Slepian-Wolf Coding for Broadcasting with Cooperative Base-Stations

Roy Timo and Michèle Wigger

Abstract—**We propose a base-station (BS) cooperation model for broadcasting a discrete memoryless source in a cellular or heterogeneous network. The model allows the receivers to use helper BSs to improve network performance, and it permits the receivers to have prior side information about the source. We establish the model's information-theoretic limits in two operational modes: In Mode 1, the helper BSs are given information about the channel codeword transmitted by the main BS, and in Mode 2 they are provided correlated side information about the source. Optimal codes for Mode 1 use *hash-and-forward coding* at the helper BSs; while, in Mode 2, optimal codes use source codes from Wyner's *helper source-coding problem* at the helper BSs. We prove the optimality of both approaches by way of a new list-decoding generalisation of [8, Thm. 6], and, in doing so, show an operational duality between Modes 1 and 2.**

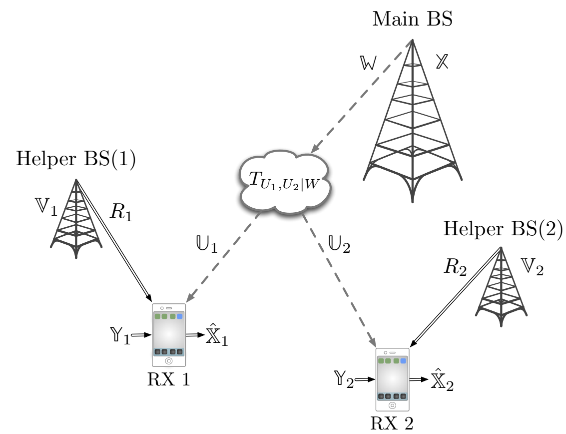

Fig. 1. Broadcasting with helper BSs and receiver side information.

I. Introduction

THE proliferation of wireless communications devices presents significant performance challenges for cellular networks, and it will require more sophisticated heterogeneous networks in the near future [1, 2]. A powerful methodology for improving performance is centered on the idea of base-station (BS) cooperation: Instead of operating independently, future BSs will coordinate encoding and decoding operations using information shared over backbone networks. The tremendous potential of BS cooperation has been widely investigated [3]–[5]; however, despite many advances, there remains significant challenges in understanding and exhausting the benefits of cooperation. Indeed, the fundamental limits of cooperation are fully understood in very few settings [4].

To help understand the full potential of BS cooperation, we consider a simple, but rather useful, broadcast model. The setup for two receivers is shown in Figure 1. A source $\mathbb{X}$ is to be reliably transmitted over a broadcast channel to many receivers, and the idea is to improve network performance by allowing the receivers to be assisted by *helper* BSs. In a future heterogenous network, for example, the helpers may be pico or femto BSs operating within the main macro cell on orthogonal channels [6]. Alternatively, the helpers may be WiFi hotspots through which traffic is diverted from a heavily loaded cellular network [7]. The purpose of this paper is to characterise the model's information-theoretic limits, and to provide architectural insights for optimal codes.

R. Timo is an Alexander von Humboldt research fellow with the Technische Universität München, e-mail roy.timo@tum.de.

M. Wigger is with Telecom ParisTech, e-mail michele.wigger@telecom-paristech.fr.

This work was supported by the city of Paris under the programme "Emergences" and the Alexander von Humboldt Foundation.

We assume that the broadcast channel from the main BS is discrete and memoryless, and the channels from the helper BSs are noiseless and rate-limited. Although this setup does not capture all modes of cooperation, it nevertheless has enough sophistication to provide insight into some important coding challenges. For example, consider the idea of augmenting traffic flow in a cellular network via a WiFi hotspot: The hotspot's radio-access technology is orthogonal to that of the cellular network, and a cellular network engineer can well approximate the WiFi link by a noiseless rate-limited channel. A natural question is then: What coding techniques at the BSs and WiFi hotspot yield the best overall performance?

Within the above framework, we consider two operational modes.

- *Mode 1:* The helper BSs are given side information about the channel codeword transmitted by the main BS.
- *Mode 2:* The helper BSs are given correlated side information about the source $\mathbb{X}$.

We will see that optimal codes for Mode 1 combine virtual-binning from *Slepian-Wolf Coding over Broadcast Channels* [8] with hash-and-forward coding for the *primitive relay channel* [9]. Optimal codes for Mode 2, on the other hand, combine virtual binning with source codes from *Wyner's helper side-information problem* [10]. We prove the optimality of both codes by way of a new list-decoding generalisation of [8, Thm. 6], and, in doing so, show an operational duality between Modes 1 and 2.

The paper is organised as follows. The BS cooperate model is defined in Section II, and our results are summarised in Section III. We introduce and solve a list-decoding broadcast problem in Sections IV through VI. Finally, we prove the BS cooperation results in Sections VII to X.

II. Preliminaries

A. Notation

We denote random variables by uppercase letters, e.g. A; their alphabets by calligraphic typeface, e.g. $\mathcal{A}$; and elements of an alphabet by lowercase letters, e.g. $a \in \mathcal{A}$. The Cartesian product of alphabets $\mathcal{A}$ and $\mathcal{B}$ is $\mathcal{A} \times \mathcal{B}$, and the n-fold Cartesian product of $\mathcal{A}$ is $\mathcal{A}^n$. When n is clear from context, we use boldface notation for a sequence of n random variables on a common alphabet, e.g. $\mathbb{A} = (A_1, A_2, \ldots, A_n) \in \mathcal{A}^n$.

B. Source and Channel Setup

The main BS is required to communicate a source

$$\mathbb{X} = (X_1, X_2, \ldots, X_{n_\text{s}})$$

over a discrete memoryless broadcast channel to K receivers with side information; the side information at receiver k, for $k \in \{1, 2, \ldots, K\}$, is denoted by

$$\mathbb{Y}_k = (Y_{k,1}, Y_{k,2}, \ldots, Y_{k,n_\text{s}}).$$

For example, $\mathbb{X}$ and $\mathbb{Y}_k$ may be the current and previous states of a mobile application, the global and local contents of a cloud storage drive, or the current and previous frames of a video feed. Alternatively, specific choices of $\mathbb{X}$ and $\mathbb{Y}_k$ lead to the bi-directional broadcast channel and complementary side information model [11]–[14]. For generality, let us only assume that the source and side information are emitted by a discrete memoryless source[1]. That is,

$$(\mathbb{X}, \mathbb{Y}_1, \mathbb{Y}_2, \ldots, \mathbb{Y}_K) := \big\{(X_i, Y_{1,i}, Y_{2,i}, \ldots, Y_{K,i})\big\}_{i=1}^{n_\text{s}}$$

is a sequence of n_s independent and identically distributed (iid) source/side-information tuples $(X, Y_1, Y_2, \ldots, Y_K)$ defined by a fixed, but arbitrary, joint probability mass function (pmf) on the Cartesian product space $\mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_K$.

Let $\mathcal{W}$ denote the broadcast channel's input alphabet and $\mathcal{U}_k$ its output alphabet at receiver k. The main BS transmits

$$\mathbb{W} := f(\mathbb{X})$$

over the broadcast channel, where $f : \mathcal{X}^{n_\text{s}} \longrightarrow \mathcal{W}^{n_\text{c}}$ is the BS's encoder and $\mathbb{W} = (W_1, W_2, \ldots, W_{n_\text{c}})$ is a codeword with n_c symbols. The ratio of channel symbols to source symbols,

$$\kappa := \frac{n_\text{c}}{n_\text{s}},$$

is called the *bandwidth expansion factor*.

Receiver k observes $\mathbb{U}_k = (U_{k,1}, U_{k,2}, \ldots, U_{k,n_\text{c}})$ from the channel. The channel outputs, across all receivers, conditionally depend on the codeword $\mathbb{W}$ via the memoryless law

$$\mathbb{P}[\mathbb{U}_1 = \mathbb{u}_1, \mathbb{U}_2 = \mathbb{u}_2, \ldots, \mathbb{U}_K = \mathbb{u}_K | \mathbb{W} = \mathbb{w}] = \prod_{i=1}^{n_\text{c}} T(u_{1,i}, u_{2,i}, \ldots, u_{K,i} | w_i),$$

where $\mathbb{w} \in \mathcal{W}^{n_\text{c}}$, $\mathbb{u}_k \in \mathcal{U}_k^{n_\text{c}}$ and $T(u_1, \ldots, u_K | w)$ is a fixed, but arbitrary, conditional probability.

[1]It is possible to extend this research to discrete ergodic sources using, for example, the methods of [15]. However, discrete memoryless sources lead to more instructive proofs with less technical and notational difficulties.

C. No Base-Station Cooperation

Momentarily suppose that there is no BS cooperation, and that the source is to be losslessly reconstructed using only the channel outputs and side information at each receiver. In this setting, reliable communication is possible if (and only if)[2] there exists a pmf P_W on $\mathcal{W}$ such that [8]

$$H(X|Y_k) < \kappa I(W; U_k), \quad \forall\ k, \tag{1}$$

where $(W, U_1, U_2, \ldots, U_K) \sim P_W(\cdot) T(\cdot|\cdot)$. The necessity and sufficiency of (1) for reliable communication is an elegant and powerful result with applications throughout network information theory; for example, consider [11]–[14] and [16]–[19]. Indeed, a new list-decoding generalisation of (1) will play a central role in this paper.

D. Base-Station Cooperation

Let us now return to the BS cooperation model. The helper BS of receiver k, denoted BS(k), obtains side information

$$\mathbb{V}_k = (V_{k,1}, V_{k,2}, \ldots, V_{k,n_\text{h}})$$

about the source $\mathbb{X}$ or the codeword $\mathbb{W}$ via a backbone network. Here $n_\text{h} = n_\text{s}$ (resp. $n_\text{h} = n_\text{c}$) when BS(k) has side information about $\mathbb{X}$ (resp. $\mathbb{W}$), and a precise definition of $\mathbb{V}_k$ will be given shortly. BS(k) sends

$$M_k := f_k(\mathbb{V}_k)$$

over a noiseless channel to receiver k, where $f_k : \mathcal{V}^{n_\text{h}} \to \{1, 2, \ldots, \lfloor 2^{n_\text{s} R_k} \rfloor\}$ is BS(k)'s encoder and R_k is its rate (in bits per source symbol[3]). Receiver k attempts to recover the source via

$$\hat{\mathbb{X}}_k := g_k(\mathbb{U}_k, \mathbb{Y}_k, M_k),$$

where $g_k : \mathcal{U}_k^{n_\text{c}} \times \mathcal{Y}_k^{n_\text{s}} \times \{1, 2, \ldots, \lfloor 2^{n_\text{s} R_k} \rfloor\} \longrightarrow \mathcal{X}^{n_\text{s}}$ is the receiver's decoder. The collection of all encoders and decoders is called an $(n_\text{s}, n_\text{c}, R_1, R_2, \ldots, R_K)$-*code*.

E. Mode 1 (helper side information about the codeword $\mathbb{W}$)

Suppose that $\mathbb{V}_k$ is the entire codeword $\mathbb{W}$ or a scalar quantised version thereof. Quantisation is appropriate, for example, when the backbone network is rate limited. More formally, let $\phi_k : \mathcal{W} \to \mathcal{V}_k$ be an arbitrary but given deterministic mapping (scalar quantiser) and

$$V_{k,i} := \phi_k(W_i), \quad \forall\ i.$$

The main problem of interest is to determine when reliable communication is achievable in the following sense.

Definition 1: Fix the bandwidth expansion factor κ, helper BS rates $\boldsymbol{R} := (R_1, R_2, \ldots, R_K)$, and scalar quantisers $\boldsymbol{\phi} := (\phi_1, \phi_2, \ldots, \phi_K)$. We say that a source/side information tuple $(X, Y_1, Y_2, \ldots, Y_K)$ is $(\kappa, \boldsymbol{R}, \boldsymbol{\phi})$-*achievable* if for any $\epsilon > 0$ there exists an $(n_\text{s}, n_\text{c}, R_1, R_2, \ldots, R_K)$-code such that

$$\frac{n_\text{c}}{n_\text{s}} = \kappa \quad \text{and} \quad \mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X}] \leq \epsilon, \quad \forall\ k, \tag{2}$$

holds for sufficiently large n_s and n_c.

[2]For the "only if" assertion: Replace the strict inequality $<$ in (1) with an inequality $\leq$ in the same direction.

[3]Here we have synchronised the rate R_k to the number of source symbols n_s. Alternatively, one could synchronise R_k to the number of channel symbols by replacing n_s with n_c in the definition of f_k.

F. Mode 2 (helper side information about the source $\mathbb{X}$)

Suppose that $\mathbb{V}_k$ is directly correlated with the source and side information. That is, assume $(\mathbb{X}, \mathbb{Y}_1, \mathbb{Y}_2, \ldots, \mathbb{Y}_K, \mathbb{V}_1, \mathbb{V}_2, \ldots, \mathbb{V}_K)$ is emitted by an arbitrary discrete memoryless source and thus is a sequence of n_s iid tuples $(X, Y_1, Y_2, \ldots, Y_K, V_1, V_2, \ldots, V_K)$. We are interested in the following definition of achievability.

Definition 2: Fix the bandwidth expansion factor κ and helper BS rates $\boldsymbol{R} := (R_1, R_2, \ldots, R_K)$. We say that a source/side information tuple $(X, Y_1, Y_2, \ldots, Y_K, V_1, V_2, \ldots, V_K)$ is $(\kappa, \boldsymbol{R})$-*achievable* if for any $\epsilon > 0$ there exists an $(n_\mathrm{s}, n_\mathrm{c}, R_1, R_2, \ldots, R_K)$-code such that (2) holds for sufficiently large n_s and n_c.

III. Information-Theoretic Limits of BS Cooperation

A. Mode 1

Theorem 1: Fix the helper BS rates $\boldsymbol{R}$, bandwidth expansion factor κ and quantisers $\boldsymbol{\phi}$. A source/side-information tuple $(X, Y_1, Y_2, \ldots, Y_K)$ is $(\kappa, \boldsymbol{R}, \boldsymbol{\phi})$-achievable if there exists a pmf P_W on $\mathcal{W}$ such that for all k

$$H(X|Y_k) < \kappa I(W; U_k) + \min\big\{R_k, \kappa I(W; V_k|U_k)\big\}, \quad (3)$$

where $(W, U_1, U_2, \ldots, U_K) \sim P_W(\cdot)T(\cdot|\cdot)$ and $V_k = \phi_k(W)$. Conversely: If the source/side-information tuple $(X, Y_1, Y_2, \ldots, Y_K)$ is $(\kappa, \boldsymbol{R}, \boldsymbol{\phi})$-achievable, then there exists a pmf P_W on $\mathcal{W}$ such that (3) holds with inequality $(\leq)$ for all k.

Theorem 1 is proved in Sections VII and VIII.

