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.

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 $X$ 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 heterogeneous 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.

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 $X$.

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. $X$; 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. $\mathbf{A} = (A_1, A_2, \ldots, A_n) \in \mathcal{A}^n$.

B. Source and Channel Setup

The main BS is required to communicate a source

$\mathcal{X} = (X_1, X_2, \ldots, X_{n_c})$

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

$\mathcal{Y}_k = (Y_{k,1}, Y_{k,2}, \ldots, Y_{k,n_c})$.

For example, $\mathcal{X}$ and $\mathcal{Y}_k$ 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 $\mathcal{X}$ and $\mathcal{Y}_k$ 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,

$$(\mathcal{X}, \mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_K) := \{(X_1, Y_{1,1}, Y_{1,2}, \ldots, Y_{1,i}), \ldots, (X_n, Y_{n,1}, Y_{n,2}, \ldots, Y_{n,i})\}_{i=1}^{n_c}$$

is a sequence of $n_c$ independent and identically distributed (iid) source/side-information tuples $(X, Y_{1,1}, Y_{1,2}, \ldots, Y_{1,i})$ 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

$w := f(\mathcal{X})$

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

$\kappa := \frac{n_c}{n_s}$,

is called the bandwidth expansion factor.

Receiver $k$ observes $U_k = (U_{k,1}, U_{k,2}, \ldots, U_{k,n_c})$ from the channel. The channel outputs, across all receivers, conditionally depend on the codeword $w$ via the memoryless law

$$P[U_1 = u_1, U_2 = u_2, \ldots, U_K = u_K|\mathcal{W} = w]$$

$$= \prod_{i=1}^{n_c} T(u_{1,i}, u_{2,i}, \ldots, u_{K,i}|w_i),$$

where $w \in \mathcal{W}^{n_c}, u_k \in \mathcal{U}_k^{n_c}$ and $T(u_{1,i}, \ldots, u_{K,i}|w_i)$ 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]

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

where $(W, U_1, U_2, \ldots, U_K) \sim P_W(\cdot|\cdot)$. 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 $B(k)$, obtains side information $\mathcal{V}_k = (V_{k,1}, V_{k,2}, \ldots, V_{k,n_c})$ about the source $\mathcal{X}$ or the codeword $\mathcal{W}$ via a backbone network. Here $n_m = n_s$ (resp. $n_k = n_c$) when BS($k$) has side information about $\mathcal{X}$ (resp. $\mathcal{W}$), and a precise definition of $\mathcal{V}_k$ will be given shortly. BS($k$) sends

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

over a noiseless channel to receiver $k$, where $f_k : \mathcal{V}_k^{n_m} \rightarrow \{1, 2, \ldots, [2^{n_mR_k}]\}$ is BS($k$)’s encoder and $R_k$ is its rate (in bits per source symbol$^1$). Receiver $k$ attempts to recover the source via $\hat{X}_k := g_k(U_k, \mathcal{Y}_k, M_k)$, where $g_k : \mathcal{U}_k^{n_k} \times \mathcal{Y}_k^{n_c} \times \{1, 2, \ldots, [2^{n_kR_k}]\} \rightarrow \mathcal{X}^{n_c}$ is the receiver’s decoder. The collection of all encoders and decoders is called an $(n_c, n_k, R_1, R_2, \ldots, R_K)$-code.

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

Suppose that $\mathcal{V}_k$ is the entire codeword $\mathcal{W}$ or a scalar quantised version thereof. Quantisation is appropriate, for example, when the backbone network is rate limited. More formally, let $\phi_k : \mathcal{W} \rightarrow \mathcal{V}_k$ 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 $\kappa$, helper BS rates $R := (R_1, R_2, \ldots, R_K)$, and scalar quantisers $\phi := (\phi_1, \phi_2, \ldots, \phi_K)$. We say that a source/side information tuple $(X, Y_1, Y_2, \ldots, Y_K)$ is $(\kappa, R, \phi)$-achievable if for any $\epsilon > 0$ there exists an $(n_c, n_k, R_1, R_2, \ldots, R_K)$-code such that

$$\frac{n_c}{n_k} = \kappa$$

and

$$P[\hat{X}_k \neq X] \leq \epsilon, \quad \forall k,$$

holds for sufficiently large $n_k$ 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 $V_k$ is directly correlated with the source and side information. That is, assume $(X, Y_1, Y_2, \ldots, Y_K, V_1, V_2, \ldots, V_K)$ is emitted by an arbitrary discrete memoryless source and thus is a sequence of $n$ 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 $\kappa$ and helper BS rates $R := (R_1, R_2, \ldots, R_K)$. We say that a source/side-information tuple $(X, Y_1, Y_2, \ldots, Y_K, V_1, V_2, \ldots, V_K)$ is $(\kappa, R, \phi)$-achievable if for any $\epsilon > 0$ there exists an $(n_\epsilon, n_c, R_1, R_2, \ldots, R_K)$-code such that (2) holds for sufficiently large $n_\epsilon$ and $n_c$.

III. INFORMATION-THEORETIC LIMITS OF BS COOPERATION

A. Mode 1

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

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

where $(W, U_1, U_2, \ldots, U_K) \sim P_W(\cdot)T(\cdot\cdot)\cdot$ and $V_k = \phi_k(W)$. Conversely: If the source/side-information tuple $(X, Y_1, Y_2, \ldots, Y_K)$ is $(\kappa, R, \phi)$-achievable, then there exists a pmf $P_W$ on $W$ 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 $R$ and bandwidth expansion factor $\kappa$. A source/side-information tuple $(X, Y_1, Y_2, \ldots, Y_K, V_1, V_2, \ldots, V_K)$ is $(\kappa, R)$-achievable if there exists a pmf $P_W$ on $W$ 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 $(W, U_1, U_2, \ldots, U_K) \sim P_W(\cdot)T(\cdot\cdot)\cdot$. Conversely: If the source/side-information tuple $(X, Y_1, Y_2, \ldots, Y_K, V_1, V_2, \ldots, V_K)$ is $(\kappa, R)$-achievable, then there exists a pmf $P_W$ on $W$ 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 $|V_k|$.

