Error Exponents for the Relay Channel

Vincent Y. F. Tan

Institute for Infocomm Research (I²R), A*STAR, Email: tanyfv@i2r.a-star.edu.sg
ECE Dept., National University of Singapore (NUS), Email: vtan@nus.edu.sg

Abstract—Achievable error exponents for the relay channel are derived using the method of types. In particular, two block-Markov coding schemes are analyzed: partial decode-forward and compress-forward. The derivations require combinations of the techniques in the proofs of the packing lemma for the error exponent of channel coding and the covering lemma for the error exponent of source coding with a fidelity criterion.

Index Terms—Relay channel, Error exponent, Partial decode-forward, Compress-forward

I. INTRODUCTION

We derive achievable error exponents for the discrete memoryless relay channel. This channel, introduced in [1], is a point-to-point communication system consisting of a sender X_1, a receiver Y_3 and a relay with input Y_2 and output X_2. The capacity is not known in general but there exists several coding schemes that are optimal for certain classes of relay channels, e.g., degraded. These coding schemes, introduced in the seminal work by Cover-El Gamal [2] include decode-forward (DF), partial decode-forward (PDF) and compress-forward (CF). Using PDF, the capacity is lower bounded as

$$C \geq \max \min\{I(X_1 X_2; Y_3), I(UY_2|X_2)+I(X_1; Y_3|X_2 U)\} \tag{1}$$

where the maximization is over all $P_{UX_1X_2}$. DF is a special case of PDF in which $U = X_1$ and instead of decoding part of the message as in PDF, the relay decodes the entire message. In CF, a more complicated coding scheme, the relay sends a description of Y_2 to the receiver. It uses Y_3 as side information à la Wyner-Ziv [3, Ch. 11] to reduce the rate of the description. One form of the CF lower bound writes

$$C \geq \max \min\{I(X_1; \hat{Y}_2 Y_3|X_2),$$
$$I(X_1 X_2; Y_3) - I(Y_2; \hat{Y}_2|X_1 X_2 Y_3)\} \tag{2}$$

where the maximization is over P_{X_1}, P_{X_2} and $P_{\hat{Y}_2|X_2 Y_2}$. Both PDF and CF involve *block-Markov coding* in which the channel is used $N = nb$ times over b blocks, each involving an independent message to be sent and the relay codeword in block j depends statistically on the message from block $j-1$.

In addition to capacities, in information theory, error exponents are also of tremendous interest. They quantify the exponential rate of decay of the error probability when the rate of the code is below capacity. Such results allow us to provide rough bounds on the blocklength needed to achieve a certain rate. By using maximum mutual information (MMI) decoding [4], [5], the error exponent we derive for PDF is universally attainable, i.e., the decoder does not need to know the channel statistics. Our two main contributions here are the

derivations of error exponents for PDF and CF (though for CF, the decoder is not universal). For CF, a key technical contribution is the taking into account of the conditional correlation between $\hat{Y}_2$ and X_1 (given X_2) using a technique introduced in [6].

A. Related Work

The work that is most closely related to this paper is [7] in which the authors derived the error exponent for sliding-window DF based on Gallager's Chernoff-bounding techniques [8]. We generalize their result to PDF and we use MMI decoding [4]. For PDF, our work leverages on the proofs of the various forms of the packing lemmas for multiuser channels in Haroutunian et al. [9]. For CF, since it is related to Wyner-Ziv, we leverage on the work of Kelly-Wagner [10] who derived an achievable exponent for Wyner-Ziv. In a similar vein, [11] and [12] derived lower bounds for the error exponents of Gel'fand-Pinsker and content identification respectively. We also note that Ngo et al. [13] presented an achievable error exponent for amplify-forward for the AWGN relay channel but does not take into account block-Markov coding.

II. NOTATIONS AND SYSTEM MODEL

We adopt the notation from Csiszár and Körner [5]. Random variables (e.g., X) and their realizations (e.g., x) are in capital and small letters respectively. All random variables take values on finite sets, denoted in calligraphic font (e.g., $\mathcal{X}$). For a sequence $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$, its *type* is the distribution $P(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{x = x_i\}$. The set of types with denominator n supported on alphabet $\mathcal{X}$ is denoted as $\mathscr{P}_n(\mathcal{X})$. The *type class* of P is denoted as $\mathcal{T}_P := \{x^n \in \mathcal{X}^n : x^n \text{ has type } P\}$. For $x^n \in \mathcal{T}_P$, the set of sequences $y^n \in \mathcal{Y}^n$ such that (x^n, y^n) has joint type $P \times V$ is the V-*shell* $\mathcal{T}_V(x^n)$. Let $\mathscr{V}_n(\mathcal{Y}; P)$ be the family of stochastic matrices $V : \mathcal{X} \to \mathcal{Y}$ for which the V-shell of a sequence of type $P \in \mathscr{P}_n(\mathcal{X})$ is not empty. Information-theoretic quantities are denoted in the usual way. For example $I(X; Y)$ and $I(P, V)$ denote the mutual information where the latter expression makes clear that the joint distribution of (X, Y) is $P \times V$. In addition, $\hat{I}(x^n \wedge y^n)$ is the empirical mutual information of (x^n, y^n), i.e., if $x^n \in \mathcal{T}_P$ and $y^n \in \mathcal{T}_V(x^n)$, then, $\hat{I}(x^n \wedge y^n) = I(P, V)$. Finally, $|a|^+ := \max\{a, 0\}$ and $[a] := \{1, \ldots, \lceil a \rceil\}$.

