decoder broadcasts the same ℓ bits of feedback to both encoders. A bit sent at time i must be a function of the encoder outputs received in time steps $1, \ldots, i-1$. (See Figure 6(a).) We bound the impact of feedback by studying a MASC with a *cooperation facilitator (CF)*.[8] The CF broadcasts the same ℓ-bit function of the sources to both encoders prior to their encoding operations. (See Figure 6(b).) Since the MASC network has no channel noise, feedback from the decoder cannot convey more information than feedback from the CF. As a result, we bound

---

[8]The CF is introduced for multiple access channel coding in [36] and extended to source and network coding in [37].
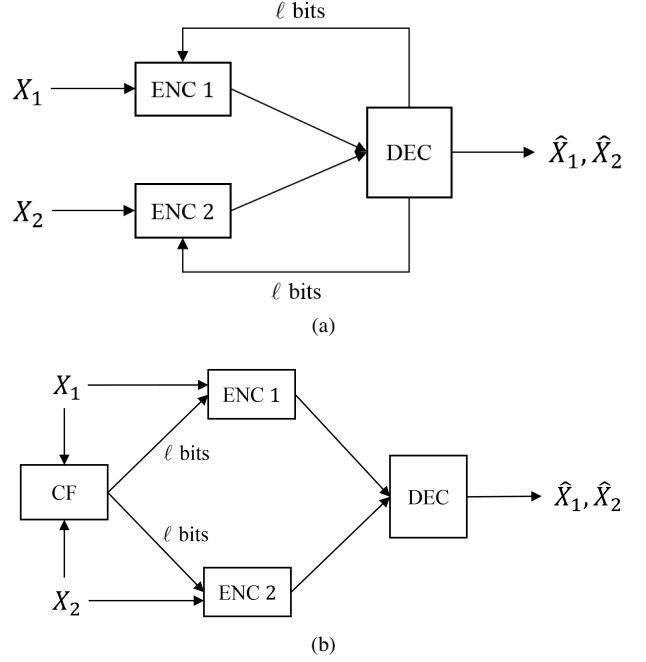

Fig. 6: The (a) FB-MASC and (b) CF-MASC.

the impact of feedback by bounding the impact of cooperation, which is easier to work with in our analysis.

We begin by defining the CF-MASC and its rate region.

Definition 9 (CF-MASC). *An* (L, M_1, M_2, ϵ) *CF-MASC for random variables* (X_1, X_2) *on* $\mathcal{X}_1 \times \mathcal{X}_2$ *comprises a CF function* L, *two encoding functions* f_1 *and* f_2, *and a decoding function* g *given by*

$$\mathsf{L} : \mathcal{X}_1 \times \mathcal{X}_2 \to [L]$$
$$\mathsf{f}_1 : [L] \times \mathcal{X}_1 \to [M_1]$$
$$\mathsf{f}_2 : [L] \times \mathcal{X}_2 \to [M_2]$$
$$\mathsf{g} : [M_1] \times [M_2] \to \mathcal{X}_1 \times \mathcal{X}_2,$$

with error probability

$$\mathbb{P}[\mathsf{g}(\mathsf{f}_1(\mathsf{L}(X_1, X_2), X_1), \mathsf{f}_2(\mathsf{L}(X_1, X_2), X_2)) \neq (X_1, X_2)] \leq \epsilon.$$

Definition 10 (Block CF-MASC). *An* $(n, L, M_1, M_2, \epsilon)$ *MASC is a CF-MASC for random variables* (X_1^n, X_2^n) *on* $\mathcal{X}_1^n \times \mathcal{X}_2^n$.

The code's finite blocklength rates are defined by

$$R_1 = \frac{1}{n} \log M_1, \ R_2 = \frac{1}{n} \log M_2. \qquad (212)$$

Definition 11 ((n, ℓ, ϵ)-CF rate region). *A rate pair* (R_1, R_2) *is* (n, ℓ, ϵ)*-CF achievable if there exists an* $(n, L, M_1, M_2, \epsilon)$ *CF-MASC with* $M_1 \leq \exp(nR_1)$, $M_2 \leq \exp(nR_2)$, *and* $L \leq \exp(\ell)$. *The* (n, ℓ, ϵ)*-CF rate region* $\mathscr{R}_{\mathrm{CF}}^*(n, \ell, \epsilon)$ *is defined as the closure of the set of all* (n, ℓ, ϵ)*-CF achievable rate pairs.*

We use $\mathscr{R}_{\mathrm{FB}}^*(n, \ell, \epsilon)$ to denote the feedback-MASC (FB-MASC) rate region, which is defined as the closure of the set of all (n, ϵ)-achievable rate pairs when the same ℓ bits of feedback from the decoder are available to both encoders.

Since the CF sees the source vectors while the decoder sees a coded description of those vectors (using a deterministic

code), an ℓ-bit CF can implement any function used to determine the decoder's ℓ-bit feedback. As a result, any rate point that is achievable by an ℓ-bit FB-MASC is also achievable by an ℓ-bit CF-MASC. Therefore, for any $0 < \epsilon < 1$ and $\ell < \infty$,

$$\mathscr{R}_{\mathrm{FB}}^*(n, \ell, \epsilon) \subseteq \mathscr{R}_{\mathrm{CF}}^*(n, \ell, \epsilon). \tag{213}$$

Theorem 23 bounds CF-MASC (and FB-MASC) performance, showing that for any $\ell < \infty$, the third-order rate region for ℓ-bit CF-MASCs cannot exceed the corresponding MASC rate region. Hence finite feedback does not enlarge the third-order (n, ϵ) MASC rate region. This result generalizes to scenarios with more than two encoders.

Theorem 23 (CF-MASC Converse)**.** *Consider stationary, memoryless sources with single-letter distribution $P_{X_1 X_2}$ satisfying* (139) *and* (140)*. For any $0 < \epsilon < 1$ and $\ell < \infty$,*

$$\mathscr{R}_{\mathrm{CF}}^*(n, \ell, \epsilon) \subseteq \mathscr{R}_{\mathrm{out}}^*(n, \epsilon). \tag{214}$$

Thus $\mathscr{R}_{\mathrm{CF}}^*(n, \ell, \epsilon)$ and $\overline{\mathscr{R}}^*(n, \epsilon)$ share the same outer bound.

Proof. Appendix J. ∎

Remark 13. The same proof can be used to show that allowing ℓ to grow as $o(\log \log n)$ does not change the first three terms in the optimal characterization of the (n, ϵ)-MASC.

Remark 14. For dependent sources, the optimal third-order MASC sum rate equals the optimal third-order sum rate with full cooperation. (See the discussion in Section V-D1, above.). Since even an infinite amount of decoder feedback is weaker than full cooperation, an infinite amount of feedback does not improve the third-order sum rate in this case.

VI. RANDOM ACCESS SOURCE CODE (RASC)

An RASC is a generalization of an MASC for networks where the set of participating encoders is unknown to both the encoders and the decoder *a priori*. We begin by defining the problem and describing our proposed communication strategy.

A. Definitions and Coding Strategy

Let $K < \infty$ be the maximal number of active encoders. We associate each encoder with a unique source from the set of sources indexed by $[K]$. Each encoder chooses whether to be *active* or *silent*. Only sources associated with active encoders are compressed and reconstructed. By assumption, the decision to remain silent is independent of the observed source instance. Given the joint distribution $P_{\mathbf{X}_{[K]}}$ on countable alphabet $\mathcal{X}_{[K]}$, when ordered set $\mathcal{T} \in \mathcal{P}([K])$ of $[K]$ is active, the marginal on the transmitted sources is

$$P_{\mathbf{X}_{\mathcal{T}}}(\mathbf{x}_{\mathcal{T}}) = \sum_{\mathbf{x}_{[K] \setminus \mathcal{T}} \in \mathcal{X}_{[K] \setminus \mathcal{T}}} P_{\mathbf{X}_{[K]}}(\mathbf{x}_{[K]}), \ \forall \mathbf{x}_{\mathcal{T}} \in \mathcal{X}_{\mathcal{T}}. \tag{215}$$

Thus, each encoder's state has no effect on the statistical relationship among sources observed by other encoders.

As in the random access *channel code* from [18], our proposed RASC organizes communication into epochs. At the beginning of each epoch, each encoder independently decides its activity state; that activity state remains unchanged until the end of the epoch. Thus, the active encoder set $\mathcal{T}$ is fixed in
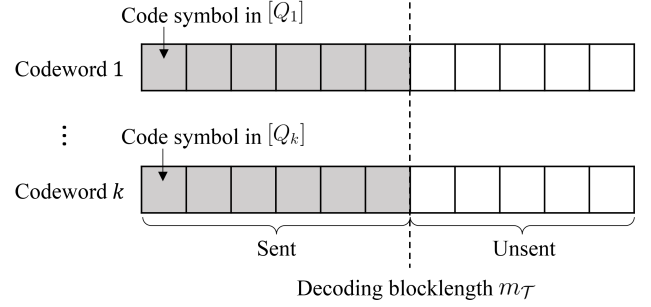

Fig. 7: Coding scheme in one epoch with $\mathcal{T} = [k]$.

each epoch. Each active encoder $i \in \mathcal{T}$ observes source output $X_i \in \mathcal{X}_i$ and independently maps it to a codeword comprised of a sequence of *code symbols* from alphabet $[Q_i]$. The $|\mathcal{T}|$ codewords are sent simultaneously to the decoder. Since set $\mathcal{T}$ is unknown *a priori*, the encoder behavior cannot vary with $\mathcal{T}$. The decoder sees $\mathcal{T}$ and decides a time $m_{\mathcal{T}}$, called the *decoding blocklength*, at which to jointly decode all received partial codewords. The set of potential decoding blocklengths $\mathcal{M} \triangleq \{m_{\mathcal{T}} : \mathcal{T} \in \mathcal{P}([K])\}$ is part of the code design; it is known to all encoders and to the decoder.

Figure 7 illustrates our coding scheme in one epoch when $\mathcal{T} = [k]$. Each encoder $i \in \mathcal{T}$ sends a single code symbol per time step. At each time $m \in \{m' \in \mathcal{M} : m' < m_{\mathcal{T}}\}$, the decoder sends a "0" to indicate that it is not yet ready to decode; at time $m = m_{\mathcal{T}}$, the decoder sends a "1," ending one epoch and starting the next. The decoder then reconstructs source vector $\mathbf{X}_{\mathcal{T}}$ using the first $m_{\mathcal{T}}$ code symbols from each active encoder. To avoid wasting time in an epoch with no active encoders, we include decoding time $m_{\emptyset} = 1$ in set $\mathcal{M}$. The decoder sends at most 2^K bits of feedback, and encoders need only listen for decoder feedback at the times in set $\mathcal{M}$.

To formalize the above strategy, fix $K \geq 1$. Define vectors

$$\overline{\boldsymbol{\epsilon}}_K \triangleq (\epsilon_{\mathcal{T}}, \mathcal{T} \in \mathcal{P}([K])) \tag{216}$$

$$\overline{\mathbf{m}}_K \triangleq (m_{\mathcal{T}}, \mathcal{T} \in \mathcal{P}([K]) \cup \{\emptyset\}) \tag{217}$$

with $m_{\emptyset} = 1$ and $m_{\max} \triangleq \max\{m_{\mathcal{T}} : \mathcal{T} \in \mathcal{P}([K])\}$.

Definition 12 (RASC)**.** *An $(\overline{\mathbf{m}}_K, \mathbf{Q}_{[K]}, \overline{\boldsymbol{\epsilon}}_K)$ RASC for sources $X_{[K]}$ on source alphabet $\mathcal{X}_{[K]}$ comprises a collection of encoding and decoding functions*

$$\mathsf{f}_i : \mathcal{X}_i \to [Q_i]^{m_{\max}}, \ i \in [K], \tag{218}$$

$$\mathsf{g}_{\mathcal{T}} : \prod_{i \in \mathcal{T}} [Q_i]^{m_{\mathcal{T}}} \to \mathcal{X}_{\mathcal{T}}, \ \mathcal{T} \in \mathcal{P}([K]), \tag{219}$$

where f_i is the encoding function for source X_i and $\mathsf{g}_{\mathcal{T}}$ is the decoding function for active coder set $\mathcal{T}$. For each $\mathcal{T} \in \mathcal{P}([K])$, source vector $\mathbf{X}_{\mathcal{T}}$ is decoded at time $m_{\mathcal{T}}$ with error probability $\mathbb{P}\big[\mathsf{g}_{\mathcal{T}}\big(\mathsf{f}_i(X_i)_{[m_{\mathcal{T}}]}, i \in \mathcal{T}\big) \neq \mathbf{X}_{\mathcal{T}}\big] \leq \epsilon_{\mathcal{T}}$, where $\mathsf{f}_i(x_i)_{[m]}$ denotes the first m code symbols of $\mathsf{f}_i(x_i)$.

Definition 13 particularizes Definition 12 to the block setting.

Definition 13 (Block RASC)**.** *An $(n, \overline{\mathbf{m}}_K, \mathbf{Q}_{[K]}, \overline{\boldsymbol{\epsilon}}_K)$ RASC is an RASC for an n-block of source outcomes. The parameter n, called the encoding blocklength does not vary with $\mathcal{T}$.*
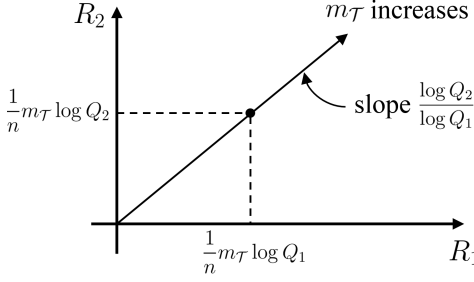
Fig. 8: The relationship between decoding blocklength $m_{\mathcal{T}}$, code symbol alphabet sizes (Q_1, Q_2), and source coding rate vector $\mathbf{R}_{\mathcal{T}}$, illustrated for $\mathcal{T} = \{1, 2\}$.

An $(\overline{\mathbf{m}}_K, \mathbf{Q}_{[K]}, \overline{\boldsymbol{\epsilon}}_K)$ RASC behaves, for each $\mathcal{T}$, like a $\big((Q_i^{m_{\mathcal{T}}}, i \in \mathcal{T}), \epsilon_{\mathcal{T}}\big)$ MASC (see Definition 5) with a finite number $|\{m \in \mathcal{M} : m \leq m_{\mathcal{T}}\}|$ of feedback bits. However, the RASC is *one code*. Its descriptions are nested (i.e., for each $x_i \in \mathcal{X}_i$, if $m_{\mathcal{T}'} < m_{\mathcal{T}}$, then $\mathsf{f}_i(x_i)_{[m_{\mathcal{T}'}]}$ is a prefix of $\mathsf{f}_i(x_i)_{[m_{\mathcal{T}}]}$). It simultaneously satisfies the error constraints for all $\mathcal{T} \in \mathcal{P}([K])$. And, since the code symbol alphabet sizes $\mathbf{Q}_{[K]}$ are fixed, its rate vectors are coupled. See Figure 8.

The following definitions build toward the non-asymptotic fundamental limit of RASCs.

Definition 14 (n-Valid and $(n, \overline{\boldsymbol{\epsilon}}_K)$-Rate sets). *A collection* $(\mathbf{R}_{\mathcal{T}})_{\mathcal{T} \in \mathcal{P}([K])}$ *of rate vectors is n-valid if $\exists \big(\overline{\mathbf{m}}_K, \mathbf{Q}_{[K]}\big)$ s.t.*

$$\mathbf{R}_{\mathcal{T}} = \frac{1}{n}(m_{\mathcal{T}} \log Q_i, i \in \mathcal{T}), \forall \mathcal{T} \in \mathcal{P}([K]). \quad (220)$$

The set $\mathcal{R}_{\mathrm{valid}}(n)$ is the set of n-valid rate collections. The collection is $(n, \overline{\boldsymbol{\epsilon}}_K)$-achievable if there exists an $\big(n, \overline{\mathbf{m}}_K, \mathbf{Q}_{[K]}, \overline{\boldsymbol{\epsilon}}_K\big)$ RASC. The $(n, \overline{\boldsymbol{\epsilon}}_K)$-rate set $\mathcal{R}^*(n, \overline{\boldsymbol{\epsilon}}_K)$ is the set of $(n, \overline{\boldsymbol{\epsilon}}_K)$-achievable rate collections.

B. Background

While the concept of an RASC is new, the RASC problem is related to the universal MASC problem. Like a universal MASC, the RASC is designed for an unknown distribution from a known collection of possible distributions. In this case, the possible distributions are $\{P_{\mathbf{X}_{\mathcal{T}}} : \mathcal{T} \in \mathcal{P}([K])\}$. The RASC differs, however, from universal MASCs since even the set of active encoders is unknown *a priori*.

A short summary of prior universal MASCs follows.

1) For a fixed-rate MASC and finite source alphabets, universal decoding can be realized using type methods. (See [13], [38], [39].) Such strategies achieve optimal performance only when the source's MASC rate region matches the code's fixed rate.
2) Oohama [40] and Jaggi and Effros [41] study the effect of limited encoder cooperation on the asymptotically universally achievable rate region. Rate-zero cooperation between encoders suffices to achieve universality in the asymptotic regime. Oohama characterizes the optimal error exponents in [40].
3) Yang et al. [42] study a block MASC with progressive encoding; the code uses zero-rate feedback to universally achieve the asymptotic MASC rate region. Sarvotham et

al. [43] propose a variable-rate block sequential coding scheme with zero-rate feedback for binary symmetric sources, showing that at blocklength n and target error probability ϵ, the backoff from the asymptotic MASC rate due to universality is $O\big(\frac{1}{\sqrt{n}} Q^{-1}(\epsilon)\big)$.
4) In [22], Draper introduces a rateless MASC with single-bit feedback. Draper's algorithm asymptotically achieves the optimal coding rates for sources with unknown joint distributions but known finite alphabet sizes. See [44] for a practical rateless MASC.