B. Mode 2

Theorem 2: Fix the helper BS rates $\boldsymbol{R}$ and bandwidth expansion factor κ. A source/side-information tuple $(X, Y_1, Y_2, \ldots, Y_K, V_1, V_2, \ldots, V_K)$ is $(\kappa, \boldsymbol{R})$-achievable if there exists a pmf P_W on $\mathcal{W}$ and K auxiliary random variables $(A_1, A_2, \ldots, A_K)$ such that for all k we have the Markov chain $(X, Y_k) \leftrightarrow V_k \leftrightarrow A_k$,

$$R_k > I(V_k; A_k|Y_k) \quad (4a)$$

and

$$H(X|A_k, Y_k) < \kappa I(W; U_k), \quad (4b)$$

where $(W, U_1, U_2, \ldots, U_K) \sim P_W(\cdot)T(\cdot|\cdot)$. Conversely: If the source/side-information tuple $(X, Y_1, Y_2, \ldots, Y_K, V_1, V_2, \ldots, V_K)$ is $(\kappa, \boldsymbol{R})$-achievable, then there exists a pmf P_W on $\mathcal{W}$ and K auxiliary random variables $(A_1, A_2, \ldots, A_K)$ such that (4a) and (4b) hold with inequalities (in the same directions as the strict inequalities) and $(X, Y_k) \leftrightarrow V_k \leftrightarrow A_k$ for all k.

Theorem 2 is proved in Sections IX and X.

Remark 1: When computing Theorem 2, we can assume that the alphabet of A_k has a cardinality of at most $|\mathcal{V}_k|$.

C. Discussion

Consider Mode 1 and Theorem 1. If the helper rates are all set to zero, then (3) becomes

$$H(X|Y_k) < \kappa I(W; U_k), \quad \forall\ k, \quad (5)$$

and we recover the setup of (1). Now suppose that for a given pmf P_W and scalar quantisers $\boldsymbol{\phi}$ we have $R_k > \kappa H(V_k|U_k)$ for all k. If $(\mathbb{V}_k, \mathbb{U}_k)$ behaves like a discrete memoryless source, then BS(k) can reliably send $\mathbb{V}_k$ to receiver k using a Slepian-Wolf code of rate R_k [22]. The receiver effectively has the combined channel output $(\mathbb{U}_k, \mathbb{V}_k)$. Since (3) simplifies to

$$H(X|Y_k) < \kappa I(W; U_k, V_k), \quad \forall\ k,$$

we again return to the result in (1), where the k-th channel output U_k is replaced by (U_k, V_k).

For other helper rates, we note the similarity of (3) to Kim's capacity theorem [9, Thm. 1] for the primitive relay channel. Intuitively, the right hand side of (3) is the maximum rate at which information can be sent to receiver k. This intuition, however, should be treated with care because, for example, the classical Shannon approach of *strictly* separating source and channel coding is suboptimal[4]. Nonetheless, it is natural to wonder whether Kim's simple timesharing proof of [9, Thm. 1] can be modified to prove Theorem 1. While we do not take the timesharing approach in this paper, D. Gündüz has noticed that it may indeed be possible to give such a proof of Theorem 1 using the *semiregular encoding* and *backward decoding* techniques developed in [19, App. B] (these techniques, for example, give an alternative proof of the no-cooperation case shown in (1)).

Modes 1 and 2 are both special cases of the more general *relay network* problem treated in [19, Sec. V], and Theorems 1 and 2 solve two previously unknown special cases of that problem. The achievability result [19, Thm. 3] (which holds for a much broader class of relay problems than Modes 1 and 2) is not optimal in the case of Mode 2. Loosely speaking, [19, Thm. 3] is built on *decode and forward* relaying protocols (using random binning) in which a selected subset of strong[5] relays decode and forward the source $\mathbb{X}$. In Mode 2, there is no mechanism for the main BS to communicate with helper BS(k), so helper BS(k) is weak in the sense that it cannot decode $\mathbb{X}$ whenever $H(X|V_k) > 0$. The achievability proof of Theorem 2 utilises a *quantise and forward* strategy in which helper BS(k) forwards partial information about the source $\mathbb{X}$ via a 'quantised' version of its side information $\mathbb{V}_k$. The quantisation is specified by the auxiliary random variable A_k in a similar way to the quantisation in Wyner's helper source-coding problem [10], [23, p. 575] or the Wyner-Ziv rate-distortion problem [24]. Intuitively, the decode and forward scheme in [19, Thm. 3] is suboptimal in Mode 2 for much the same reason as *random binning* at the helper is suboptimal in the *helper source-coding* problem. It may be possible to improve the general achievability result [19, Thm. 3] by considering *partial* decode and forward protocols, where selected weak relays decode and forward partial information about $\mathbb{X}$. For example, this partial information might be specified via Gács-Körner common information in the sense of [40].

The single-letter characterisations in Theorems 1 and 2 depend only on the marginal source and channel distributions,

[4] To see why *strict* source-channel separation fails, set $R_k = 0$ for all k and consider the examples in [8].

[5] Relays with high quality channels / side information.

instead of the complete joint source-channel distribution[6] — the latter being more typical in the joint source-channel coding literature, e.g., see [38]. The separation of source and channel variables in Theorems 1 and 2 is reminiscent of *operational separation* described in [8] and can be similarly understood[7]. Indeed, in both modes we will see that it is optimal to separate the source, channel and helper codebooks as well as the encoders, but joint decoding across all three codebooks is required. In particular, the approach taken in this paper is to first require that receiver k decodes a list of likely source sequences using a joint source-channel decoder on its channel output $\mathbb{U}_k$ and side information $\mathbb{Y}_k$. The receiver then determines the correct source sequence, within this list, using the helper BS's message and codebook. The list decoding approach is particularly useful because it highlights an operation duality between Modes 1 and 2: The helper BS's task in both modes is to help the receiver resolve the correct source sequence from the receiver's list.

Comparing Theorems 1 and 2: Increasing the helper rates in Theorem 2 allows larger 'quantisation rates' and reductions in the left hand side of (4b). In contrast, increasing the helper rates in Theorem 1 improves the 'relay capacity' and increases the right hand side of (3). We will see that these properties are dual consequences of the same random-coding idea.

Finally, we note that Theorems 1 and 2 are existential statements that do not give constructive arguments for low-complexity codes. That being said, however, the single-letter expressions and (as we will see) the structure of the random-coding achievability proofs give insight into the architecture of good low-complexity codes. For example, the hash-and-forward technique used in Mode 1 is similar to distributed source coding using LDPC codes [25]. Similarly, in Mode 2, preliminary work suggests that (nonlinear) trellis codes and rate-distortion codes perform well for quantising $\mathbb{V}_k$ [26]. Finally, recent work [39] suggests that repeat-accumulate codes can be useful for Slepian-Wolf coding over broadcast channels.

D. Mixed modes

Suppose that some helper BSs have information about the codeword $\mathbb{W}$, while others have information about the source $\mathbb{X}$ — a mix of Modes 1 and 2. Let $\mathcal{K}_1$ and $\mathcal{K}_2$ denote the index sets of Mode 1 and 2 helper BSs respectively. It can be argued from Theorems 1 and 2 that a source/side-information tuple is achievable if there exists a pmf P_W on $\mathcal{W}$ and $|\mathcal{K}_2|$ auxiliary random variables $\{A_k;\ k\in\mathcal{K}_2\}$ such that

$$H(X|Y_k) < \kappa I(W;U_k) + \min\big\{R_k, \kappa I(W;V_k|U_k)\big\} \tag{6}$$

for all $k\in\mathcal{K}_1$, and

$$R_k > I(V_k;A_k|Y_k) \tag{7a}$$

$$H(X|A_k,Y_k) < \kappa I(W;U_k) \tag{7b}$$

and $A_k \leftrightarrow V_k \leftrightarrow (X,Y_k)$ for all $k\in\mathcal{K}_2$. Conversely: If the source/side-information tuple is achievable, then there exists a pmf P_W on $\mathcal{W}$ and $|\mathcal{K}_2|$ auxiliary random variables $\{A_k;\ k\in\mathcal{K}_2\}$ such that (6) and (7) hold with inequalities (in the same directions as the strict inequalities) and $A_k \leftrightarrow V_k \leftrightarrow (X,Y_k)$.

E. Broadcast capacity with helpers

Consider Mode 1 for the bandwidth-matched case $n_{\mathrm{s}} = n_{\mathrm{c}} = n$, and fix a positive rate R^*. Suppose that there is no side information and the main BS is required to broadcast a discrete rate R^* message M to the receivers, where M is uniformly distributed on $\{1,2,\ldots,\lfloor 2^{nR^*}\rfloor\}$. For example, in Theorem 1 suppose that 2^{R^*} is an integer, $\kappa=1$, $Y_k=$ constant and $M=\mathbb{X}$, where $\mathbb{X}$ is iid with a uniform distribution on $\{1,2,\ldots,2^{R^*}\}$. Then $H(X|Y_k)=H(X)=R^*$ for all k.

Given helper rates $\boldsymbol{R}$, we can define the *helper capacity* $C(\boldsymbol{R})$ to be the supremum of all achievable message rates R^*; that is, those rates R^* for which there exists a sequence of codes with vanishing probability of decoding error. It can be argued from Theorem 1 that

$$C(\boldsymbol{R}) = \max_{P_W}\min_{k}\Big[I(W;U_k) + \min\big\{R_k, I(W;V_k|U_k)\big\}\Big], \tag{8}$$

where the maximisation is taken over all pmfs P_W on $\mathcal{W}$.

If the channel outputs are defined over a common alphabet, say $\mathcal{U}_k=\mathcal{U}$ for all k, then (8) is a type of compound channel capacity with relays. Indeed, one recovers the compound channel capacity theorem [20, 21] upon setting $R_k=0$ in (8).

F. Bidirectional broadcast channel with helpers

Consider Mode 1 with two receivers for the bandwidth matched case $n_{\mathrm{s}}=n_{\mathrm{c}}=n$, and fix positive rates R_1^* and R_2^*. Recall the bidirectional setup of [11]: The main BS has two independent uniformly distributed messages M_1 and M_2 on $\{1,2,\ldots,\lfloor 2^{nR_1^*}\rfloor\}$ and $\{1,2,\ldots,\lfloor 2^{nR_2^*}\rfloor\}$ respectively; receiver 1 has M_1 as side information and requires M_2; and receiver 2 has M_2 as side information and requires M_1. For example, in Theorem 1 suppose that $\kappa=1$, $2^{R_1^*}$ and $2^{R_2^*}$ are integers, $M_1=\mathbb{X}_1=\mathbb{Y}_1$ and $M_2=\mathbb{X}_2=\mathbb{Y}_2$, where $\mathbb{X}_1$ and $\mathbb{X}_2$ are independent with iid uniform distributions on $\{1,2,\ldots,2^{R_1^*}\}$ and $\{1,2,\ldots,2^{R_2^*}\}$ respectively. Then, setting $\mathbb{X}=(\mathbb{X}_1,\mathbb{X}_2)$ gives

$$H(X|Y_1)=H(X_1,X_2|X_1)=H(X_2)=R_2^* \quad\text{and}$$
$$H(X|Y_2)=H(X_1,X_2|X_2)=H(X_1)=R_1^*.$$

For fixed helper rates (R_1,R_2), we can define the *helper capacity region* $\mathcal{C}(R_1,R_2)$ to be closure of the set of all (R_1,R_2)-achievable rate pairs (R_1^*,R_2^*). It can be argued from Theorem 1 that $\mathcal{C}(R_1,R_2)$ is equal to the set of all (R_1^*,R_2^*) for which there exists a pmf P_W on $\mathcal{W}$ such that

$$R_1^* \le I(W;U_2) + \min\big\{R_2, I(W;V_2|U_2)\big\}$$
$$R_2^* \le I(W;U_1) + \min\big\{R_1, I(W;V_1|U_1)\big\}.$$

[6] All of the entropy and mutual information functions in Theorems 1 and 2 depend on either the source variables or the channel variables, but not both.

[7] More detailed discussions on the various types of source-channel separation can be found in [8, 16, 19].

G. Example for Theorem 2

Consider Theorem 2 with $\kappa = 1$, and choose $(\rho_1, \rho_2, \ldots, \rho_K) \in [0, 1/2]^K$. Suppose that the source is uniform and binary, $X \sim$ Bern(1/2); there is no receiver side information, $Y_k =$ constant; and helper BS(k)'s side information is

$$V_k := X \oplus Z_k, \qquad \text{(modulo 2)}, \tag{9}$$

where $Z_k := \text{Bern}(\rho_k)$ is independent additive binary noise.

The source/side-information tuple $(X, V_1, V_2, \ldots, V_K)$ is $(1, \boldsymbol{R})$-achievable if there exists a pmf P_W on $\mathcal{W}$ and constants $(\alpha_1, \alpha_2, \ldots, \alpha_K) \in [0, 1/2]^K$ such that

$$R_k > 1 - h(\alpha_k) \quad \text{and} \quad h(\alpha_k \star \rho_k) < \kappa I(W; U_k) \tag{10}$$

holds for all k. Here

$$h(a) := \begin{cases} -a \log_2 a - (1-a) \log_2 (1-a), & a \in (0, 1/2], \\ 0, & a = 0. \end{cases}$$

denotes the binary entropy function and

$$a \star b := a(1-b) + (1-a)b, \quad 0 \leq a, b \leq 1.$$

Conversely: If the source/side-information tuple $(X, V_1, V_2, \ldots, V_K)$ is $(1, \boldsymbol{R})$-achievable, then there exists a pmf P_W on $\mathcal{W}$ and $(\alpha_1, \alpha_2, \ldots, \alpha_K) \in [0, 1/2]^K$ such that (10) holds with inequalities (in the same directions as the strict inequalities).

The above example is an application of Wyner's *binary helper source coding problem* [10] (see also [35]–[37] and [29, Thm. 10.2]). To see why (10) holds, consider the following: Let $(A_1, A_2, \ldots, A_K)$ be any tuple of auxiliary random variables satisfying the conditions of Theorem 2. We first notice that

$$\begin{aligned} H(X|V_k) &\overset{\text{a}}{=} h(\rho_k) \\ &\overset{\text{b}}{\leq} H(X|A_k) \overset{\text{c}}{=} H(X|A_k, Y_k) \overset{\text{d}}{\leq} 1, \quad \forall\, k, \end{aligned} \tag{11}$$

where step (a) follows from (9); (b) notes that $X \leftrightarrow V_k \leftrightarrow A_k$ forms a Markov chain and applies the data processing lemma; (c) follows because Y_k is constant by definition; and (d) follows because X is binary. From (11), it follows that we can find constants $\alpha_k \in [0, 1/2]$, for all k, such that

$$H(X|A_k, Y_k) = H(X|A_k) = h(\alpha_k \star \rho_k). \tag{12a}$$

In addition, we have

$$\begin{aligned} I(V_k; A_k|Y_k) &\overset{\text{a}}{=} I(V_k; A_k) \\ &\overset{\text{b}}{=} 1 - H(V_k|A_k) \overset{\text{c}}{\geq} 1 - h(\alpha_k), \end{aligned} \tag{12b}$$

where (a) follows because Y_k is a constant and (b) follows because $V_k \sim$ Bern(1/2). Step (c) invokes Mrs Gerber's Lemma [35, 36] (see [29, p. 19]) to upper bound $H(V_k|A_k)$ by $h(\alpha_k)$: First recall that V_k is binary, A_k is arbitrary and $X = V_k \oplus Z_k$, where $Z_k \sim \text{Bern}(\rho_k)$ is independent of V_k by definition. In addition, Z_k is independent of (V_k, A_k) since $X \leftrightarrow V_k \leftrightarrow A_k$ implies

$$0 = I(A_k; V_k \oplus Z_k | V_k) = I(A_k; Z_k | V_k)$$

and therefore $I(V_k, A_k; Z_k) = I(V_k; Z_k) + I(A_k; Z_k|V_k) = 0$. Mrs Gerber's Lemma [29, p. 19] gives

$$H(X|A_k) \geq h(h^{-1}(H(V_k|A_k)) \star \rho_k), \tag{13}$$

where $h^{-1} : [0, 1] \to [0, 1/2]$ denotes the inverse of $h(\cdot)$. The binary entropy function $h(\cdot)$ is increasing on $[0, 1/2]$, so (12a) and (13) together imply $h(\alpha_k) \geq H(V_k|A_k)$.