Definition 1. *A 3-node discrete memoryless relay channel (DM-RC) is a tuple* $(\mathcal{X}_1 \times \mathcal{X}_2, W, \mathcal{Y}_2 \times \mathcal{Y}_3)$ *where* $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}_2 \times \mathcal{Y}_3$ *is a stochastic matrix. The sender (node 1) wishes to communicate a message* M *to the receiver (node 3) with the help of the relay node (node 2).*

Definition 2. *A $(2^{nR}, n)$ code for the DM-RC consists of a message set $\mathcal{M} = [2^{nR}]$, an encoder $f : \mathcal{M} \to \mathcal{X}_1^n$ that assigns a codeword to each message, a sequence of relay encoders $g_i : \mathcal{Y}_2^{i-1} \to \mathcal{X}_2, i \in [n]$ each assigning a symbol to each past received sequence and a decoder $\varphi : \mathcal{Y}_3^n \to \mathcal{M}$ that assigns an estimate of the message to each channel output.*

We assume that M is uniformly on $\mathcal{M}$ and the channel is memoryless. The average error probability is $\mathbb{P}(\varphi(Y_3^n) \neq M)$.

As in [7], for both PDF and CF, we will use block-Markov coding to send a message M representing NR_{eff} bits of information. We use the channel $N = nb$ times and this total blocklength is partitioned into b blocks each of blocklength n. The number of blocks b is fixed and regarded as a constant (does not grow with n). The message is split into $b - 1$ sub-messages M_j, each representing nR bits of information. Thus, the effective rate is $R_{\text{eff}} := \frac{b-1}{b}R$. Under this setup, we wish to determine the error exponent under PDF and CF.

Definition 3. *The b-block-error exponent is defined as*

$$E_b(R_{\text{eff}}) = \sup \left\{ \liminf_{N \to \infty} -\frac{1}{N} \log \mathbb{P}(\hat{M} \neq M) \right\}$$

where $N = nb$, M is uniform on $\mathcal{M} := [2^{NR_{\text{eff}}}]$ and the supremum is over all block-Markov coding schemes with b-blocks for the DM-RC.

The usual reliability function $E(R)$ is lower bounded by $\sup_{b \in \mathbb{N}} E_b(\frac{b-1}{b}R)$. In the following, we will study two schemes that provide lower bounds to $E_b(R_{\text{eff}})$.

III. PARTIAL DECODE-FORWARD

We warm up by deriving the error exponent for PDF. In PDF, the relay decodes part of the message in each block. By using the technique to prove the packing lemmas in Haroutunian *et al.* [9] and MMI decoding [4], we obtain:

Theorem 1. *Fix $b \in \mathbb{N}$, auxiliary alphabet $\mathcal{U}$ and distribution $Q_{X_2} \times Q_{U|X_2} \times Q_{X_1|U,X_2} \in \mathscr{P}(\mathcal{X}_2 \times \mathcal{U} \times \mathcal{X}_1)$. We have*

$$E_b(R_{\text{eff}}) \geq \frac{1}{b} \left[\max_{R'+R''=R} \min\{F(R'), G(R'), \tilde{G}(R'')\} \right], \quad (3)$$

where $F(R'), G(R')$ and $\tilde{G}(R'')$ are exponents defined as

$$F(R') := \min_{V : \mathcal{U} \times \mathcal{X}_2 \to \mathcal{Y}_2} D(V \| W_{Y_2|UX_2} | Q_{UX_2})$$
$$+ |I(Q_{U|X_2}, V|Q_{X_2}) - R'|^+,$$

$$G(R') := \min_{V : \mathcal{U} \times \mathcal{X}_2 \to \mathcal{Y}_3} D(V \| W_{Y_3|UX_2} | Q_{UX_2})$$
$$+ |I(Q_{U,X_2}, V) - R'|^+, \quad (4)$$

$$\tilde{G}(R'') := \min_{V : \mathcal{U} \times \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}_3} D(V \| W_{Y_3|UX_1X_2} | Q_{UX_2X_1})$$
$$+ |I(Q_{X_1|UX_2}, V|Q_{UX_2}) - R''|^+.$$

Note that $W_{Y_2|UX_2}$, $W_{Y_3|UX_2}$ and $W_{Y_3|UX_1X_2}$ are virtual channels induced by W and Q_{X_2}, $Q_{U|X_2}$ and $Q_{X_1|U,X_2}$.