C. Asymptotics: Third-Order Performance of the RASC

In this section, we analyze the performance of an $\big(n, \overline{\mathbf{m}}_K, \mathbf{Q}_{[K]}, \overline{\boldsymbol{\epsilon}}_K\big)$ RASC for stationary, memoryless sources. Results include both achievability and converse characterizations of the $(n, \overline{\boldsymbol{\epsilon}}_K)$-rate set $\mathcal{R}^*(n, \overline{\boldsymbol{\epsilon}}_K)$ under the assumption that the single-letter joint source distribution $P_{\mathbf{X}_{[K]}}$ satisfies

$$\mathbb{E}\Big[V_c\big(\mathbf{X}_{\hat{\mathcal{T}}} | \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}}}\big)\Big] > 0 \quad \forall \hat{\mathcal{T}} \subseteq \mathcal{T} \subseteq [K], \hat{\mathcal{T}}, \mathcal{T} \neq \emptyset \quad (221)$$

$$T\big(\mathbf{X}_{\hat{\mathcal{T}}} | \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}}}\big) < \infty \qquad \forall \hat{\mathcal{T}} \subseteq \mathcal{T} \subseteq [K], \hat{\mathcal{T}}, \mathcal{T} \neq \emptyset. \quad (222)$$

$$\mathbb{E}\Big[T_c^2\big(\mathbf{X}_{\hat{\mathcal{T}}} | \mathbf{X}_{\mathcal{T} \setminus \hat{\mathcal{T}}}\big)\Big] < \infty \quad \forall \hat{\mathcal{T}} \subset \mathcal{T} \subseteq [K], \hat{\mathcal{T}}, \mathcal{T} \neq \emptyset. \quad (223)$$

Constraints (221)–(223) enable us to use Berry-Esseen bounds. The resulting characterization is tight up to the third-order term. While the existence of an $\big(n, \overline{\mathbf{m}}_K, \mathbf{Q}_{[K]}, \overline{\boldsymbol{\epsilon}}_K\big)$ RASC implies the existence of an $\big(n, (Q_i^{m_{\mathcal{T}}}, i \in \mathcal{T}), \epsilon_{\mathcal{T}}\big)$ MASC for each $\mathcal{T} \in \mathcal{P}([K])$, the existence of individual MASCs does not imply the existence of a single RASC that simultaneously satisfies the error probability constraints for all possible configurations of active encoders. Indeed, the existence of a single RASC that simultaneously performs as well (up to the third-order term) as the optimal MASC for each $\mathcal{T} \in \mathcal{P}([K])$ is one of our most surprising results.

Define the inner and outer bounding sets

$$\mathcal{R}_{\mathrm{in}}^*(n, \overline{\boldsymbol{\epsilon}}_K) \triangleq \Big\{ (\mathbf{R}_{\mathcal{T}})_{\mathcal{T} \in \mathcal{P}([K])} \in \mathcal{R}_{\mathrm{valid}}(n) :$$
$$\mathbf{R}_{\mathcal{T}} \in \mathscr{R}_{\mathrm{in}, \mathcal{T}}^*(n, \epsilon_{\mathcal{T}}) \ \forall \mathcal{T} \in \mathcal{P}([K]) \Big\} \quad (224)$$

$$\mathcal{R}_{\mathrm{out}}^*(n, \overline{\boldsymbol{\epsilon}}_K) \triangleq \Big\{ (\mathbf{R}_{\mathcal{T}})_{\mathcal{T} \in \mathcal{P}([K])} \in \mathcal{R}_{\mathrm{valid}}(n) :$$
$$\mathbf{R}_{\mathcal{T}} \in \mathscr{R}_{\mathrm{out}, \mathcal{T}}^*(n, \epsilon_{\mathcal{T}}) \ \forall \mathcal{T} \in \mathcal{P}([K]) \Big\}, \quad (225)$$

where $\mathscr{R}_{\mathrm{in}, \mathcal{T}}^*(n, \epsilon)$ and $\mathscr{R}_{\mathrm{out}, \mathcal{T}}^*(n, \epsilon)$ are the third-order MASC bounding sets for distribution $P_{\mathbf{X}_{\mathcal{T}}}$. (See (194) and (195).)

Theorem 24 (Third-order RASC performance). *For any $K < \infty$, consider stationary, memoryless sources specified by a single-letter joint distribution $P_{\mathbf{X}_{[K]}}$ satisfying (221)–(223). For any $0 < \overline{\boldsymbol{\epsilon}}_K < 1$,*

$$\mathcal{R}_{\mathrm{in}}^*(n, \overline{\boldsymbol{\epsilon}}_K) \subseteq \mathcal{R}^*(n, \overline{\boldsymbol{\epsilon}}_K) \subseteq \mathcal{R}_{\mathrm{out}}^*(n, \overline{\boldsymbol{\epsilon}}_K). \quad (226)$$

The converse and achievability proofs follow.

Proof of Theorem 24: converse. As shown in Section V-E (Theorem 23), even with *a priori* knowledge of the encoder set $\mathcal{T} \in \mathcal{P}([K])$ and 2^K bits of feedback, a MASC for the encoders in set $\mathcal{T}$ cannot achieve performance outside of the third-order MASC outer bounding set $\mathscr{R}_{\mathrm{out}, \mathcal{T}}^*(n, \epsilon_{\mathcal{T}})$. ∎

The achievability part of Theorem 24 provides a sufficient condition for the existence of a *single* RASC that is simultaneously good for all $\mathcal{T} \in \mathcal{P}([K])$. To prove this, we first derive an achievability result assuming that the encoders and decoder share the common randomness used to generate a random code (Theorem 26). Unfortunately, the existence of a random code ensemble with expected error probability satisfying the error probability constraint for each $\mathcal{T} \in \mathcal{P}([K])$ does not guarantee the existence of a *single* deterministic code satisfying those constraints simultaneously. We therefore take a different approach, which, unexpectedly, combines a converse bound on error probability and a random coding argument to show achievability.

The following refinement of the random coding argument provides a bound on the probability (with respect to the random code choice) that the error probability of a randomly chosen code exceeds a certain threshold. The code of interest here can be any type of source or channel code.

Lemma 25. *Let $\mathcal{C}$ be any class of codes with a corresponding error probability $P_e(\mathsf{c})$ for each $\mathsf{c} \in \mathcal{C}$. Let*

$$\epsilon^*(\mathcal{C}) = \min_{\mathsf{c} \in \mathcal{C}} P_e(\mathsf{c}) \tag{227}$$

denote the error probability of the best code in $\mathcal{C}$. Then any random code ensemble[9] C defined over $\mathcal{C}$ satisfies

$$\mathbb{P}[P_e(\mathsf{C}) > \epsilon] \le \frac{\mathbb{E}[P_e(\mathsf{C})] - \epsilon^*(\mathcal{C})}{\epsilon - \epsilon^*(\mathcal{C})}, \ \forall \epsilon > \epsilon^*(\mathcal{C}). \tag{228}$$

Proof. Let Y be any non-negative random variable and define $y_{\min} \triangleq \operatorname{ess\,inf} Y$; that is, $y_{\min}$ is the largest constant $y \in \mathcal{Y}$ for which $Y \ge y$ almost surely. By Markov's inequality,

$$\begin{aligned}\mathbb{P}[Y \ge y] &= \mathbb{P}[Y - y_{\min} \ge y - y_{\min}] \\ &\le \frac{\mathbb{E}[Y] - y_{\min}}{y - y_{\min}} \quad \forall y > y_{\min}.\end{aligned}$$

Taking $Y = P_e(\mathsf{C})$ and $y = \epsilon$ yields the desired result. ∎

In the regime of interest $\mathbb{E}[P_e(\mathsf{C})] < \epsilon$. Therefore, the right side of (228) is decreasing as a function of $\epsilon^*(\mathcal{C})$, and replacing $\epsilon^*(\mathcal{C})$ by any converse on $\epsilon^*(\mathcal{C})$ yields a valid achievability bound. Thus Lemma 25 provides a means to leverage a converse to prove achievability.

Given any RASC c, for each $\mathcal{T} \in \mathcal{P}([K])$ let $P_{e,\mathcal{T}}(\mathsf{c})$ denote the error probability of code c under active encoder set $\mathcal{T}$. The RASC achievability proof applies Lemma 25 with error probability $P_{e,\mathcal{T}}(\mathsf{c})$ for each $\mathcal{T} \in \mathcal{P}([K])$. Before proceeding to that proof, we use Theorem 26, below, to define a random code ensemble and calculate its expected error probability.

Theorem 26 (Random code). *For any $K < \infty$, consider a source distribution $P_{\mathbf{X}_{[K]}}$ defined on countable alphabet $\mathcal{X}_{[K]}$. There exists a random code ensemble C defined on the set of all RASCs with decoding blocklengths $\overline{\mathbf{m}}_K$ and code alphabets*

$\mathbf{Q}_{[K]}$ *for which the following inequalities hold simultaneously for all $\mathcal{T} \in \mathcal{P}([K])$:*

$$\mathbb{E}[P_{e,\mathcal{T}}(\mathsf{C})] \le \mathbb{E}\left[\min\left\{1, \sum_{\hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})} \exp\left(-m_{\mathcal{T}} \cdot \overline{Q}(\hat{\mathcal{T}})\right) A_{\hat{\mathcal{T}}}\right\}\right], \tag{229}$$

where

$$\overline{Q}(\hat{\mathcal{T}}) \triangleq \sum_{i \in \hat{\mathcal{T}}} \log Q_i \tag{230}$$

$$\begin{aligned}A_{\hat{\mathcal{T}}} \triangleq \mathbb{E}\big[\exp\left(\imath(\bar{\mathbf{X}}_{\hat{\mathcal{T}}}|\mathbf{X}_{\mathcal{T}\setminus\hat{\mathcal{T}}})\right)\cdot \\ \mathbb{1}\{\imath(\bar{\mathbf{X}}_{\hat{\mathcal{T}}}|\mathbf{X}_{\mathcal{T}\setminus\hat{\mathcal{T}}}) \le \imath(\mathbf{X}_{\hat{\mathcal{T}}}|\mathbf{X}_{\mathcal{T}\setminus\hat{\mathcal{T}}})\}\big|\mathbf{X}_{\mathcal{T}}\big] \tag{231}\end{aligned}$$

and the expectation in (231) *is with respect to the conditional distribution*

$$P_{\bar{\mathbf{X}}_{\hat{\mathcal{T}}}|\mathbf{X}_{\mathcal{T}}} = P_{\mathbf{X}_{\hat{\mathcal{T}}}|\mathbf{X}_{\mathcal{T}\setminus\hat{\mathcal{T}}}}. \tag{232}$$

Proof. We construct the random code ensemble C as follows.

Random Encoding Map: For every $i \in [K]$, draw encoder outputs $\mathsf{F}_i(x_i)$ for all $x_i \in \mathcal{X}_i$ i.i.d. uniformly at random from $[Q_i]^{m_{\max}}$, where $m_{\max} \triangleq \max\{m_{\mathcal{T}} : \mathcal{T} \in \mathcal{P}([K])\}$.

Maximum Likelihood Decoder: For any $m \in [m_{\max}]$, $x_i \in \mathcal{X}_i$, and $i \in [K]$, denote the first m symbols of $\mathsf{F}_i(x_i)$ by $\mathsf{F}_i(x_i)_{[m]}$. For each $\mathcal{T} \in \mathcal{P}([K])$, the maximum likelihood decoder $\mathsf{g}_{\mathcal{T}}$ for $\mathcal{T}$ observes the first $m_{\mathcal{T}}$ symbols from the encoders in $\mathcal{T}$, here denoted by

$$\mathsf{F}(\mathbf{x}_{\mathcal{T}})_{[m_{\mathcal{T}}]} \triangleq \big(\mathsf{F}_i(x_i)_{[m_{\mathcal{T}}]}\big)_{i \in \mathcal{T}}, \tag{233}$$

and, for each $\mathbf{c}_{\mathcal{T}} = (\mathbf{c}_i)_{i \in \mathcal{T}} \in \prod_{i \in \mathcal{T}}[Q_i]^{m_{\mathcal{T}}}$, produces the output

$$\mathsf{g}_{\mathcal{T}}(\mathbf{c}_{\mathcal{T}}) = \arg \min_{\substack{\mathbf{x}_{\mathcal{T}} \in \mathcal{X}_{\mathcal{T}}: \\ \mathsf{F}(\mathbf{x}_{\mathcal{T}})_{[m_{\mathcal{T}}]} = \mathbf{c}_{\mathcal{T}}}} \imath(\mathbf{x}_{\mathcal{T}}). \tag{234}$$

Expected Error Analysis: The expected error probability $\mathbb{E}[P_{e,\mathcal{T}}(\mathsf{C})]$ over the random code ensemble is bounded above by the probability of event

$$\begin{aligned}\mathcal{E}_{\mathcal{T}} \triangleq \{\exists \bar{\mathbf{x}}_{\mathcal{T}} \in \mathcal{X}_{\mathcal{T}}\setminus\{\mathbf{X}_{\mathcal{T}}\} : \\ \imath(\bar{\mathbf{x}}_{\mathcal{T}}) \le \imath(\mathbf{X}_{\mathcal{T}}), \mathsf{F}(\bar{\mathbf{x}}_{\mathcal{T}})_{[m_{\mathcal{T}}]} = \mathsf{F}(\mathbf{X}_{\mathcal{T}})_{[m_{\mathcal{T}}]}\}. \tag{235}\end{aligned}$$

It follows that

$$\mathbb{P}[\mathcal{E}_{\mathcal{T}}]$$

$$= \mathbb{P}\left[\bigcup_{\bar{\mathbf{x}}_{\mathcal{T}} \in \mathcal{X}_{\mathcal{T}}\setminus\{\mathbf{X}_{\mathcal{T}}\}} \{\imath(\bar{\mathbf{x}}_{\mathcal{T}}) \le \imath(\mathbf{X}_{\mathcal{T}}), \right.$$
$$\left. \mathsf{F}(\bar{\mathbf{x}}_{\mathcal{T}})_{[m_{\mathcal{T}}]} = \mathsf{F}(\mathbf{X}_{\mathcal{T}})_{[m_{\mathcal{T}}]}\}\right] \tag{236}$$

$$= \mathbb{P}\left[\bigcup_{\hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})} \left\{\bigcup_{\substack{\bar{\mathbf{x}}_{\hat{\mathcal{T}}} \in \mathcal{X}_{\hat{\mathcal{T}}}: \\ \bar{\mathbf{x}}_i \ne \mathbf{X}_i \ \forall i \in \hat{\mathcal{T}}}} \{\imath(\bar{\mathbf{x}}_{\hat{\mathcal{T}}}, \mathbf{X}_{\mathcal{T}\setminus\hat{\mathcal{T}}}) \le \imath(\mathbf{X}_{\mathcal{T}}), \right.\right.$$
$$\left.\left. \mathsf{F}(\bar{\mathbf{x}}_{\hat{\mathcal{T}}})_{[m_{\mathcal{T}}]} = \mathsf{F}(\mathbf{X}_{\hat{\mathcal{T}}})_{[m_{\mathcal{T}}]}\}\right\}\right] \tag{237}$$

$$= \mathbb{P}\left[\bigcup_{\hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})} \left\{ \bigcup_{\substack{\bar{\mathbf{x}}_{\hat{\mathcal{T}}} \in \mathcal{X}_{\hat{\mathcal{T}}} : \\ \bar{\mathbf{x}}_i \neq \mathbf{X}_i \ \forall i \in \hat{\mathcal{T}}}} \left\{ \imath(\bar{\mathbf{x}}_{\hat{\mathcal{T}}}|\mathbf{X}_{\mathcal{T}\setminus\hat{\mathcal{T}}}) \leq \imath(\mathbf{X}_{\hat{\mathcal{T}}}|\mathbf{X}_{\mathcal{T}\setminus\hat{\mathcal{T}}}), \right. \right. \right.$$
$$\left. \left. \left. \mathsf{F}(\bar{\mathbf{x}}_{\hat{\mathcal{T}}})_{[m_{\mathcal{T}}]} = \mathsf{F}(\mathbf{X}_{\hat{\mathcal{T}}})_{[m_{\mathcal{T}}]} \right\} \right\} \right] \tag{238}$$

$$\leq \mathbb{E}\left[\min\left\{ 1, \sum_{\hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})} \exp\left(-m_{\mathcal{T}} \cdot \overline{Q}(\hat{\mathcal{T}})\right) \right. \right.$$
$$\left. \left. \cdot \sum_{\bar{\mathbf{x}}_{\hat{\mathcal{T}}} \in \mathcal{X}_{\hat{\mathcal{T}}}} \mathbb{1}\left\{ \imath(\bar{\mathbf{x}}_{\hat{\mathcal{T}}}|\mathbf{X}_{\mathcal{T}\setminus\hat{\mathcal{T}}}) \leq \imath(\mathbf{X}_{\hat{\mathcal{T}}}|\mathbf{X}_{\mathcal{T}\setminus\hat{\mathcal{T}}}) \right\} \right\} \right], \tag{239}$$

and (239) is equal to the right-hand-side of (229). Here, (237) considers the case where source symbols in set $\hat{\mathcal{T}}$ are decoded incorrectly for each $\hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})$. The derivation of (239) from (238) follows the argument in (122)–(124). Specifically, since each component of $\bar{\mathbf{x}}_{\hat{\mathcal{T}}}$ differs from the corresponding component of $\mathbf{X}_{\hat{\mathcal{T}}}$ and since the encoder output for each is drawn independently and uniformly at random from $[Q_i]^{m_{\max}}$,

$$\mathbb{P}\left[\mathsf{F}(\bar{\mathbf{x}}_{\hat{\mathcal{T}}})_{[m_{\mathcal{T}}]} = \mathsf{F}(\mathbf{X}_{\hat{\mathcal{T}}})_{[m_{\mathcal{T}}]} \big| \mathbf{X}_{\mathcal{T}}\right]$$
$$= \prod_{i \in \hat{\mathcal{T}}} \frac{1}{Q_i^{m_{\mathcal{T}}}} \tag{240}$$
$$= \exp\left(-m_{\mathcal{T}} \cdot \overline{Q}(\hat{\mathcal{T}})\right) \tag{241}$$

for any $\bar{\mathbf{x}}_{\hat{\mathcal{T}}} \in \mathcal{X}_{\hat{\mathcal{T}}} \setminus \{\mathbf{X}_{\hat{\mathcal{T}}}\}$. ∎

We now prove the achievability part of Theorem 24 by applying Lemma 25 to the random code in Theorem 26.