The above discussion shows that (12) holds for any choice of auxiliary random variables satisfying the conditions of Theorem 2. To complete the example, we need only find auxiliary random variables for which (12) holds with equality. To this end, simply let A_k be the output of a binary symmetric channel with input V_k and crossover probability α_k to obtain

$$H(X|A_k) = h(\alpha \star \rho_k) \quad \text{and} \quad I(V_k; A_k) = 1 - h(\alpha_k), \quad \forall\, k.$$

IV. Slepian-Wolf Coding over Broadcast Channels with List decoding

It is useful to consider a list-decoding extension to (1) before proving Theorems 1 and 2. In this section, suppose that there is *no* BS cooperation and the receivers employ list decoding.

A. Setup and Main Result

Let $\Omega(L) := \{\mathcal{L} \subseteq \mathcal{X}^{n_\text{s}} : |\mathcal{L}| = L\}$ denote the collection of all subsets of $\mathcal{X}^{n_\text{s}}$ with cardinality L. An $(n_\text{s}, n_\text{c}, L_1, L_2, \ldots, L_K)$ *list code* is a collection of $(K+1)$ maps $(f, g_1, g_2, \ldots, g_K)$, where

$$f : \mathcal{X}^{n_\text{s}} \longrightarrow \mathcal{W}^{n_\text{c}}$$

is the encoder at the transmitter and

$$g_k : \mathcal{U}_k^{n_\text{c}} \times \mathcal{Y}_k^{n_\text{s}} \longrightarrow \Omega(L_k)$$

is the list decoder at receiver k. Upon observing the channel output $\mathbb{U}_k$ and side information $\mathbb{Y}_k$, receiver k computes the list

$$\mathcal{L}_k := g_k(\mathbb{U}_k, \mathbb{Y}_k).$$

An error is declared at receiver k if $\mathbb{X} \notin \mathcal{L}_k$.

If (1) holds, then [8, Thm. 6] guarantees the existence of a sequence of list codes with $|\mathcal{L}_k| = 1$ and $\mathbb{P}[\mathbb{X} \notin \mathcal{L}_k] \to 0$ for all k. On the other hand: If (1) does not hold, then $|\mathcal{L}_k|$ must grow exponentially in n_s to ensure $\mathbb{P}[\mathbb{X} \notin \mathcal{L}_k] \to 0$. We are concerned with the smallest such exponent.

Definition 3: Fix the bandwidth expansion factor κ and list exponents $\boldsymbol{D} = (D_1, D_2, \ldots, D_K)$, with $D_k \geq 0$, $\forall k$. We say that the pair $(\kappa, \boldsymbol{D})$ is *achievable* if for any $\epsilon > 0$ there exists a $(n_\text{s}, n_\text{c}, L_1, L_2, \ldots, L_K)$ list code such that

$$\frac{n_\text{c}}{n_\text{s}} = \kappa, \tag{14a}$$

$$L_k \leq 2^{n_\text{s} D_k} \quad \text{and} \quad \mathbb{P}\left[\mathbb{X} \notin \mathcal{L}_k\right] \leq \epsilon, \ \forall\, k, \tag{14b}$$

where n_s and n_c are sufficiently large.

The next lemma is proved in Sections V and VI.

Lemma 3: $(\kappa, \boldsymbol{D})$ is achievable if there exists a pmf P_W on $\mathcal{W}$ such that

$$D_k > \max\left\{H(X|Y_k) - \kappa I(W; U_k),\ 0\right\}, \quad \forall\, k, \tag{15}$$

where $(W, U_1, \ldots, U_K) \sim P_W(\cdot)T(\cdot|\cdot)$. Conversely: If $(\kappa, \boldsymbol{D})$ is achievable, then there exists a pmf P_W on $\mathcal{W}$ such that (15) holds with an inequality.

Lemma 3 is quite intuitive: The best exponent of receiver k's list size can be larger, but not smaller, than the equivocation in X given Y_k minus the information conveyed over the channel.

Remark 2: Definition 3 is a lossy generalisation of the setup for (1). The standard (per-letter / average distortion) generalisation of (1) is called *"Wyner-Ziv Coding over broadcast channels"* [17], and it is a formidable open problem that includes Heegard and Berger's rate-distortion function [14, 27, 28] as well as the broadcast capacity region [29].

Remark 3: Definition 3 and Lemma 3 are related to Chia's recent list-decoding result [30, Prop. 1] for Heegard and Berger's rate-distortion problem [27]. For example, suppose that $\kappa = 1$ and we replace the memoryless BC $T(\cdot|\cdot)$ in our model with a noiseless source-coding 'index' channel, with alphabet $\{1, 2, \ldots, \lfloor 2^{n_s R_s} \rfloor\}$. In this case, the mutual information $I(W; U_k)$ transforms to the source-coding rate R_s and Lemma 3 reduces to [30, Prop. 1]

$$R_s > \max_k \big\{ H(X|Y_k) - D_k \big\}.$$

Remark 4: Lemma 3 is consistent with Tuncel's result for unique decoding (1) in the following sense. Suppose that we are interested in unique decoding and hence the all-zero list exponent vector $\boldsymbol{D} = (0, 0, \ldots, 0)$. The reverse (converse) assertion of Lemma 3 shows that $(\kappa, \boldsymbol{D})$ is achievable *only if*

$$H(X|Y_k) \leq \kappa I(W; U_k), \quad \forall k. \tag{16}$$

The forward (achievability) assertion of Lemma 3, unfortunately, does not include the all-zero list exponent. It does, however, say the following: Any arbitrarily small positive list exponent $\boldsymbol{D}$ is achievable if (16) holds.

B. Discussion: List Decoding and the Operational Duality of Theorems 1 and 2

It turns out that the following approach to BS cooperation is optimal in both modes: Use a good list code on the broadcast channel, and task BS(k) with helping receiver k determine which element of its decoded list $\mathcal{L}_k$ is equal to the source $\mathbb{X}$. This list will, with high probability, include $\mathbb{X}$ and have $|\mathcal{L}_k| \approx 2^{n_s D_k}$ elements. To resolve receiver k's uncertainty, BS(k) needs to encode its side information $\mathbb{V}_k$ at a rate R_k that is proportional to the list exponent D_k. In both modes, the smallest achievable rate R_k is fundamentally determined by Lemma 3. Theorems 1 and 2 are duals in the operational sense that changing from Mode 1 to Mode 2 (or, vice versa) does not change the underlying coding problem — it only changes BS(k)'s approach to the problem. The side information $\mathbb{V}_k$ in Mode 2 is directly correlated with the source $\mathbb{X}$, and, in this setting, it is optimal for BS(k) to use a good source code from Wyner's 'helper' source coding problem [10]. In Mode 1, on the other hand, the side information $\mathbb{V}_k$ is a scalar quantised version of the channel codeword, and it is optimal for BS(k) to use a version of Kim's 'random-hashing' for the relay channel [9]. The remainder of the paper is devoted to proving Lemma 3 and Theorems 1 and 2.

V. Proof of Lemma 3 — Converse

Fix $\epsilon > 0$, and suppose that we have a $(n_s, n_c, L_1, L_2, \ldots, L_K)$ list code such that (14) holds. As before, let $\mathbb{W} = f(\mathbb{X})$ and $\mathbb{U}_k = (U_{k,1}, U_{k,2}, \ldots, U_{k,n_c})$ denote the transmitted codeword and the channel output at receiver k.

The first step mirrors that of [8, Thm. 6]. Consider the j-th symbol W_j of $\mathbb{W} = (W_1, W_2, \ldots, W_{n_c})$, and let P_{W_j} denote its pmf. Construct a *timeshared* pmf $P_{\tilde{W}}$ on $\mathcal{W}$ by setting

$$P_{\tilde{W}}(w) := \frac{1}{n_c} \sum_{j=1}^{n_c} P_{W_j}(w), \quad w \in \mathcal{W}. \tag{17}$$

Let $(\tilde{W}, \tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_K) \sim P_{\tilde{W}}(\cdot) T(\cdot|\cdot)$. We have

$$\begin{aligned} n_c I(\tilde{W}; \tilde{U}_k) &\overset{\text{a}}{\geq} \sum_{i=1}^{n_c} I(W_i; U_{k,i}) \\ &\overset{\text{b}}{\geq} I(\mathbb{W}; \mathbb{U}_k) \\ &\overset{\text{c}}{\geq} I(\mathbb{X}; \mathbb{U}_k | \mathbb{Y}_k) \\ &\overset{\text{d}}{=} n_s H(X|Y_k) - H(\mathbb{X}|\mathbb{Y}_k, \mathbb{U}_k). \end{aligned} \tag{18}$$

Notes: (a) use Jensen's inequality [23, Thm. 2.7.4]; (b)

$$\begin{aligned} &\sum_{i=1}^{n_c} I(W_i; U_{k,i}) \\ &\quad = \sum_{i=1}^{n_c} \Big(H(U_{k,i}) - H(U_{k,i}|W_i) \Big) \\ &\quad \geq H(\mathbb{U}_k) - \sum_{i=1}^{n_c} H(U_{k,i}|W_i) \\ &\quad \overset{*}{=} H(\mathbb{U}_k) - \sum_{i=1}^{n_c} H(U_{k,i}|\mathbb{W}, U_{k,1}, U_{k,2}, \ldots, U_{k,i-1}) \\ &\quad = H(\mathbb{U}_k) - H(\mathbb{U}_k|\mathbb{W}), \end{aligned}$$

where $(*)$ follows because the broadcast channel is memoryless and therefore $U_{k,i} \leftrightarrow W_i \leftrightarrow (\mathbb{W}, U_{k,1}, U_{k,2}, \ldots, U_{k,i-1})$ forms a Markov chain; (c) $(\mathbb{X}, \mathbb{Y}_k) \leftrightarrow \mathbb{W} \leftrightarrow \mathbb{U}_k$ forms a Markov chain; and (d) the source is iid.

We now use a list-decoding version of Fano's inequality, e.g., see [30, Lem. 1] or [31, Lem. 1]:

$$\begin{aligned} H(\mathbb{X}|\mathbb{Y}_k, \mathbb{U}_k) &\leq \log |\mathcal{L}_k| + 1 \\ &\quad + (n_s \log |\mathcal{X}| - \log |\mathcal{L}_k|) \mathbb{P}\left[\mathbb{X} \notin \bigcap_{k=1}^{K} \mathcal{L}_k \right]. \end{aligned}$$

By (14b), and since for any k, $\mathbb{P}[\mathbb{X} \notin \mathcal{L}_k] \geq \mathbb{P}[\mathbb{X} \notin \cap_{k'} \mathcal{L}_{k'}]$, this inequality implies

$$H(\mathbb{X}|\mathbb{Y}_k, \mathbb{U}_k) \leq n_s \big(D_k + \varepsilon(n_s, \epsilon) \big), \tag{19}$$

where

$$\varepsilon(n_s, \epsilon) := \frac{1}{n_s} + \epsilon(1 + \log |\mathcal{X}| - D_k - \epsilon).$$

Combining (14a), (18) and (19), we have

$$\kappa I(\tilde{W}; U_k) \geq H(X|Y_k) - D_k - \varepsilon(n_s, \epsilon).$$

To complete the converse: Take any positive and vanishing sequence $\{\epsilon\} \to 0$. Consider the corresponding sequence of list codes (with increasing blocklengths n_{s} and n_{c}) and time-shared pmfs $\{P_{\tilde{W}}\}$. Since $\mathcal{W}$ is a finite alphabet, by the Bolzano-Weierstrass theorem, $\{P_{\tilde{W}}\}$ will contain a convergent subsequence with respect to the variational distance. Let $P^*_{\tilde{W}}$ denote the limit of the convergent subsequence and $\tilde{W}^* \sim P^*_{\tilde{W}}$. We then have $\kappa I(\tilde{W}^*; U_k) \geq H(X|Y_k) - D_k$ by the continuity of mutual information [32, Sec. 2.3]. ■

VI. Proof of Lemma 3 — Achievability

We restrict attention to the bandwidth matched case ($\kappa = 1$ and $n_{\mathrm{s}} = n_{\mathrm{c}} = n$), to help simplify notation and elucidate the main ideas of the achievability proof. Extending this proof to the bandwidth mismatched case is relatively easy, because we will use separate source and channel codebooks and the error probability bounds depend only on the marginal source and channel distributions.

A. Notation and Letter-Typical Sets

For any given random variable ω and set Ω, let us denote the indicator function for the event that ω falls in Ω by

$$\mathbb{1}\{\omega \in \Omega\} := \begin{cases} 1 & \text{if } \omega \in \Omega \\ 0 & \text{otherwise.} \end{cases}$$

The proof will use *letter typical* sets [33]. Consider a pair of random variables $(A, B) \sim P_{A,B}$ on $\mathcal{A} \times \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are finite alphabets. Let P_A denote the marginal pmf of A. For $\epsilon > 0$ and a positive integer n, the *typical set* of P_A is

$$\mathcal{T}^n_\epsilon(P_A) := \Big\{ \mathbb{a} \in \mathcal{A}^n : \Big| \frac{1}{n} N(a'|\mathbb{a}) - P_A(a') \Big| \leq \epsilon P_A(a'),\ \forall a' \in \mathcal{A} \Big\},$$

where $N(a'|\mathbb{a})$ represents the number of occurrences of a' in the sequence $\mathbb{a}$. The *jointly typical set* of $P_{A,B}$ is

$$\mathcal{T}^n_\epsilon(P_{A,B}) := \Big\{ (\mathbb{a}, \mathbb{b}) \in \mathcal{A}^n \times \mathcal{B}^n : \Big| \frac{1}{n} N(a', b'|\mathbb{a}, \mathbb{b}) - P_{A,B}(a', b') \Big| \leq \epsilon P_{A,B}(a', b'),\ \forall (a', b') \Big\}.$$

The *conditionally typical set* of $P_{A,B}$ given $\mathbb{b} \in \mathcal{B}^n$ is

$$\mathcal{T}^n_\epsilon(P_{A,B}|\mathbb{b}) := \big\{ \mathbb{a} \in \mathcal{A}^n : (\mathbb{a}, \mathbb{b}) \in \mathcal{T}^n_\epsilon(P_{A,B}) \big\}.$$

The proof will frequently use the property that joint typicality implies marginal typicality,

$$(\mathbb{a}, \mathbb{b}) \in \mathcal{T}^n_\epsilon(P_{A,B}) \Rightarrow \mathbb{a} \in \mathcal{T}^n_\epsilon(P_A) \text{ and } \mathbb{b} \in \mathcal{T}^n_\epsilon(P_B),$$

and the following lemmas. Let

$$\mu_A := \min_{a \in \mathrm{supp}(P_A)} P_A(a),$$

and

$$\mu_{A,B} := \min_{(a,b) \in \mathrm{supp}(P_{A,B})} P_{A,B}(a, b),$$

where supp$(\cdot)$ denotes the support set of the indicated distribution.