Clearly, for a fixed $R' + R'' = R$, if

$$R' < \min\{I(U; Y_2|X_2), I(UX_2; Y_3)\},$$
$$R'' < I(X_1; Y_3|UX_2),$$

then $F(R'), G(R')$ and $\tilde{G}(R'')$ are positive. Hence, the error probability decays exponentially fast if R satisfies the PDF lower bound (1). Note that $F(R')$ is the exponent at the relay and $G(R')$ and $\tilde{G}(R'')$ are the exponents at the decoder. Setting $U = X_1$ recovers DF for which the exponent is provided in [7]. Note that (3) indicates a tradeoff between rate and error probability: as b increases, R_{eff} increases but the error exponent decreases. We omit the proof of Theorem 1 here because it is fairly standard. We refer the reader to [14].

IV. COMPRESS-FORWARD (CF)

In this section, we state and prove an achievable error exponent for CF. CF is more complicated than PDF because the relay does vector quantization on channel outputs Y_2^n and forwards the description to the destination. This quantized version of the channel output is $\hat{Y}_2^n$ and the error here is analyzed using techniques that Marton used to derive the error exponent for rate-distortion [5, Ch. 9]. The receiver decodes both the bin index and the message. This combination of covering and packing leads to a more involved analysis that needs to leverage on ideas in [10] where the error exponent for Wyner-Ziv was derived. It also leverages on a technique [6] to analyze the error when two indices are to be simultaneously decoded given a channel output. At a high level, we operate on a conditional type-by-conditional type basis for the covering step at the relay. We also use an α-decoding rule [15] for decoding the messages and the bin indices at the receiver.

A. Basic Definitions

We find it convenient to define several quantities upfront. For CF, the following types and conditional types will be kept fixed and hence can be optimized over eventually: input distributions $Q_{X_1} \in \mathscr{P}_n(\mathcal{X}_1)$, $Q_{X_2} \in \mathscr{P}_n(\mathcal{X}_2)$ and test channel $Q_{\hat{Y}_2|Y_2X_2} \in \mathscr{V}_n(\hat{\mathcal{Y}}_2; Q_{Y_2|X_2}Q_{X_2})$ for some (adversarial) channel realization $Q_{Y_2|X_2} \in \mathscr{V}_n(\mathcal{Y}_2; Q_{X_2})$.

1) Auxiliary Channels: Let the auxiliary channel $W_{Q_{X_1}} : \mathcal{X}_2 \to \mathcal{Y}_2$ be defined as

$$W_{Q_{X_1}}(y_2, y_3|x_2) := \sum_{x_1} W(y_2, y_3|x_1, x_2)Q_{X_1}(x_1).$$

This is simply the original relay channel averaged over Q_{X_1}. With a slight abuse of notation, we denote its $\mathcal{Y}_2$- and $\mathcal{Y}_3$-marginals using the same notation. Define another auxiliary channel $W_{Q_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2X_2}} : \mathcal{X}_1 \times \mathcal{X}_2 \to \hat{\mathcal{Y}}_2 \times \mathcal{Y}_3$ as

$$W_{Q_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2X_2}}(\hat{y}_2, y_3|x_1, x_2)$$
$$:= \sum_{y_2} W(y_3|x_1, x_2, y_2)Q_{\hat{Y}_2|Y_2X_2}(\hat{y}_2|y_2, x_2)Q_{Y_2|X_2}(y_2|x_2).$$

This is simply the original relay channel averaged over both channel realization $Q_{Y_2|X_2}$ and test channel $Q_{\hat{Y}_2|Y_2X_2}$:

2) Other Channels and Distributions: For any two channels $Q_{Y_2|X_2}, \tilde{Q}_{Y_2|X_2} : \mathcal{X}_2 \to \mathcal{Y}_2$, define two $\hat{y}_2$-modified channels

$$Q_{\hat{Y}_2|X_2}(\hat{y}_2|x_2) := \sum_{y_2} Q_{\hat{Y}_2|Y_2,X_2}(\hat{y}_2|y_2, x_2)Q_{Y_2|X_2}(y_2|x_2),$$

$$\tilde{Q}_{\hat{Y}_2|X_2}(\hat{y}_2|x_2) := \sum_{y_2} Q_{\hat{Y}_2|Y_2,X_2}(\hat{y}_2|y_2, x_2)\tilde{Q}_{Y_2|X_2}(y_2|x_2).$$

Implicit in these definitions are $Q_{\hat{Y}_2|Y_2,X_2}$, $Q_{Y_2|X_2}$ and $\tilde{Q}_{Y_2|X_2}$ but these dependencies are suppressed for the sake of brevity. For any $V : \mathcal{X}_1 \times \mathcal{X}_2 \times \hat{\mathcal{Y}}_2 \to \mathcal{Y}_3$, let the induced conditional distributions $V_{Q_{X_1}} : \mathcal{X}_2 \times \hat{\mathcal{Y}}_2 \to \mathcal{Y}_3$ and $Q_{\hat{Y}_2|X_2} \times V : \mathcal{X}_1 \times \mathcal{X}_2 \to \hat{\mathcal{Y}}_2 \times \mathcal{Y}_3$ be defined as

$$V_{Q_{X_1}}(y_3|x_2,\hat{y}_2) := \sum_{x_1} V(y_3|x_1,x_2,\hat{y}_2)Q_{X_1}(x_1),$$

$$(\tilde{Q}_{\hat{Y}_2|X_2} \times V)(\hat{y}_2,y_3|x_1,x_2) := V(y_3|x_1,x_2,\hat{y}_2)\tilde{Q}_{\hat{Y}_2|X_2}(\hat{y}_2|x_2).$$