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 $\phi$ we have $R_k > \kappa H(V_k|U_k)$ for all $k$. If $(V_k, U_k)$ behaves like a discrete memoryless source, then BS$(k)$ can reliably send $V_k$ to receiver $k$ using a Slepian-Wolf code of rate $R_k$ [22]. The receiver effectively has the combined channel output $(U_k, V_k)$. 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 suboptimal4. 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 strong5 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 $V_k$. 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,

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

5Relays with high quality channels $i$ 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 \(U_k\) and side information \(Y_k\). 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 in 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 \(Y_k\) [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 \(W\), while others have information about the source \(X\) — 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 \(W\) 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 \(A_k \leftrightarrow V_k \leftrightarrow (X, Y_k)\).

\[ H(X|Y_k) < \kappa I(W; U_k) + \min \{R_k, \kappa I(W; V_k|U_k)\} \]  

(6)

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

\[ R_k > I(V_k; A_k|Y_k) \]  

(7a)

\[ H(X|A_k, Y_k) < \kappa I(W; U_k) \]  

(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 \(W\) 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 \(n_s = n_c = n\), 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 \(\{1, 2, \ldots, 2^nR^*\}\). For example, in Theorem 1 suppose that \(2^nR^*\) is an integer, \(\kappa = 1\), \(Y_k = \text{constant and } M = \mathcal{X}\), where \(\mathcal{X}\) is iid with a uniform distribution on \(\{1, 2, \ldots, 2^n\}\). Then \(H(X|Y_k) = H(X) = R^*\) for all \(k\).

Given helper rates \(R\), we can define the helper capacity \(C(R)\) 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

\[ C(R) = \max \min_{P_W} I(W; U_k) + \min \{R_k, I(W; V_k|U_k)\}, \]  

(8)

where the maximisation is taken over all pmfs \(P_W\) on \(W\).

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 \(n_s = n_c = n\), 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 \(\{1, 2, \ldots, 2^nR_1^*\}\) and \(\{1, 2, \ldots, 2^nR_2^*\}\) 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 \(\kappa = 1\), \(2^nR_1^*\) and \(2^nR_2^*\) are integers, \(M_1 = \mathcal{X}_1 = \mathcal{Y}_1\) and \(M_2 = \mathcal{X}_2 = \mathcal{Y}_2\), where \(\mathcal{X}_1\) and \(\mathcal{X}_2\) are independent with iid uniform distributions on \(\{1, 2, \ldots, 2^nR_1^*\}\) and \(\{1, 2, \ldots, 2^nR_2^*\}\) respectively. Then, setting \(\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)\) gives

\[ H(X|Y_1) = H(X_1, X_2|X_1) = H(X_2) = R_2^* \]  

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 \(C(R_1, R_2)\) to be the 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 \(C(R_1, R_2)\) is equal to the set of all \((R_1^*, R_2^*)\) for which there exists a pmf \(P_{W}\) on \(W\) such that

\[ R_1^* \leq I(W; U_2) + \min \{R_2, I(W; V_2|U_2)\} \]

and

\[ R_2^* \leq I(W; U_1) + \min \{R_1, I(W; V_1|U_1)\}. \]
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, $X \sim \text{Bern}(1/2)$; there is no receiver side information, $Y_k =$ constant; and helper BS(k)'s side information is

$$V_k := X \oplus Z_k,$$

(modulo 2),

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 $(1, R)$-achievable if there exists a pmf $P_W$ on $W$ and constants $(\alpha_1, \alpha_2, \ldots, \alpha_K) \in [0,1/2]^K$ such that

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

(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

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

Conversely: If the source/side-information tuple $(X, V_1, V_2, \ldots, V_K)$ is $(1, R)$-achievable, then there exists a pmf $P_W$ on $W$ 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

$$H(X|V_k) \underset{b}{\geq} h(\rho_k) \leq H(X|A_k) \overset{d}{=} H(X|A_k, Y_k) \overset{d}{\leq} 1, \quad \forall k,$$

(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

$$H(X|A_k, Y_k) = H(X|A_k) = h(\alpha_k \ast \rho_k).$$

(12a)

In addition, we have

$$I(V_k; A_k|Y_k) \overset{b}{=} I(V_k; A_k) \overset{b}{\geq} 1 - H(V_k|A_k) \overset{c}{\geq} 1 - h(\alpha_k),$$

(12b)

where (a) follows because $Y_k$ is a constant; and (b) follows because $V_k \sim \text{Bern}(1/2)$. 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

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

(13)

where $h^{-1} : [0,1] \rightarrow [0,1/2]$ 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 $\alpha_k$ to obtain

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

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

$$f : \mathcal{X}^n \rightarrow \mathcal{Y}^n$$

is the encoder at the transmitter and

$$g_k : \mathcal{U}_k^n \times \mathcal{Y}_k^n \rightarrow \Omega(L_k)$$

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

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

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

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

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

$$\frac{n_c}{n_s} = \kappa, \quad (14a)$$

$$L_k \leq 2^{n_s D_k} \quad \text{and} \quad P[\mathcal{X} \notin \mathcal{L}_k] \leq \epsilon, \quad \forall k, \quad (14b)$$

where $n_s$ and $n_c$ are sufficiently large.

The next lemma is proved in Sections V and VI.

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

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

(15)

where $(W, U_1, \ldots, U_K) \sim P_W(\cdot|T(\cdot))$. Conversely: If $(\kappa, D)$ is achievable, then there exists a pmf $P_W$ on $W$ 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 $\{1,2,\ldots,\lfloor 2^nR_k \rfloor \}$. In this case, the mutual information $I(W;U_k)$ transforms to the source-coding rate $R_k$ and Lemma 3 reduces to [30, Prop. 1]

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

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 $D = (0,0,\ldots,0)$. The reverse (converse) assertion of Lemma 3 shows that $(\kappa, D)$ is achievable only if

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

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 $D$ 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 $L_k$ is equal to the source $X$. This list will, with high probability, include $X$ and have $|L_k| \approx 2^{nD_k}$ elements. To resolve receiver $k$'s uncertainty, $BS(k)$ needs to encode its side information $Y_k$ 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 dual 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 $Y_k$ in Mode 2 is directly correlated with the source $X$, 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 $Y_k$ 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 $(s, n_s, L_1, L_2, \ldots, L_K)$ list code such that (14) holds. As before, let $W = f(X)$ and $U_k = (U_{k,1}, U_{k,2}, \ldots, U_{k,n_s})$ 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 $W = (W_1, W_2, \ldots, W_{n_s})$, and let $P_{W_j}$ denote its pmf. Construct a timeshared pmf $P_{\tilde{W}}$ on $W$ by setting

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

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

$$n_s I(\tilde{W}; \tilde{U}_k) \geq \frac{b}{H(\tilde{U}_k)} \geq I(W; U_k) \geq I(X; U_k|Y_k) \geq d \approx n_s H(X|Y_k) - H(X|Y_k).$$  