Lemma 4: If $\mathbb{A} := (A_1, A_2, \ldots, A_n)$ is generated iid with P_A, $0 < \epsilon \leq \mu_A$ and $\mathbb{a} \in \mathcal{T}^n_\epsilon(P_A)$, then [33, Thm. 1.1]

$$2^{-nH(A)(1+\epsilon)} \leq \mathbb{P}[\mathbb{A} = \mathbb{a}] \leq 2^{-nH(A)(1-\epsilon)}$$

and

$$1 - 2|\mathcal{A}| \exp(-n\epsilon^2 \mu_A) \leq \mathbb{P}[\mathbb{A} \in \mathcal{T}^n_\epsilon(P_A)] \leq 1.$$

Lemma 5: If $\mathbb{A} := (A_1, A_2, \ldots, A_n)$ is generated iid with P_A, $0 < \epsilon_1 < \epsilon \leq \mu_{AB}$ and $\mathbb{b} \in \mathcal{T}^n_{\epsilon_1}(P_B)$, then [33, Thm. 1.3]

$$\mathbb{P}\big[\mathbb{A} \in \mathcal{T}^n_\epsilon(P_{A,B}|\mathbb{b})\big] \leq 2^{-n(I(A;B) - 2\epsilon H(A))}$$

and

$$\mathbb{P}\big[\mathbb{A} \in \mathcal{T}^n_\epsilon(P_{A,B}|\mathbb{b})\big] \geq (1 - \zeta_n) 2^{-n(I(A;B) + 2\epsilon H(A))},$$

where[8]

$$\zeta_n := 2|\mathcal{A}||\mathcal{B}| \exp\left(-2n(1-\epsilon_1) \frac{(\epsilon - \epsilon_1)^2}{1 + \epsilon_1} \mu^2_{AB} \right).$$

B. Distributions and Typicality Constants

Pick any pmf P_W on $\mathcal{W}$. Let

$$X \sim P_X, \quad (X, Y_k) \sim P_{X,Y_k} \quad \text{and} \quad (W, U_k) \sim P_{W,U_k}$$

denote the pmfs of the indicated variables. Fix any arbitrarily small constants $\epsilon, \epsilon_1, \delta$ and δ_1 satisfying

$$0 < \delta_1 < \delta < \min_k \mu_{W,U_k} \quad \text{and} \quad 0 < \epsilon_1 < \epsilon < \min_k \mu_{X,Y_k}, \tag{20a}$$

with

$$\epsilon < \frac{\min_k \mu_{W,U_k}}{2H(X) \ln 2} \delta^2. \tag{20b}$$

C. Code Construction and Encoding

The encoder mirrors that of [8, Thm. 6]. Randomly generate a source codebook $\mathcal{C}_X$ with

$$M = \lfloor 2^{nH(X)(1+\epsilon)} \rfloor \tag{21}$$

codewords, each of length n, by selecting symbols from $\mathcal{X}$ in an iid fashion using P_X:

$$\mathcal{C}_X := \Big\{ \mathbb{X}(m) = \big(X_1(m), X_2(m), \ldots, X_n(m)\big) \Big\}_{m=1}^{M}.$$

In the same way, generate a channel codebook $\mathcal{C}_W$ with M codewords of length n using P_W:

$$\mathcal{C}_W := \Big\{ \mathbb{W}(m) = \big(W_1(m), W_2(m), \ldots, W_n(m)\big) \Big\}_{m=1}^{M}.$$

Upon observing the source $\mathbb{X}$, the transmitter searches through the source codebook $\mathcal{C}_X$ for the smallest index m such that $\mathbb{X} = \mathbb{X}(m)$. If successful, the transmitter sends the corresponding channel codeword $\mathbb{W}(m)$; and, if unsuccessful, it sends $\mathbb{W}$ generated iid $\sim P_W$.

[8] Here we use I. Sason's correction to [33, Thm. 1.3], see [34, pp. 140–154].

D. List Decoding at Receiver k

The decoder (and error analysis) differ from [8, Thm. 6]. Upon observing the channel output $\mathbb{U}_k$ and side information $\mathbb{Y}_k$, receiver k outputs the list

$$\mathcal{L}_k := \Big\{\mathbb{X}(m) \in \mathcal{C}_X : \big(\mathbb{X}(m), \mathbb{Y}_k\big) \in \mathcal{T}_\epsilon^n(P_{X,Y_k}) \text{ and } \big(\mathbb{W}(m), \mathbb{U}_k\big) \in \mathcal{T}_\delta^n(P_{W,U_k})\Big\}. \quad (20)$$

An error is declared at receiver k if the source is not in the list $\mathbb{X} \notin \mathcal{L}_k$ or the list is too large

$$|\mathcal{L}_k| > 2^{nD_k}.$$

E. Error Analysis: Decoding error event $\mathcal{E}$

Denote the event of an error at any receiver by

$$\mathcal{E} := \bigcup_{k=1}^{K} \Big(\big\{\mathbb{X} \notin \mathcal{L}_k\big\} \cup \big\{|\mathcal{L}_k| > 2^{nD_k}\big\}\Big). \quad (21)$$

By the union bound,

$$\mathbb{P}[\mathcal{E}] \le \sum_{k=1}^{K} \Big(\mathbb{P}\big[\mathbb{X} \notin \mathcal{L}_k\big] + \mathbb{P}\big[|\mathcal{L}_k| > 2^{nD_k}\big]\Big). \quad (22)$$

In the following subsections, we show that the average error probability $\mathbb{P}[\mathcal{E}]$ satisfies

$$\mathbb{P}[\mathcal{E}] \le b\ 2^{-an}, \quad (23)$$

for some finite $a, b > 0$, whenever ϵ and δ satisfy (20) and

$$D_k > \max\big\{H(X|Y_k) - I(W;U_k), 0\big\}, \quad \forall\ k.$$

Therefore, for any $\epsilon^* > 0$ there exists an $(n, L_1, L_2, \dots, L_K)$ list code such that $\mathbb{P}[\mathbb{X} \notin \mathcal{L}_k] \le \epsilon^*$ and $|\mathcal{L}_k| \le 2^{nD_k}$ for all k.

The remainder of this section is devoted to proving (23). The derivation of the bound is a little tedious and the reader needs only (23) to proceed to the achievability proofs of Theorems 1 and 2 in Sections VIII and X respectively.

F. Error Analysis: Probability $\mathbb{X}$ is not in the source codebook

The probability that the source is not in the source codebook $\mathbb{P}[\mathbb{X} \notin \mathcal{C}_X]$ is bounded by

$$\mathbb{P}[\mathbb{X} \notin \mathcal{C}_X] \le b_1\ 2^{-a_1 n}, \quad (24)$$

where

$$a_1 := \min\big\{\epsilon_1^2 \mu_X, (H(X) \cdot (\epsilon - \epsilon_1))\big\} / \ln 2$$

and $b_1 := 2|\mathcal{X}| + 1$ are both positive by (20).

The steps leading to (24) are

$$\begin{aligned}
&\mathbb{P}[\mathbb{X} \notin \mathcal{C}_X] \\
&\le \mathbb{P}\big[\mathbb{X} \notin \mathcal{T}_{\epsilon_1}^n(P_X)\big] + \mathbb{P}\big[\mathbb{X} \notin \mathcal{C}_X \big| \mathbb{X} \in \mathcal{T}_{\epsilon_1}^n(P_X)\big] \\
&\overset{\text{a}}{\le} 2|\mathcal{X}|e^{-n\epsilon_1^2\mu_X} + \mathbb{P}\left[\bigcap_{m=1}^{M} \{\mathbb{X}(m) \ne \mathbb{X}\} \middle| \mathbb{X} \in \mathcal{T}_{\epsilon_1}^n(P_X)\right] \\
&\overset{\text{b}}{=} 2|\mathcal{X}|e^{-n\epsilon_1^2\mu_X} + \prod_{m=1}^{M} \Big(1 - \mathbb{P}\big[\mathbb{X}(m) = \mathbb{X} \big| \mathbb{X} \in \mathcal{T}_{\epsilon_1}^n(P_X)\big]\Big) \\
&\overset{\text{c}}{\le} 2|\mathcal{X}|e^{-n\epsilon_1^2\mu_X} + \Big(1 - 2^{-nH(X)(1+\epsilon_1)}\Big)^M \\
&\overset{\text{d}}{\le} 2|\mathcal{X}|e^{-n\epsilon_1^2\mu_X} + \exp\big(-M2^{-nH(X)(1+\epsilon_1)}\big) \\
&\overset{\text{e}}{\le} 2|\mathcal{X}|e^{-n\epsilon_1^2\mu_X} + \exp(-2^{nH(X)(\epsilon-\epsilon_1)}).
\end{aligned} \quad (25)$$

Notes:

a. apply Lemma 4;
b. the codewords in $\mathcal{C}_X$ are independent;

$$\begin{aligned}
\mathbb{P}\big[\mathcal{S}_1\big] &\overset{\text{a}}{=} \mathbb{P}\Big[\{\mathbb{X} \notin \mathcal{L}_k\} \cap \{(\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}_{\epsilon_1}^n(P_{XY_k})\} \cap \Big(\bigcup_{m=1}^{M} \{\mathbb{X} \ne \mathbb{X}(m'), \forall\ m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\}\Big)\Big] \\
&\overset{\text{b}}{\le} \sum_{m=1}^{M} \mathbb{P}\Big[\{\mathbb{X} \notin \mathcal{L}_k\} \cap \{(\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}_{\epsilon_1}^n(P_{XY_k})\} \cap \{\mathbb{X} \ne \mathbb{X}(m'),\ \forall\ m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\}\Big] \\
&\overset{\text{c}}{\le} \sum_{m=1}^{M} \mathbb{P}\Big[\{\mathbb{X} \ne \mathbb{X}(m'),\ \forall\ m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \Big| (\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}_{\epsilon_1}^n(P_{X,Y_k})\Big] \\
&\qquad \mathbb{P}\Big[\mathbb{X} \notin \mathcal{L}_k \Big| \{(\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}_{\epsilon_1}^n(P_{XY_k})\} \cap \{\mathbb{X} \ne \mathbb{X}(m'),\ \forall\ m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\}\Big] \\
&\overset{\text{d}}{\le} \sum_{m=1}^{M} \mathbb{P}\Big[\mathbb{X} = \mathbb{X}(m) \Big| \mathbb{X} \in \mathcal{T}_{\epsilon_1}^n(P_X)\Big]\ \mathbb{P}\Big[(\mathbb{W}, \mathbb{U}_k) \notin \mathcal{T}_\delta^n(P_W, U_k) \Big| \{(\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}_{\epsilon_1}^n(P_{XY_k})\} \\
&\qquad \cap \{\mathbb{X} \ne \mathbb{X}(m'), \forall\ m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \cap \{\mathbb{W} = \mathbb{W}(m)\}\Big] \\
&\overset{\text{e}}{=} \sum_{m=1}^{M} \mathbb{P}\Big[\mathbb{X} = \mathbb{X}(m) \Big| \mathbb{X} \in \mathcal{T}_{\epsilon_1}^n(P_X)\Big]\ \mathbb{P}\Big[(\mathbb{W}, \mathbb{U}_k) \notin \mathcal{T}_\delta^n(P_{W,U_k})\Big] \\
&\overset{\text{f}}{\le} M\ 2|\mathcal{W}||\mathcal{U}_k| 2^{-nH(X)(1-\epsilon_1)}\ 2^{-n\frac{\delta^2 \mu_{W,U_k}}{\ln 2}} \\
&\overset{\text{g}}{\le} 2|\mathcal{W}||\mathcal{U}_k|\ 2^{+2\epsilon nH(X)} 2^{-n\frac{\delta^2 \mu_{W,U_k}}{\ln 2}}.
\end{aligned} \quad (28)$$

c. use Lemma 4 with $\mathbb{X}(m)$ iid $\sim P_X$;
d. use the inequality

$$(1-c)^M \le e^{-cM}, \quad \forall\, M \ge 1,\, c \in [0,1]; \text{ and}$$

e. bound M via (21).

The bound in (24) follows since $H(X)(\epsilon - \epsilon_1) > 0$ from (20).

G. Error Analysis: Probability $\mathbb{X}$ is not in receiver k's list $\mathcal{L}_k$

Consider the probability that the source $\mathbb{X}$ is not in receiver k's list $\mathcal{L}_k$. We have

$$\mathbb{P}\big[\mathbb{X} \notin \mathcal{L}_k\big] \le \mathbb{P}[\mathbb{X} \notin \mathcal{C}_X] + \mathbb{P}[(\mathbb{X}, \mathbb{Y}_k) \notin \mathcal{T}_{\epsilon_1}(P_{X,Y_k})] + \mathbb{P}\big[\mathcal{S}_1\big], \quad (26)$$

where

$$\mathcal{S}_1 := \{\mathbb{X} \notin \mathcal{L}_k\} \cap \{(\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}^n_{\epsilon_1}(P_{XY_k})\} \cap \{\mathbb{X} \in \mathcal{C}_X\}.$$

The probability $\mathbb{P}[\mathcal{S}_1]$ is bounded from above by

$$\mathbb{P}\big[\mathcal{S}_1\big] \le 2|\mathcal{W}||\mathcal{U}_k|\; 2^{-a_2 n} \quad (27)$$

where

$$a_2 := \delta^2 \mu_{W,U_k} / \ln 2 - 2\epsilon_1 H(X)$$

is positive by (20). The steps leading to (27) are described above in (28). Notes for (28):

a. expand the event that the source $\mathbb{X}$ appears in $\mathcal{C}_X$;
b. union bound;
c. Bayes' law and $\mathbb{P}[(\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}^n_\epsilon(P_{XY_k})] \le 1$;
d. conditioned on $(\mathbb{X}, \mathbb{Y}_k)$ typical, $\mathbb{X} = \mathbb{X}(m)$ and $\mathbb{W} = \mathbb{W}(m)$, the error $\mathbb{X} \notin \mathcal{L}_k$ occurs if and only if $(\mathbb{W}, \mathbb{U}_k)$ are not jointly typical;
e. the source and channel codebooks are independent, and all channel codewords are constructed in the same way;
f. Lemma 4; and
g. bound the codebook cardinality M as in (21).

Combining (26) and (27) with (24) and Lemma 4, we have

$$\mathbb{P}[\mathbb{X} \notin \mathcal{L}_k] \le b_3\; 2^{-a_3 n}, \quad (29)$$

for some finite $a_3, b_3 > 0$.