Proof of Theorem 24: achievability. The probability that random RASC C has error probability $P_{e,\mathcal{T}}(\mathsf{C})$ greater than $\epsilon_{\mathcal{T}}$ for some possible set $\mathcal{T} \in \mathcal{P}([K])$ of active encoders is

$$\mathbb{P}\left[\bigcup_{\mathcal{T} \in \mathcal{P}([K])} \{P_{e,\mathcal{T}}(\mathsf{C}) > \epsilon_{\mathcal{T}}\} \right] \leq \sum_{\mathcal{T} \in \mathcal{P}([K])} \mathbb{P}[P_{e,\mathcal{T}}(\mathsf{C}) > \epsilon_{\mathcal{T}}]. \tag{242}$$

To bound each term $\mathbb{P}[P_{e,\mathcal{T}}(\mathsf{C}) > \epsilon_{\mathcal{T}}]$ using Lemma 25, we next bound the expected error probability $\mathbb{E}[P_{e,\mathcal{T}}(\mathsf{C})]$ and the error probability $\epsilon^*(\mathcal{C}_{\mathcal{T}})$ for the best code in $\mathcal{C}_{\mathcal{T}}$, where $\mathcal{C}_{\mathcal{T}}$ is the set of $\left(n, (Q_i^{m_{\mathcal{T}}}, i \in \mathcal{T}), \epsilon_{\mathcal{T}}\right)$ MASCs with $m_{\mathcal{T}}$ set as in (248) below.

To find $\mathbb{E}[P_{e,\mathcal{T}}(\mathsf{C})]$, we apply Theorem 26 to our stationary, memoryless sources with n-symbol distribution $P_{\mathbf{X}_{[K]}^n} = P_{\mathbf{X}_{[K]}}^n$. Given any $\mathcal{T} \in \mathcal{P}([K])$ and $\hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})$, let

$$I_{\mathcal{T},\hat{\mathcal{T}}} \triangleq \imath\left(\mathbf{X}_{\hat{\mathcal{T}}}^n | \mathbf{X}_{\mathcal{T}\setminus\hat{\mathcal{T}}}^n\right). \tag{243}$$

Under moment assumptions (221)–(223), one can generalize the argument in (146)–(186) to $|\mathcal{T}|$ active encoders to obtain

$$\mathbb{E}[P_{e,\mathcal{T}}(\mathsf{C})] \leq \mathbb{P}\left[\bigcup_{\hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})} \left\{ I_{\mathcal{T},\hat{\mathcal{T}}} > m_{\mathcal{T}}\overline{Q}(\hat{\mathcal{T}}) + \frac{\log n}{2} \right. \right. \tag{244}$$
$$\left. \left. - \log\left(\bar{K}_{\mathcal{T},\hat{\mathcal{T}}}\left(2^{|\mathcal{T}|}-1\right)\right) \right\} \right] + \sum_{\hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})} \left(\frac{K_{\mathcal{T},\hat{\mathcal{T}}}}{\sqrt{n}} + \frac{S_{\mathcal{T},\hat{\mathcal{T}}}}{n} \right),$$

where $\bar{K}_{\mathcal{T},\hat{\mathcal{T}}}$, $K_{\mathcal{T},\hat{\mathcal{T}}}$ and $S_{\mathcal{T},\hat{\mathcal{T}}}$ are finite positive constants.

Fix any $\mathbf{Q}_{[K]}$. By the definition of $\overline{\mathbf{R}}_{\mathcal{T}}$ in (189) and the relation in (220), we see that

$$\overline{\mathbf{R}}_{\mathcal{T}} = \frac{1}{n}\left(m_{\mathcal{T}}\overline{Q}(\hat{\mathcal{T}}), \hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})\right). \tag{245}$$

For brevity, define constant vector

$$\mathbf{C}_{\mathcal{T}} \triangleq \left(\log\left(\bar{K}_{\mathcal{T},\hat{\mathcal{T}}}\left(2^{|\mathcal{T}|}-1\right)\right), \hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})\right). \tag{246}$$

and the almost-constant error thresholds

$$\epsilon_{\mathcal{T}}' \triangleq \epsilon_{\mathcal{T}} - \frac{B}{\sqrt{n}} - \sum_{\hat{\mathcal{T}} \in \mathcal{P}(\mathcal{T})} \left(\frac{K_{\mathcal{T},\hat{\mathcal{T}}}}{\sqrt{n}} + \frac{S_{\mathcal{T},\hat{\mathcal{T}}}}{n} \right) \tag{247}$$

where B is the Bentkus constant (81) for the vector of information densities (243). We choose the decoding blocklength $m_{\mathcal{T}}$ as

$$m_{\mathcal{T}} = \min\left\{ m_{\mathcal{T}} : \overline{\mathbf{R}}_{\mathcal{T}} \in \overline{\mathscr{R}}_{\mathcal{T}}^*(n, \epsilon_{\mathcal{T}}' - \delta_{\mathcal{T}}) + \frac{1}{n}\mathbf{C}_{\mathcal{T}} \right\}, \tag{248}$$

where $\delta_{\mathcal{T}}$ (which may be a function of n) satisfying $0 \leq \delta_{\mathcal{T}} < \epsilon_{\mathcal{T}}$ will be determined in the sequel, and $\overline{\mathscr{R}}_{\mathcal{T}}^*(n, \epsilon)$ is defined in (193). Applying Lemma 12 to (244) with $m_{\mathcal{T}}$ in (248) yields

$$\mathbb{E}[P_{e,\mathcal{T}}(\mathsf{C})] \leq \epsilon_{\mathcal{T}} - \delta_{\mathcal{T}}. \tag{249}$$

To lower-bound $\epsilon^*(\mathcal{C})$, for each n and ϵ define

$$m_{\mathcal{T}}^*(n, \epsilon) \triangleq \min\{m_{\mathcal{T}} : \mathbf{R}_{\mathcal{T}} \in \mathscr{R}_{\mathcal{T}}^*(n, \epsilon)\} \tag{250}$$
$$\geq \min\{m_{\mathcal{T}} : \mathbf{R}_{\mathcal{T}} \in \mathscr{R}_{\text{out},\mathcal{T}}^*(n, \epsilon)\}, \tag{251}$$

where $\mathscr{R}_{\mathcal{T}}^*(n, \epsilon)$ is the (n, ϵ)-MASC rate region (see Remark 12), $\mathscr{R}_{\text{out},\mathcal{T}}^*(n, \epsilon)$ is defined in (195), and (251) is by the converse (Theorem 23). By Lemma 13-2, one can always choose $\Delta_{\mathcal{T}} = O\left(\frac{1}{\sqrt{n}}\right)$ such that for n sufficiently large

$$\mathscr{R}_{\text{out},\mathcal{T}}^*(n, \epsilon_{\mathcal{T}}' - \delta_{\mathcal{T}} - \Delta_{\mathcal{T}}) \subseteq \overline{\mathscr{R}}_{\mathcal{T}}^*(n, \epsilon_{\mathcal{T}}' - \delta_{\mathcal{T}}) + \frac{1}{n}\mathbf{C}_{\mathcal{T}}. \tag{252}$$

It follows that

$$m_{\mathcal{T}}^*(n, \epsilon_{\mathcal{T}}' - \delta_{\mathcal{T}} - \Delta_{\mathcal{T}}) \geq m_{\mathcal{T}}. \tag{253}$$

Equation (253) and the converse (Theorem 23) imply that the minimal error probability over $\mathcal{C}_{\mathcal{T}}$ satisfies

$$\epsilon^*(\mathcal{C}_{\mathcal{T}}) \geq \epsilon_{\mathcal{T}}' - \delta_{\mathcal{T}} - \Delta_{\mathcal{T}}. \tag{254}$$

Plugging (249) and (254) into Lemma 25 and noting the monotonicity of the bound in Lemma 25 gives

$$\mathbb{P}[P_{e,\mathcal{T}}(\mathsf{C}) > \epsilon_{\mathcal{T}}] \leq \frac{\mathbb{E}[P_{e,\mathcal{T}}(\mathsf{C})] - \epsilon^*(\mathcal{C}_{\mathcal{T}})}{\epsilon_{\mathcal{T}} - \epsilon^*(\mathcal{C}_{\mathcal{T}})} \tag{255}$$
$$\leq \frac{\epsilon_{\mathcal{T}} - \epsilon_{\mathcal{T}}' + \Delta_{\mathcal{T}}}{\epsilon_{\mathcal{T}} - \epsilon_{\mathcal{T}}' + \delta_{\mathcal{T}}}, \tag{256}$$

We may choose $\delta_{\mathcal{T}} = O\left(\frac{1}{\sqrt{n}}\right)$ to ensure that the right-hand side of (256) is as small a constant as desired. Specifically, we choose constants $(\lambda_{\mathcal{T}})_{\mathcal{T} \in \mathcal{P}([K])}$ to satisfy

$$\sum_{\mathcal{T} \in \mathcal{P}([K])} \frac{1}{\lambda_{\mathcal{T}} + 1} < 1, \tag{257}$$
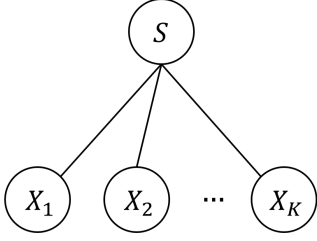
Fig. 9: A graphical model of a common distributed sensing scenario.

and put

$$\delta_{\mathcal{T}} = \lambda_{\mathcal{T}}(\epsilon_{\mathcal{T}} - \epsilon'_{\mathcal{T}} + \Delta_{\mathcal{T}}). \tag{258}$$

With (256) and (258), we bound the right-hand side of (242) as

$$\sum_{\mathcal{T} \in \mathcal{P}([K])} \mathbb{P}[P_{e,\mathcal{T}}(\mathsf{C}) > \epsilon_{\mathcal{T}}] \leq \sum_{\mathcal{T} \in \mathcal{P}([K])} \frac{1}{\lambda_{\mathcal{T}} + 1} < 1, \tag{259}$$

which implies the existence of a deterministic $\left(n, \overline{\mathbf{m}}_K, \mathbf{Q}_{[K]}, \overline{\boldsymbol{\epsilon}}_K\right)$ RASC with $m_{\mathcal{T}}$ in (248), $\overline{Q}(\hat{\mathcal{T}})$ in (245), and $\mathbf{R}_{\mathcal{T}}$ in (248). ∎

Remark 15. When parameters $\left(n, \mathbf{Q}_{[K]}, \overline{\boldsymbol{\epsilon}}_K\right)$ are fixed, increasing $\lambda_{\mathcal{T}}$ yields larger decoding blocklengths $m_{\mathcal{T}}$. Therefore, the choice of $(\lambda_{\mathcal{T}})_{\mathcal{T} \in \mathcal{P}([K])}$ to satisfy (257) controls the RASC performance trade-off across different active encoder sets. This trade-off affects the performance of the RASC in the fourth- or higher-order terms.

D. RASC for Permutation-Invariant Sources

A *permutation-invariant*[10] source is defined by the constraint

$$P_{\mathbf{X}_{[K]}}\left(\mathbf{x}_{[K]}\right) = P_{\mathbf{X}_{[K]}}\left(\mathbf{x}_{\pi([K])}\right) \tag{260}$$

for all permutations π on $[K]$ and all $\mathbf{x}_{[K]} \in \mathcal{X}_{[K]}$. For example, given any P_S and $P_{X|S}$, the marginal $P_{\mathbf{X}_{[K]}}$ of $P_{\mathbf{X}_{[K]}S} = (P_{X|S})^K P_S$ satisfies (260). Such "hidden variable" models have applications in statistics, science, and economics, where latent variables (e.g., the health of the world economy or the state of the atmosphere) influence observables (e.g., stock prices or climates). Figure 9 shows an example with K sensors reading measurements of a common hidden state S.

Permutation-invariant source models interest us both because of their wide applicability and because they present an opportunity for code simplification through *identical encoding*, where all encoders employ the same encoding map. For any permutation-invariant source, (215) and (260) imply that $\mathcal{X}_i = \mathcal{X}$ for all $i \in [K]$ and, for any $\mathcal{T} \in \mathcal{P}([K])$ with $|\mathcal{T}| = k$,

$$P_{\mathbf{X}_{\mathcal{T}}} = P_{\mathbf{X}_{[k]}}. \tag{261}$$

Thus, $P_{\mathbf{X}_{\mathcal{T}}}$ is permutation-invariant for every $\mathcal{T}$ and the joint source distribution depends on the number of active encoders but not their identities. Assuming that we further employ the same error probability ϵ_k for all $\mathcal{T} \in \mathcal{P}([K])$ with $|\mathcal{T}| = k$, we can fix a single decoding blocklength for each number $k \in [K]$

---

[10]Polyanskiy [17] introduces a similar notion of permutation invariance for multiple access channel coding in [17].

of active encoders and use identical encoders at all transmitters, allowing us to accommodate an arbitrarily large number of encoders without designing a unique encoder for each. A similar phenomenon arises for RA channel coding [18].

In analyzing RASC performance with identical encoders on a permutation-invariant source, we assume in addition to (221) and (222) that no two sources are identical, i.e.,

$$\mathbb{P}\left[\bigcup_{i,j \in [K], \, i \neq j} \{X_i = X_j\}\right] < 1. \tag{262}$$

This is important since using identical encoders on identical sources yields identical descriptions, in which case descriptions from multiple encoders are no better than descriptions from a single encoder. Under these assumptions, Theorem 24 continues to hold. In the analysis, we modify the decoder to output the most probable source vector $\mathbf{x}_{\mathcal{T}} \in \mathcal{X}_{\mathcal{T}}$ that contains no repeated symbols (see the proof of Theorem 26), treating the case where $\mathbf{X}_{\mathcal{T}}$ contains repeated symbols as an error. In the asymptotic analysis for stationary, memoryless sources, the probability of this error event is bounded by

$$\mathbb{P}\left[\bigcup_{i,j \in [K], \, i \neq j} \{\mathbf{X}_i^n = \mathbf{X}_j^n\}\right]$$
$$\leq \left(\mathbb{P}\left[\bigcup_{i,j \in [K], \, i \neq j} \{X_i = X_j\}\right]\right)^n, \tag{263}$$

which decays exponentially in n by (262). Therefore, under the assumption in (262), identical encoding does not incur a first-, second-, or third-order performance penalty.

VII. CONCLUDING REMARKS

This paper studies finite-blocklength lossless source coding in three scenarios.

We derive a new non-asymptotic achievability (RCU) bound (Theorem 4) and use it to show that for point-to-point coding on stationary, memoryless sources, random code design with maximum likelihood decoding achieves the same coding rate up to the third-order as the optimal code from [6]. The RCU bound generalizes to the MASC scenario (Theorem 18).

A new HT converse (Theorem 19) extends the channel coding meta-converse [5] to an MASC and suggests the possibility of using composite hypothesis testing to derive converses for other multi-terminal scenarios. Our analysis of composite hypothesis testing provides general tools (Lemmas 8, 9, and 10) for use in other related problems. Just as the meta-converse for channel coding recovers previously known converses, our HT converse recovers Han's MASC converse [9, Lemma. 7.2.2]. Just as the HT converse for lossy source coding [23, Th. 8] is equivalent to the LP-based converse for that setting (see [15, Cor. 3]), our MASC HT converse is equivalent to the MASC LP-based converse [15, Th. 12].

We give the first third-order characterization of the MASC rate region for stationary, memoryless sources, tightening prior second-order characterizations from [13] and [14] and replacing the $2^k - 1$ thresholds used there to decode for k users

by a maximum likelihood decoder that chooses the jointly most probable source realizations consistent with the received codewords. We show that for rate points converging to a non-corner point on the asymptotic sum-rate boundary, separate encoding does not compromise the performance in lossless data compression up to the third-order term. Numerical comparison of the new HT converse and the optimal performance of point-to-point source coding in Figure 4 allows one to bound from below the small gap between joint and separate encoding, which is not captured in the first three terms of the asymptotic expansion. For independent sources, there are no non-corner points, and MASC separate encoding incurs a positive penalty in the second-order term relative to joint encoding with a point-to-point code. When two sources have the same marginals, this penalty equals the penalty for using two independent blocklength-n codes rather than a single blocklength-$2n$ point-to-point code for encoding $2n$ samples.

Our proposed RASC works universally for all possible encoder activity patterns. The nested structure of the RASC demonstrates that there is no need for the encoders to know the set of active encoders *a priori*. The third-order-optimal MASC performance is achievable even when the only information the encoders receive is the acknowledgment that tells them when to stop transmitting (Theorem 24).

Our refinement of the traditional random coding argument (Lemma 25 and (242)) uses bounds on the minimal (converse) and expected (achievability) error probabilities for *each* possible active encoder set to show the existence of a single code that is good for *all* possible active encoder sets. This argument is likely to be useful for other information-theoretic problems.

APPENDIX A
PROOF OF THEOREM 3

Following [5, Eq. (68)], note that for $z > 0$ and $\gamma > 0$

$$\exp\left\{-\left|\log\frac{\gamma}{z}\right|_+\right\} = 1\{z > \gamma\} + \frac{z}{\gamma}1\{z \leq \gamma\}. \quad (A.1)$$

Let $z = \frac{1}{P_X(X)}$ and $\gamma = M$. Then taking the expectation of both sides of (A.1) with respect to P_X gives

$$\mathbb{E}\left[\exp\left\{-|\log M - \imath(X)|_+\right\}\right]$$
$$= \mathbb{P}[\imath(X) > \log M] + \frac{1}{M}\mathbb{U}[\imath(X) \leq \log M], \quad (A.2)$$

where $\mathbb{P}[\cdot]$ denotes a probability with respect to P_X and $\mathbb{U}[\cdot]$ denotes a mass with respect to the counting measure U_X on $\mathcal{X}$, which assigns unit weight to each $x \in \mathcal{X}$. In light of (A.2), we can prove (23) by demonstrating the existence of an (M, ϵ) code for which the right-hand side of (A.2) exceeds ϵ. We prove a slightly stronger result, showing that there exists an (M, ϵ) code with a threshold decoder such that

$$\epsilon \leq \mathbb{P}[\imath(X) > \log\gamma] + \frac{1}{M}\mathbb{U}[\imath(X) \leq \log\gamma] \quad (A.3)$$

for all $\gamma > 0$. Setting $\gamma = M$ in (A.3) yields the desired bound.