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

$$\sum_{i=1}^{n_s} I(W_i; U_{k,i}) = \frac{1}{n_s} \left( \sum_{i=1}^{n_s} H(U_{k,i}) - H(U_{k,i}|W_i) \right)$$

$$\geq H(U_k) - \frac{1}{n_s} \sum_{i=1}^{n_s} H(U_{k,i}|W_i)$$

$$\geq H(U_k) - \frac{1}{n_s} \sum_{i=1}^{n_s} H(U_{k,i}|W, U_{k,1}, U_{k,2}, \ldots, U_{k,i-1})$$

$$= H(U_k) - H(U_k|W),$$

where (a) follows because the broadcast channel is memoryless and therefore $U_{k,i} \leftrightarrow W_i \leftrightarrow (W, U_{k,1}, U_{k,2}, \ldots, U_{k,i-1})$ forms a Markov chain; (c) $(X, Y_k) \leftrightarrow (W, U_k)$ 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]:

$$H(X|Y_k, U_k) \leq \log |L_k| + 1$$

$$+ (n_s \log |X| - \log |L_k|)P \left[ X \notin \bigcup_{k=1}^{K} L_k \right].$$

By (14b), and since for any $k, P[X \notin L_k] \geq P[X \notin \bigcap_{k \neq k'} L_{k'}]$, this inequality implies

$$H(X|Y_k, U_k) \leq n_s (D_k + \varepsilon(n_s, \epsilon)),$$  

where

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

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 \( n_k \) and \( n_k \)) and time-shared pmfs \( \{P_{W_k}\} \). Since \( W \) is a finite alphabet, by the Bolzano-Weierstrass theorem, \( \{P_{W_k}\} \) will contain a convergent subsequence with respect to the variational distance. Let \( P_{W_k}^* \) denote the limit of the convergent subsequence and \( \hat{W}^* \sim P_{W_k}^* \). We then have \( \kappa I(W^*;U_k) \geq H(X|Y_k) - D_k \) by the continuity of mutual information [32, Sec. 2.3].

VI. PROOF OF LEMMA 3 — ACHIEVABILITY

We restrict attention to the bandwidth matched case \( (k = 1 \) and \( n_k = n = n) \), 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 \( \omega \) and set \( \Omega \), let us denote the indicator function for the event that \( \omega \) falls in \( \Omega \) 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 \( A \times B \), where \( A \) and \( B \) 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

\[
T^n_\epsilon(P_A) := \left\{ a \in A^n : \left| \frac{1}{n} N(a'|a) - P_A(a') \right| \leq \epsilon P_A(a'), \forall a' \in A \right\},
\]

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

\[
T^n_\epsilon(P_{A,B}) := \left\{ (a, b) \in A^n \times B^n : \left| \frac{1}{n} N(a', b'|a, b) - P_{A,B}(a', b') \right| \leq \epsilon P_{A,B}(a', b'), \forall (a', b') \right\}.
\]

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

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

and the following lemmas. Let

\[
\mu_A := \min_{a \in \text{supp}(P_A)} P_A(a),
\]

and

\[
\mu_{A,B} := \min_{(a, b) \in \text{supp}(P_{A,B})} P_{A,B}(a, b),
\]

where \( \text{supp}(\cdot) \) denotes the support set of the indicated distribution.

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

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

and

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

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

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

and

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

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

B. Distributions and Typicality Constants

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

\[
X \sim P_X, \ (X, Y_k) \sim P_{X,Y_k} \text{ and } (W, U_k) \sim P_{W,U_k}
\]

denote the pmfs of the indicated variables. Fix any arbitrarily small constants \( \epsilon, \epsilon_1, \delta \) and \( \delta_1 \) satisfying

\[
0 < \delta_1 < \delta < \min_k \mu_{W,U_k}
\]

and \( 0 < \epsilon_1 < \epsilon < \min_k \mu_{X,Y_k} \),

with

\[
\epsilon < \frac{\min_k \mu_{W,U_k}}{2 H(X) \ln 2} \delta^2.
\]

C. Code Construction and Encoding

The encoder mirrors that of [8, Thm. 6]. Randomly generate a source codebook \( C_X \) with

\[
M = \left[ 2^{nH(X)(1+\epsilon)} \right]
\]

codewords, each of length \( n \), by selecting symbols from \( X \) in an iid fashion using \( P_X \):

\[
C_X := \left\{ \mathbb{x}(m) = (X_1(m), X_2(m), \ldots, X_n(m)) \right\}_{m=1}^M.
\]

In the same way, generate a channel codebook \( C_W \) with \( M \) codewords of length \( n \) using \( P_W \):

\[
C_W := \left\{ \mathbb{w}(m) = (W_1(m), W_2(m), \ldots, W_n(m)) \right\}_{m=1}^M.
\]

Upon observing the source \( \mathbb{x} \), the transmitter searches through the source codebook \( C_X \) 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 \( \sim P_W \).

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 $U_k$ and side information $Y_k$, receiver $k$ outputs the list

$$L_k := \{X(m) \in C_X : (X(m), Y_k) \in T^n_e(P_{XY_k}) \text{ and } (\bar{X}(m), U_k) \in T^n_d(P_{W,U_k})\}. \quad (20)$$

An error is declared at receiver $k$ if the source is not in the list $X \not\in L_k$ or the list is too large

$$|L_k| > 2^{nD_k}.$$  

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

Denote the event of an error at any receiver by

$$\mathcal{E} := \bigcup_{k=1}^{K} \left( \{X \not\in L_k\} \cup \{|L_k| > 2^{nD_k}\} \right). \quad (21)$$

By the union bound,

$$P[\mathcal{E}] \leq \sum_{k=1}^{K} \left( P[X \not\in L_k] + P[|L_k| > 2^{nD_k}] \right). \quad (22)$$

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

$$P[\mathcal{E}] \leq b 2^{-a_1 n}, \quad (23)$$

for some finite $a, b > 0$, whenever $\epsilon$ and $\delta$ satisfy (20) and

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

Therefore, for any $\epsilon^* > 0$ there exists an $(n, L_1, L_2, \ldots, L_K)$ list code such that $P[X \not\in L_k] \leq \epsilon^*$ and $|L_k| \leq 2^{nD_k}$ 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 $X$ is not in the source codebook