3) Sets of Distributions and α-Decoder: Define the set of joint types $P_{X_1X_2\hat{Y}_2Y_3}$ with marginals consistent with Q_{X_1}, Q_{X_2} and $Q_{\hat{Y}_2|X_2}$ as

$$\mathscr{P}_n(Q_{X_1},Q_{X_2},Q_{\hat{Y}_2|X_2}) := \{P_{X_1X_2\hat{Y}_2Y_2} \in \mathscr{P}_n(\mathcal{X}_1 \times \mathcal{X}_2 \times \hat{\mathcal{Y}}_2 \times \mathcal{Y}_3)$$
$$: (P_{X_1}, P_{X_2}, P_{\hat{Y}_2|X_2}) = (Q_{X_1}, Q_{X_2}, Q_{\hat{Y}_2|X_2})\}.$$

We will use the notation $\mathscr{P}(Q_{X_1}, Q_{X_2}, Q_{\hat{Y}_2|X_2})$ (without subscript n) to mean the same set without the restriction to types but all distributions in $\mathscr{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \hat{\mathcal{Y}}_2 \times \mathcal{Y}_3)$. For any four sequences $(x_1^n, x_2^n, \hat{y}_2^n, y_3^n)$, define the function α as

$$\alpha(x_1^n, \hat{y}_2^n, y_3^n|x_2^n) := D(V\|W_{Q_{Y_2|X_2},Q_{\hat{Y}_2|Y_2X_2}}|P) + H(V|P),$$

where P is the joint type of $(x_1^n, x_2^n, \hat{y}_2^n)$ and $V : \mathcal{X}_1 \times \mathcal{X}_2 \times \hat{\mathcal{Y}}_2 \to \mathcal{Y}_3$ is the conditional type of y_3^n given $(x_1^n, x_2^n, \hat{y}_2^n)$. Roughly speaking, to decode the bin index and message, we will maximize α over bin indices, messages and conditional types $Q_{Y_2|X_2}$. This is analogous to maximum-likelihood decoding [15]. See (10). Define the set of conditional types

$$\mathscr{K}_n(Q_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2X_2}) := \{V \in \mathscr{V}_n(\mathcal{Y}_3; Q_{X_1}Q_{X_2}Q_{\hat{Y}_2|X_2}) :$$
$$\alpha(Q_{X_1}Q_{X_2}Q_{\hat{Y}_2|X_2}, V)$$
$$\geq \alpha(Q_{X_1}Q_{X_2}Q_{\hat{Y}_2|X_2}, W_{Q_{Y_2|X_2},Q_{\hat{Y}_2|Y_2X_2}})\}.$$

Intuitively, the conditional types in $\mathscr{K}_n(Q_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2X_2})$ are those corresponding to sequences y_3^n that lead to an error as the likelihood computed with respect to V is larger than that for the true averaged channel $W_{Q_{Y_2|X_2},Q_{\hat{Y}_2|Y_2X_2}}$. We will use the notation $\mathscr{K}(Q_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2X_2})$ (without subscript n) to mean the same set without the restriction to conditional types but all conditional distributions from $\mathcal{X}_1 \times \mathcal{X}_2 \times \hat{\mathcal{Y}}_2$ to $\mathcal{Y}_3$.

B. Error Exponent for Compress-Forward

Theorem 2. *Fix $b \in \mathbb{N}$ and "Wyner-Ziv rate" $R_2 \geq 0$, distributions $Q_{X_1} \in \mathscr{P}(\mathcal{X}_1)$ and $Q_{X_2} \in \mathscr{P}(\mathcal{X}_2)$ and auxiliary alphabet $\hat{\mathcal{Y}}_2$. We have*

$$E_b(R_{\text{eff}}) \geq \frac{1}{b}\min\{G_1(R, R_2), G_2(R, R_2)\}$$

where the constituent exponents are defined as

$$G_1(R, R_2) := \min_{V:\mathcal{X}_2 \to \mathcal{Y}_3} D(V\|W_{Q_{X_1}}|Q_{X_2}) + |I(Q_{X_2}, V) - R_2|^+$$
$$G_2(R, R_2) := \min_{Q_{Y_2|X_2}:\mathcal{X}_2 \to \mathcal{Y}_2} D(Q_{Y_2|X_2}\|W_{Q_{X_1}}|Q_{X_2})$$
$$+ \max_{Q_{\hat{Y}_2|Y_2X_2}:\mathcal{Y}_2 \times \mathcal{X}_2 \to \hat{\mathcal{Y}}_2} J(R, R_2, Q_{\hat{Y}_2|Y_2X_2}, Q_{Y_2|X_2}).$$