Fix $\gamma > 0$. For each $x \in \mathcal{X}$, randomly and independently draw each encoder output $\mathsf{F}(x)$ from the uniform distribution on $[M]$. Define the threshold decoder

$$g(c) = \begin{cases} x, & \text{if } \exists\,\text{unique } x \in \mathcal{X} \\ & \quad \text{s.t. } \mathsf{F}(x) = c,\ \imath(x) \leq \log\gamma \\ \text{error}, & \text{otherwise.} \end{cases} \quad (A.4)$$

We capture all errors using a union of error events

$$\mathcal{E}_1 \triangleq \{\imath(X) > \log\gamma\} \quad (A.5)$$
$$\mathcal{E}_2 \triangleq \{\exists\,\bar{x} \in \mathcal{X}\backslash\{X\} \text{ s.t. } \mathsf{F}(\bar{x}) = \mathsf{F}(X),\ \imath(\bar{x}) \leq \log\gamma\}. \quad (A.6)$$

By the random coding argument and the union bound, there exists an (M, ϵ) code such that

$$\epsilon \leq \mathbb{P}[\mathcal{E}_1 \cup \mathcal{E}_2] \leq \mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2]. \quad (A.7)$$

Here,

$$\mathbb{P}[\mathcal{E}_1] = \mathbb{P}[\imath(X) > \log\gamma] \quad (A.8)$$

$$\mathbb{P}[\mathcal{E}_2] = \mathbb{P}\left[\bigcup_{\bar{x}\in\mathcal{X}\backslash\{X\}}\{\mathsf{F}(\bar{x}) = \mathsf{F}(X),\ \imath(\bar{x}) \leq \log\gamma\}\right] \quad (A.9)$$

$$\leq \sum_{\bar{x}\in\mathcal{X}\backslash\{X\}}\mathbb{P}[\mathsf{F}(\bar{x}) = \mathsf{F}(X)]1\{\imath(\bar{x}) \leq \log\gamma\} \quad (A.10)$$

$$\leq \frac{1}{M}\sum_{\bar{x}\in\mathcal{X}}1\{\imath(\bar{x}) \leq \log\gamma\} \quad (A.11)$$

$$= \frac{1}{M}\mathbb{U}[\imath(X) \leq \log\gamma], \quad (A.12)$$

where (A.10) applies the union bound and (A.11) holds since the encoder outputs are i.i.d. and uniformly distributed. ∎

APPENDIX B
PROOF OF LEMMA 9

The proof extends the proof of [5, Eq. (102)] (e.g., [24]). We show that for any test $P_{Z|X}$ that decides between P vs. $\{Q_j\}_{j=1}^k$,

$$\mathbb{P}[Z = 1] - \sum_{j=1}^k \gamma_j\mathbb{Q}_j[Z = 1]$$

$$\leq \mathbb{P}\left[\bigcap_{j\in[k]}\left\{\frac{P(X)}{Q_j(X)} > \gamma_j\right\}\right], \quad (B.1)$$

where $\gamma_j \geq 0$, $j \in [k]$ are arbitrary constants. Then Lemma 9 follows immediately by definition of $\beta_\alpha(P, \{Q_j\}_{j=1}^k)$.

To prove (B.1), fix a $\gamma_j \geq 0$ for each $j \in [k]$. We then have

$$\mathbb{P}[Z = 1] - \sum_{j=1}^{k} \gamma_j \mathbb{Q}_j[Z = 1]$$

$$= \sum_{x \in \mathcal{X}} P_{Z|X}(1|x)\left(P(x) - \sum_{j=1}^{k} \gamma_j Q_j(x) \right) \tag{B.2}$$

$$\leq \sum_{x \in \mathcal{X}} P_{Z|X}(1|x)\left(P(x) - \sum_{j=1}^{k} \gamma_j Q_j(x) \right)$$
$$1\left\{ P(x) > \sum_{j=1}^{k} \{\gamma_j Q_j(x)\} \right\} \tag{B.3}$$

$$\leq \sum_{x \in \mathcal{X}} P_{Z|X}(1|x) P(x) 1\left\{ P(x) > \sum_{j=1}^{k} \gamma_j Q_j(x) \right\} \tag{B.4}$$

$$= \mathbb{P}\left[Z = 1, P(X) > \sum_{j=1}^{k} \gamma_j Q_j(X) \right] \tag{B.5}$$

$$\leq \mathbb{P}\left[P(X) > \sum_{j=1}^{k} \gamma_j Q_j(X) \right] \tag{B.6}$$

$$\leq \mathbb{P}\left[\bigcap_{j \in [k]} \{ P(X) > \gamma_j Q_j(X) \} \right], \tag{B.7}$$

where (B.4) follows from the non-negativity of probability and each γ_j. The proof is complete since (B.7) equals the right-hand-side of (B.1). ∎

APPENDIX C
PROOF OF LEMMA 10

For any test $P_{Z|X}$ deciding between P vs. $\{Q_j\}_{j=1}^{k}$, we show that

$$\mathbb{P}[Z = 1] - \sum_{j=1}^{k} \gamma_j \mathbb{Q}_j[Z = 1]$$

$$\leq 1 - \sum_{x \in \mathcal{X}} \min\left\{ P(x), \sum_{j=1}^{k} \gamma_j Q_j(x) \right\}, \tag{C.1}$$

where $\gamma_j \geq 0$, $j \in [k]$ are arbitrary constants. Fix a $\gamma_j \geq 0$ for each $j \in [k]$. For notational brevity, define sets

$$\mathcal{X}^{(<)} \triangleq \left\{ x \in \mathcal{X} : P(x) < \sum_{j=1}^{k} \gamma_j Q_j(x) \right\} \tag{C.2}$$

$$\mathcal{X}^{(=)} \triangleq \left\{ x \in \mathcal{X} : P(x) = \sum_{j=1}^{k} \gamma_j Q_j(x) \right\} \tag{C.3}$$

$$\mathcal{X}^{(>)} \triangleq \left\{ x \in \mathcal{X} : P(x) > \sum_{j=1}^{k} \gamma_j Q_j(x) \right\}. \tag{C.4}$$

For any test $P_{Z|X}$, we have

$$\mathbb{P}[Z = 1] - \sum_{j=1}^{k} \gamma_j \mathbb{Q}_j[Z = 1]$$
$$+ \sum_{x \in \mathcal{X}} \min\left\{ P(x), \sum_{j=1}^{k} \gamma_j Q_j(x) \right\}$$

$$= \sum_{x \in \mathcal{X}} \left(P_{Z|X}(1|x)\left(P(x) - \sum_{j=1}^{k} \gamma_j Q_j(x) \right) \right.$$
$$\left. + \min\left\{ P(x), \sum_{j=1}^{k} \gamma_j Q_j(x) \right\} \right) \tag{C.5}$$

$$= \sum_{x \in \mathcal{X}^{(<)}} \left(P_{Z|X}(1|x)\left(P(x) - \sum_{j=1}^{k} \gamma_j Q_j(x) \right) + P(x) \right)$$
$$+ \sum_{x \in \mathcal{X}^{(>)}} \left(P_{Z|X}(1|x)\left(P(x) - \sum_{j=1}^{k} \gamma_j Q_j(x) \right) \right.$$
$$\left. + \sum_{j=1}^{k} \gamma_j Q_j(x) \right) + \sum_{x \in \mathcal{X}^{(=)}} P(x) \tag{C.6}$$

$$\leq \sum_{x \in \mathcal{X}^{(<)}} P(x) + \sum_{x \in \mathcal{X}^{(>)}} P(x) + \sum_{x \in \mathcal{X}^{(=)}} P(x) \tag{C.7}$$

$$= 1. \tag{C.8}$$

The equality in (C.7) is achieved by test

$$P_{Z|X}(1|x) = \begin{cases} 1 & \text{for } x \in \mathcal{X}^{(>)} \\ 0 & \text{for } x \in \mathcal{X}^{(<)} \\ \lambda & \text{for } x \in \mathcal{X}^{(=)} \end{cases} \tag{C.9}$$

for any $\lambda \in [0,1]$. Rearranging (C.8) yields (C.1). Choosing the unique $\lambda \in [0,1]$ to satisfy $\mathbb{P}[Z = 1] = \alpha$, we obtain Lemma 10 by the definition of $\beta_\alpha\big(P, \{Q_j\}_{j=1}^{k}\big)$. ∎

APPENDIX D
PROOF OF LEMMA 12

Recall that T is composed of the r normalized eigenvectors corresponding to the non-zero eigenvalues of covariance matrix V and $\mathbf{U}_i = \mathsf{T}\mathbf{W}_i$, where $\mathbf{W}_i \in \mathbb{R}^r$ for $i = 1, \ldots, n$. Thus $\mathsf{V} = \mathsf{T}\mathsf{V}_r\mathsf{T}^T$, where $\mathsf{V}_r \triangleq \mathrm{Cov}[\mathbf{W}_1]$ is non-singular.

For each $\mathbf{z} \in \mathbb{R}^d$, define

$$\mathscr{A}_r(\mathbf{z}) \triangleq \{ \mathbf{x} \in \mathbb{R}^r : \mathsf{T}\mathbf{x} \leq \mathbf{z} \}, \tag{D.1}$$

which is a convex subset of $\mathbb{R}^r$. Let $\mathbf{Z}_r \sim \mathcal{N}(\mathbf{0}, \mathsf{V}_r) \in \mathbb{R}^r$. Applying [13, Cor. 8] to the i.i.d. random vectors $\mathbf{W}_1, \ldots, \mathbf{W}_n$, we obtain

$$\sup_{\mathbf{z} \in \mathbb{R}^d} \left| \mathbb{P}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{W}_i \in \mathscr{A}_r(\mathbf{z}) \right] - \mathbb{P}[\mathbf{Z}_r \in \mathscr{A}_r(\mathbf{z})] \right|$$
$$\leq \frac{400 r^{1/4} \beta_r}{\lambda_{\min}(\mathsf{V}_r)^{3/2} \sqrt{n}}, \tag{D.2}$$

which is equivalent to (81) by the definition of $\mathscr{A}_r(\mathbf{z})$. ∎

APPENDIX E
PROOF OF LEMMA 13

For simplicity, we assume that V is non-singular. When V is singular, a similar analysis can be applied with V replaced by V_r defined in Lemma 12.

Let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathsf{V})$ be a d-dimensional multivariate Gaussian with covariance matrix V. Recall from (77) that $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ is defined as

$$\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) \triangleq \{\mathbf{z} \in \mathbb{R}^d : \mathbb{P}[\mathbf{Z} \leq \mathbf{z}] \geq 1 - \epsilon\}. \quad \text{(E.1)}$$

By the definition of $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$ and the definition of $\Phi(\mathsf{V}; \mathbf{z})$ in (5), $\Phi(\mathsf{V}; \mathbf{z}) = 1 - \epsilon$ if and only if $\mathbf{z}$ lies on the boundary of $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$, and $\Phi(\mathsf{V}; \mathbf{z}) > 1 - \epsilon$ if and only if $\mathbf{z}$ lies in the interior of $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$.

Proof of Lemma 13. To prove (82), consider any $D_1 > 0$ and $\delta \geq 0$. Since $\Phi(\mathsf{V}; \mathbf{z})$ is continuously differentiable everywhere provided that V is non-singular, we can apply the multivariate Taylor's theorem to expand $\Phi(\mathsf{V}; \mathbf{z} + D_1 \delta \mathbf{1})$ as

$$\Phi(\mathsf{V}; \mathbf{z} + D_1 \delta \mathbf{1}) = \Phi(\mathsf{V}; \mathbf{z}) + D_1 \delta \sum_{i=1}^{d} \frac{\partial \Phi(\mathsf{V}; \mathbf{z})}{\partial z_i} + \xi(\mathbf{z}, D_1 \delta).$$
$$\text{(E.2)}$$

The second-order residual term $\xi(\mathbf{z}, D_1 \delta)$ can be bounded as

$$|\xi(\mathbf{z}, D_1 \delta)| \leq \frac{\xi_{\max}}{2} (d \cdot D_1 \delta)^2, \quad \text{(E.3)}$$

where

$$\xi_{\max} \triangleq \max_{\delta' \in [0, D_1 \delta]} \left\| \nabla^2 \Phi(\mathsf{V}; \mathbf{z} + \delta' \mathbf{1}) \right\|_{\max} \quad \text{(E.4)}$$

and $\|\cdot\|_{\max}$ denotes the max norm of a matrix.

Denote

$$D' \triangleq \sum_{i=1}^{d} \frac{\partial \Phi(\mathsf{V}; \mathbf{z})}{\partial z_i}. \quad \text{(E.5)}$$

Since $\Phi(\mathsf{V}; \mathbf{z})$ is increasing in any coordinate of $\mathbf{z}$, $D' > 0$. Then, for any $\mathbf{z} \in \mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$, we have

$$\Phi(\mathsf{V}; \mathbf{z} + D_1 \delta \mathbf{1}) \geq \Phi(\mathsf{V}; \mathbf{z}) + D' D_1 \delta - \frac{\xi_{\max}}{2}(d \cdot D_1 \delta)^2 \quad \text{(E.6)}$$

$$\geq 1 - \epsilon + \delta \left(D' D_1 - \frac{\xi_{\max}}{2} d^2 D_1^2 \delta \right). \quad \text{(E.7)}$$

We note that for any finite positive D_1, $\xi_{\max}$ approaches $\|\nabla^2 \Phi(\mathsf{V}; \mathbf{z})\|_{\max}$ as $\delta \to 0$. Thus, for any finite positive D_1 that satisfies $D' D_1 > 1$, there exists some $\delta_1 > 0$ such that for all $0 \leq \delta < \delta_1$,

$$D' D_1 - \frac{\xi_{\max}}{2} d^2 D_1^2 \delta \geq 1, \quad \text{(E.8)}$$

which yields

$$\Phi(\mathsf{V}; \mathbf{z} + D_1 \delta \mathbf{1}) \geq 1 - \epsilon + \delta. \quad \text{(E.9)}$$

By the definitions of $\Phi(\mathsf{V}; \mathbf{z})$ and $\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon)$, (E.9) implies

$$\mathbf{z} + D_1 \delta \mathbf{1} \in \mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon - \delta), \quad \text{(E.10)}$$

and consequently

$$\mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon) + D_1 \delta \mathbf{1} \subseteq \mathscr{Q}_{\mathrm{inv}}(\mathsf{V}, \epsilon - \delta), \quad \text{(E.11)}$$

which proves (82).

Eq. (83) can be proved in a similar way. ∎

APPENDIX F
EQUIVALENCE BETWEEN HT AND LP-BASED CONVERSES FOR THE MASC

In this appendix, we establish the equivalence between the HT converse and the LP-based converse by showing that the bounds in (108) and (138) are equivalent. According to [15, Eq. (31)], (108) is equivalent to the following converse

$$\epsilon \geq \sup_{\eta_1, \eta_2, \eta_3 \in \mathcal{Z}} \left\{ \sum_{\substack{x_1 \in \mathcal{X}_1 \\ x_2 \in \mathcal{X}_2}} \min \left\{ P_{X_1 X_2}(x_1, x_2), \sum_{j=1}^{3} \eta_j(x_1, x_2) \right\} \right.$$
$$- M_1 \sum_{x_2 \in \mathcal{X}_2} \max_{\hat{x}_1 \in \mathcal{X}_1} \min \{ P_{X_1 X_2}(\hat{x}_1, x_2), \eta_1(\hat{x}_1, x_2) \}$$
$$- M_2 \sum_{x_1 \in \mathcal{X}_1} \max_{\hat{x}_2 \in \mathcal{X}_2} \min \{ P_{X_1 X_2}(x_1, \hat{x}_2), \eta_2(x_1, \hat{x}_2) \}$$
$$\left. - M_1 M_2 \max_{\substack{\hat{x}_1 \in \mathcal{X}_1 \\ \hat{x}_2 \in \mathcal{X}_2}} \min \{ P_{X_1 X_2}(\hat{x}_1, \hat{x}_2), \eta_3(\hat{x}_1, \hat{x}_2) \} \right\}, \quad \text{(F.1)}$$

where the supremum is over

$$\mathcal{Z} \triangleq \{ z : \mathcal{X}_1 \times \mathcal{X}_2 \to [0, \infty) \}. \quad \text{(F.2)}$$

Therefore, we show that (F.1) is equivalent to (138).

We first demonstrate that (F.1) implies (138). Set

$$\eta_i = \gamma_i Q_{X_1 X_2}^{(i)}$$

for any σ-finite $Q_{X_1 X_2}^{(i)}$ and $\gamma_i \geq 0$, $i \in [3]$. Since

$$\min \{ P_{X_1 X_2}(\hat{x}_1, x_2), \eta_1(\hat{x}_1, x_2) \} \leq \eta_1(\hat{x}_1, x_2)$$
$$\min \{ P_{X_1 X_2}(x_1, \hat{x}_2), \eta_2(x_1, \hat{x}_2) \} \leq \eta_2(x_1, \hat{x}_2)$$
$$\min \{ P_{X_1 X_2}(\hat{x}_1, \hat{x}_2), \eta_3(\hat{x}_1, \hat{x}_2) \} \leq \eta_3(\hat{x}_1, \hat{x}_2)$$

in (F.1), we obtain (138).