The probability that the source is not in the source codebook $P[X \not\in C_X]$ is bounded by

$$P[X \not\in C_X] \leq b_1 \exp \left( -\frac{2^n \epsilon_1 n}{2 \ln 2} \right), \quad (24)$$

where

$$a_1 := \min \left\{ \epsilon_1 n \mu_X, (H(X) : (\epsilon - \epsilon_1)) / \ln 2 \right\}.$$  

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

The steps leading to (24) are

$$P[X \not\in C_X] \leq P[X \not\in T^n_e(P_X)] + P[X \not\in C_X | X \in T^n_e(P_X)]$$

$$\leq 2|X| \exp \left( -n \epsilon_1 \mu_X \right) + P \left[ \bigcap_{m=1}^{N} (X(m) \neq \bar{X}(m)) \right] \quad (25)$$

Notes:

- a. apply Lemma 4;
- b. the codewords in $C_X$ are independent.

\[
\begin{align*}
P[S_1] & \leq P \left[ \bigcup_{m=1}^{M} \left\{ X(m) \in T^n_e(P_{XY}) \cap \bigcup_{m=1}^{M} \{ X \neq X(m'), \forall m' < m \} \cap \{ X = X(m) \} \right\} \right] \\
& \leq \sum_{m=1}^{M} P \left[ \{ X \neq L_k \} \cap \{ (X, Y_k) \in T^n_e(P_{XY}) \} \cap \{ X \neq X(m'), \forall m' < m \} \cap \{ X = X(m) \} \right] \\
& \leq \sum_{m=1}^{M} P \left[ \{ X \neq X(m), \forall m' < m \} \cap \{ X = X(m) \} \cap \{ (X, Y_k) \in T^n_e(P_{XY}) \} \right] \\
& \leq \sum_{m=1}^{M} P \left[ \{ X = X(m) \} \cap \{ X \in T^n_e(P_X) \} \cap \{ (W, U_k) \not\in T^n_d(P_{W,U}) \} \right] \\
& \leq \sum_{m=1}^{M} P \left[ \{ X = X(m) \} \cap \{ X \in T^n_e(P_X) \} \cap \{ (W, U_k) \not\in T^n_d(P_{W,U}) \} \right] \\
& \leq M 2|W| |U_k| 2^{-nH(X)(1-\epsilon_1)} 2^{-n \frac{3^2 \mu_W U_k}{m^2}} \\
& \leq 2|W| |U_k| 2^{-2nH(X)} 2^{-n \frac{3^2 \mu_W U_k}{m^2}}.
\end{align*}
\]
c. use Lemma 4 with $X(m)$ iid $\sim P_X$;
d. use the inequality
\[(1-c)^M \leq e^{-cM}, \quad \forall M \geq 1, \ c \in [0, 1]; \text{ and}
\]
e. bound $M$ via (21).

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

**G. Error Analysis: Probability $X$ is not in receiver $k$’s list $L_k$**

Consider the probability that the source $X$ is not in receiver $k$’s list $L_k$. We have
\[
P[X \notin L_k] \leq P[X \notin C_X] + P[(X, Y_k) \notin T_{\epsilon_1}(P_{X,Y_k})] + P[S_1],
\]
(26)
where
\[S_1 := \{(X, Y_k) \notin T_{\epsilon_1}(P_{X,Y_k}) \} \cap \{X \notin C_X\}.
\]
The probability $P[S_1]$ is bounded from above by
\[
P[S_1] \leq 2|W||U_k| 2^{-a_2 n}
\]
(27)
where
\[a_2 := d^2 \mu_{W,U_k}/\ln 2 - 2\epsilon_1 H(X)
\]
is positive by (20). The steps leading to (27) are described above in (28). Notes for (28):
a. expand the event that the source $X$ appears in $C_X$;
b. union bound;
c. Bayes’ law and $P[(X, Y_k) \in T_{\epsilon_1}^n(P_{X,Y_k})] \leq 1$;
d. conditioned on $(X, Y_k)$ typical, $X = X(m)$ and $W = W(m)$, the error $X \notin L_k$ occurs if and only if $(W, U_k)$ 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
\[
P[X \notin L_k] \leq b_3 2^{-a_3 n},
\]
(29)
for some finite $a_3, b_3 > 0$.

**H. Error Analysis: Probability receiver $k$’s list $L_k$ is too large**

Now consider the probability that the size of list $L_k$ is too large. We start with
\[
P[|L_k| > 2^{nD_k}] \leq P[X \notin C_X] + P[(X, Y_k) \notin T_{\epsilon_1}] + P[(W, U_k) \notin T_{\epsilon_1}^n] + P[|L_k| > 2^{nD_k} | S_2],
\]
(30)
where
\[S_2 := \{X \notin C_X\} \cap \{(X, Y_k) \notin T_{\epsilon_1}\} \cap \{(W, U_k) \notin T_{\epsilon_1}^n\}
\]
and $T_{\epsilon_1}^n(P_{X,Y_k})$ and $T_{\epsilon_1}^n(P_{W,U_k})$ have been abbreviated by $T_{\epsilon_1}^n$ and $T_{\epsilon_1}^n$ respectively. Apply Markov’s inequality to the rightmost probability in (30) to get
\[
P[|L_k| > 2^{nD_k} | S_2] \leq 2^{-nD_k} E[|L_k| | S_2],
\]
(31)
where the expectation is understood to be
\[
E[|L_k| | S_2] := \sum_l l \cdot P[|L_k| = l | S_2].
\]

We now expand the above expectation over $X \in C_X$ (the $M$ possible encodings of $X$) to get
\[
E[|L_k| | S_2] = \sum_{m=1}^{M} E[|L_k| | S_2 \cap \{X \neq X(m'), \forall m' < m\} \cap \{X = X(m)\}]
\]
\[
\cdot P[\{X \neq X(m'), \forall m' < m\} \cap \{X = X(m)\} | S_2].
\]
(32)
Consider the expectation on the right hand side of (32). Let $S_{2,m} := S_2 \cap \{X \neq X(m'), \forall m' < m\} \cap \{X = X(m)\}$.