H. Error Analysis: Probability receiver k's list $\mathcal{L}_k$ is too large

Now consider the probability that *the size of list $\mathcal{L}_k$ is too large*. We start with

$$\mathbb{P}\big[|\mathcal{L}_k| > 2^{nD_k}\big] \le \mathbb{P}\big[\mathbb{X} \notin \mathcal{C}_X\big] + \mathbb{P}\big[(\mathbb{X}, \mathbb{Y}_k) \notin \mathcal{T}^n_{\epsilon_1}\big] + \mathbb{P}\big[(\mathbb{W}, \mathbb{U}_k) \notin \mathcal{T}^n_{\delta_1}\big] + \mathbb{P}\big[|\mathcal{L}_k| > 2^{nD_k}\big|\mathcal{S}_2\big], \quad (30)$$

where

$$\mathcal{S}_2 := \big\{\mathbb{X} \in \mathcal{C}_X\big\} \cap \big\{(\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}^n_{\epsilon_1}\big\} \cap \big\{(\mathbb{W}, \mathbb{U}_k) \in \mathcal{T}^n_{\delta_1}\big\},$$

and $\mathcal{T}^n_{\epsilon_1}(P_{X,Y_k})$ and $\mathcal{T}^n_{\delta_1}(P_{W,U_k})$ have been abbreviated by $\mathcal{T}^n_{\epsilon_1}$ and $\mathcal{T}^n_{\delta_1}$ respectively. Apply Markov's inequality to the rightmost probability in (30) to get

$$\mathbb{P}\big[|\mathcal{L}_k| > 2^{nD_k}\big|\mathcal{S}_2\big] \le 2^{-nD_k}\; \mathbb{E}\big[|\mathcal{L}_k|\big|\mathcal{S}_2\big], \quad (31)$$

where the expectation is understood to be

$$\mathbb{E}\big[|\mathcal{L}_k|\big|\mathcal{S}_2\big] := \sum_l l \cdot \mathbb{P}\big[|\mathcal{L}_k| = l\big|\mathcal{S}_2\big].$$

We now expand the above expectation over $\mathbb{X} \in \mathcal{C}_X$ (the M possible encodings of $\mathbb{X}$) to get

$$\begin{aligned}&\mathbb{E}\big[|\mathcal{L}_k|\big|\mathcal{S}_2\big] \\ &= \sum_{m=1}^{M} \mathbb{E}\Big[|\mathcal{L}_k|\Big|\mathcal{S}_2 \cap \{\mathbb{X} \ne \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\}\Big] \\ &\quad \cdot \mathbb{P}\Big[\{\mathbb{X} \ne \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\}\Big|\mathcal{S}_2\Big]. \quad (32)\end{aligned}$$

Consider the expectation on the right hand side of (32). Let

$$\mathcal{S}_{2,m} := \mathcal{S}_2 \cap \{\mathbb{X} \ne \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\}.$$

We have

$$\begin{aligned}&\mathbb{E}\big[|\mathcal{L}_k|\big|\mathcal{S}_{2,m}\big] \\ &= \mathbb{E}\left[\sum_{\tilde{m}=1}^{M} \mathbb{1}\{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}^n_\epsilon\}\mathbb{1}\{(\mathbb{W}(\tilde{m}), \mathbb{U}_k) \in \mathcal{T}^n_\delta\}\middle|\mathcal{S}_{2,m}\right] \\ &= \sum_{\tilde{m}=1}^{M} \mathbb{P}\Big[\{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}^n_\epsilon\} \cap \{(\mathbb{W}(\tilde{m}), \mathbb{U}_k) \in \mathcal{T}^n_\delta\}\Big|\mathcal{S}_{2,m}\Big] \\ &= \sum_{\tilde{m}=1}^{M} \mathbb{P}\big[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}^n_\epsilon\big|\mathcal{S}_{2,m}\big] \\ &\quad \cdot \mathbb{P}\big[(\mathbb{W}(\tilde{m}), \mathbb{U}_k) \in \mathcal{T}^n_\delta\big|\mathcal{S}_{2,m} \cap \{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}^n_\epsilon\}\big], \quad (33)\end{aligned}$$

where $\mathcal{T}^n_\epsilon(P_{X,Y_k})$ and $\mathcal{T}^n_\delta(P_{W,U_k})$ have been abbreviated by $\mathcal{T}^n_\epsilon$ and $\mathcal{T}^n_\delta$ respectively.

The event $\mathcal{S}_{2,m}$ implies that the source $\mathbb{X}$ is equal to the m-th source codeword $\mathbb{X}(m)$ and $\mathbb{W} = \mathbb{W}(m)$ is sent over the channel. We now bound the two probabilities on the right hand side of (33) separately for each of the three cases $1 \le \tilde{m} < m$, $\tilde{m} = m$ and $m < \tilde{m} \le M$.

Case 1 ($1 \le \tilde{m} < m$)*:* The first probability in (33) is bounded by

$$\begin{aligned}&\mathbb{P}\big[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}^n_\epsilon\big|\mathcal{S}_{2,m}\big] \\ &\overset{\text{a}}{=} \mathbb{P}\big[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}^n_\epsilon\big|\{(\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}^n_{\epsilon_1}\} \cap \{\mathbb{X} \ne \mathbb{X}(\tilde{m})\}\big] \\ &\overset{\text{b}}{\le} \frac{\mathbb{P}\big[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}^n_\epsilon(P_{X,Y_k})\big|\mathbb{Y}_k \in \mathcal{T}^n_{\epsilon_1}(P_{Y_k})\big]}{\mathbb{P}\big[\mathbb{X} \ne \mathbb{X}(\tilde{m})\big|\mathbb{X} \in \mathcal{T}^n_{\epsilon_1}(P_X)\big]} \\ &\overset{\text{c}}{\le} \alpha_n\; 2^{-n(I(X;Y_k) - 2\epsilon H(X))}. \quad (34)\end{aligned}$$

Notes:

a. codewords and codebook are generated independently;
b. Bayes' law and the trivial bound $\mathbb{P}\big[\mathbb{X} \ne \mathbb{X}(\tilde{m})\big|\{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}^n_\epsilon(P_{X,Y_k})\} \cap \{(\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}^n_{\epsilon_1}\}\big] \le 1$; and
c. apply Lemmas 4 and 5 respectively to the denominator and numerator in step (b) and set

$$\alpha_n := \frac{\exp(2^{nH(X)(1-\epsilon_1)})}{\exp(2^{nH(X)(1-\epsilon_1)}) - 1}. \quad (35)$$

Similarly, by Lemma 5, the rightmost probability in (33) is bounded by

$$\mathbb{P}\big[(\mathbb{W}(\tilde{m}), \mathbb{U}_k) \in \mathcal{T}^n_\delta\big|\mathcal{S}_{2,m} \cap \{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}^n_\epsilon\}\big]$$

$$\leq 2^{-n(I(W;U_k)-2\delta H(W))}. \quad (36)$$

Case 2 ($\tilde{m} = m$): Bound both probabilities in (33) by one.

Case 3 ($m < \tilde{m} \leq M$): Conditioned on $\mathcal{S}_{2,m}$, the encoder has only considered the codewords $\mathbb{X}(1), \ldots, \mathbb{X}(m)$. Thus, even conditional on $\mathcal{S}_{2,m}$, the codewords thereafter $\mathbb{X}(m+1), \ldots, \mathbb{X}(M)$ are independent iid $\sim P_X$ sequences. From Lemma 5,

$$\mathbb{P}\big[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}_\epsilon^n \big| \mathcal{S}_{2,m}\big] \leq 2^{-n(I(X;Y_k)-2\epsilon H(X))}, \quad (37)$$

Similarly,

$$\mathbb{P}\big[(\mathbb{W}(\tilde{m}), \mathbb{U}_k) \in \mathcal{T}_\delta^n \big| \mathcal{S}_{2,m} \cap \{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}_\epsilon^n\}\big] \leq 2^{-n(I(W;U_k)-2\delta H(W))}. \quad (38)$$

Collectively, (21) and (33) to (38) imply

$$\mathbb{E}\big[|\mathcal{L}_k| \big| \mathcal{S}_{2,m}\big] \leq 1 + \alpha_n 2^{n(H(X|Y_k)-I(W;U_k))} 2^{n(3\epsilon H(X)+2\delta H(W))}. \quad (39)$$

Combine (30), (31) and (39) to get

$$\begin{aligned}\mathbb{P}\big[|\mathcal{L}_k| > 2^{nD_k}\big] &\leq \mathbb{P}[\mathbb{X} \notin \mathcal{C}_X] + \mathbb{P}[(\mathbb{X}, \mathbb{Y}_k) \notin \mathcal{T}_\epsilon^n] + \mathbb{P}[(\mathbb{W}, \mathbb{U}_k) \notin \mathcal{T}_\delta^n] \\ &\quad + \alpha_n 2^{-n(D_k - H(X|Y_k) + I(W;U_k))} 2^{n(3\epsilon H(X)+2\delta H(W))} \\ &\quad + 2^{-nD_k}. \end{aligned} \quad (40)$$

Lemma 4 and (24) imply

$$\mathbb{P}\big[|\mathcal{L}_k| > 2^{nD_k}\big] \leq b_4 \ 2^{-a_4 n}, \quad (41)$$

for some finite $a_4, b_4 > 0$ whenever

$$D_k > \max\{H(X|Y_k) - I(W;U_k), 0\} + 3\epsilon H(X) + 2\delta H(W)$$

and $\epsilon, \epsilon_1, \delta$ and δ_1 satisfy (20). The result follows because ϵ and δ can be chosen arbitrarily small and $H(X)$ and $H(W)$ are finite. ■

VII. Proof of Theorem 1 — Converse

Fix $\epsilon > 0$. Consider any $(n_\text{s}, n_\text{c}, R_1, R_2, \ldots, R_K)$-code with $\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X}] \leq \epsilon$ for all k. Recall the timeshared pmf $P_{\tilde{W}}$ on $\mathcal{W}$, defined in (17). Let

$$(\tilde{W}, \tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_K) \sim P_{\tilde{W}}(\cdot) \ T(\cdot|\cdot)$$

and $\tilde{V}_k = \phi_k(\tilde{W})$. Mirroring the steps of Section V:

$$\begin{aligned} n_\text{c} I(\tilde{W}; \tilde{U}_k, \tilde{V}_k) &\geq I(\mathbb{W}; \mathbb{U}_k, \mathbb{V}_k) \overset{\text{a}}{\geq} I(\mathbb{W}; \mathbb{U}_k, M_k) \\ &\overset{\text{b}}{\geq} I(\mathbb{X}; \mathbb{U}_k, M_k | \mathbb{Y}_k) \\ &= n_\text{s} H(X|Y_k) - H(\mathbb{X}|\mathbb{Y}_k, \mathbb{U}_k, M_k), \end{aligned} \quad (42)$$

where (a) and (b) use $M_k \leftrightarrow (\mathbb{U}_k, \mathbb{V}_k) \leftrightarrow \mathbb{W}$ and $(\mathbb{X}, \mathbb{Y}_k) \leftrightarrow \mathbb{W} \leftrightarrow (\mathbb{U}_k, M_k)$. Similarly,

$$\begin{aligned} n_\text{c} I(\tilde{W}; \tilde{U}_k) + n_\text{s} R_k &\geq I(\mathbb{W}; \mathbb{U}_k) + H(M_k|\mathbb{U}_k) \\ &\geq I(\mathbb{W}; \mathbb{U}_k, M_k) \\ &\geq n_\text{s} H(X|Y_k) - H(\mathbb{X}|\mathbb{U}_k, \mathbb{Y}_k, M_k). \end{aligned} \quad (43)$$

After applying Fano's inequality [23, Thm. 2.10.1] to $H(\mathbb{X}|\mathbb{U}_k, \mathbb{Y}_k, M_k)$ in (42) and (43), the converse follows in the same way as the closing of Section V. ■

VIII. Proof of Theorem 1 — Achievability

We now present an achievability proof for the bandwidth matched case, where $\kappa = 1$ and $n_\text{s} = n_\text{c} = n$. The mismatched bandwidth case follows by similar arguments. Our approach to the proof combines the list decoder of Section VI with hash-and-forward coding at the helpers.

A. Code Construction

Fix a pmf P_W on $\mathcal{W}$ and let us assume that for all k

$$H(X|Y_k) < I(W;U_k) + \min\big\{R_k, I(W;V_k|U_k)\big\} \quad (44)$$

and

$$H(X|Y_k) \geq I(W;U_k). \quad (45)$$

The assumption above (44) matches that in Theorem 1, and (45) ensures that every receiver requires a positive helper rate to reliably decode the source $\mathbb{X}$. At the end of the proof, we will relax (45) to include situations where some receivers don't require a positive helper rate, i.e., $H(X|Y_k) < I(W;U_k)$ for some k.

Generate a random list code, as described in Section VI, with the parameters described above, and let $\mathcal{C}_X$ and $\mathcal{C}_W$ denote the source and channel codebooks respectively. Fix $\epsilon, \epsilon_1, \delta$ and δ_1 arbitrarily small, but always satisfying (20). For each receiver k, choose any list exponent D_k satisfying

$$H(X|Y_k) - I(W;U_k) < D_k < I(W;V_k|U_k) - 4\delta H(W), \quad (46)$$

and set the helper rate to be

$$R_k = D_k + \epsilon_\text{h} \quad (47)$$

for some arbitrarily small $\epsilon_\text{h} > 0$. Notice that it is always possible to choose D_k in (46) because (44) and (45) imply $H(X|Y_k) < I(W;U_k,V_k)$ and $I(W;V_k|U_k) > 0$; we can choose δ arbitrarily small; and $H(W)$ is finite.

We construct a random codebook for helper BS(k): The codebook is generated by applying the map ϕ_k symbol-by-symbol to each codeword $\mathbb{W}(m) \in \mathcal{C}_W$; that is,

$$\mathcal{C}_{V_k} := \bigcup_{m=1}^{M} \big\{\phi_k(\mathbb{W}(m))\big\},$$

where

$$\phi_k(\mathbb{W}(m)) = \big(\phi_k(W_1(m)), \phi_k(W_2(m)), \ldots, \phi_k(W_n(m))\big)$$

is a slight abuse of notation.

Uniformly at random place each codeword in $\mathcal{C}_{V_k}$ into one of $\lceil 2^{nR_k} \rceil$ bins. Uniquely label each bin with an index from the set $\{1, 2, \ldots, \lceil 2^{nR_k} \rceil\}$, and let $f_k(\mathbb{v})$ denote the bin index of the codeword $\mathbb{v} \in \mathcal{C}_{V_k}$. Denote the set of all codewords in the b-th bin by $\mathcal{B}_k(b) := \big\{\mathbb{v} \in \mathcal{C}_{V_k} : f_k(\mathbb{v}) = b\big\}$ for $b \in \{1, 2, \ldots, \lceil 2^{nR_k} \rceil\}$.

B. Encoding and Decoding

The list encoder and decoders operate as before, see Sections VI-C and VI-D. Helper BS(k) looks for $\mathbb{V}_k = \phi_k(\mathbb{W})$ in $\mathcal{C}_{V_k}$ and, if successful, sends the bin index $B = f_k(\mathbb{V}_k)$ to receiver k. If unsuccessful, the helper sends an index with an independent and uniform distribution.