To prove the other direction, we substitute $z_1 = \gamma_1 Q_{X_1 X_2}^{(1)}$, $z_2 = \gamma_2 Q_{X_1 X_2}^{(2)}$, and $z_3 = \gamma_3 Q_{X_1 X_2}^{(3)}$ in the right-hand side of (138) to obtain

$$\epsilon \geq \sup_{z_1, z_2, z_3 \in \mathcal{Z}} \left\{ \sum_{\substack{x_1 \in \mathcal{X}_1 \\ x_2 \in \mathcal{X}_2}} \min \left\{ P_{X_1 X_2}(x_1, x_2), \sum_{j=1}^{3} z_j(x_1, x_2) \right\} \right.$$
$$- M_1 \sum_{x_2 \in \mathcal{X}_2} \max_{\hat{x}_1 \in \mathcal{X}_1} z_1(\hat{x}_1, x_2) - M_2 \sum_{x_1 \in \mathcal{X}_1} \max_{\hat{x}_2 \in \mathcal{X}_2} z_2(x_1, \hat{x}_2)$$
$$\left. - M_1 M_2 \max_{\substack{\hat{x}_1 \in \mathcal{X}_1 \\ \hat{x}_2 \in \mathcal{X}_2}} z_3(\hat{x}_1, \hat{x}_2) \right\}, \quad \text{(F.3)}$$

Take a supremum of the right-hand side of (F.3) over $\eta_1, \eta_2, \eta_3 \in \mathcal{Z}$. Since (F.3) does not contain η_1, η_2, η_3, this does not change anything. Now, weaken (i.e., lower-bound) the inner supremum over $z_1, z_2, z_3 \in \mathcal{Z}$ by setting

$$z_j(x_1, x_2) = \min \{ P_{X_1 X_2}(x_1, x_2), \eta_j(x_1, x_2) \}. \quad \text{(F.4)}$$

Observing that

$$\min\left\{P_{X_1X_2}(x_1,x_2), \sum_{j=1}^{3}\min\{P_{X_1X_2}(x_1,x_2), \eta_j(x_1,x_2)\}\right\}$$

$$= \min\left\{P_{X_1X_2}(x_1,x_2), \sum_{j=1}^{3}\eta_j(x_1,x_2)\right\}, \qquad \text{(F.5)}$$

we see that the result of our weakening is exactly the right-hand side of (F.1), as desired. ∎

APPENDIX G
MASCs FOR SOURCES WITH LESS REDUNDANCY

Applying Lemma 7 to get the asymptotic achievability result in Theorem 20 requires that all $V(X_1, X_2)$, $V(X_1|X_2)$, and $V(X_2|X_1)$ are strictly positive (as an implication of assumption (139)). Thus, the analysis in Section V-D breaks down when any of these varentropies is equal to zero. (We refer to such a source as being *less redundant*.) In this appendix, we analyze the performance of the MASC for less redundant sources. Specifically, we consider a pair of stationary, memoryless sources and analyze the following three cases:

1) all *three* varentropies are equal to zero;
2) exactly *two* of the varentropies are equal to zero;
3) exactly *one* of the varentropies is equal to zero.

We continue to assume that the joint distribution $P_{X_1X_2}$ satisfies (140) and (141). For those cases in which $V(X_2|X_1) > 0$, we continue to assume $\mathbb{E}[V_c(X_2|X_1)] > 0$. Likewise, if $V(X_1|X_2) > 0$, we continue to assume $\mathbb{E}[V_c(X_1|X_2)] > 0$.

In point-to-point almost-lossless source coding, the optimal code for a non-redundant source is easy to find (see Remark 2). When the encoders are required to operate independently in a MASC, we know no easy way to find the optimal codes in general. In Section A below, we give characterizations of the (n, ϵ)-rate region in the three general cases listed above using the techniques developed in Section V-D. Then, in Section B, we restrict attention to the case where $P_{X_1X_2}(x_1, x_2) > 0$ for every $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$; under this condition, the optimal codes can be found and analyzed directly.

A. General Characterizations of the (n, ϵ)-Rate Region

We first list our results in the three general cases below.

Case 1): Suppose that $V(X_1|X_2) = 0$, $V(X_2|X_1) = 0$, and $V(X_1, X_2) = 0$. For any δ_1, δ_2, $\delta_{12} > 0$, let

$$\hat{\mathscr{R}}_{\text{in}}^{(1)}(n, \delta_1, \delta_2, \delta_{12}) \triangleq \left\{(R_1, R_2) \in \mathbb{R}^2 : \right.$$

$$R_1 \geq H(X_1|X_2) + \frac{1}{n}\log\frac{1}{\delta_1}$$

$$R_2 \geq H(X_2|X_1) + \frac{1}{n}\log\frac{1}{\delta_2}$$

$$\left. R_1 + R_2 \geq H(X_1, X_2) + \frac{1}{n}\log\frac{1}{\delta_{12}}\right\}. \qquad \text{(G.1)}$$

Define

$$\mathscr{R}_{\text{in}}^{(1)}(n, \epsilon) \triangleq \bigcup_{\substack{\delta_1, \delta_2, \delta_{12} > 0 \\ \delta_1 + \delta_2 + \delta_{12} = \epsilon}} \hat{\mathscr{R}}_{\text{in}}^{(1)}(n, \delta_1, \delta_2, \delta_{12}) \qquad \text{(G.2)}$$

$$\mathscr{R}_{\text{out}}^{(1)}(n, \epsilon) \triangleq \left\{(R_1, R_2) \in \mathbb{R}^2 : \right.$$

$$R_1 \geq H(X_1|X_2) - \frac{1}{n}\log\frac{1}{1-\epsilon}$$

$$R_2 \geq H(X_2|X_1) - \frac{1}{n}\log\frac{1}{1-\epsilon}$$

$$\left. R_1 + R_2 \geq H(X_1, X_2) - \frac{1}{n}\log\frac{1}{1-\epsilon}\right\}. \qquad \text{(G.3)}$$

Theorem 27. *When $V(X_1|X_2) = 0$, $V(X_2|X_1) = 0$, and $V(X_1, X_2) = 0$, the (n, ϵ)-rate region $\mathscr{R}^*(n, \epsilon)$ satisfies*

$$\mathscr{R}_{\text{in}}^{(1)}(n, \epsilon) \subseteq \mathscr{R}^*(n, \epsilon) \subseteq \mathscr{R}_{\text{out}}^{(1)}(n, \epsilon). \qquad \text{(G.4)}$$

As in the point-to-point scenario, there are no second-order dispersion terms or $-\frac{\log n}{2n}$ third-order terms in the characterization of $\mathscr{R}^*(n, \epsilon)$ in this case. For any n and ϵ, the achievable region $\mathscr{R}_{\text{in}}^{(1)}(n, \epsilon)$ has a curved boundary due to the trade-off in the $O(\frac{1}{n})$ fourth-order terms, while the converse region $\mathscr{R}_{\text{out}}^{(1)}(n, \epsilon)$ has three linear boundaries.

Case 2): There are three possible cases where exactly two of the three varentropies are equal to zero. Here, we suppose that $V(X_1|X_2) > 0$ while $V(X_2|X_1) = V(X_1, X_2) = 0$. The other two cases can be analyzed in the same way. Let B_1 denote the Berry-Esseen constant for the random variable $\imath(X_1|X_2)$, and let S_2, K_1, $\bar{K}_1$ be the finite positive constants defined in (181), (158), and (176), respectively. For any δ_1, δ_2, $\delta_{12} > 0$, let

$$\hat{\mathscr{R}}_{\text{in}}^{(2)}(n, \delta_1, \delta_2, \delta_{12}) \triangleq \left\{(R_1, R_2) \in \mathbb{R}^2 : \right.$$

$$R_1 \geq H(X_1|X_2)$$

$$+ \sqrt{\frac{V(X_1|X_2)}{n}}Q^{-1}\left(\delta_1 - \frac{B_1 + K_1}{\sqrt{n}} - \frac{S_2}{n}\right)$$

$$- \frac{\log n}{2n} + \frac{1}{n}\log\frac{\bar{K}_1}{1 - \delta_2 - \delta_{12}}$$

$$R_2 \geq H(X_2|X_1) + \frac{1}{n}\log\frac{1}{\delta_2}$$

$$\left. R_1 + R_2 \geq H(X_1, X_2) + \frac{1}{n}\log\frac{1}{\delta_{12}}\right\}. \qquad \text{(G.5)}$$

Define

$$\mathscr{R}_{\text{in}}^{(2)}(n, \epsilon) \triangleq \bigcup_{\substack{\delta_1, \delta_2, \delta_{12} > 0 \\ \delta_1 + \delta_2 + \delta_{12} = \epsilon}} \hat{\mathscr{R}}_{\text{in}}^{(2)}(n, \delta_1, \delta_2, \delta_{12}) \qquad \text{(G.6)}$$

$$\mathscr{R}_{\text{out}}^{(2)}(n, \epsilon) \triangleq \left\{(R_1, R_2) \in \mathbb{R}^2 : \right.$$

$$R_1 \geq H(X_1|X_2) + \sqrt{\frac{V(X_1|X_2)}{n}}Q^{-1}\left(\epsilon + \frac{B_1 + 1}{\sqrt{n}}\right)$$

$$- \frac{\log n}{2n}$$

$$R_2 \geq H(X_2|X_1) - \frac{1}{n}\log\frac{1}{1-\epsilon}$$

$$\left. R_1 + R_2 \geq H(X_1, X_2) - \frac{1}{n}\log\frac{1}{1-\epsilon}\right\}. \qquad \text{(G.7)}$$

Theorem 28. *When $V(X_1|X_2) > 0$, $V(X_2|X_1) = 0$, and $V(X_1, X_2) = 0$, the (n, ϵ)-rate region $\mathscr{R}^*(n, \epsilon)$ satisfies*

$$\mathscr{R}^{(2)}_{\text{in}}(n, \epsilon) \subseteq \mathscr{R}^*(n, \epsilon) \subseteq \mathscr{R}^{(2)}_{\text{out}}(n, \epsilon). \qquad \text{(G.8)}$$

The achievable region $\mathscr{R}^{(2)}_{\text{in}}(n, \epsilon)$ has a curved boundary due to the trade-off in δ_1, δ_2, and δ_{12}. If we let

$$\delta_1 = \epsilon - \frac{2}{\sqrt{n}}, \ \delta_2 = \frac{1}{\sqrt{n}}, \ \delta_{12} = \frac{1}{\sqrt{n}}, \qquad \text{(G.9)}$$

then it is apparent that the dispersion corresponding to R_1 is $V(X_1|X_2)$ with a $-\frac{\log n}{2n}$ third-order term, while the dispersions of R_2 and $R_1 + R_2$ are zero.

Case 3): Similar to Case 2), there are three possible cases where exactly one of the three varentropies is equal to zero. Here, we consider the case where $V(X_1|X_2) = 0$ while $V(X_2|X_1) > 0$ and $V(X_1, X_2) > 0$. Let S_1, K_2, $\bar{K}_2$, and K_{12} be the finite positive constants defined in (180), (159), (177), and (160), respectively, and let B be the Bentkus constant (81) for the vector (I_2, I_{12}). For any $\delta \in (0, \epsilon)$, let

$$\hat{\mathscr{R}}^{(3)}_{\text{in}}(n, \delta) \triangleq \left\{ (R_1, R_2) \in \mathbb{R}^2 : R_1 \geq H(X_1|X_2) + \frac{1}{n} \log \frac{1}{\delta} \right.$$

$$\begin{bmatrix} R_2 \\ R_1 + R_2 \end{bmatrix} \in \begin{bmatrix} H(X_2|X_1) \\ H(X_1, X_2) \end{bmatrix} + \frac{1}{\sqrt{n}} \mathcal{Q}_{\text{inv}} \left(\mathsf{V}_2, \epsilon - \delta - \frac{C_{\text{in}}}{\sqrt{n}} \right)$$

$$\left. - \frac{\log n}{2n} \mathbf{1} + \frac{1}{n} \log \frac{1}{1 - \delta} \mathbf{1} + \frac{1}{n} \begin{bmatrix} \log 2\bar{K}_2 \\ \log 2K_{12} \end{bmatrix} \right\}, \qquad \text{(G.10)}$$

where $C_{\text{in}} \triangleq K_2 + K_{12} + B + \frac{S_1}{\sqrt{n}}$, and V_2 is the covariance matrix of the random vector $(\imath(X_2|X_1), \imath(X_1, X_2))$. Define

$$\mathscr{R}^{(3)}_{\text{in}}(n, \epsilon) \triangleq \bigcup_{\delta \in (0, \epsilon)} \hat{\mathscr{R}}^{(3)}_{\text{in}}(n, \delta) \qquad \text{(G.11)}$$

$$\mathscr{R}^{(3)}_{\text{out}}(n, \epsilon) \triangleq \left\{ (R_1, R_2) \in \mathbb{R}^2 : \right.$$

$$R_1 \geq H(X_1|X_2) + \frac{1}{n} \log \frac{1}{1 - \epsilon}$$

$$\begin{bmatrix} R_2 \\ R_1 + R_2 \end{bmatrix} \in \begin{bmatrix} H(X_2|X_1) \\ H(X_1, X_2) \end{bmatrix} + \frac{1}{\sqrt{n}} \mathcal{Q}_{\text{inv}} \left(\mathsf{V}_2, \epsilon + \frac{B + 2}{\sqrt{n}} \right)$$

$$\left. - \frac{\log n}{2n} \mathbf{1} \right\}. \qquad \text{(G.12)}$$

Theorem 29. *When $V(X_1|X_2) = 0$, $V(X_2|X_1) > 0$, and $V(X_1, X_2) > 0$, the (n, ϵ)-rate region $\mathscr{R}^*(n, \epsilon)$ satisfies*

$$\mathscr{R}^{(3)}_{\text{in}}(n, \epsilon) \subseteq \mathscr{R}^*(n, \epsilon) \subseteq \mathscr{R}^{(3)}_{\text{out}}(n, \epsilon). \qquad \text{(G.13)}$$

For any n and ϵ, the achievable region $\mathscr{R}^{(3)}_{\text{in}}(n, \epsilon)$ has a curved boundary that is characterized by the trade-off between a separate bound on R_1 and a region in $\mathbb{R}^2$ that bounds $(R_2, R_1 + R_2)$ jointly. The converse region $\mathscr{R}^{(3)}_{\text{out}}(n, \epsilon)$ is the intersection of a region with a linear boundary that bounds R_1 only and a region with a curved boundary that bounds $(R_2, R_1 + R_2)$ jointly. If we let

$$\delta = \frac{1}{\sqrt{n}}, \qquad \text{(G.14)}$$

then it is apparent that the dispersion corresponding to R_2 and $R_1 + R_2$ is given by V_2 with a $-\frac{\log n}{2n}$ third-order term, while the dispersion of R_1 is zero.

A less redundant stationary, memoryless source has some useful properties. When $V(X_1, X_2) = 0$,

$$P_{X_1^n X_2^n}(x_1^n, x_2^n) \in \{0, \exp(-nH(X_1, X_2))\}, \qquad \text{(G.15)}$$

for every $(x_1^n, x_2^n) \in \mathcal{X}_1^n \times \mathcal{X}_2^n$; in other words, (X_1, X_2) is uniformly distributed over its support in $\mathcal{X}_1 \times \mathcal{X}_2$. When $V(X_1|X_2) = 0$,

$$P_{X_1^n|X_2^n}(x_1^n|x_2^n) = \begin{cases} \exp(-nH(X_1|X_2)), \\ \qquad \text{if } P_{X_1^n X_2^n}(x_1^n, x_2^n) > 0 \\ 0, \quad \text{otherwise}; \end{cases}$$
$$\qquad \text{(G.16)}$$

in other words, X_1 is uniformly distributed over its conditional support for each $x_2 \in \mathcal{X}_2$. When $V(X_2|X_1) = 0$, a result analogous to (G.16) holds. These properties do not reduce the difficulty of characterizing the optimal MASCs in general. As a result, we continue to employ the random coding techniques from Section V-D in our analysis here. For the achievability argument, we invoke the MASC RCU bound (Theorem 18); for the converse, we appeal to a modified version of [9, Lemma 7.2.2], as stated below.

Lemma 30 (Modified [9, Lemma 7.2.2]). *Any $(n, \exp(nR_1), \exp(nR_2), \epsilon')$ MASC satisfies*

$$\epsilon' \geq \mathbb{P} \left[\left\{ \frac{1}{n} I_1 \geq R_1 + \gamma_1 \right\} \cup \left\{ \frac{1}{n} I_2 \geq R_2 + \gamma_2 \right\} \cup \qquad \text{(G.17)} \right.$$

$$\left. \left\{ \frac{1}{n} I_{12} \geq R_1 + R_2 + \gamma_{12} \right\} \right]$$

$$- \min \left\{ \mathbb{P} \left[\frac{1}{n} I_1 \geq R_1 + \gamma_1 \right], \exp(-n\gamma_1) \right\}$$

$$- \min \left\{ \mathbb{P} \left[\frac{1}{n} I_2 \geq R_2 + \gamma_2 \right], \exp(-n\gamma_2) \right\}$$

$$- \min \left\{ \mathbb{P} \left[\frac{1}{n} I_{12} \geq R_1 + R_2 + \gamma_{12} \right], \exp(-n\gamma_{12}) \right\},$$

for any γ_1, γ_2, $\gamma_{12} > 0$, where I_1, I_2 and I_{12} are defined in (146)–(148).

We next prove Theorems 27, 28, and 29.