We have
\[
E[|L_k| | S_{2,m}] = \sum_{m=1}^{M} E[|L_k| | S_{2,m} \cap \{(X(m), Y_k) \in T_{\epsilon_1}^n\}] \cdot P[\{(X(m), Y_k) \in T_{\epsilon_1}^n\} | S_{2,m}]
\]
\[
= \sum_{m=1}^{M} E[|L_k| | S_{2,m} \cap \{(X(m), Y_k) \in T_{\epsilon_1}^n\}] \cdot P[\{(X(m), Y_k) \in T_{\epsilon_1}^n\} | S_{2,m}]
\]
\[
= \sum_{m=1}^{M} E[|L_k| | S_{2,m} \cap \{(X(m), Y_k) \in T_{\epsilon_1}^n\}] \cdot P[\{(X(m), Y_k) \in T_{\epsilon_1}^n\} | S_{2,m}]
\]
\[
\leq \alpha_n 2^{-n(I(X(Y)k) - 2\epsilon_1 H(X))}.
\]
(34)

Notes:
a. codewords and codebook are generated independently;
b. Bayes’ law and the trivial bound $P[X \neq X(m)] \leq 1$; and
c. apply Lemmas 4 and 5 respectively to the denominator and numerator in step (b) and set
\[
\alpha_n := \frac{\exp(2nH(X)(1-\epsilon_1))}{\exp(2nH(X)(1-\epsilon_1)) - 1}.
\]
(35)

Similarly, by Lemma 5, the rightmost probability in (33) is bounded by
\[
P[\{(W(\tilde{m}), U_k) \in T_{\epsilon_1}^n| S_{2,m} \cap \{(\tilde{X}(\tilde{m}), Y_k) \in T_{\epsilon_1}^n\}]
\]
\[ \leq 2^{-n(I(W;U_k) - 2\delta H(W))}. \]  
(36)

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

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

\[ \mathbb{P}[(X(\tilde{m}), Y_k) \in T^e_{\rho}|S_{2,m}] \leq 2^{-n(I(X;Y_k) - 2\epsilon H(X))}. \]  
(37)

Similarly,

\[ \mathbb{P}[(X(\tilde{m}), U_k) \in T^e_{\rho}|S_{2,m} \cap \{(X(\tilde{m}), Y_k) \in T^e_{\rho}\}] \leq 2^{-n(I(W;U_k) - 2\delta H(W))}. \]  
(38)

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

\[ \mathbb{E}[|L_k| |S_{2,m}] \leq 1 + \alpha_n 2^n(H(X|Y_k) - I(W;U_k))2^n(3\epsilon H(X) + 2\delta H(W)), \]  
(39)

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

\[ \mathbb{P}[|L_k| > 2^{nD_k}] \leq \mathbb{P}[X \notin C_X] + \mathbb{P}[(X; Y_k) \notin T^n_e] + \mathbb{P}[(X; U_k) \notin T^n_e] + \alpha_n 2^{-n(D_k - H(X|Y_k) + I(W;U_k))2^n(3\epsilon H(X) + 2\delta H(W))} + 2^{-nD_k}. \]  
(40)

Lemma 4 and (24) imply

\[ \mathbb{P}[|L_k| > 2^{nD_k}] \leq b_d 2^{-a_n}, \]  
(41)

for some finite \(a_d, b_d > 0\) 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 \(\delta_1\) satisfy (20). The result follows because \(\epsilon\) and \(\delta\) can be chosen arbitrarily small and \(H(X)|H(W)\) are finite. \[\Box\]

**VII. PROOF OF THEOREM 1 — CONVERSE**

Fix \(\epsilon > 0\). Consider any \((n_k, n_c, R_1, R_2, \ldots, R_K)\)-code with \(\mathbb{P}[X_k = 0] \leq \epsilon\) for all \(k\). Recall the timeshared pmf \(P_W\) on \(\mathcal{W}\), defined in (17). Let

\[ (\tilde{W}, \tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_K) \sim P_W(\cdot) \mathcal{T}(\cdot) \]

and \(\tilde{V}_k = f_k(\tilde{W})\). Mimicking the steps of Section V:

\[ n_c I(\tilde{W}; \tilde{U}_k, \tilde{V}_k) \geq I(Y_k; U_k, \tilde{V}_k) \geq I(Y_k; U_k, M_k) \]  
(42)

where (a) and (b) use \(M_k \leftrightarrow (U_k, Y_k) \leftrightarrow \mathcal{W}\) and \((X, Y) \leftrightarrow \mathcal{W} \leftrightarrow (U_k, M_k)\). Similarly,

\[ n_c I(\tilde{W}; \tilde{U}_k) + n_c R_k \]  
(43)

After applying Fano’s inequality [23, Thm. 2.10.1] to \(H(X|U_k, \tilde{U}_k, M_k)\) 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 \(n_c = n_c = n\). 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 \(P_\mathcal{W}\) on \(\mathcal{W}\) and let us assume that for all \(k\)

\[ H(X|Y_k) < I(W;U_k) + \min \left\{ R_k, I(W;V_k|U_k) \right\} \]  
(44)

and

\[ H(X|Y_k) \geq I(W;U_k). \]  
(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 \(X\). 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 \(C_X\) and \(C_\mathcal{W}\) denote the source and channel codebooks respectively. Fix \(\epsilon, \epsilon_1, \delta\) and \(\delta_1\) arbitrarily small, but always satisfying (20). For each receiver \(k\), choose any list exponent \(D_k\) satisfying

\[ H(X|Y_k) - I(W;U_k) < D_k \]  
(46)

and set the helper rate to be

\[ R_k = D_k + \epsilon_h \]  
(47)

for some arbitrarily small \(\epsilon_h > 0\). 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 \(\delta\) arbitrarily small; and \(H(W)\) is finite.

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

\[ C_{V_k} := \bigcup_{m=1}^{M} \{ f_k(\mathcal{W}(m)) \}, \]

where

\[ \phi_k(\mathcal{W}(m)) = (\phi_k(W_1(m)), \phi_k(W_2(m)), \ldots, \phi_k(W_n(m))) \]

is a slight abuse of notation.