The quantity $J(R, R_2, Q_{\hat{Y}_2|Y_2X_2}, Q_{Y_2|X_2})$ that constitutes $G_2(R, R_2)$ is defined as

$$J(R, R_2, Q_{\hat{Y}_2|Y_2X_2}, Q_{Y_2|X_2}) := \min_{P_{X_1X_2\hat{Y}_2Y_3} \in \mathscr{P}(Q_{X_1},Q_{X_2},Q_{\hat{Y}_2|X_2})}$$
$$D(P_{\hat{Y}_2Y_3|X_1X_2}\|W_{Q_{Y_2|X_2},Q_{\hat{Y}_2|Y_2,X_2}}|Q_{X_1}Q_{X_2})$$
$$+ \min_{\tilde{Q}_{Y_2|X_2}:\mathcal{X}_2 \to \mathcal{Y}_2} \min_{V \in \mathscr{K}(Q_{Y_2|X_2},Q_{\hat{Y}_2|Y_2X_2})}$$
$$\min_{l=1,2} \psi_l(V, \tilde{Q}_{Y_2|X_2}, R, R_2, P_{X_1X_2\hat{Y}_2Y_3}) \quad (5)$$

where the functions $\psi_l, l = 1, 2$ are defined as

$$\psi_1(V, \tilde{Q}_{Y_2|X_2}, R, R_2, P_{X_1X_2\hat{Y}_2Y_3})$$
$$:= |I(Q_{X_1}, \tilde{Q}_{\hat{Y}_2|X_2} \times V|Q_{X_2}) - R|^+ \quad (6)$$
$$\psi_2(V, \tilde{Q}_{Y_2|X_2}, R, R_2, P_{X_1X_2\hat{Y}_2Y_3})$$
$$:= |I(\tilde{Q}_{\hat{Y}_2|X_2}, V_{Q_{X_1}}|Q_{X_2}) + |I(Q_{X_1}, \tilde{Q}_{\hat{Y}_2|X_2} \times V) - R|^+$$
$$- |I(\tilde{Q}_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2,X_2}|Q_{X_2}) - R_2|^+|^+. \quad (7)$$

C. Remarks on the Error Exponent for Compress-Forward

In this Section, we dissect the main features of the CF error exponent presented in Theorem 2.

We are free to choose the input distributions Q_{X_1} and Q_{X_2}, though these will be n-types for finite $n \in \mathbb{N}$. We also have the freedom to choose any "Wyner-Ziv rate" $R_2 \geq 0$.

In CF [2], the relay transmits a description $\hat{y}_2^n(j)$ of its received sequence $y_2^n(j)$ via a covering step. This explains the final mutual information term in (7), namely $I(Y_2; \hat{Y}_2|X_2)$. Since covering results in super-exponential decay in the error, this does not affect the overall exponent. See (11).

The exponent $G_1(R, R_2)$ is analogous to $G(R')$ in (4). This represents the error rate in the estimation of X_2's index given Y_3^n using MMI decoding. However, in the CF proof, we do not use the packing lemma. Rather we construct a random code and show that on expectation, the error probability decays exponentially fast with the exponent given by $G_1(R, R_2)$.

In $G_2(R, R_2)$, $Q_{Y_2|X_2}$ is the realization of the conditional type of $y_2^n(j)$ given $x_2^n(j)$. The divergence term $D(Q_{Y_2|X_2}\|W_{Q_{X_1}}|Q_{X_2})$ represents the deviation from the true channel behavior $W_{Q_{X_1}}$. We can optimize for the conditional distribution $Q_{\hat{Y}_2|Y_2,X_2}$ explaining the inner maximization over $Q_{\hat{Y}_2|Y_2,X_2}$ and outer minimization over $Q_{Y_2|X_2}$. This is a game-theoretic-type result similar to [10]–[12].

The first part of J given by ψ_1 in (6) represents incorrect decoding of the index of X_1^n (message M_j) as well as the conditional type $Q_{Y_2|X_2}$ given that the bin index of the description $\hat{Y}_2^n$ is decoded correctly. The second part of J given by ψ_2 in (7) represents the incorrect decoding the bin index of $\hat{Y}_2^n$, the index of X_1^n (message M_j) as well as the conditional type $Q_{Y_2|X_2}$. We see the different sources of "errors" in (5): a minimization over the different types of channel behavior represented by $P_{X_1X_2\hat{Y}_2Y_3}$ and another minimization over conditional types $\tilde{Q}_{Y_2|X_2}$. Subsequently, the error in α-decoding of the message and the bin index of the description sequence is represented by the minimization over $V \in \mathscr{K}(Q_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2X_2})$.

The choice of the "Wyner-Ziv rate" R_2 allows us to operate in two distinct regimes. This can be seen from the two different cases involving R_2 in (7). The number of Wyner-Ziv bins is designed to be $\doteq \exp(nI(\tilde{Q}_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2,X_2}|Q_{X_2}))$, where the choice of $Q_{\hat{Y}_2|Y_2,X_2}$ depends on the realized conditional type $Q_{Y_2|X_2}$. Thus, when $R_2 \leq I(\tilde{Q}_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2,X_2}|Q_{X_2})$, we do additional binning as there are more bins than description sequences. If R_2 is larger than $I(\tilde{Q}_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2,X_2}|Q_{X_2})$, no binning is required.