Proof of Theorem 27. Achievability: We employ the RCU bound in (155). To evaluate the terms in (155), note that the uniformity over the distribution's support that results from $V(X_1, X_2) = V(X_1|X_2) = V(X_2|X_1) = 0$ implies that for any (x_1^n, x_2^n) such that $P_{X_1^n X_2^n}(x_1^n, x_2^n) > 0$,

$$A_1 M_1 = \exp(nH(X_1|X_2)) \quad \text{a.s.} \qquad \text{(G.18)}$$

Similar equalities hold for A_2 and A_{12}, and for any $(R_1, R_2) \in \mathscr{R}^{(1)}_{\text{in}}(n, \epsilon)$, (155) gives

$$\epsilon' \leq \delta_1 + \delta_2 + \delta_{12} = \epsilon, \qquad \text{(G.19)}$$

implying that such a rate pair (R_1, R_2) is achievable. Therefore, the (n, ϵ)-rate region satisfies

$$\mathscr{R}^*(n, \epsilon) \supseteq \mathscr{R}^{(1)}_{\text{in}}(n, \epsilon). \qquad \text{(G.20)}$$

Converse: Consider any (R_1, R_2) with $R_1 < H(X_1|X_2) - \frac{1}{n}\log\frac{1}{1-\epsilon}$. Since the bound in (G.17) holds for any γ_1, γ_2, $\gamma_{12} > 0$, we take

$$\gamma_1 = H(X_1|X_2) - R_1 > \frac{1}{n}\log\frac{1}{1-\epsilon}, \qquad \text{(G.21)}$$

which, under the given uniformity, implies

$$\mathbb{P}\left[\frac{1}{n}I_1 \geq R_1 + \gamma_1\right] = 1. \qquad \text{(G.22)}$$

We take γ_2 and γ_{12} sufficiently large so that

$$R_2 + \gamma_2 > H(X_2|X_1) \qquad \text{(G.23)}$$
$$R_1 + R_2 + \gamma_{12} > H(X_1, X_2) \qquad \text{(G.24)}$$

and hence

$$\mathbb{P}\left[\frac{1}{n}I_2 \geq R_2 + \gamma_2\right] = \mathbb{P}\left[\frac{1}{n}I_{12} \geq R_1 + R_2 + \gamma_{12}\right] = 0. \qquad \text{(G.25)}$$

Under these conditions, (G.17) gives

$$\epsilon' \geq 1 - \exp(-n\gamma_1) > 1 - (1-\epsilon) = \epsilon. \qquad \text{(G.26)}$$

Therefore, any achievable rate pair (R_1, R_2) must satisfy

$$R_1 \geq H(X_1|X_2) - \frac{1}{n}\log\frac{1}{1-\epsilon}. \qquad \text{(G.27)}$$

The same analysis applies to R_2 and $R_1 + R_2$. We then conclude that any achievable rate pair (R_1, R_2) must satisfy $(R_1, R_2) \in \mathscr{R}_{\text{out}}^{(1)}(n, \epsilon)$. Thus,

$$\mathscr{R}^*(n, \epsilon) \subseteq \mathscr{R}_{\text{out}}^{(1)}(n, \epsilon). \qquad \text{(G.28)}$$

∎

Proof of Theorem 28. Achievability: Take $(R_1, R_2) \in \mathscr{R}_{\text{in}}^{(2)}(n, \epsilon)$ (G.6) satisfying the inequalities in (G.5) with some δ_1, δ_2, $\delta_{12} > 0$ such that $\delta_1 + \delta_2 + \delta_{12} = \epsilon$. We again employ the RCU bound from (155). Since $\mathbb{E}[V_c(X_1|X_2)] > 0$, we use (174) to bound A_1. The terms A_2 and A_{12} are constants (cf. (G.18)). With these observations, we weaken (156) as

$$\epsilon' \leq \mathbb{E}[A_1 \mathbb{1}\{A_1 \leq 1\}] + A_2 + A_{12} + \mathbb{P}\left[\bar{\bar{A}}_1 > 1 - A_2 - A_{12}\right]$$
$$+ \mathbb{P}[\mathcal{S}_2^c] \qquad \text{(G.29)}$$
$$\leq \mathbb{E}[A_1 \mathbb{1}\{A_1 \leq 1\}] + \delta_2 + \delta_{12} + \mathbb{P}\left[\bar{\bar{A}}_1 > 1 - \delta_2 - \delta_{12}\right]$$
$$+ \mathbb{P}[\mathcal{S}_2^c] \qquad \text{(G.30)}$$
$$\leq \delta_1 + \delta_2 + \delta_{12}, \qquad \text{(G.31)}$$

where (G.30) is by our choice of (R_1, R_2), and (G.31) applies (169), the Berry-Esseen inequality (Theorem 6), and (179) to bound the three probability terms. Therefore, (R_1, R_2) is achievable at blocklength n and error probability ϵ, implying

$$\mathscr{R}^*(n, \epsilon) \supseteq \mathscr{R}_{\text{in}}^{(2)}(n, \epsilon). \qquad \text{(G.32)}$$

Converse: We next apply Lemma 30 to derive a converse result. Recall that under our assumptions $V(X_2|X_1) = V(X_1, X_2) = 0$, $\imath(X_2|X_1) = H(X_2|X_1)$ and $\imath(X_1, X_2) = H(X_1, X_2)$ almost surely. Consider any (R_1, R_2) such that

$R_2 < H(X_2|X_1) - \frac{1}{n}\log\frac{1}{1-\epsilon}$. Since the bound in (G.17) holds for any γ_1, γ_2, $\gamma_{12} > 0$, we can take

$$\gamma_2 = H(X_2|X_1) - R_2 > \frac{1}{n}\log\frac{1}{1-\epsilon} \qquad \text{(G.33)}$$

so that

$$\mathbb{P}\left[\frac{1}{n}I_2 \geq R_2 + \gamma_2\right] = 0. \qquad \text{(G.34)}$$

By this choice of γ_2, $1 - \epsilon - \exp(-n\gamma_2) > 0$. Thus, we can take γ_1 and γ_{12} sufficiently large such that

$$\exp(-n\gamma_1) + \exp(-n\gamma_{12}) < 1 - \epsilon - \exp(-n\gamma_2). \quad \text{(G.35)}$$

By the above choices of γ_1, γ_2, and γ_{12}, (G.17) gives

$$\epsilon' \geq 1 - \exp(-n\gamma_1) - \exp(-n\gamma_2) - \exp(-n\gamma_{12}) > \epsilon. \qquad \text{(G.36)}$$

Therefore, any achievable rate pair (R_1, R_2) must satisfy

$$R_2 \geq H(X_2|X_1) - \frac{1}{n}\log\frac{1}{1-\epsilon}. \qquad \text{(G.37)}$$

The same analysis applies to $R_1 + R_2$, and we conclude that any achievable rate pair (R_1, R_2) must also satisfy

$$R_1 + R_2 \geq H(X_1, X_2) - \frac{1}{n}\log\frac{1}{1-\epsilon}. \qquad \text{(G.38)}$$

Given (G.37) and (G.38), we re-evaluate the bound in (G.17) by taking

$$\gamma_1 = \frac{\log n}{2n}, \quad \gamma_2 > \frac{1}{n}\log\frac{1}{1-\epsilon}, \quad \gamma_{12} > \frac{1}{n}\log\frac{1}{1-\epsilon}. \text{ (G.39)}$$

Under these conditions, we have

$$\mathbb{P}\left[\frac{1}{n}I_2 \geq R_2 + \gamma_2\right] = \mathbb{P}\left[\frac{1}{n}I_{12} \geq R_1 + R_2 + \gamma_{12}\right] = 0, \text{ (G.40)}$$

and the bound in (G.17) becomes

$$\epsilon' \geq \mathbb{P}\left[\frac{1}{n}I_1 \geq R_1 + \frac{\log n}{2n}\right] - \frac{1}{\sqrt{n}}. \qquad \text{(G.41)}$$

Then, by the Berry-Esseen inequality (Theorem 6), taking

$$R_1 = H(X_1|X_2) + \sqrt{\frac{V(X_2|X_1)}{n}}Q^{-1}\left(\epsilon + \frac{B_1 + 1}{\sqrt{n}}\right) - \frac{\log n}{2n} \qquad \text{(G.42)}$$

in (G.41) yields $\epsilon' \geq \epsilon$. Therefore, any achievable rate pair (R_1, R_2) must satisfy $(R_1, R_2) \in \mathscr{R}_{\text{out}}^{(2)}(n, \epsilon)$. Thus,

$$\mathscr{R}^*(n, \epsilon) \subseteq \mathscr{R}_{\text{out}}^{(2)}(n, \epsilon). \qquad \text{(G.43)}$$

∎

Proof of Theorem 29. Achievability: Take any $(R_1, R_2) \in \mathscr{R}_{\text{in}}^{(3)}(n, \epsilon)$ satisfying the inequalities in (G.10) with some $\delta \leq \epsilon$. We employ the RCU bound in (155). Since $\mathbb{E}[V_c(X_2|X_1)] > 0$ and $V(X_1, X_2) > 0$, we use (175) and (163) to bound A_2 and A_{12}, respectively; A_1 is the constant in (G.18). With these observations, we weaken (156) as

$$\epsilon' \leq A_1 + \mathbb{E}[A_2 \mathbb{1}\{A_2 \leq 1\}] + \mathbb{E}[A_{12} \mathbb{1}\{A_{12} \leq 1\}]$$
$$+ \mathbb{P}\left[2\bar{\bar{A}}_2 > 1 - A_1 \cup 2\bar{A}_{12} > 1 - A_1\right] + \mathbb{P}[\mathcal{S}_1^c] \quad \text{(G.44)}$$
$$\leq \delta + \mathbb{E}[A_2 \mathbb{1}\{A_2 \leq 1\}] + \mathbb{E}[A_{12} \mathbb{1}\{A_{12} \leq 1\}]$$
$$+ \mathbb{P}\left[2\bar{\bar{A}}_2 > 1 - \delta \cup 2A_{12} > 1 - \delta\right] + \mathbb{P}[\mathcal{S}_1^c] \qquad \text{(G.45)}$$
$$\leq \epsilon, \qquad \text{(G.46)}$$

where (G.45) is by our choice of (R_1, R_2), and (G.46) applies (170), (171), Lemma 12 (multidimensional Berry-Esseen Theorem), and (179) to bound the four probability terms. Therefore, (R_1, R_2) is achievable at blocklength n and error probability ϵ, implying that

$$\mathscr{R}^*(n, \epsilon) \supseteq \mathscr{R}_{\text{in}}^{(3)}(n, \epsilon). \tag{G.47}$$

Converse: We employ Lemma 30 to derive a converse. Recall that in this case, $\imath(X_1|X_2) = H(X_1|X_2)$ almost surely. Consider any (R_1, R_2) such that $R_1 < H(X_1|X_2) - \frac{1}{n} \log \frac{1}{1-\epsilon}$. Since the bound in (G.17) holds for any γ_1, γ_2, $\gamma_{12} > 0$, we can set

$$\gamma_1 = H(X_1|X_2) - R_1 > \frac{1}{n} \log \frac{1}{1-\epsilon} \tag{G.48}$$

so that

$$\mathbb{P}\left[\frac{1}{n} I_1 \geq R_1 + \gamma_1\right] = 1. \tag{G.49}$$

By this choice of γ_1, $1 - \epsilon - \exp(-n\gamma_1) > 0$. Thus, we can take γ_2 and γ_{12} sufficiently large such that

$$\exp(-n\gamma_2) + \exp(-n\gamma_{12}) < 1 - \epsilon - \exp(-n\gamma_1). \tag{G.50}$$

By the above choices of γ_1, γ_2, and γ_{12}, (G.17) gives

$$\epsilon' \geq 1 - \exp(-n\gamma_1) - \exp(-n\gamma_2) - \exp(-n\gamma_{12}) > \epsilon. \tag{G.51}$$

Therefore, any achievable rate pair (R_1, R_2) must satisfy

$$R_1 \geq H(X_1|X_2) - \frac{1}{n} \log \frac{1}{1-\epsilon}. \tag{G.52}$$

Given that (G.52) holds, we re-evaluate the bound in (G.17) by taking

$$\gamma_1 > \frac{1}{n} \log \frac{1}{1-\epsilon}, \quad \gamma_2 = \frac{\log n}{2n}, \quad \gamma_{12} = \frac{\log n}{2n}. \tag{G.53}$$

Under these conditions, the bound in (G.17) becomes

$$\epsilon' \geq \mathbb{P}\left[\left\{\frac{1}{n} I_2 \geq R_2 + \frac{\log n}{2n}\right\} \cup \left\{\frac{1}{n} I_{12} \geq R_1 + R_2 + \frac{\log n}{2n}\right\}\right] - \frac{2}{\sqrt{n}}. \tag{G.54}$$

Applying Lemmas 12 and 13 to (G.54), we conclude that any (R_1, R_2) in the (n, ϵ)-rate region must satisfy $(R_1, R_2) \in \mathscr{R}_{\text{out}}^{(3)}(n, \epsilon)$. Thus,

$$\mathscr{R}^*(n, \epsilon) \subseteq \mathscr{R}_{\text{out}}^{(3)}(n, \epsilon). \tag{G.55}$$

∎

B. Two Special Cases

The analysis in Section G-A above applies to any stationary, memoryless source with single-letter distribution $P_{X_1 X_2}$ that satisfies (140). In such a general setting, it is hard to find an optimal code. However, there are some special cases in which the optimal codes for a less redundant source can be characterized.

To enable the following analysis on these special cases, we assume that $P_{X_1 X_2}(x_1, x_2) > 0$ for every $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$. Under this assumption, $V(X_1, X_2) = 0$ if and only if

$V(X_1|X_2) = V(X_2|X_1) = 0$. As a result, the three cases discussed in Section A reduce to only *two* possible scenarios:

1) $V(X_1, X_2) = V(X_1|X_2) = V(X_2|X_1) = 0$;
2) $V(X_1, X_2) > 0$, and either $V(X_1|X_2) = 0$ or $V(X_2|X_1) = 0$.

Note that X_1 and X_2 are independent in both of these scenarios.

We first summarize the results below.

Special Case 1):

Theorem 31. *Suppose that $V(X_1|X_2) = 0$, $V(X_2|X_1) = 0$, and $V(X_1, X_2) = 0$. If $P_{X_1 X_2}$ satisfies $P_{X_1 X_2}(x_1, x_2) > 0$ for every $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, then*

$$\mathscr{R}^*(n, \epsilon) = \mathscr{R}_{\text{out}}^{(1)}(n, \epsilon), \tag{G.56}$$

where $\mathscr{R}_{\text{out}}^{(1)}(n, \epsilon)$ is defined in (G.3).

This scenario is a special example of Case 1) discussed in Section A above. The (n, ϵ)-rate region here coincides with the converse region $\mathscr{R}_{\text{out}}^{(1)}(n, \epsilon)$ presented in (G.3) for general source distributions. See Figure 10(a) for a comparison among $\mathscr{R}_{\text{in}}^{(1)}(n, \epsilon)$, $\mathscr{R}_{\text{out}}^{(1)}(n, \epsilon)$, and $\mathscr{R}^*(n, \epsilon)$ in this special case.

Special Case 2): With $V(X_1, X_2) > 0$, we here assume that $V(X_1|X_2) = 0$ and $V(X_2|X_1) > 0$. The other case can be analyzed similarly. For any $\delta \in [0, \epsilon)$, we define

$$\hat{\mathscr{R}}_{\text{in}}^s(n, \delta) \triangleq \left\{(R_1, R_2) \in \mathbb{R}^2 : \tag{G.57}\right.$$

$$R_1 \geq H(X_1) - \frac{1}{n} \log \frac{1}{1-\delta}$$

$$\left. R_2 \geq H(X_2) + \sqrt{\frac{V(X_2)}{n}} Q^{-1}\left(\frac{\epsilon - \delta}{1 - \delta}\right) - \frac{\log n}{2n} + \xi_{\text{in}}(\epsilon, \delta, n)\right\}$$

$$\hat{\mathscr{R}}_{\text{out}}^s(n, \delta) \triangleq \left\{(R_1, R_2) \in \mathbb{R}^2 : \tag{G.58}\right.$$

$$R_1 \geq H(X_1) - \frac{1}{n} \log \frac{1}{1-\delta}$$

$$R_2 \geq H(X_2) + \sqrt{\frac{V(X_2)}{n}} Q^{-1}\left(\frac{\epsilon - \delta}{1 - \delta}\right) - \frac{\log n}{2n}$$

$$\left. - \xi_{\text{out}}(\epsilon, \delta, n)\right\},$$

where the functions $\xi_{\text{in}}(\epsilon, \delta, n)$ and $\xi_{\text{out}}(\epsilon, \delta, n)$ are characterized as follows. For any fixed δ, $\xi_{\text{out}}(\epsilon, \delta, n) = O(\frac{1}{n})$ and $\xi_{\text{in}}(\epsilon, \delta, n) = O(\frac{1}{n})$. For any fixed n, both $\xi_{\text{out}}(\epsilon, \delta, n)$ and $\xi_{\text{in}}(\epsilon, \delta, n)$ blow up as δ approaches ϵ. (These bounds are applications of the point-to-point results in Theorem 1.) Also define

$$\mathscr{R}_{\text{in}}^s(n, \epsilon) \triangleq \bigcup_{\delta \in [0, \epsilon)} \hat{\mathscr{R}}_{\text{in}}^s(n, \delta) \tag{G.59}$$

$$\mathscr{R}_{\text{out}}^s(n, \epsilon) \triangleq \bigcup_{\delta \in [0, \epsilon)} \hat{\mathscr{R}}_{\text{out}}^s(n, \delta). \tag{G.60}$$

Theorem 32. *Suppose that $V(X_1|X_2) = 0$, $V(X_2|X_1) > 0$, and $V(X_1, X_2) > 0$. If $P_{X_1 X_2}$ satisfies $P_{X_1 X_2}(x_1, x_2) > 0$ for every $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, then*

$$\mathscr{R}_{\text{in}}^s(n, \epsilon) \subseteq \mathscr{R}^*(n, \epsilon) \subseteq \mathscr{R}_{\text{out}}^s(n, \epsilon). \tag{G.61}$$

This scenario is a special example of Case 3) discussed in Section A of this appendix. The (n, ϵ)-rate region characterized in (G.61) is sandwiched between the achievable region presented in (G.11) and the converse region presented in (G.12). To compare these regions, we note that the bounds on $R_1 + R_2$ in (G.11) and (G.12) become inactive in this special scenario where X_1 and X_2 are independent. As a result, the achievable region in (G.11) becomes

$$\mathscr{R}_{\text{in}}^{(3)}(n, \epsilon) = \bigcup_{\delta \in (0, \epsilon)} \left\{ (R_1, R_2) \in \mathbb{R}^2 : \right. \tag{G.62}$$

$$R_1 \geq H(X_1) + \frac{1}{n} \log \frac{1}{\delta}$$

$$R_2 \geq H(X_2) + \sqrt{\frac{V(X_2)}{n}} Q^{-1} \left(\epsilon - \delta - \frac{C_{\text{in}}}{\sqrt{n}} \right)$$

$$\left. - \frac{\log n}{2n} + \frac{1}{n} \log \frac{1}{1 - \delta} \right\},$$

and the converse region in (G.12) becomes

$$\mathscr{R}_{\text{out}}^{(3)}(n, \epsilon) = \left\{ (R_1, R_2) \in \mathbb{R}^2 : \right. \tag{G.63}$$

$$R_1 \geq H(X_1) - \frac{1}{n} \log \frac{1}{1 - \epsilon}$$

$$\left. R_2 \geq H(X_2) + \sqrt{\frac{V(X_2)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} - O\left(\frac{1}{n}\right) \right\}.$$

As δ approaches ϵ, the boundary of the (n, ϵ)-rate region given in (G.59) approaches the line $R_1 = H(X_1) - \frac{1}{n} \log \frac{1}{1-\epsilon}$, which matches the vertical segment of the boundary of the converse region $\mathscr{R}_{\text{out}}^{(3)}(n, \epsilon)$. See Figure 10(b) for a comparison of $\mathscr{R}_{\text{in}}^{(3)}(n, \epsilon)$, $\mathscr{R}_{\text{out}}^{(3)}(n, \epsilon)$, and $\mathscr{R}^*(n, \epsilon)$ in this case.