Uniformly at random place each codeword in \(C_{V_k}\) into one of \([2^{n_{R_k}}]\) bins. Uniquely label each bin with an index from the set \(\{1, 2, \ldots, [2^{n_{R_k}}]\}\), and let \(f_k(v)\) denote the bin index of the codeword \(v \in C_{V_k}\). Denote the set of all codewords in the \(b\)-th bin by \(S_b(b) := \{ v \in C_{V_k} : f_k(v) = b \}\) for \(b \in \{1, 2, \ldots, [2^{n_{R_k}}]\}\).
B. Encoding and Decoding

The list encoder and decoders operate as before, see Sections VI-C and VI-D. Helper BS(k) looks for \( \forall k = \phi_k(v) \) in \( \mathcal{C}_V \) and, if successful, sends the bin index \( B = f_k(\mathcal{V}_k) \) 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:

\[
\mathcal{L}_k^*: = \{ v \in \mathcal{C}_V : \exists \mathcal{X}(m) \in \mathcal{L}_k \text{ with } v = \phi_k(\mathcal{V}(m)) \}.
\]

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

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

C. Error Analysis

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

\[
P[\hat{\mathcal{X}}_k \neq \mathcal{X}] \leq P[\hat{\mathcal{X}}_k \neq \mathcal{V}_k] + P[\{\hat{\mathcal{X}}_k = \mathcal{V}_k\} \cap \{\hat{\mathcal{X}}_k \neq \mathcal{X}\}],
\]

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

\[
P[\hat{\mathcal{X}}_k \neq \mathcal{V}_k] \leq P[\mathcal{E}_k] + P[\mathcal{X} \notin \mathcal{C}_X] + P[\hat{\mathcal{X}}_k \neq \mathcal{V}_k | \mathcal{E}_k \cap \{\mathcal{X} \in \mathcal{C}_X\}].
\] (49)

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

\[
P[\hat{\mathcal{X}}_k \neq \mathcal{V}_k | \mathcal{E}_k \cap \{\mathcal{X} \in \mathcal{C}_X\}] = \sum_{b=1}^{[2^{nR_k}]} P[f_k(\mathcal{V}_k) = b \cap \hat{\mathcal{X}}_k \neq \mathcal{V}_k | \mathcal{E}_k \cap \{\mathcal{X} \in \mathcal{C}_X\}]
\]

(50)

We now fix a bin \( b \) and derive

\[
P[\hat{\mathcal{X}}_k \neq \mathcal{V}_k | \{f_k(\mathcal{V}_k) = b\} \cap \mathcal{E}_k \cap \{\mathcal{X} \in \mathcal{C}_X\}] = \sum_{v \in \mathcal{L}_k^*} P[\{f_k(\mathcal{V}_k) = b\} \cap \mathcal{E}_k \cap \{\mathcal{X} \in \mathcal{C}_X\}]
\]

\[
\mbox{if } v \neq \phi_k(v)
\]

(57)

\[
P[\hat{\mathcal{X}}_k \neq \mathcal{X}] \leq \gamma_n^{-2} \cdot \left(2^n - \frac{1}{2} + \lambda_n\right)
\]

(59)
\[ \mathbb{P}[\{\hat{W}(m), U_k, \hat{V}_k\} \in T_{\delta} | S_{3,m}] \]
\[ \leq \mathbb{P}[\{\hat{W}(\tilde{m}), U_k, \tilde{V}_k\} \in T_{\delta} | \{\hat{W}(m), U_k, \tilde{V}_k\} \in T_{\delta} \} \cap \{\{\hat{W}(\tilde{m}), U_k, \tilde{V}_k\} \in T_{\delta} \} \cap \{\{\hat{W}(m), U_k, \tilde{V}_k\} \in T_{\delta} \} \cap \{\{\hat{W}(\tilde{m}), U_k, \tilde{V}_k\} \in T_{\delta} \} \]}

The right-hand side of (51) is independent of \( P = 2 \).

c.

d.

\[ v \mid L \text{ into } v \mid l \]

1. \( \mathbb{P}[\mathcal{L}_l = l |_| S_{3,m}] \left( \sum_{m=1}^{M} \mathbb{P}[\{X \neq X(m'), \forall m' < m \} \cap \{X = X(m)\} | S_{3} \cap \mathcal{L}_k = l] \right) = 1. \]}

\[ \sum_{l:|l| \leq 2^{nD_k}} \mathbb{P}[\mathcal{L}_l = l |_| S_{3,m}] \left( \sum_{m=1}^{M} \mathbb{P}[\{X \neq X(m'), \forall m' < m \} \cap \{X = X(m)\} | S_{3} \cap \mathcal{L}_k = l] \right) = 1. \]}

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

\[ \mathbb{P}[\hat{X}_k \neq X | S_{3,m}] \leq \gamma_n 2^{-n(I(W; V_k | U_k) - D_k + 4\delta H(W))} \]

where

\[ \gamma_n := \frac{\exp\left(2n(1 - \delta_1) W^2_{W,U_k} \right)}{\exp\left(2n(1 - \delta_1) W^2_{W,U_k} \right) - 2W^{||U_k||}} \]

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

a. Write \( \mathbb{P}[\hat{X}_k \neq X | S_{3,m}] \) 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 \( S_3 \).

b. Let

\[ S_{3,m} := \{X \neq X(m'), \forall m' < m \} \cap \{X = X(m)\} \cap \mathcal{L}_k = l \}

and note that \( X \in C_X \), after conditioning on \( S_3 \).

c. An error may only occur if there is some other index \( \tilde{m} \neq m \) such that \( \hat{X}(\tilde{m}) \in \mathcal{L}_k \) and \( (\hat{W}(\tilde{m}), U_k, \tilde{V}_k) \) is jointly typical. Here we note that after conditioning on \( S_{3,m} \) the following holds with probability one: the source \( X \) is equal to the \( m \)-th codeword \( X(m) \) in the source codebook \( C_X \); the \( m \)-th channel codeword is transmitted \( W = W(m) \); the source and side information \( (X, Y_k) \) are \( \epsilon \)-jointly typical; \( (W, U_k, V_k) \) are strictly typical; and \( \hat{V} = \hat{V} \). We have also abbreviated \( T_{\epsilon}(P_{X,Y_k}) \) and \( T_{\delta_1}(P_{W,U_k,V_k}) \) as \( T_{\epsilon} \) and \( T_{\delta_1} \) respectively.

d. Apply the union bound.

e. The rightmost probability in step (d) is bounded by

\[ \mathbb{P}[\{\hat{W}(\tilde{m}), U_k, \tilde{V}_k\} \in T_{\delta} | S_{3,m}] \leq \gamma_n 2^{-n(I(W; V_k | U_k) - 4\delta H(W))} \]

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 \(|l| \leq 2^{nD_k} \),

\[ \sum_{m \notin l} \mathbb{P}[\{\hat{W}(m), U_k, \hat{V}_k\} \in T_{\delta} | S_{3,m}] \]

(58)
\[ \leq \gamma n 2^n D_k 2^{-n(I(W; V_k | U_k) - 4\delta H(W))}. \]

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

Notes for (58):
e. This step follows from the independence of the source and channel codebooks, the independence of codewords within each codebook, conditioning on \( S_i, m \) and \( \{L_k = l\} \) being equivalent to
\[(X(m'), \gamma_k) \in T_c \text{ and } (X(m'), \Upsilon_k) \in T_\delta, \forall m' \in l, \]
and
\[(X(m'), \gamma_k) \notin T_c \text{ or } (X(m'), \Upsilon_k) \notin T_\delta, \forall m' \notin l. \]
e. Bayes' rule.
e. Apply Lemma 5 to (e.2).