The list decoder at receiver k outputs $\mathcal{L}_k$, see (20), from which the receiver computes a new list of V_k-codewords:

$$\mathcal{L}_k^* := \Big\{ \mathbb{v} \in \mathcal{C}_{V_k} : \exists\ \mathbb{X}(m) \in \mathcal{L}_k \text{ with } \mathbb{v} = \phi_k\big(\mathbb{W}(m)\big) \Big\}.$$

If there is a unique codeword $\mathbb{v}'$ in the intersection of the list $\mathcal{L}_k^*$ and the bin $\mathcal{B}_k(B)$, then receiver k sets $\hat{\mathbb{V}}_k := \mathbb{v}'$. Otherwise, receiver k generates $\hat{\mathbb{V}}_k$ iid $\sim P_{V_k}$.

Finally, receiver k looks for a unique source codeword $\mathbb{X}(m') \in \mathcal{L}_k$ such that $(\mathbb{W}(m'), \mathbb{U}_k, \hat{\mathbb{V}}_k) \in \mathcal{T}_\epsilon(P_{W,U_k,V_k})$. If successful, the receiver outputs $\hat{\mathbb{X}}_k := \mathbb{X}(m')$; otherwise, it selects a codeword $\mathbb{X}(m)$ uniformly at random from $\mathcal{C}_X$.

C. Error Analysis

To bound the probability that receiver k decodes in error, $\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X}]$, it is useful to start with

$$\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X}] \leq \mathbb{P}[\hat{\mathbb{V}}_k \neq \mathbb{V}_k] + \mathbb{P}[\{\hat{\mathbb{V}}_k = \mathbb{V}_k\} \cap \{\hat{\mathbb{X}}_k \neq \mathbb{X}\}]. \quad (48)$$

We may bound the probability that receiver k incorrectly decodes $\mathbb{V}_k$ by conditioning on the list error event $\mathcal{E}$, defined in (21), and the encoder error $\{\mathbb{X} \notin \mathcal{C}_X\}$ as follows:

$$\begin{aligned}\mathbb{P}[\hat{\mathbb{V}}_k \neq \mathbb{V}_k] &\leq \mathbb{P}[\mathcal{E}] + \mathbb{P}[\mathbb{X} \notin \mathcal{C}_X] \\ &\quad + \mathbb{P}\big[\hat{\mathbb{V}}_k \neq \mathbb{V}_k \big| \mathcal{E}^c \cap \{\mathbb{X} \in \mathcal{C}_X\}\big]. \quad (49)\end{aligned}$$

Upper bounds for $\mathbb{P}[\mathcal{E}]$ and $\mathbb{P}[\mathbb{X} \notin \mathcal{C}_X]$ are given in (23) and (24) respectively. Let us now rewrite the conditional probability in (49) using the law of total probability as

$$\begin{aligned}&\mathbb{P}\Big[\hat{\mathbb{V}}_k \neq \mathbb{V}_k \Big| \mathcal{E}^c \cap \{\mathbb{X} \in \mathcal{C}_X\}\Big] \\ &= \sum_{b=1}^{\lceil 2^{nR_k} \rceil} \mathbb{P}\Big[\{f_k(\mathbb{V}_k) = b\} \cap \{\hat{\mathbb{V}}_k \neq \mathbb{V}_k\} \Big| \mathcal{E}^c \cap \{\mathbb{X} \in \mathcal{C}_X\}\Big] \\ &= \sum_{b=1}^{\lceil 2^{nR_k} \rceil} \mathbb{P}\Big[f_k(\mathbb{V}_k) = b \Big| \mathcal{E}^c \cap \{\mathbb{X} \in \mathcal{C}_X\}\Big] \\ &\qquad \cdot \mathbb{P}\Big[\hat{\mathbb{V}}_k \neq \mathbb{V}_k \Big| \{f_k(\mathbb{V}_k) = b\} \cap \mathcal{E}^c \cap \{\mathbb{X} \in \mathcal{C}_X\}\Big]. \quad (50)\end{aligned}$$

We now fix a bin b and derive

$$\begin{aligned}&\mathbb{P}\Big[\hat{\mathbb{V}}_k \neq \mathbb{V}_k \Big| \{f_k(\mathbb{V}_k) = b\} \cap \mathcal{E}^c \cap \{\mathbb{X} \in \mathcal{C}_X\}\Big] \\ &\overset{\text{a}}{=} \mathbb{P}\Big[\bigcup_{\substack{\mathbb{v} \in \mathcal{L}_k^* \\ \mathbb{v} \neq \phi_k(\mathbb{W})}} \{f_k(\mathbb{v}) = b\} \Big| \{f_k(\mathbb{V}_k) = b\} \cap \mathcal{E}^c \cap \{\mathbb{X} \in \mathcal{C}_X\}\Big]\end{aligned}$$

$$\begin{aligned}\mathbb{P}[\hat{\mathbb{X}} \neq \mathbb{X} | \mathcal{S}_3] &\overset{\text{a}}{=} \sum_{l:\ |l| \leq 2^{nD_k}} \mathbb{P}\big[\mathcal{L}_k = l \big| \mathcal{S}_3\big]\ \mathbb{P}\big[\hat{\mathbb{X}}_k \neq \mathbb{X} \big| \mathcal{S}_3 \cap \{\mathcal{L}_k = l\}\big] \\ &\overset{\text{b}}{=} \sum_{l:\ |l| \leq 2^{nD_k}} \mathbb{P}\big[\mathcal{L}_k = l \big| \mathcal{S}_3\big] \left(\sum_{m=1}^{M} \mathbb{P}\big[\{\mathbb{X} \neq \mathbb{X}(m'),\ \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \big| \right. \\ &\qquad\qquad \left. \mathcal{S}_3 \cap \{\mathcal{L}_k = l\}\big]\ \mathbb{P}\big[\hat{\mathbb{X}}_k \neq \mathbb{X} \big| \mathcal{S}_{3,m}\big] \right) \\ &\overset{\text{c}}{=} \sum_{l:\ |l| \leq 2^{nD_k}} \mathbb{P}\big[\mathcal{L}_k = l \big| \mathcal{S}_3\big] \left(\sum_{m=1}^{M} \mathbb{P}\big[\{\mathbb{X} \neq \mathbb{X}(m'),\ \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \big| \right. \\ &\qquad\qquad \left. \mathcal{S}_3 \cap \{\mathcal{L}_k = l\}\big]\ \mathbb{P}\Big[\bigcup_{\substack{\tilde{m} \in l \\ \tilde{m} \neq m}} \{(\mathbb{W}(\tilde{m}), \mathbb{U}_k, \hat{\mathbb{V}}_k) \in \mathcal{T}_\delta\} \Big| \mathcal{S}_{3,m}\Big] \right) \\ &\overset{\text{d}}{\leq} \sum_{l:\ |l| \leq 2^{nD_k}} \mathbb{P}\big[\mathcal{L}_k = l \big| \mathcal{S}_3\big] \left(\sum_{m=1}^{M} \mathbb{P}\big[\{\mathbb{X} \neq \mathbb{X}(m'),\ \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \big| \right. \\ &\qquad\qquad \left. \mathcal{S}_3 \cap \{\mathcal{L}_k = l\}\big] \sum_{\substack{\tilde{m} \in l \\ \tilde{m} \neq m}} \mathbb{P}\big[(\mathbb{W}(\tilde{m}), \mathbb{U}_k, \hat{\mathbb{V}}_k) \in \mathcal{T}_\delta \big| \mathcal{S}_{3,m}\big] \right) \\ &\overset{\text{e}}{\leq} \sum_{l:\ |l| \leq 2^{nD_k}} \mathbb{P}\big[\mathcal{L}_k = l \big| \mathcal{S}_3\big] \left(\sum_{m=1}^{M} \mathbb{P}\big[\{\mathbb{X} \neq \mathbb{X}(m'),\ \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \big| \right. \\ &\qquad\qquad \left. \mathcal{S}_3 \cap \{\mathcal{L}_k = l\}\big] \sum_{\substack{\tilde{m} \in l \\ \tilde{m} \neq m}} \gamma_n\ 2^{-n(I(W;V_k|U_k) - 4\delta H(W))} \right) \quad (57) \\ &\overset{\text{f}}{\leq} \gamma_n\ 2^{nD_k} 2^{-n(I(W;V_k|U_k) - 4\delta H(W))}\end{aligned}$$

$$\begin{aligned}
&\mathbb{P}\big[(\mathbb{W}(\tilde{m}), \mathbb{U}_k, \hat{\mathbb{V}}_k) \in \mathcal{T}_\delta \big| \mathcal{S}_{3,m}\big] \\
&\quad \overset{\text{e.1}}{=} \mathbb{P}\Big[(\mathbb{W}(\tilde{m}), \mathbb{U}_k, \mathbb{V}_k) \in \mathcal{T}_\delta \Big| \{\mathbb{W} = \mathbb{W}(m)\} \cap \{(\mathbb{W}(\tilde{m}), \mathbb{U}_k) \in \mathcal{T}_\delta\} \cap \{(\mathbb{W}, \mathbb{U}_k, \mathbb{V}_k) \in \mathcal{T}_{\delta_1}\}\Big] \\
&\quad \overset{\text{e.2}}{\leq} \frac{\mathbb{P}\Big[(\mathbb{W}(\tilde{m}), \mathbb{U}_k, \mathbb{V}_k) \in \mathcal{T}_\delta \Big| \{\mathbb{W} = \mathbb{W}(m)\} \cap \{(\mathbb{W}, \mathbb{U}_k, \mathbb{V}_k) \in \mathcal{T}_{\delta_1}\}\Big]}{\mathbb{P}\Big[(\mathbb{W}(\tilde{m}), \mathbb{U}_k) \in \mathcal{T}_\delta \Big| \{\mathbb{W} = \mathbb{W}(m)\} \cap \{(\mathbb{W}, \mathbb{U}_k, \mathbb{V}_k) \in \mathcal{T}_{\delta_1}\}\Big]} \\
&\quad \overset{\text{e.3}}{\leq} \gamma_n \; \frac{2^{-n(I(W;U_k,V_k) - 2\delta H(W))}}{2^{-n(I(W;U_k) + 2\delta H(W)))}}.
\end{aligned} \tag{58}$$

$$\sum_{l:|l| \leq 2^{nD_k}} \mathbb{P}\big[\mathcal{L}_k = l \big| \mathcal{S}_3\big] \left(\sum_{m=1}^{M} \mathbb{P}\big[\{\mathbb{X} \neq \mathbb{X}(m'),\ \forall\, m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \big| \mathcal{S}_3 \cap \{\mathcal{L}_k = l\}\big] \right) = 1. \tag{59}$$

$$\begin{aligned}
&\overset{\text{b}}{\leq} \sum_{\substack{\mathbb{v} \in \mathcal{L}_k^* \\ \mathbb{v} \neq \phi_k(\mathbb{W})}} \mathbb{P}\Big[f_k(\mathbb{v}) = b \Big| \{f_k(\mathbb{V}_k) = b\} \cap \mathcal{E}^c \cap \{\mathbb{X} \in \mathcal{C}_X\}\Big] \\
&\overset{\text{c}}{=} \sum_{\substack{\mathbb{v} \in \mathcal{L}_k^* \\ \mathbb{v} \neq \phi_k(\mathbb{W})}} \frac{1}{\lceil 2^{nR_k} \rceil} \\
&\overset{\text{d}}{\leq} 2^{-n(R_k - D_k)} \\
&\overset{\text{f}}{=} 2^{-n\epsilon_{\text{h}}}.
\end{aligned} \tag{51}$$

Notes on (51):

a. receiver k decodes $\mathbb{V}_k$ in error if and only if there is another $\mathbb{v} \in \mathcal{L}_k^*$ assigned to the same bin as the correct $\mathbb{v}$-codeword;
b. the union bound;
c. the codewords in $\mathcal{C}_{V_k}$ are thrown uniformly at random into $\lceil 2^{nR_k} \rceil$ bins;
d. $|\mathcal{L}_{V_k}| \leq 2^{nD_k}$, since we condition on $\mathcal{E}^c$; and
f. substitute the choice of helper rate R_k in (47).

The right-hand side of (51) is independent of b, so (50) gives

$$\mathbb{P}\big[\hat{\mathbb{V}}_k \neq \mathbb{V}_k \big| \mathcal{E}^c \cap \{\mathbb{X} \in \mathcal{C}_X\}\big] \leq 2^{-n\epsilon_{\text{h}}}. \tag{52}$$

Combining (49), (23), (24) and (52) gives

$$\mathbb{P}[\hat{\mathbb{V}}_k \neq \mathbb{V}_k] \leq b_5 \; 2^{-a_5 n}, \tag{53}$$

for some finite $a_5, b_5 > 0$.

We now turn to the rightmost probability in (48). We have

$$\mathbb{P}[\{\hat{\mathbb{V}}_k = \mathbb{V}_k\} \cap \{\hat{\mathbb{X}}_k \neq \mathbb{X}\}] \leq \mathbb{P}[\mathcal{S}_3^c] + \mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X} | \mathcal{S}_3], \tag{54}$$

where

$$\begin{aligned}
\mathcal{S}_3 := \{\mathbb{X} \in \mathcal{C}_X\} \cap \{(\mathbb{X}, \mathbb{Y}_k) \in \mathcal{T}_{\epsilon_1}(P_{X,Y_k})\} \cap \{\hat{\mathbb{V}}_k = \mathbb{V}_k\} \\
\cap \{(\mathbb{W}, \mathbb{U}_k, \mathbb{V}_k) \in \mathcal{T}_{\delta_1}(P_{W,U_k,V_k})\} \cap \{|\mathcal{L}_k| \leq 2^{nD_k}\}.
\end{aligned}$$

An upper bound on the probability $\mathbb{P}[\mathcal{S}_3^c]$ in (54) follows easily from previous bounds:

$$\begin{aligned}
\mathbb{P}[\mathcal{S}_3^c] &\leq \mathbb{P}[\mathbb{X} \notin \mathcal{C}_X] + \mathbb{P}[(\mathbb{X}, \mathbb{Y}_k) \notin \mathcal{T}_{\epsilon_1}] + \mathbb{P}[\hat{\mathbb{V}}_k \neq \mathbb{V}_k] \\
&\qquad + \mathbb{P}[(\mathbb{W}, \mathbb{U}_k, \mathbb{V}_k) \notin \mathcal{T}_{\delta_1}] + \mathbb{P}[|\mathcal{L}_k| > 2^{nD_k}] \\
&\overset{*}{\leq} b_6 \; 2^{-a_6 n},
\end{aligned} \tag{55}$$

where (*) holds for some finite $a_6, b_6 > 0$ by (24), (41), (53) and Lemma 4.