We next give proofs for Theorems 31 and 32.

Proof of Theorem 31. When $V(X_1|X_2) = V(X_2|X_1) = V(X_1, X_2) = 0$, (X_1, X_2) is uniformly distributed over $\mathcal{X}_1 \times \mathcal{X}_2$, which restricts $\mathcal{X}_1$ and $\mathcal{X}_2$ to be finite and X_1 and X_2 to be independent. The MASC problem reduces to independent (point-to-point) almost-lossless source coding problems for the two sources with a single compound error probability. As a result, the optimal MASC with blocklength n and code sizes (M_1, M_2) has an error probability given by

$$1 - \min\left\{1, \frac{M_1}{|\mathcal{X}_1|^n}\right\} \cdot \min\left\{1, \frac{M_2}{|\mathcal{X}_2|^n}\right\}. \tag{G.64}$$

Therefore, for any $0 < \epsilon < 1$, there exists an (n, M_1, M_2, ϵ) MASC if and only if

$$\min\left\{1, \frac{M_1}{|\mathcal{X}_1|^n}\right\} \cdot \min\left\{1, \frac{M_2}{|\mathcal{X}_2|^n}\right\} \geq 1 - \epsilon. \tag{G.65}$$

In this case, $H(X_1) = \log |\mathcal{X}_1|$ and $H(X_2) = \log |\mathcal{X}_2|$.
- For $R_1 < H(X_1)$ and $R_2 < H(X_2)$, (G.65) becomes

$$M_1 M_2 \geq (1 - \epsilon) |\mathcal{X}_1|^n |\mathcal{X}_2|^n, \tag{G.66}$$

which is equivalent to

$$R_1 + R_2 \geq H(X_1) + H(X_2) - \frac{1}{n} \log \frac{1}{1 - \epsilon}. \tag{G.67}$$
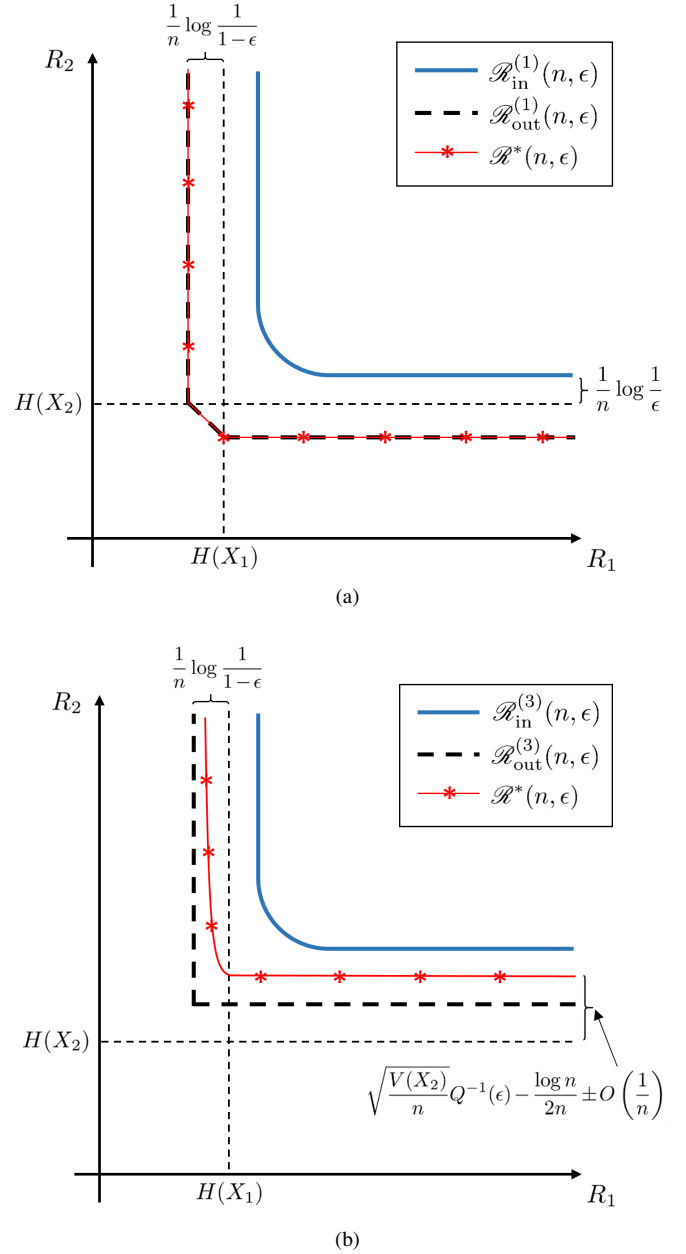

(a)

(b)

Fig. 10: Schematic illustrations of the MASC rate regions for a less redundant source. The drawing in (a) illustrates the achievable and converse regions in Case 1) $(V(X_1, X_2) = V(X_1|X_2) = V(X_2|X_1) = 0)$ and the (n, ϵ)-rate region $\mathscr{R}^*(n, \epsilon)$ when $P_{X_1 X_2}$ is assumed to have no zeros (Special Case 1)). The drawing in (b) illustrates the achievable and converse regions in Case 3) $(V(X_1|X_2) = 0, V(X_1, X_2), V(X_2|X_1) > 0)$ and the (n, ϵ)-rate region $\mathscr{R}^*(n, \epsilon)$ when $P_{X_1 X_2}$ is assumed to have no zeros (Special Case 2)).

- For $R_1 \geq H(X_1)$, (G.65) becomes

$$M_2 \geq (1 - \epsilon) |\mathcal{X}_2|^n, \tag{G.68}$$

which is equivalent to

$$R_2 \geq H(X_2) - \frac{1}{n} \log \frac{1}{1 - \epsilon}. \tag{G.69}$$

- For $R_2 \geq H(X_2)$, (G.65) gives

$$R_1 \geq H(X_1) - \frac{1}{n} \log \frac{1}{1-\epsilon}. \tag{G.70}$$

For all $0 < \epsilon < 1$ and $n \geq 1$,

$$\mathscr{R}^*(n,\epsilon) = \left\{ (R_1, R_2) \in \mathbb{R}^2 : \tag{G.71} \right.$$
$$R_1 \geq H(X_1) - \frac{1}{n} \log \frac{1}{1-\epsilon}$$
$$R_2 \geq H(X_2) - \frac{1}{n} \log \frac{1}{1-\epsilon}$$
$$\left. R_1 + R_2 \geq H(X_1) + H(X_2) - \frac{1}{n} \log \frac{1}{1-\epsilon} \right\} = \mathscr{R}_{\text{out}}^{(1)}(n,\epsilon).$$

∎

Proof of Theorem 32. When $V(X_1|X_2) = 0$ and $V(X_2|X_1), V(X_1, X_2) > 0$, X_1 is uniformly distributed over $\mathcal{X}_1$, which implies that $\mathcal{X}_1$ is finite and $H(X_1) = \log |\mathcal{X}_1|$. In contrast, X_2 is not uniform over $\mathcal{X}_2$. Moreover, X_1 and X_2 are independent. The MASC problem in this case can also be resolved via independent point-to-point source coding for each of the two sources. The optimal code with blocklength n and code sizes (M_1, M_2) encodes M_1 arbitrary symbols in $\mathcal{X}_1^n$ and a cardinality-M_2 subset of $\mathcal{X}_2^n$ that has the largest probability with respect to $P_{X_2^n}$. As a result, for any $0 < \epsilon < 1$, there exists an (M_1, M_2, ϵ) MASC if and only if

$$(1-\delta)\delta' \geq 1 - \epsilon, \tag{G.72}$$

where $\delta = 1 - \min\left\{1, \frac{M_1}{|\mathcal{X}_1|^n}\right\}$ is the total marginal probability of symbols that are not encoded in $\mathcal{X}_1^n$, and δ' is the total marginal probability (with respect to $P_{X_2^n}$) of the encoded symbols in $\mathcal{X}_2^n$. Eq. (G.72) implicitly requires $\delta \in [0, \epsilon]$ and $\delta' \in [1-\epsilon, 1]$.

- For $\delta = 0$, we have

$$R_1 \geq H(X_1). \tag{G.73}$$

In this case, (G.72) gives

$$1 - \delta' \leq \epsilon. \tag{G.74}$$

We can apply the point-to-point almost-lossless source coding results from Theorem 1 to obtain

$$H(X_2) + \sqrt{\frac{V(X_2)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} - O\left(\frac{1}{n}\right)$$
$$\leq R_2$$
$$\leq H(X_2) + \sqrt{\frac{V(X_2)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} + O\left(\frac{1}{n}\right). \tag{G.75}$$

- For $0 < \delta \leq \epsilon$, we have

$$R_1 = H(X_1) - \frac{1}{n} \log \frac{1}{\delta}. \tag{G.76}$$

In this case, (G.72) gives

$$1 - \delta' \leq \frac{\epsilon - \delta}{1 - \delta}. \tag{G.77}$$

We can also apply the point-to-point results to get

$$H(X_2) + \sqrt{\frac{V(X_2)}{n}} Q^{-1}\left(\frac{\epsilon - \delta}{1 - \delta}\right) - \frac{\log n}{2n} - \xi_{\text{out}}(\epsilon, \delta, n)$$
$$\leq R_2 \tag{G.78}$$
$$\leq H(X_2) + \sqrt{\frac{V(X_2)}{n}} Q^{-1}\left(\frac{\epsilon - \delta}{1 - \delta}\right) - \frac{\log n}{2n} + \xi_{\text{in}}(\epsilon, \delta, n),$$

where for any fixed δ, $\xi_{\text{out}}(\epsilon, \delta, n) = O\left(\frac{1}{n}\right)$ and $\xi_{\text{in}}(\epsilon, \delta, n) = O\left(\frac{1}{n}\right)$; for any fixed n, both $\xi_{\text{out}}(\epsilon, \delta, n)$ and $\xi_{\text{in}}(\epsilon, \delta, n)$ blow up as δ approaches ϵ (see Theorem 1 for the case where ϵ approaches 0). ∎

APPENDIX H
PROOF OF COROLLARY 21

1) When X_1 and X_2 are dependent, our choice of $\mathbf{R} = (R_1, R_2)$ in (197)–(199) implies that

$$R_1 \geq H(X_1|X_2) + \delta_2 - \frac{\log n}{2n} \tag{H.1}$$
$$R_2 \geq H(X_2|X_1) + \delta_1 - \frac{\log n}{2n}. \tag{H.2}$$

Define

$$\mathbf{a} \triangleq \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \triangleq \overline{\mathbf{R}} - \overline{\mathbf{H}} + \frac{\log n}{2n} \mathbf{1}. \tag{H.3}$$

We have

$$a_1 = R_1 - H(X_1|X_2) + \frac{\log n}{2n} \geq \delta_2 \tag{H.4}$$
$$a_2 = R_2 - H(X_2|X_1) + \frac{\log n}{2n} \geq \delta_1 \tag{H.5}$$
$$a_3 = \sqrt{\frac{V(X_1, X_2)}{n}} Q^{-1}\left(\epsilon - \frac{G}{\sqrt{n}}\right). \tag{H.6}$$

Let $\mathbf{Z} \triangleq (Z_1, Z_2, Z_3) \sim \mathcal{N}(\mathbf{0}, \mathsf{V})$ be a multivariate Gaussian in $\mathbb{R}^3$, where V is the entropy dispersion matrix (see Definition 8). Then

$$\mathbb{P}\left[\mathbf{Z} \leq \sqrt{n}\mathbf{a}\right] \tag{H.7}$$
$$\geq 1 - \left(\mathbb{P}\left[Z_1 > a_1\sqrt{n}\right] + \mathbb{P}\left[Z_2 > a_2\sqrt{n}\right] + \mathbb{P}\left[Z_3 > a_3\sqrt{n}\right]\right),$$

where (H.7) holds by the union bound. It follows that

$$\mathbb{P}\left[Z_1 > a_1\sqrt{n}\right] = \mathbb{P}\left[Z_1 \geq a_1\sqrt{n}\right] \tag{H.8}$$
$$= Q\left(\frac{a_1\sqrt{n}}{\sqrt{V(X_1|X_2)}}\right) \tag{H.9}$$
$$\leq e^{-na_1^2/(2V(X_1|X_2))} \tag{H.10}$$
$$\leq e^{-n\delta_2^2/(2V(X_1|X_2))}, \tag{H.11}$$

where (H.10) applies the Chernoff bound of the Q-function, and (H.11) holds since $a_1 \geq \delta_2 > 0$. Similarly,

$$\mathbb{P}\left[Z_2 > a_2\sqrt{n}\right] \leq e^{-n\delta_1^2/(2V(X_2|X_1))}. \tag{H.12}$$

In contrast,

$$\mathbb{P}\left[Z_3 > a_3\sqrt{n}\right] = \epsilon - \frac{G}{\sqrt{n}}. \tag{H.13}$$

Plugging (H.11)–(H.13) into (H.7), we conclude that for all n sufficiently large such that

$$e^{-n\delta_2^2/(2V(X_1|X_2))}+e^{-n\delta_1^2/(2V(X_2|X_1))} \leq \frac{G}{\sqrt{n}}, \quad \text{(H.14)}$$

the bound

$$\mathbb{P}\big[\mathbf{Z} \leq \sqrt{n}\mathbf{a}\big] \geq 1-\epsilon \quad \text{(H.15)}$$

holds. Therefore, $\sqrt{n}\mathbf{a} \in \mathscr{Q}_{\mathrm{inv}}(\mathsf{V},\epsilon)$, and hence $\overline{\mathbf{R}} \in \overline{\mathscr{R}}^*(n,\epsilon)$ (142).

2) Recall vector $\mathbf{a}$ defined in (H.3). With $R_1 = H(X_1)$,

$$a_1 = H(X_1)-H(X_1|X_2)+\frac{\log n}{2n} \quad \text{(H.16)}$$

$$a_2 = R_2-H(X_2|X_1)+\frac{\log n}{2n} \quad \text{(H.17)}$$

$$a_3 = R_2-H(X_2|X_1)+\frac{\log n}{2n}. \quad \text{(H.18)}$$

Note that

$$\mathbb{P}\big[\mathbf{Z} \leq \sqrt{n}\mathbf{a}\big] \quad \text{(H.19)}$$
$$\geq \mathbb{P}\big[Z_2 \leq a_2\sqrt{n}, Z_3 \leq a_3\sqrt{n}\big] -\mathbb{P}\big[Z_1 > a_1\sqrt{n}\big].$$

Since $H(X_1)-H(X_1|X_2) > 0$, $\mathbb{P}[Z_1 > a_1\sqrt{n}]$ decays exponentially in n. Therefore, by the definition of r^* in (202) and a first-order multivariate Taylor bound, $G > 0$ in (200) can be chosen so that the right side of (H.19) is equal to $1-\epsilon$, which implies that $\overline{\mathbf{R}} \in \overline{\mathscr{R}}^*(n,\epsilon)$ (142).