For the analysis of the bin and message indices, if we simply apply the packing lemmas in [5], [9], [15], this would result in a suboptimal rate vis-à-vis CF. This is because the conditional correlation of X_1 and $\hat{Y}_2$ given X_2 would not be taken into account. Thus, we need to analyze this error exponent more carefully using techniques introduced in [6] for the multiple-access channel. Note that the first two mutual informations (ignoring the $|\cdot|^+$) in (7) can be written as

$$I(\hat{Y}_2; Y_3|X_2) + I(X_1; \hat{Y}_2 Y_3|X_2)$$
$$= H(X_1|X_2) + H(\hat{Y}_2|X_2) + H(Y_3|X_2) - H(X_1 \hat{Y}_2 Y_3|X_2),$$

demonstrating the symmetry between X_1 and $\hat{Y}_2$ when they are decoded jointly at Y_3.

From the exponents in Theorem 2, it is clear upon eliminating R_2 (if R_2 is chosen small enough so that Wyner-Ziv binning is necessary) that we recover the CF lower bound in (2). Indeed, if ψ_1 is active in the minimization in (5), the first term in (2) is positive if and only if the error exponent G_2 is positive for some choice of distributions Q_{X_1}, Q_{X_2} and $Q_{\hat{Y}_2|Y_2,X_2}$. Also, if ψ_2 is active in the minimization in (5) and R_2 is chosen sufficiently small, G_2 is positive if

$$R < R_2 + I(X_1; \hat{Y}_2 Y_3|X_2) + I(\hat{Y}_2; Y_3|X_2) - I(Y_2; \hat{Y}_2|X_2)$$
$$< I(X_1; \hat{Y}_2 Y_3|X_2) + I(\hat{Y}_2; X_2 Y_3) - I(Y_2; \hat{Y}_2|X_2)$$
$$= I(X_1 X_2; Y_3) - I(\hat{Y}_2; Y_2|X_1 X_2 Y_3). \quad (8)$$

where we used the Markov chain $\hat{Y}_2 - (X_1, X_2) - (Y_2, Y_3)$ [3, pp. 402]. Equation (8) matches the second term in (2).

D. Proof Sketch of Theorem 2

We only provide a sketch here. See [14] for details.

Proof: **Random Codebook Generation:** Fix types $Q_{X_1} \in \mathscr{P}_n(\mathcal{X}_1)$ and $Q_{X_2} \in \mathscr{P}_n(\mathcal{X}_2)$ as well as rates $R, R_2 \geq 0$. For each $j \in [b]$, generate a random codebook in the following manner. Randomly and independently generate $\exp(nR)$ codewords $x_1^n(m_j) \sim \text{Unif}[\mathcal{T}_{Q_{X_1}}]$. Randomly and independently generate $\exp(nR_2)$ codewords $x_2^n(l_{j-1}) \sim \text{Unif}[\mathcal{T}_{Q_{X_2}}]$. Now for every $Q_{Y_2|X_2} \in \mathscr{V}_n(\mathcal{Y}_2; Q_{X_2})$ fix a different test channel $Q_{\hat{Y}_2|Y_2,X_2}(Q_{Y_2|X_2}) \in \mathscr{V}_n(\hat{\mathcal{Y}}_2; Q_{Y_2|X_2} Q_{X_2})$. For every $Q_{Y_2|X_2} \in \mathscr{V}_n(\mathcal{Y}_2; Q_{X_2})$ and every $x_2^n(l_{j-1})$ construct a conditional type-dependent codebook $\mathcal{B}(Q_{Y_2|X_2}, l_{j-1}) \subset \hat{\mathcal{Y}}_2^n$ of integer size $|\mathcal{B}(Q_{Y_2|X_2}, l_{j-1})|$ whose rate satisfies

$$\tilde{R}_2(Q_{Y_2|X_2}) := I(Q_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2,X_2}(Q_{Y_2|X_2})|Q_{X_2}) + \nu_n,$$

where $\nu_n \in \Theta(\frac{\log n}{n})$. Each sequence in $\mathcal{B}(Q_{Y_2|X_2}, l_{j-1})$ is indexed as $\hat{y}_2^n(k_j|l_{j-1})$ and is drawn independently according

to $\text{Unif}[\mathcal{T}_{Q_{\hat{Y}_2|X_2}}(x_2^n(l_{j-1}))]$ where $Q_{\hat{Y}_2|X_2}$ is the marginal induced by $Q_{Y_2|X_2}$ and $Q_{\hat{Y}_2|Y_2,X_2}(Q_{Y_2|X_2})$. Depending on the choice of R_2, do one of the following:

- If $R_2 \leq \tilde{R}_2(Q_{Y_2|X_2})$, partition the conditional type-dependent codebook $\mathcal{B}(Q_{Y_2|X_2}, l_{j-1})$ into $\exp(nR_2)$ equal-sized bins $\mathcal{B}_{l_j}(Q_{Y_2|X_2}, l_{j-1})$ indexed by $l_j \in [\exp(nR_2)]$ (Wyner-Ziv binning).
- If $R_2 > \tilde{R}_2(Q_{Y_2|X_2})$, assign each element of $\mathcal{B}(Q_{Y_2|X_2}, l_{j-1})$ a unique index in $[\exp(nR_2)]$.