The rightmost probability in (54) is bounded by

$$\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X} | \mathcal{S}_3] \leq \gamma_n \; 2^{-n(I(W;V_k|U_k) - D_k + 4\delta H(W))}, \tag{56}$$

where

$$\gamma_n := \frac{\exp\left(2n(1-\delta_1)\frac{(\delta-\delta_1)^2}{1+\delta_1}\mu_{W,U_k}^2\right)}{\exp\left(2n(1-\delta_1)\frac{(\delta-\delta_1)^2}{1+\delta_1}\mu_{W,U_k}^2\right) - 2|\mathcal{W}||\mathcal{U}_k|}.$$

The steps leading to (56) are shown above in (57). Notes:

a. Write $\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X} | \mathcal{S}_3]$ as an expectation over all possible realisations of decoder k's list $\mathcal{L}_k$. We note that $|\mathcal{L}_k| \leq 2^{nD_k}$ with probability one, after conditioning on $\mathcal{S}_3$.
b. Let
$$\begin{aligned}
\mathcal{S}_{3,m} := \{\mathbb{X} \neq \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \\
\cap\, \mathcal{S}_3 \cap \{\mathcal{L} = l\}.
\end{aligned}$$
and note that $\mathbb{X} \in \mathcal{C}_X$, after conditioning on $\mathcal{S}_3$.
c. An error may only occur if there is some other index $\tilde{m} \neq m$ such that $\mathbb{X}(\tilde{m}) \in \mathcal{L}_k$ and $(\mathbb{W}(\tilde{m}), \mathbb{U}_k, \hat{\mathbb{V}}_k)$ is jointly typical. Here we note that after conditioning on $\mathcal{S}_{3,m}$ the following holds with probability one: the source $\mathbb{X}$ equals the m-th codeword $\mathbb{X}(m)$ in the source codebook $\mathcal{C}_X$; the m-th channel codeword is transmitted $\mathbb{W} = \mathbb{W}(m)$; the source and side information $(\mathbb{X}, \mathbb{Y}_k)$ are ϵ-jointly typical; $(\mathbb{W}, \mathbb{U}_k, \mathbb{V}_k)$ are δ-jointly typical; and $\hat{\mathbb{V}} = \mathbb{V}$. We have also abbreviated $\mathcal{T}_{\epsilon_1}(P_{X,Y_k})$ and $\mathcal{T}_{\delta_1}(P_{W,U_k,V_k})$ as $\mathcal{T}_{\epsilon_1}$ and $\mathcal{T}_{\delta_1}$ respectively.
d. Apply the union bound.
e. The rightmost probability in step (d) is bounded by
$$\begin{aligned}
\mathbb{P}\big[(\mathbb{W}(\tilde{m}), \mathbb{U}_k, \hat{\mathbb{V}}_k) \in \mathcal{T}_\delta \big| \mathcal{S}_{3,m}\big] \\
\leq \gamma_n \; 2^{-n(I(W;V_k|U_k) - 4\delta H(W))}.
\end{aligned}$$
The steps leading to this bound are shown above in (58). (See below for detailed notes on each step in (58).)
f. For each list l that satisfies $|l| \leq 2^{nD_k}$,
$$\sum_{\substack{\tilde{m} \in l \\ \tilde{m} \neq m}} \gamma_n 2^{-n(I(W;V_k|U_k) - 4\delta H(W))}$$

$$\leq \gamma_n 2^{nD_k} 2^{-n(I(W;V_k|U_k)-4\delta H(W))}.$$

Step (f) now follows from (59) above.

Notes for (58):

e.1. This step follows from the independence of the source and channel codebooks, the independence of codewords within each codebook, conditioning on $\mathcal{S}_{3,m}$ and $\{\mathcal{L}_k = l\}$ being equivalent to

$$(\mathbb{X}(m'), \mathbb{Y}_k) \in \mathcal{T}_\epsilon \text{ and } (\mathbb{W}(m'), \mathbb{U}_k) \in \mathcal{T}_\delta, \ \forall m' \in l,$$

and

$$(\mathbb{X}(m'), \mathbb{Y}_k) \notin \mathcal{T}_\epsilon \text{ or } (\mathbb{W}(m'), \mathbb{U}_k) \notin \mathcal{T}_\delta, \ \forall m' \notin l.$$

e.2. Bayes' rule.

e.3. Apply Lemma 5 to (e.2).

Thus,

$$\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X} | \mathcal{S}_3] \leq b_7 \ 2^{-a_7 n}, \tag{60}$$

for some $b_7 > 0$ and $a_7 := I(W; V_k|U_k) - D_k - 4\delta H(W)$, where $a_7 > 0$ by (46). Whenever (44) and (45) both hold, the achievability achievability of Theorem 1 follows from (48) and (53), (54), (55), and (60).

To complete the achievability proof of Theorem 1, we need only relax the assumption (45) and suppose that $H(X|Y_k) < I(W; U_k)$ for one or more receivers k. Such receivers do not require a positive helper rate or list exponent (i.e., we can set $R_k = 0$ and $D_k = 0$), and we can instead impose unique decoding. Indeed, the error analysis in [8, Sec. IV] shows that the probability of error $\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X}]$ at such receivers decays exponentially in n. (The error analysis in [8, Sec. IV] is valid because we use the same random source and channel codebooks.) ■

IX. Proof of Theorem 2 — Converse

Fix $\epsilon > 0$. Consider any $(n_\text{s}, n_\text{c}, R_1, R_2, \ldots, R_K)$-code with $\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X}] \leq \epsilon$ for all k. Following the now familiar path of defining $(\tilde{W}, \tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_K) \sim P_{\tilde{W}}(\cdot)T(\cdot|\cdot)$, with the timeshared pmf $P_{\tilde{W}}$ given in (17), we have

$$\begin{aligned}
n_\text{c} I(\tilde{W}; \tilde{U}_k) &\overset{\text{a}}{\geq} I(\mathbb{W}; \mathbb{U}_k) \overset{\text{b}}{=} I(\mathbb{X}, \mathbb{Y}_k, M_k, \mathbb{W}; \mathbb{U}_k) \\
&\geq H(\mathbb{X}|M_k, \mathbb{Y}_k) - H(\mathbb{X}|M_k, \mathbb{Y}_k, \mathbb{U}_k) \\
&\overset{\text{c}}{\geq} \sum_{i=1}^{n_\text{s}} H(X_i|M_k, \mathbb{Y}_k, X_1^{i-1}) - n_\text{s}\varepsilon(n_\text{s}) \\
&\overset{\text{d}}{\geq} \sum_{i=1}^{n_\text{s}} H(X_i|M_k, \mathbb{Y}_k, X_1^{i-1}, V_{k,1}^{i-1}) - n_\text{s}\varepsilon(n_\text{s}) \\
&\overset{\text{e}}{=} \sum_{i=1}^{n_\text{s}} H(X_i|M_k, \mathbb{Y}_k, V_{k,1}^{i-1}) - n_\text{s}\varepsilon(n_\text{s}) \\
&\overset{\text{f}}{=} \sum_{i=1}^{n_\text{s}} H(X_i|A_{k,i}, Y_{k,i}) - n_\text{s}\varepsilon(n_\text{s}).
\end{aligned} \tag{61}$$

Notes:

a. Jensen's inequality;

b. $(\mathbb{X}, \mathbb{Y}_k, M_k) \leftrightarrow \mathbb{W} \leftrightarrow \mathbb{U}_k$ forms a Markov chain;

c. Fano's inequality, where $\varepsilon(n) \to 0$, and the shorthand notation $X_1^{i-1} = (X_1, X_2, \ldots, X_{i-1})$;

d. conditioning reduces entropy and the notation $V_{k,1}^{i-1} = (V_{k,1}, V_{k,2}, \ldots, V_{k,i-1})$;

e. $X_i \leftrightarrow (M_k, V_{k,1}^{i-1}, \mathbb{Y}_k) \leftrightarrow X_1^{i-1}$ forms a Markov chain (see below for details); and

f. substitutes $A_{k,i} := (M_k, Y_{k,1}^{i-1}, Y_{k,i+1}^n, V_{k,1}^{i-1})$.

To see that $X_i \leftrightarrow (M_k, V_{k,1}^{i-1}, \mathbb{Y}_k) \leftrightarrow X_1^{i-1}$ forms a Markov chain in step (e) above, we first notice that

$$\big(X_i, M_k, Y_{k,i}^n\big) \leftrightarrow V_{k,1}^{i-1} \leftrightarrow \big(X_1^{i-1}, Y_1^{i-1}\big) \tag{62}$$

forms a Markov chain because the source and side information are memoryless and M_k is a function only of $\mathbb{V}_k$. The chain (62) implies $X_i \leftrightarrow \big(M_k, V_{k,1}^{i-1}, Y_{k,i}^n\big) \leftrightarrow \big(X_1^{i-1}, Y_1^{i-1}\big)$, which, in turn, implies $X_i \leftrightarrow \big(M_k, V_{k,1}^{i-1}, \mathbb{Y}_k\big) \leftrightarrow X_1^{i-1}$.

The bound for helper rate R_k follows a similar argument to that of [23, Sec. 15.8]. Specifically,

$$\begin{aligned}
n_\text{s} R_k &\geq H(M_k) \geq I(\mathbb{V}_k; M_k|\mathbb{Y}_k) \\
&= \sum_{i=1}^{n_\text{s}} I(V_{k,i}; M_k|\mathbb{Y}_k, V_{k,1}^{i-1}) \\
&\overset{\text{a}}{=} \sum_{i=1}^{n_\text{s}} I(V_{k,i}; M_k, Y_{k,1}^{i-1}, Y_{k,i+1}^n, V_{k,1}^{i-1}|Y_{k,i}) \\
&\overset{\text{b}}{=} \sum_{i=1}^{n_\text{s}} I(V_{k,i}; A_{k,i}|Y_{k,i}),
\end{aligned}$$

where step (a) follows because the source is memoryless and (b) substitutes $A_{k,i}$.

The source and the side information are iid and M_k is only a function of $\mathbb{V}_k$, so

$$(X_i, Y_i) \leftrightarrow V_{k,i} \leftrightarrow (M_k, \mathbb{V}_k, Y_{k,1}^{i-1}, Y_{k,i+1}^n). \tag{63}$$

The Markov chain (63) implies $(X_i, Y_i) \leftrightarrow V_{k,i} \leftrightarrow A_{k,i}$, and the converse follows from standard timesharing arguments, e.g. see [23, p. 578]. ■

X. Proof of Theorem 2 — Achievability

The proof combes the list decoder of Section VI with a 'helper' source code at BS(k).

A. Code Construction

Fix a pmf P_W on $\mathcal{W}$ and auxiliary random variables $(A_1, A_2, \ldots, A_K)$ satisfying $A_k \leftrightarrow V_k \leftrightarrow (X, Y_k)$. Let us assume that

$$H(X|A_k, Y_k) < I(W; U_k) \quad \forall\ k \tag{64}$$

and

$$R_k > I(A_k; V_k|Y_k) \quad \forall\ k. \tag{65}$$

As in Section VIII (the achievability proof Theorem 1), let us also assume that (45) holds so that every receiver requires a positive helper rate.

Fix constants ϵ, ϵ_1, δ and δ_1 satisfying (20), and choose $0 < \epsilon_1 < \epsilon_{\text{h}1} < \epsilon_\text{h} < \mu_{A_k,X,Y_k}$. Generate a random list code, as described in Section VI, with the parameters described above, and let $\mathcal{C}_X$ and $\mathcal{C}_W$ denote the source and channel codebooks respectively. Fix the list exponents to be

$$D_k = H(X|Y_k) - I(W; U_k) + \rho, \quad \forall\ k,$$

for any

$$\rho > 3\epsilon H(X) + 2\delta H(W). \quad (66)$$

Let P_{A_k} denote the marginal distribution of A_k. Randomly generate a source codebook for BS(k), with codewords of length n, by selecting symbols from $\mathcal{A}_k$ iid $\sim P_{A_k}$:

$$\mathcal{C}_{A_k} := \Big\{ \mathbb{A}_k(j,j') = \big(A_{k,1}(j,j'), A_{k,2}(j,j'), \ldots, A_{k,n}(j,j')\big) \Big\}$$

where we call j the *bin index* and

$$j = 1,2,\ldots,\lfloor 2^{nR_k} \rfloor \quad \text{and} \quad j' = 1,2,\ldots,\lfloor 2^{n(I(A_k;Y_k)-\epsilon_{\text{h}1})} \rfloor.$$

B. Encoding and Decoding

The list encoder and decoders operate as before, see Sections VI-C and VI-D. Helper BS(k) searches through the A_k-codebook $\mathcal{C}_{A_k}$ for a pair (J, J') such that $\big(\mathbb{A}_k(J,J'), \mathbb{V}_k\big) \in \mathcal{T}_{\epsilon_{\text{h}1}}$. If successful, BS($k$) sends the smallest such bin index J to receiver k. If unsuccessful, the helper sends an index J with an independent and uniform distribution over all possible bin indices.

Receiver k first attempts to decode $\mathbb{A}_k(J,J')$ by looking for a unique $\hat{J}$ in the J-th bin such that $\big(\mathbb{A}_k(J,\hat{J}), \mathbb{Y}_k\big) \in \mathcal{T}_{\epsilon_{\text{h}1}}$. If successful, receiver k sets $\hat{\mathbb{A}}_k = \mathbb{A}_k(J,\hat{J})$. Otherwise, it randomly selects $\hat{\mathbb{A}}_k$ iid $\sim P_{A_k}$.

The list decoder at receiver k outputs $\mathcal{L}_k$, see (20). Receiver k looks for a unique $\mathbb{X}(m') \in \mathcal{L}_k$ such that $\big(\mathbb{X}(m'), \mathbb{Y}_k, \hat{\mathbb{A}}_k\big) \in \mathcal{T}_\epsilon$. If successful, receiver k outputs $\hat{\mathbb{X}}_k := \mathbb{X}(m')$. Otherwise, it randomly generates $\hat{\mathbb{X}}_k$ using P_X.

C. Error Analysis

We first bound the probability of error at receiver k by

$$\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X}] \leq \mathbb{P}[\mathcal{S}_4^c] + \mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X} | \mathcal{S}_4],$$

where

$$\mathcal{S}_4 := \{\mathbb{X} \in \mathcal{C}_X\} \cap \{\hat{\mathbb{A}}_k = \mathbb{A}_k\} \cap \{|\mathcal{L}_k| \leq 2^{nD_k}\} \cap \{(\mathbb{X}, \mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_{\epsilon_1}\} \cap \{(\mathbb{W}, \mathbb{U}_k) \in \mathcal{T}_{\delta_1}\},$$

and we have abbreviated the typical sets $\mathcal{T}_{\epsilon_1}(P_{X,Y_k,A_k})$ and $\mathcal{T}_{\delta_1}(P_{W,U_k})$ as $\mathcal{T}_{\epsilon_1}$ and $\mathcal{T}_{\delta_1}$, respectively.

We have

$$\mathbb{P}\big[\mathcal{S}_4^c\big] \leq b_7 \; 2^{-a_7 n}, \quad (67)$$

for some finite $a_7, b_7 > 0$. To see (67), apply the union bound to $\mathbb{P}\big[\mathcal{S}_4^c\big]$; use (25) to bound $\mathbb{P}[\mathbb{X} \notin \mathcal{C}_X]$; use (40), (41) and (66) to bound $\mathbb{P}[|\mathcal{L}_k| > 2^{nD_k}]$; and use Lemmas 4 and 5 to bound $\mathbb{P}[(\mathbb{W}, \mathbb{U}_k) \notin \mathcal{T}_{\delta_1}^n]$. The final two probabilities, $\mathbb{P}[\hat{\mathbb{A}}_k \neq \mathbb{A}]$ and $\mathbb{P}[(\mathbb{X}, \mathbb{Y}_k, \mathbb{A}_k) \notin \mathcal{T}_{\epsilon_1}^n]$, also tend to zero exponentially in n by Lemmas 4 and 5; see, for example, Kramer's achievabiltiy proof of the Wyner-Ziv theorem [33, Sec. 5.3]. Finally, the conditional probability $\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X} | \mathcal{S}_4]$ tends to zero exponentially in n, as shown below in (68).