Thus,
\[ P[X_k \neq \hat{X}|S_3] \leq b_7 2^{-a_7 n}, \]
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 achievable 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 \( P[X_k \neq \hat{X}] \) 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 \((n_k, n_e, R_1, R_2, \ldots, R_K)\)-code with \( P[X_k \neq \hat{X}] \leq \epsilon \) for all \( k \). Following the familiar path of defining \((W, \hat{U}_1, \hat{U}_2, \ldots, \hat{U}_K) \sim P_\hat{W}(.)T(\cdot), \) with the timeshared pmf \( P_\hat{W} \) given in (17), we have
\[ n_k I(\hat{W}; \hat{U}_k) \geq I(\hat{W}; \gamma_k, M_k, W; U_k) \geq H(X|M_k, Y_k) - H(X|M_k, Y_k, U_k) \]
\[ = \sum_{i=1}^{n_e} H(X_i|M_k, \gamma_k, X_i^{i-1}) - n_e \epsilon(n_e) \]
\[ = n_e \sum_{i=1}^{n_e} H(X_i|M_k, \gamma_k, Y_k, V_{k,i}^{i-1}) - n_e \epsilon(n_e) \]
\[ = \sum_{i=1}^{n_e} \sum_{j=1}^{n_e} H(X_i|M_k, \gamma_k, Y_k, V_{k,i}^{i-1}) - n_e \epsilon(n_e) \]
\[ = \sum_{i=1}^{n_e} H(X_i|M_k, \gamma_k, Y_k, V_{k,i}^{i-1}) - n_e \epsilon(n_e). \]

A. Code Construction

Fix a pmf \( P_W \) on \( W \) and auxiliary random variables \((A_1, A_2, \ldots, A_K)\) satisfying \( A_k \leftrightarrow V_k \leftrightarrow (X_i, Y_k) \). Let us assume that
\[ H(X|A_k, Y_k) < I(W; U_k) \quad \forall k \]
and
\[ R_k > I(A_k; V_k|Y_k) \quad \forall k. \]

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 \( \epsilon, \epsilon_1, \delta \) and \( \delta_1 \) satisfying (20), and choose \( 0 < \epsilon_1 < \epsilon_1 < \epsilon_1 < \mu_{A_k, X_i, Y_i} \). Generate a random list code, as described in Section VI, with the parameters described above, and let \( C_X \) and \( C_W \) 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
\[ \rho > 3cH(X) + 2\delta H(W). \] (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 \( A_k \) iid \( \sim P_{A_k} \):
\[
C_{A_k} := \left\{ \hat{A}_k(j, j') = (A_{k,1}(j, j'), A_{k,2}(j, j'), \ldots, A_{k,n}(j, j')) \right\}
\]
where we call \( j \) the bin index and \( j = 1, 2, \ldots, \left[ 2^nR_k \right] \) and \( j' = 1, 2, \ldots, \left[ 2^{n(I(A_k;Y_k) - \epsilon_k)} \right] \).

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 \( C_{A_k} \) for a pair \((J, J')\) such that \((\hat{A}_k(J, J'), Y_k) \in \mathcal{T}_{S_1}\). 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 \( \hat{A}_k(J, J') \) by looking for a unique \( \hat{X}_k = \hat{A}_k(J, J') \) in \( \mathcal{T}_{S_1} \). If successful, receiver \( k \) sets \( \hat{X}_k = \hat{A}_k(J, J) \); otherwise, it randomly selects \( \hat{A}_k \) iid \( \sim P_{A_k} \).

The list decoder at receiver \( k \) outputs \( \mathcal{L}_k \), see (20). Receiver \( k \) looks for a unique \( \hat{X}(m') \in \mathcal{L}_k \) such that \((\hat{X}(m'), Y_k, \hat{A}_k) \in \mathcal{T}_e \). If successful, receiver \( k \) outputs \( \hat{X}_k = \hat{X}(m') \); otherwise, it randomly generates \( \hat{X}_k \) using \( P_X \).

C. Error Analysis

We first bound the probability of error at receiver \( k \) by
\[
P[\hat{X}_k \neq X] \leq P[S_4^e] + P[\hat{X}_k \neq X|S_4],
\]
where
\[
S_4 := \{X \in \mathcal{C}_X \cap \{\hat{A}_k = A_k\} \cap \{\mathcal{L}_k \leq 2^nD_k\} \cap \{\mathcal{X}(X, Y_k, \hat{A}_k) \in \mathcal{T}_{e_1}\} \}
\]
and we have abbreviated the typical sets \( \mathcal{T}_{e_1}(P_{X,Y_k,A_k}) \) and \( \mathcal{T}_{e_1}(P_{W,U_k}) \) as \( \mathcal{T}_{e_1} \) and \( \mathcal{T}_{e_1} \), respectively.

We have
\[
P[S_4^e] \leq b_7 2^{-a_7n},
\] (67)
for some finite \( a_7, b_7 > 0 \). To see (67), apply the union bound to \( P[S_4^e] \); use (25) to bound \( P[X \notin \mathcal{C}_X] \); use (40), (41) and (66) to bound \( P[|\mathcal{L}_k| > 2^nD_k] \); and use Lemmas 4 and 5 to bound \( P[|\mathcal{W}, U_k| \notin \mathcal{T}_{e_1}] \). The final two probabilities, \( P[\hat{A}_k \neq A] \) and \( P[(X, Y_k, \hat{A}_k) \notin \mathcal{T}_{e_1}] \), also tend to zero exponentially in \( n \) by Lemmas 4 and 5; see, for example, Kramer’s achievability proof of the Wyner-Ziv theorem [33, Sec. 5.3]. Finally, the conditional probability \( P[\hat{X}_k \neq X|S_4] \) tends to zero exponentially in \( n \), as shown below in (68).