Transmitter Encoding: The encoder transmits $x_1^n(m_j)$ in block $j \in [b]$.

Relay Encoding: At the end of block $j \in [b]$, the relay encoder has $x_2^n(l_{j-1})$ (by convention $l_0 := 1$) and its input sequence $y_2^n(j)$. It computes the conditional type $Q_{Y_2|X_2} \in \mathscr{V}_n(\mathcal{Y}_2; Q_{X_2})$. Then it searches in $\mathcal{B}(Q_{Y_2|X_2}, l_{j-1})$ for a description sequence

$$\hat{y}_2^n(\hat{k}_j|l_{j-1}) \in \mathcal{T}_{Q_{\hat{Y}_2|Y_2,X_2}(Q_{Y_2|X_2})}(y_2^n(j), x_2^n(l_{j-1})). \quad (9)$$

If more than one such sequence exists, choose one uniformly at random in $\mathcal{B}(Q_{Y_2|X_2}, l_{j-1})$ from those satisfying (9). If none exists, choose uniformly at random from $\mathcal{B}(Q_{Y_2|X_2}, l_{j-1})$. Identify the bin index $\hat{l}_j$ of $\hat{y}_2^n(\hat{k}_j|l_{j-1})$ and send $x_2^n(\hat{l}_j)$.

Decoding: At the end of block $j+1$, the receiver has channel output $y_3^n(j+1)$. It does MMI decoding [4] as follows:

$$\hat{l}_j := \arg\max_{l_j \in [\exp(nR_2)]} \hat{I}(x_2^n(l_j) \wedge y_3^n(j+1)).$$

Having identified $\hat{l}_{j-1}, \hat{l}_j$ above, find message $\hat{m}_j$, index $\hat{k}_j$ and conditional type $\hat{Q}_{Y_2|X_2}^{(j)} \in \mathscr{V}_n(\mathcal{Y}_2; Q_{X_2})$ satisfying

$$(\hat{m}_j, \hat{k}_j, \hat{Q}_{Y_2|X_2}^{(j)}) = \arg\max_{(m_j, k_j, Q_{Y_2|X_2}): \hat{y}_2^n(k_j|\hat{l}_{j-1}) \in \mathcal{B}_{\hat{l}_j}(Q_{Y_2|X_2}, \hat{l}_{j-1})}$$
$$\alpha(x_1^n(m_j), \hat{y}_2^n(k_j|\hat{l}_{j-1}), y_3^n(j)|x_2^n(\hat{l}_{j-1})). \quad (10)$$

Declare that $\hat{m}_j$ was sent.

Analysis of Error Probability: We now analyze the error probability. Assume that $M_j = 1$ and let L_{j-1}, L_j and K_j be indices chosen by the relay in block j. Note that [3, Thm. 16.4]

$$\mathbb{P}(\hat{M}_j \neq M_j \text{ for any } j \in [b-1]) \leq (b-1)(\epsilon_R + 2\epsilon_{D,1} + \epsilon_{D,2}),$$

where ϵ_R is the error event that there is no description sequence $\hat{y}_2^n(\hat{k}_j|l_{j-1})$ in the bin $\mathcal{B}(Q_{Y_2|X_2}, l_{j-1})$ that satisfies (9), $\epsilon_{D,1} := \mathbb{P}(\hat{L}_j \neq L_j)$ is the error probability in decoding the wrong l_j bin index, and $\epsilon_{D,2} := \mathbb{P}(\hat{M}_j \neq 1|L_j, L_{j-1}$ decoded correctly$)$ is the error probability in decoding the message incorrectly.

Covering Error ϵ_R: For ϵ_R, we follow the proof idea in [10, Lem. 2]. For every realized conditional type $Q_{Y_2|X_2}$,

$$\epsilon_R \leq \mathbb{P}(\mathcal{F}|Y_2^n = y_2^n, X_2^n = x_2^n, y_2^n \in \mathcal{T}_{Q_{Y_2|X_2}}(x_2^n)),$$

where $\mathcal{F}$ is the event that *every* sequence $\hat{y}_2^n(k_j|l_{j-1}) \in \mathcal{B}(Q_{Y_2|X_2}, l_{j-1})$ does *not* satisfy (9). Now we use the independence of the codewords in $\mathcal{B}(Q_{Y_2|X_2}, l_{j-1})$ and properties of types to assert that with the right choice of $\nu_n \in \Theta(\frac{\log n}{n})$,

$$\epsilon_R \leq e^{-(n+1)^2}, \qquad \forall n \in \mathbb{N}. \quad (11)$$

Thus, ϵ_{R} decays super-exponentially fast.

First Packing Error $\epsilon_{\mathrm{D},1}$**:** For $\epsilon_{\mathrm{D},1}$, we leverage on techniques to prove the random coding error exponent for channel coding [5, Ch. 10]. This is standard so we omit the proof.