Notes for (68):

a. Write $\mathbb{P}[\hat{\mathbb{X}}_k \neq \mathbb{X} | \mathcal{S}_4]$ as an expectation over all possible realisations of decoder k's list $\mathcal{L}_k$. Here we note that $|\mathcal{L}_k| \leq 2^{nD_k}$ with probability one, after conditioning on $\mathcal{S}_4$.

b. Write the second conditional probability in step (a) as an expectation over all possible encodings of $\mathbb{X}$. Here we note that $\mathbb{X} \in \mathcal{C}_X$ with probability one, after conditioning on $\mathcal{S}_4$.

c. In the rightmost conditional probability in step (b), the error event $\{\hat{\mathbb{X}}_k \neq \mathbb{X}\}$ is equivalent to the following: There exists an index $\tilde{m}$ in decoder k's list $\mathcal{L}_k$, which is different to the correct index m and such that $(\mathbb{X}(\tilde{m}), \mathbb{Y}_k, \hat{\mathbb{A}}_k)$ is jointly typical. Here we note that the correct index m is in decoder k's list with probability one, after conditioning on $\mathcal{S}_4$. We have also slightly abused notation and written the union over all indices $m' \in \mathcal{L}_k$, but $\mathcal{L}_k$ is a list of source codewords, see (20).

d. Apply the union bound to (c).

e. If the index $\tilde{m}$ is smaller than m, $\tilde{m} < m$, then the rightmost conditional probability in step (d) is bounded from above by (69), which is given below. It can also be shown that (69) holds for indices $\tilde{m} > m$. To see this note that $\gamma_n > 1$ and the righthand side of step (e.1) in (69) simplifies to

$$\mathbb{P}\big[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_\epsilon \big| \{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}_\epsilon\} \cap \{(\mathbb{X}, \mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_{\epsilon_1}\}\big].$$

The bound then follows from Lemma 5.

f. Substitute $D_k = H(X|Y_k) - I(W;U_k) + \rho$.

Notes for (69):

e.1. The first step is a consequence of the independence of the source and channel codebooks, the independence of codewords within each codebook, and $\{\mathcal{L}_k = l\}$ is equivalent to

$$(\mathbb{X}(m'), \mathbb{Y}_k) \in \mathcal{T}_\epsilon \text{ and } (\mathbb{W}(m'), \mathbb{U}_k) \in \mathcal{T}_\delta, \quad \forall m' \in l,$$

and

$$(\mathbb{X}(m'), \mathbb{Y}_k) \notin \mathcal{T}_\epsilon \text{ or } (\mathbb{W}(m'), \mathbb{U}_k) \notin \mathcal{T}_\delta, \quad \forall m' \notin l.$$

e.2. Apply Bayes' law twice and use the upper bound

$$\mathbb{P}\big[\{\mathbb{X} \neq \mathbb{X}(\tilde{m})\} \cap \{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}_\epsilon\} \big| \{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_\epsilon\} \cap \{(\mathbb{X}, \mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_{\epsilon_1}\}\big] \leq 1$$

e.3. Use Lemmas 4 and 5 to lower bound

$$1 - \mathbb{P}\big[\mathbb{X} = \mathbb{X}(\tilde{m}) \big| \{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}_\epsilon\} \cap \{(\mathbb{X}, \mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_{\epsilon_1}\}\big].$$

Use Lemma 5 to bound the numerator and denominator of the rightmost term in step (e.2).

e.4. Set

$$\gamma_n := \frac{1}{1 - \frac{1}{1-\zeta_n} 2^{-n(H(X|Y_k) - 3\epsilon H(X))}},$$

where

$$\zeta_n := 2|\mathcal{X}||\mathcal{Y}_k| \exp\Big(-2n(1-\epsilon_1)\frac{(\epsilon-\epsilon_1)^2}{1+\epsilon_1}\mu_{X,Y_k}^2 \Big)$$

We notice that $\gamma_n \to 1$ from above whenever $3\epsilon H(X) < H(X|Y_k)$ and $\epsilon > \epsilon_1$.

The proof now follows from (69), because $H(X|A_k, Y_k) < I(W; U_k)$, we can choose ϵ and ρ arbitrarily small, and $H(X)$ is finite. ■

$$\begin{aligned}
\mathbb{P}\left[\hat{\mathbb{X}}_k \neq \mathbb{X} \middle| \mathcal{S}_4\right] &\overset{\text{a}}{=} \sum_{l:\ |l| \leq 2^{nD_k}} \mathbb{P}\left[\mathcal{L}_k = l \middle| \mathcal{S}_4\right]\ \mathbb{P}\left[\hat{\mathbb{X}}_k \neq \mathbb{X} \middle| \mathcal{S}_4 \cap \{\mathcal{L} = l\}\right] \\
&\overset{\text{b}}{=} \sum_{l:\ |l| \leq 2^{nD_k}} \mathbb{P}\left[\mathcal{L}_k = l \middle| \mathcal{S}_4\right] \left(\sum_{m=1}^{M} \mathbb{P}\left[\{\mathbb{X} \neq \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \middle| \mathcal{S}_4 \cap \{\mathcal{L}_k = l\}\right] \right. \\
&\qquad \left. \cdot\, \mathbb{P}\left[\hat{\mathbb{X}}_k \neq \mathbb{X} \middle| \{\mathbb{X} \neq \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \cap \mathcal{S}_4 \cap \{\mathcal{L}_k = l\}\right] \right) \\
&\overset{\text{c}}{=} \sum_{l:\ |l| \leq 2^{nD_k}} \mathbb{P}\left[\mathcal{L}_k = l \middle| \mathcal{S}_4\right] \left(\sum_{m=1}^{M} \mathbb{P}\left[\{\mathbb{X} \neq \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \middle| \mathcal{S}_4 \cap \{\mathcal{L}_k = l\}\right] \right. \\
&\qquad \left. \cdot\, \mathbb{P}\left[\bigcup_{\substack{\tilde{m} \in l \\ \tilde{m} \neq m}} \{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k, \hat{\mathbb{A}}_k) \in \mathcal{T}_\epsilon\} \middle| \{\mathbb{X} \neq \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \cap \mathcal{S}_4 \cap \{\mathcal{L}_k = l\}\right] \right) \\
&\overset{\text{d}}{\leq} \sum_{l:\ |l| \leq 2^{nD_k}} \mathbb{P}\left[\mathcal{L}_k = l \middle| \mathcal{S}_4\right] \left(\sum_{m=1}^{M} \mathbb{P}\left[\{\mathbb{X} \neq \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \middle| \mathcal{S}_4 \cap \{\mathcal{L}_k = l\}\right] \right. \\
&\qquad \left. \cdot \left(\sum_{\substack{\tilde{m} \in l \\ \tilde{m} \neq m}} \mathbb{P}\left[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k, \hat{\mathbb{A}}_k) \in \mathcal{T}_\epsilon \middle| \{\mathbb{X} \neq \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \cap \mathcal{S}_4 \cap \{\mathcal{L}_k = l\}\right] \right) \right) \\
&\overset{\text{e}}{\leq} \gamma_n\ 2^{-n(I(X;A_k|Y_k) - D_k - 4\epsilon H(X)} \\
&\overset{\text{f}}{=} \gamma_n\ 2^{-n(I(W;U_k) - H(X|A_k,Y_k) - 4\epsilon H(X) - \rho)}.
\end{aligned} \tag{68}$$

$$\begin{aligned}
&\mathbb{P}\left[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k, \hat{\mathbb{A}}_k) \in \mathcal{T}_\epsilon \middle| \{\mathbb{X} \neq \mathbb{X}(m'), \forall m' < m\} \cap \{\mathbb{X} = \mathbb{X}(m)\} \cap \mathcal{S}_4 \cap \{\mathcal{L}_k = l\}\right] \\
&\overset{\text{e.1}}{=} \mathbb{P}\left[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_\epsilon \middle| \{\mathbb{X} \neq \mathbb{X}(\tilde{m})\} \cap \{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}_\epsilon\} \cap \{(\mathbb{X}, \mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_{\epsilon_1}\}\right] \\
&\overset{\text{e.2}}{\leq} \left(\frac{1}{\mathbb{P}\left[\mathbb{X} \neq \mathbb{X}(\tilde{m}) \middle| \{(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}_\epsilon\} \cap \{(\mathbb{X}, \mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_{\epsilon_1}\}\right]} \right) \\
&\qquad \left(\frac{\mathbb{P}\left[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_\epsilon \middle| (\mathbb{Y}_k, \mathbb{A}_k) \in \mathcal{T}_{\epsilon_1}\right]}{\mathbb{P}\left[(\mathbb{X}(\tilde{m}), \mathbb{Y}_k) \in \mathcal{T}_\epsilon \middle| \mathbb{Y}_k \in \mathcal{T}_{\epsilon_1}\right]} \right) \\
&\overset{\text{e.3}}{\leq} \left(\frac{1}{1 - \frac{1}{1-\zeta_n} 2^{-n(H(X|Y_k) - 3\epsilon H(X))}} \right) \left(\frac{2^{-n(I(X;Y_k,A_k) - 2\epsilon H(X))}}{2^{-n(I(X;Y_k) + 2\epsilon H(X))}} \right) \\
&\overset{\text{e.4}}{=} \gamma_n\ 2^{-n(I(X;A_k|Y_k) - 4\epsilon H(X))}
\end{aligned} \tag{69}$$

ACKNOWLEDGEMENTS

The authors would like to thank Gerhard Kramer, the associate editor and the anonymous reviewers for their thoughtful comments on the paper.

REFERENCES

[1] A. Damnjanovic, J. Montojo, W. Yongbin, J. Tingfang, L. Tao, M. Vajapeyam, Y. Taesang, S. Osok, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun*, vol. 18, no. 3, pp. 10 – 21, 2011.

[2] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. A. Thomas, J. G. Andrews, P. Xia, H. S. Jo, H. S. Dhillon, and T. D. Novlan, "Heterogeneous cellular networks: from theory to practice," *IEEE Commun. Magazine*, vol. 50, no. 6, pp. 54 – 64, 2012.

[3] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 56–61, 2006.

[4] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and Y. Wei, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380 – 1408, 2010.

[5] J. G. Andrews, "Interference cancellation for cellular systems: a contemporary overview," *IEEE Wireless Commun.*, vol. 12, no. 2, pp. 19 – 29, 2005.

[6] V. Chandrasekhar and J. G. Andrews, "Spectrum allocation in tiered cellular networks," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3059 – 3068, 2009.

[7] S. Singh, H. S. Dhillon, and J. A. Andrews, "Offloading in heterogeneous networks: modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. (to appear), 2014.

[8] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1469 –1482, 2006.

[9] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inform. Theory*, vol. 54, no. 3, pp. 1328 –1329, 2008.

[10] A. Wyner, "On source coding with side-information at the decoder," *IEEE Trans. Inform. Theory*, vol. 21, no. 3, pp. 294–300, 1975.

[11] T. J. Oechtering, C. Schnurr, I. Bjelakovic, and H. Boche, "Broadcast capacity region of two-phase bidirectional relaying," *IEEE Trans. Inform. Theory*, vol. 54, no. 1, pp. 454–458, 2008.

[12] G. Kramer and S. Shamai, "Capacity for classes of broadcast channels with receiver side information," in *proc. IEEE Inform. Theory Workshop*, Lake Tahoe, California, 2007.

[13] A. Wyner, J. Wolf, and F. Willems, "Communicating via a processing broadcast satellite," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1243–1249, 2002.

[14] R. Timo, A. Grant, and G. Kramer, "Lossy broadcasting with complementary side information," *IEEE Trans. Inform. Theory*, vol. 59, no. 1, pp. 104 – 131, 2013.

[15] T. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. 21, no. 2, pp. 226–228, 1975.

[16] D. Gunduz, E. Erkip, A. Goldsmith, and H. V. Poor, "Source and channel coding for correlated sources over multiuser channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 3927–3944, 2009.

[17] J. Nayak, E. Tuncel, and D. Gunduz, "Wyner-Ziv coding over broadcast channels: digital schemes," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1782–1799, 2010.

[18] R. Timo, G. Lecher, L. Ong, and S. Johnson, "Multi-way relay networks: orthogonal uplink, source-channel separation and code design," *IEEE Trans. Commun.*, vol. 61, no. 2, pp. 753 – 768, 2013.

[19] D. Gündüz, E. Erkip, A. Goldsmith, and H. V. Poor, "Reliable joint source-channel cooperative transmission over relay networks," *IEEE Trans. Inform. Theory*, vol. 59, no. 4, 2013.

[20] D. Blackwell, L. Breiman, and A. J. Thomasian, "Capacity of a class of channels," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1229 – 1241, 1959.

[21] A. Lapidoth and E. Telatar, "The compound channel capacity of a class of finite-state channels," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 973–983, 1998.

[22] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[23] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 2006.

[24] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[25] C. Chen, D.-k. He, and A. Jagmohan, "The equivalence between Slepian-Wolf coding and channel coding under density evolution," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2534–2540, 2009.

[26] K. M. Morshed, I. Land, and G. Lechner, "Lossless compression of uniform binary sources with coded side-information," *proc. Australian Commun. Theory Workshop*, Sydney, Australia, 2014.

[27] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inform. Theory*, vol. 31, no. 6, pp. 727–734, 1985.

[28] R. Timo, T. Chan, and A. Grant, "Rate distortion with side-information at many decoders," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 5240–5257, 2011.

[29] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[30] Y.-K. Chia, "On multiterminal source coding with list decoding constraints," in *Intl. Symp. Inform. Theory*, HI, USA, 2014.

[31] Y.-H. Kim, A. Sutivong, and T. M. Cover, "State amplification," *IEEE Trans. Inform. Theory*, vol. 54, no. 5, pp. 1850 – 1859, 2008.

[32] R. W. Yeung, *Information theory and network coding*. Springer, 2008.

[33] G. Kramer, "Topics in multi-user information theory," *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 45, pp. 265–444, 2008.

[34] G. Kramer, "Lecture notes on Multi-User Information Theory," LNT, Technische Universität München, 2012.

[35] A. Wyner, "A theorem on the entropy of certain binary sequences and applications: part I," *IEEE Trans. Inform. Theory*, vol. 19, no. 6, pp. 769–772, 1975.

[36] A. Wyner, "A theorem on the entropy of certain binary sequences and applications: part II," *IEEE Trans. Inform. Theory*, vol. 19, no. 6, pp. 772–777, 1975.

[37] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. 21, no. 6, pp. 629 – 637, 1975.

[38] T. Cover, A. E. Gamal and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inform. Theory*, vol. 26, no. 6, pp. 648 – 657, 1980.

[39] O. İşcan, I. Land and G. Lechner, "Repeat-Accumulate Codes for Broadcast Channels with Side Information," *proc. Intl. ITG Conf. Systems, Commun. Coding,* Hamburg, Germany, Feb., 2015.

[40] D. Marco and M. Effros, "On lossless coding with coded side information," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, 2009.