Notes for (68):

a. Write \( P[\hat{X}_k \neq X|S_4] \) 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 \( S_4 \).

b. Write the second conditional probability in step (a) as an expectation over all possible encodings of \( X \). Here we note that \( X \in \mathcal{C}_X \) with probability one, after conditioning on \( S_4 \).

c. In the rightmost conditional probability in step (b), the error event \( \hat{X}_k \neq X \) 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 \((X(\tilde{m}), Y_k, \hat{A}_k) \) is jointly typical. Here we note that the correct index \( m \) is in decoder \( k \)'s list with probability one, after conditioning on \( S_4 \). 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
\[
P\left[\mathcal{X}(\tilde{m}), Y_k, \hat{A}_k \right] \in \mathcal{T}_e \left| \left\{ \left( X(\tilde{m}), Y_k \right) \in \mathcal{T}_e \right\} \right. \cong \left\{ \left( X, Y_k, \hat{A}_k \right) \in \mathcal{T}_{e_1} \right\}.
\]
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
\[
(\mathcal{X}(m'), Y_k) \in \mathcal{T}_e \text{ and } (\mathcal{W}(m'), U_k) \in \mathcal{T}_d, \forall m' \in l,
\]
and
\[
(\mathcal{X}(m'), Y_k) \notin \mathcal{T}_e \text{ or } (\mathcal{W}(m'), U_k) \notin \mathcal{T}_d, \forall m' \notin l.
\]
e.2. Apply Bayes’ law twice and use the upper bound
\[
P\left[\mathcal{X} \neq X(\tilde{m}) \right] \cap \{X(\tilde{m}), Y_k \in \mathcal{T}_e\} \{X(\tilde{m}), Y_k, \hat{A}_k \in \mathcal{T}_{e_1}\} \leq 1
\]
e.3. Use Lemmas 4 and 5 to lower bound
\[
1 - P[\mathcal{X} = X(\tilde{m})|\{X(\tilde{m}), Y_k \in \mathcal{T}_e\} \cap \{X, Y_k, \hat{A}_k \in \mathcal{T}_{e_1}\}]
\]
Use Lemma 5 to bound the numerator and denominator of the rightmost term in step (e.2).

e.4. Set
\[
\gamma_n := \frac{1}{1 - \frac{1}{1-cH(X|Y_k)} 2^{-nH(X|Y_k) - 3cH(X)}},
\]
where
\[
\zeta_n := 2|X||Y_k| \exp \left( -2n(1 - \epsilon_1) \frac{(\epsilon - \epsilon_1)^2}{1 + \epsilon_1} 2^{X,Y_k} \right).
\]
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 \( \epsilon \) and \( \rho \) arbitrarily small, and \( H(X) \) is finite.
\[
\begin{align*}
P[\hat{\mathbf{x}}_k \neq \mathbf{x}|S_4] & \leq \sum_{l: |l| \leq 2^n D_k} P[\mathcal{L}_k = l|S_4] P[\hat{\mathbf{x}}_k \neq \mathbf{x}|S_4 \cap \{\mathcal{L} = l\}] \\
& \leq \sum_{l: |l| \leq 2^n D_k} P[\mathcal{L}_k = l|S_4] \left( \sum_{m=1}^{M} P[\mathbf{x} \neq \mathbf{x}(m'), \forall m' < m] \cap \{\mathbf{x} = \mathbf{x}(m)\} |S_4 \cap \{\mathcal{L}_k = l\} \right) \\
& \quad \cdot P[\hat{\mathbf{x}}_k \neq \mathbf{x}|\{\mathbf{x} = \mathbf{x}(m)\} \cap \{\mathbf{x} \neq \mathbf{x}(m'), \forall m' < m\} \cap \{\mathcal{L}_k = l\}] \\
& \leq \sum_{l: |l| \leq 2^n D_k} P[\mathcal{L}_k = l|S_4] \left( \sum_{m=1}^{M} P[\mathbf{x} \neq \mathbf{x}(m'), \forall m' < m] \cap \{\mathbf{x} = \mathbf{x}(m)\} |S_4 \cap \{\mathcal{L}_k = l\} \right) \\
& \quad \cdot \left( \sum_{\tilde{m} \in \mathcal{T}_e} \left\{ \left( \mathbf{x}(\tilde{m}), \mathbf{y}, \hat{\mathbf{a}} \right) \in \mathcal{T}_e \right\} \{\mathbf{x} \neq \mathbf{x}(m'), \forall m' < m\} \cap \{\mathbf{x} = \mathbf{x}(m)\} \cap \{\mathcal{L}_k = l\} \right) \\
& \leq \gamma_n 2^{-n(I(X;A_k|Y_k)-D_k-4eH(X)} \\
& \leq \gamma_n 2^{-n(I(W;U_k)-H(X|A_k,Y_k)-4eH(X)-\rho)},
\end{align*}
\]

\[
\begin{align*}
P[\mathbf{x}(\tilde{m}), \mathbf{y}, \hat{\mathbf{a}} \in \mathcal{T}_e | \{\mathbf{x} \neq \mathbf{x}(m'), \forall m' < m\} \cap \{\mathbf{x} = \mathbf{x}(m)\} \cap \{\mathcal{L}_k = l\}] & \leq P[\mathbf{x}(\tilde{m}), \mathbf{y}, \hat{\mathbf{a}} \in \mathcal{T}_e | \{\mathbf{x} = \mathbf{x}(m)\} \cap \{\mathbf{x} \neq \mathbf{x}(m'), \forall m' < m\} \cap \{\mathcal{L}_k = l\}] \\
& \leq \frac{1}{P[\mathbf{x} \neq \mathbf{x}(\tilde{m}) | \mathcal{T}_e \cap \{\mathbf{x} \neq \mathbf{x}(m'), \forall m' < m\} \cap \{\mathcal{L}_k = l\}]} \\
& \left( \frac{P[\mathbf{x}(\tilde{m}), \mathbf{y}, \hat{\mathbf{a}} \in \mathcal{T}_e | \{\mathbf{x} = \mathbf{x}(m)\} \cap \{\mathcal{L}_k = l\}]}{P[\mathbf{x}(\tilde{m}), \mathbf{y}, \hat{\mathbf{a}} \in \mathcal{T}_e | \mathcal{T}_e \cap \{\mathbf{y} \in \mathcal{T}_1\}] \right) \\
& \leq \frac{1}{\frac{1}{\gamma_n} 2^{-n(H(X|Y_k)-3eH(X))}} \left( \frac{2^{-n(I(X;A_k|Y_k)-2eH(X))}}{2^{-n(I(X;Y_k)+2eH(X))}} \right) \\
& \leq \gamma_n 2^{-n(I(X;A_k|Y_k)-4eH(X))}
\end{align*}
\]

ACKNOWLEDGEMENTS

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

REFERENCES