Conversely, for any R_2 such that $\overline{\mathbf{R}} \in \overline{\mathscr{R}}^*(n,\epsilon)$,

$$\mathbb{P}\big[\mathbf{Z} \leq \sqrt{n}\mathbf{a}\big] \geq 1-\epsilon, \quad \text{(H.20)}$$

which further implies

$$\mathbb{P}\big[Z_2 \leq a_2\sqrt{n}, Z_3 \leq a_3\sqrt{n}\big] \geq 1-\epsilon. \quad \text{(H.21)}$$

Thus, by the definition of r^*,

$$\sqrt{n}\left(R_2-H(X_2|X_1)+\frac{\log n}{2n}\right) \geq r^*, \quad \text{(H.22)}$$

which is equivalent to (201). ∎

APPENDIX I
PROOF OF COROLLARY 22

Fix any $\lambda \in [0,1]$. Define

$$\mathbf{a} \triangleq \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \triangleq \overline{\mathbf{R}}-\overline{\mathbf{H}}+\frac{\log n}{2n}\begin{bmatrix} \lambda \\ 1-\lambda \\ 1 \end{bmatrix}. \quad \text{(I.1)}$$

By the assumption that X_1 and X_2 are independent, we have

$$a_3 = a_1+a_2. \quad \text{(I.2)}$$

Denote

$$r_1 \triangleq \frac{a_1\sqrt{n}}{\sqrt{V(X_1)}}, \ r_2 \triangleq \frac{a_2\sqrt{n}}{\sqrt{V(X_2)}}. \quad \text{(I.3)}$$

Let $\mathbf{Z} \triangleq (Z_1, Z_2, Z_3) \sim \mathcal{N}(\mathbf{0}, \mathsf{V})$ be a multivariate Gaussian in $\mathbb{R}^3$, where V is the entropy dispersion matrix of the independent sources X_1 and X_2. It follows in this case that Z_1 and Z_2 are independent and $Z_3 = Z_1+Z_2$. We then have

$$\mathbb{P}\big[\mathbf{Z} \leq \sqrt{n}\mathbf{a}\big]$$
$$= \mathbb{P}\big[Z_1 \leq a_1\sqrt{n}\big]\mathbb{P}\big[Z_2 \leq a_2\sqrt{n}\big]$$
$$\quad \mathbb{P}\big[Z_3 \leq a_3\sqrt{n} \,\big|\, Z_1 \leq a_1\sqrt{n}, Z_2 \leq a_2\sqrt{n}\big] \quad \text{(I.4)}$$
$$= \mathbb{P}\big[Z_1 \leq a_1\sqrt{n}\big]\mathbb{P}\big[Z_2 \leq a_2\sqrt{n}\big] \quad \text{(I.5)}$$
$$= \Phi(r_1)\Phi(r_2). \quad \text{(I.6)}$$

Thus, for any r_1, r_2 such that

$$\Phi(r_1)\Phi(r_2) \geq 1-\epsilon, \quad \text{(I.7)}$$

$\mathbf{a} \in \frac{\mathscr{Q}_{\mathrm{inv}}(\mathsf{V},\epsilon)}{\sqrt{n}}$ and hence

$$\overline{\mathbf{R}} \in \overline{\mathbf{H}}+\frac{\mathscr{Q}_{\mathrm{inv}}(\mathsf{V},\epsilon)}{\sqrt{n}}-\frac{\log n}{2n}\begin{bmatrix} \lambda \\ 1-\lambda \\ 1 \end{bmatrix} \subseteq \overline{\mathscr{R}}^*(n,\epsilon). \quad \text{(I.8)}$$

Therefore,

$$\overline{R}_{\mathrm{sum}}^*(n,\epsilon) \leq H(X_1)+H(X_2)+ \quad \text{(I.9)}$$
$$\min_{\substack{(r_1,r_2): \\ \Phi(r_1)\Phi(r_2)\geq 1-\epsilon}} \left(\sqrt{\frac{V(X_1)}{n}}r_1+\sqrt{\frac{V(X_2)}{n}}r_2\right)-\frac{\log n}{2n}.$$

On the other hand, for any r_1, r_2 such that

$$\Phi(r_1)\Phi(r_2) < 1-\epsilon, \quad \text{(I.10)}$$

$\mathbf{a} \notin \frac{\mathscr{Q}_{\mathrm{inv}}(\mathsf{V},\epsilon)}{\sqrt{n}}$ and hence $\overline{\mathbf{R}} \notin \overline{\mathscr{R}}^*(n,\epsilon)$ (142). Thus, (I.9) holds with equality.

APPENDIX J
PROOF OF THEOREM 23

The proof employs an extension of Han's MASC converse [9, Lemma 7.2.2].

Given an (L, M_1, M_2, ϵ) CF-MASC $(\mathsf{L}, \mathsf{f}_1, \mathsf{f}_2, \mathsf{g})$, let

$$\mathcal{S} \triangleq \big\{(x_1, x_2) \in \mathcal{X}_1\times\mathcal{X}_2 :$$
$$(x_1, x_2) = \mathsf{g}(\mathsf{f}_1(\mathsf{L}(x_1, x_2), x_1), \mathsf{f}_2(\mathsf{L}(x_1, x_2), x_2))\big\}$$
$$\mathcal{S}_1(x_2) \triangleq \big\{x_1 \in \mathcal{X}_1 : (x_1, x_2) \in \mathcal{S}\big\} \ \forall x_2 \in \mathcal{X}_2$$
$$\mathcal{S}_2(x_1) \triangleq \big\{x_2 \in \mathcal{X}_2 : (x_1, x_2) \in \mathcal{S}\big\} \ \forall x_1 \in \mathcal{X}_1.$$

Then $\mathbb{P}[\mathcal{S}^c]$ equals the code's error probability, and

$$|\mathcal{S}| \leq M_1 M_2 \quad \text{(J.1)}$$
$$|\mathcal{S}_1(x_2)| \leq LM_1, \text{ for any } x_2 \in \mathcal{X}_2 \quad \text{(J.2)}$$
$$|\mathcal{S}_2(x_1)| \leq LM_2, \text{ for any } x_1 \in \mathcal{X}_1, \quad \text{(J.3)}$$

where the bound on $|\mathcal{S}|$ is the number of distinct decoder inputs and the bounds on $|\mathcal{S}_1(x_2)|$ and $|\mathcal{S}_2(x_1)|$ are the number of distinct decoder inputs under fixed values of x_2 and x_1 and an ℓ-bit CF. Fix $\gamma > 0$. Define sets

$$\mathcal{U} \triangleq \big\{(x_1, x_2) \in \mathcal{X}_1\times\mathcal{X}_2 :$$
$$\imath(x_1, x_2) \geq \log M_1+\log M_2+\gamma\big\} \quad \text{(J.4)}$$
$$\mathcal{U}_1 \triangleq \big\{(x_1, x_2) \in \mathcal{X}_1\times\mathcal{X}_2 : \imath(x_1|x_2) \geq \log(LM_1)+\gamma\big\} \quad \text{(J.5)}$$
$$\mathcal{U}_2 \triangleq \big\{(x_1, x_2) \in \mathcal{X}_1\times\mathcal{X}_2 : \imath(x_2|x_1) \geq \log(LM_2)+\gamma\big\}. \quad \text{(J.6)}$$

Then,

$$\mathbb{P}[\mathcal{U}_1 \cap \mathcal{S}] \tag{J.7}$$

$$= \mathbb{E}\left[1\left\{P_{X_1|X_2}(X_1|X_2) \leq \frac{\exp(-\gamma)}{LM_1}\right\}1\{(X_1,X_2)\in\mathcal{S}\}\right]$$

$$\leq \sum_{x_2\in\mathcal{X}_2} P_{X_2}(x_2)|\mathcal{S}_1(x_2)|\frac{\exp(-\gamma)}{LM_1} \tag{J.8}$$

$$\leq \exp(-\gamma), \tag{J.9}$$

where (J.7) follows the definition of $\mathcal{U}_1$, (J.8) applies $1\{Z \leq z\} \leq z$, and (J.9) holds by (J.2). Similarly,

$$\mathbb{P}[\mathcal{U}_2 \cap \mathcal{S}] \leq \exp(-\gamma) \tag{J.10}$$

$$\mathbb{P}[\mathcal{U} \cap \mathcal{S}] \leq \exp(-\gamma). \tag{J.11}$$

Thus,

$$\mathbb{P}[\mathcal{U}_1\cup\mathcal{U}_2\cup\mathcal{U}] \leq \mathbb{P}[\mathcal{U}_1\cap\mathcal{S}]+\mathbb{P}[\mathcal{U}_2\cap\mathcal{S}]+\mathbb{P}[\mathcal{U}\cap\mathcal{S}]+\mathbb{P}[\mathcal{S}^c] \tag{J.12}$$

$$\leq 3\exp(-\gamma)+\mathbb{P}[\mathcal{S}^c]. \tag{J.13}$$

Rearranging (J.13) gives a lower bound on the error probability $\epsilon = \mathbb{P}[\mathcal{S}^c]$. Thus, any (L, M_1, M_2, ϵ) CF-MASC must satisfy

$$\epsilon \geq \mathbb{P}[\{\imath(X_1|X_2) \geq \log(LM_1)+\gamma\}\cup$$
$$\{\imath(X_2|X_1) \geq \log(LM_2)+\gamma\}\cup$$
$$\{\imath(X_1,X_2) \geq \log(M_1M_2)+\gamma\}]$$
$$-3\exp(-\gamma). \tag{J.14}$$

Particularizing (J.14) to stationary, memoryless sources with single-letter distribution $P_{X_1X_2}$ satisfying (139) and (140) shows that any $(n, L, M_1, M_2, \epsilon)$ CF-MASC must satisfy

$$\epsilon \geq \mathbb{P}[\{I_1 \geq \log(LM_1)+\gamma\}\cup\{I_2 \geq \log(LM_2)+\ell+\gamma\}$$
$$\cup\{I_{12} \geq \log(M_1M_2)+\gamma\}]-3\exp(-\gamma) \tag{J.15}$$

$$= 1-\mathbb{P}\left[\sum_{i=1}^{n}\mathbf{U}_i < n\overline{\mathbf{R}}-n\overline{\mathbf{H}}+\gamma\mathbf{1}\right]-3\exp(-\gamma), \tag{J.16}$$

where $\gamma > 0$ is an arbitrary constant, I_1, I_2, and I_{12} are defined in (146), (147), and (148), and $\mathbf{U}_i$ is defined in (183). Let L be a finite constant that does not grow with n and let $\gamma = \frac{\log n}{2} - \log L$. Applying Lemma 12 and Lemma 13-2 to bound the probability (J.16) in a manner similar to (182)–(187), we conclude that any (n, ℓ, ϵ)-achievable rate pair (R_1, R_2) must be in $\mathscr{R}_{\text{out}}^*(n, \epsilon)$ (144). ∎

Remark 16. One could also prove Theorem 23 by extending our HT converse (Theorem 19) to the setting with a cooperation facilitator. Our Theorem 19 continues to hold with M_1 and M_2 replaced by LM_1 and LM_2 in (126) and (127), respectively ((128) remains unchanged).

ACKNOWLEDGMENT

REFERENCES

[1] S. Chen, M. Effros, and V. Kostina, "Lossless source coding in the point-to-point, multiple access, and random access scenarios," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 1692–1696.

[2] SPECTRE: Short packet communication toolbox. [Online]. Available: https://github.com/yp-mit/spectre/tree/master/lossless-sc

[3] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 2, pp. 379–423 and 623–656, July and Oct. 1948.

[4] V. Strassen, "Asymptotische abschäzungen in shannons informationstheorie," in *Proc. Trans. Third Prague Conf. Inf. Theory, Statist., Decision Funct., Random Process.*, 1964, pp. 689–723.

[5] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[6] I. Kontoyiannis and S. Verdú, "Optimal lossless data compression: Non-asymptotics and asymptotics," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 777–795, Feb. 2014.

[7] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors," *IEEE Trans. Inf. Theory*, vol. 61, no. 8, pp. 4316–4330, Aug. 2015.

[8] A. A. Yushkevich, "On limit theorems connected with the concept of entropy of Markov chains," *Uspekhi Mat. Nauk*, vol. 8, no. 5(57), pp. 177–180, 1953.

[9] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer-Verlag Berlin Heidelberg, 2003.

[10] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4619–4637, Aug. 2008.

[11] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[12] S. Miyake and F. Kanaya, "Coding theorems on correlated general sources," *IEICE Trans. on Fundamentals of Electronics*, vol. E78-A, no. 9, pp. 1063–1070, Sep. 1995.

[13] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 881–903, Feb. 2014.

[14] R. Nomura and T. S. Han, "Second-order Slepian-Wolf coding theorems for non-mixed and mixed sources," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5553–5572, Sep. 2014.

[15] S. T. Jose and A. A. Kulkarni, "Improved finite blocklength converses for Slepian–Wolf coding via linear programming," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2423–2441, Apr. 2019.

[16] P. Minero, M. Franceschetti, and D. N. C. Tse, "Random access: An information-theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 909–930, Feb. 2012.

[17] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 2523–2527.

[18] M. Effros, V. Kostina, and R. C. Yavas, "Random access channel coding in the finite blocklength regime," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1261–1265.

[19] M. V. Burnashev, "Data transmission over a discrete channel with feedback: Random transmission time," *Problems of Information Transmission*, vol. 12, no. 4, pp. 10–30, Aug. 1976.

[20] A. Tchamkerten and I. E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, May 2006.

[21] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4903–4925, Aug. 2011.

[22] S. Draper, "Universal incremental Slepian-Wolf coding," in *Proc. 42th Annual Allerton Conference on Communications, Control, and Computing*, Sep. 2004.

[23] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.

[24] S. Verdú. Ee528 - information theory, lecture notes. Princeton University, Princeton, NJ, 2007.

[25] Y. Polyanskiy and Y. Wu. Lecture notes on information theory. MIT (6.441), UIUC (ECE 563), Yale (STAT 664), 2012-2017. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/itlectures_v5.pdf

[26] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. John Wiley & Sons, 1971, vol. II.

[27] I. G. Shevtsova, "On the absolute constants in the Berry-Esseen inequality and its structural and nonuniform improvements," *Informatika i Ee Primeneniya [Informatics and its Applications]*, vol. 7, no. 1, pp. 124–125, 2013.

[28] H. Nagaoka, "Strong converse theorems in quantum information theory," in *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*. World Scientific, 2005, pp. 64–65.

[29] M. Hayashi, *Quantum Information: An introduction*. Berlin Heidelberg: Springer, Apr. 2006.

[30] A. T. Campo, G. Vazquez-Vilar, A. G. i. Fàbregas, and A. Martinez, "Converse bounds for finite-length joint source-channel coding," in *Proc. 50th Annual Allerton Conference on Communications, Control, and Computing*, Oct. 2012.

[31] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.

[32] Y. Huang and P. Moulin, "Strong large deviations for composite hypothesis testing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2014, pp. 556–560.

[33] M. Effros, V. Kostina, and R. C. Yavas, "Random access channel coding in the finite blocklength regime," *arXiv:1801.09018v3 [cs.IT]*, Jul. 2019.

[34] N. Elkayam and M. Feder, "On the calculation of the minimax-converse of the channel coding problem," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2017, pp. 1247–1251.

[35] V. Bentkus, "On the dependence of the Berry-Esseen bound on dimension," *J. Stat. Planning and Inference*, vol. 113, pp. 385–402, May 2003.

[36] P. Noorzad, M. Effros, M. Langberg, and T. Ho, "On the power of cooperation: Can a little help a lot?" in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, June/July 2014, pp. 3132–3136.

[37] M. Langberg and M. Effros, "Network coding: Is zero error always possible?" in *Proc. 49th Annual Allerton Conference on Communications, Control, and Computing*, Sep. 2011.

[38] Y. Oohama and T. S. Han, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1908–1919, Nov. 1994.

[39] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inf. Theory*, vol. 26, no. 2, pp. 155–165, Mar. 1980.

[40] Y. Oohama, "Universal coding for correlated sources with linked encoders," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 837–847, May 1996.

[41] S. Jaggi and M. Effros, "Universal linked multiple access source codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2002, p. 95.

[42] E. Yang, D. He, T. Uyematsu, and R. W. Yeung, "Universal multiterminal source coding algorithms with asymptotically zero feedback: Fixed database case," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5575–5590, Dec. 2008.

[43] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Variable-rate universal Slepian-Wolf coding with feedback," in *Proc. 39th Asilomar Conf. on Signals, Systems, and Computers*, Oct. 2005, pp. 8–12.

[44] A. W. Eckford and W. Yu, "Rateless Slepian-Wolf codes," in *Proc. 39th Asilomar Conf. on Signals, Systems, and Computers*, Oct. 2005, pp. 1757–1761.

Shuqing Chen (Graduate Student Member, IEEE) received the B.S. degree from Rice University in 2017 and the M.S. (2020) from California Institute of Technology, in electrical engineering. She worked with the Data Compression Laboratory, Caltech, from 2017 to 2019. Her research involves non-asymptotic information theory, composite hypothesis testing and multi-terminal source coding.

Michelle Effros (Fellow, IEEE) received the B.S. (Hons.), M.S., and Ph.D. degrees in electrical engineering from Stanford University, in 1989, 1990, and 1994, respectively.

In 1994, she joined the faculty at the California Institute of Technology, where she is currently the George Van Osdol Professor of electrical engineering. Her research interests include information theory, network coding, data compression, and communications. She received Stanford's Frederick Emmons Terman Engineering Scholastic Award (for excellence in engineering) in 1989, the Hughes Masters Full-Study Fellowship in 1989, the National Science Foundation Graduate Fellowship in 1990, the AT&T Ph.D. Scholarship in 1993, the NSF CAREER Award in 1995, the Charles Lee Powell Foundation Award in 1997, the Richard Feynman-Hughes Fellowship in 1997, and an Okawa Research Grant in 2000. She was cited by Technology Review as one of the world's top young innovators in 2002. She and her co-authors received the Communications Society and Information Theory Society Joint Paper Award in 2009. She became a fellow of the IEEE in 2009. She is a member of Tau Beta Pi, Phi Beta Kappa, and Sigma Xi. She served as the Editor of the IEEE Information Theory Society Newsletter 1995 to 1998 and as a Member of the Board of Governors of the IEEE Information Theory Society from 1998 to 2003 and from 2008 to 2017. She also served as President of the Information Theory Society in 2015. She was a member of the Advisory Committee and the Committee of Visitors for the Computer and Information Science and Engineering (CISE) Directorate at the National Science Foundation from 2009 to 2012 and in 2014, respectively. She served on the IEEE Signal Processing Society Image and Multi-Dimensional Signal Processing (IMDSP) Technical Committee from 2001 to 2007 and on ISAT from 2006 to 2009. She served as Associate Editor for the joint special issue on Networking and Information Theory in the *IEEE Transactions on Information Theory* and the *IEEE Transactions on Networking/ACM Transactions on Networking* and as Associate Editor for Source Coding for the *IEEE Transactions on Information Theory* from 2004 to 2007. She has served on numerous technical program committees and review boards, including serving as general co-chair for the 2009 Network Coding Workshop and technical program committee co-chair for the 2012 IEEE International Symposium on Information Theory.

Victoria Kostina (Member, IEEE) received the bachelor's degree from the Moscow institute of Physics and Technology in 2004, the master?s degree from the University of Ottawa in 2006, and the Ph.D. degree from Princeton University in 2013. She was affiliated with the Institute for Information Transmission Problems, Russian Academy of Sciences. In 2014, she joined Caltech, where she is currently a Professor of electrical engineering. Her research spans information theory, coding, control, learning, and communications. She received the Natural Sciences and Engineering Research Council of Canada master's scholarship in 2009, the Princeton Electrical Engineering Best Dissertation Award in 2013, the Simons-Berkeley Research Fellowship in 2015, and the NSF CAREER Award in 2017.