Second Packing Error $\epsilon_{\mathrm{D},2}$**:** We partition the sample space into subsets where the conditional type of relay input y_2^n given relay output x_2^n is $Q_{Y_2|X_2} \in \mathscr{V}_n(\mathcal{Y}_2; Q_{X_2})$. That is,

$$\epsilon_{\mathrm{D},2} = \sum_{Q_{Y_2|X_2} \in \mathscr{V}_n(\mathcal{Y}_2; Q_{X_2})} \mathbb{P}(Y_2^n \in \mathcal{T}_{Q_{Y_2|X_2}}(X_2^n)) \varphi_n(Q_{Y_2|X_2})$$

where $\varphi_n(Q_{Y_2|X_2})$ is defined as

$$\varphi_n(Q_{Y_2|X_2}) := \mathbb{P}(\hat{M}_j \neq 1 | L_j, L_{j-1} \text{ correct}, Y_2^n \in \mathcal{T}_{Q_{Y_2|X_2}}(X_2^n)).$$

By using properties of types,

$$\mathbb{P}(Y_2^n \in \mathcal{T}_{Q_{Y_2|X_2}}(X_2^n)) \dot{\leq} \exp\left[-n D(Q_{Y_2|X_2} \| W_{Q_{X_1}} | Q_{X_2})\right].$$

Following [6], we now bound $\varphi_n(Q_{Y_2|X_2})$ by first conditioning on joint types $P_{X_1 X_2 \hat{Y}_2 Y_3} \in \mathscr{P}_n(Q_{X_1}, Q_{X_2}, Q_{\hat{Y}_2|X_2})$

$$\varphi_n(Q_{Y_2|X_2}) \leq \sum_{P_{X_1 X_2 \hat{Y}_2 Y_3} \in \mathscr{P}_n(Q_{X_1}, Q_{X_2}, Q_{\hat{Y}_2|X_2})} \mathbb{P}(\mathcal{T}_{P_{X_1 X_2 \hat{Y}_2 Y_3}})$$

$$\times \mathbb{P}\left(\bigcup_{V \in \mathscr{K}_n(Q_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2 X_2})} \mathcal{E}_V \,\Big|\, \mathcal{T}_{P_{X_1 X_2 \hat{Y}_2 Y_3}}\right) \quad (12)$$

where $Q_{\hat{Y}_2|X_2}$ is induced by $Q_{\hat{Y}_2|Y_2 X_2}$ and $Q_{Y_2|X_2}$ and the event $\mathcal{E}_V$ is defined as

$$\mathcal{E}_V := \bigcup_{\tilde{Q}_{Y_2|X_2} \in \mathscr{V}_n(\mathcal{Y}_2; Q_{X_2}) \setminus \{Q_{Y_2|X_2}\}} \mathcal{E}_V(\tilde{Q}_{Y_2|X_2}),$$

where the constituent events are defined as

$$\mathcal{E}_V(\tilde{Q}_{Y_2|X_2}) := \bigcup_{\tilde{m}_j \neq 1} \bigcup_{\tilde{k}_j \in \mathcal{B}_{\hat{L}_j}(\tilde{Q}_{Y_2|X_2}, \hat{L}_{j-1})} \mathcal{E}_V(\tilde{Q}_{Y_2|X_2}, \tilde{m}_j, \tilde{k}_j),$$

and $\mathcal{E}_V(\tilde{Q}_{Y_2|X_2}, \tilde{m}_j, \tilde{k}_j)$ is defined as the event that $\left(X_1^n(\tilde{m}_j), X_2^n(L_{j-1}), \hat{Y}_2^n(\tilde{k}_j | L_{j-1}), Y_3^n(j)\right)$ has joint type $Q_{X_1} Q_{X_2} \tilde{Q}_{\hat{Y}_2|X_2} V$. Ineq. (12) reflects the error in α-decoding.

Now, we bound the constituent elements in (12). By a simple calculation using the method of types [5],

$$\mathbb{P}(\mathcal{T}_{P_{X_1 X_2 \hat{Y}_2 Y_3}})$$
$$\dot{\leq} \exp\left[-n D(P_{\hat{Y}_2 Y_3|X_1 X_2} \| W_{Q_{Y_2|X_2}}, Q_{\hat{Y}_2|Y_2, X_2} | Q_{X_1} Q_{X_2})\right].$$

Hence, all that remains is to bound the probability of the union in (12). There are two cases: For the case where the decoded bin index $\tilde{k}_j$ is correct (i.e., equal to K_j) but message is $\tilde{m}_j$ wrong (i.e., not equal to 1), the analysis is relatively straightforward. Thus, we consider the case where *both* decoded bin index and message are wrong.

For any conditional distribution $\tilde{Q}_{Y_2|X_2}$, define the *excess rate* $\Delta R_2(\tilde{Q}_{Y_2|X_2}) := \tilde{R}_2(\tilde{Q}_{Y_2|X_2}) - R_2$. Assume for the moment that $\Delta R_2(\tilde{Q}_{Y_2|X_2}) \geq 0$ (the other case can be dealt with similarly). Equivalently, this means that $R_2 \leq I(\tilde{Q}_{Y_2|X_2}, Q_{\hat{Y}_2|Y_2, X_2} | Q_{X_2}) + \nu_n$ so Wyner-Ziv binning *is required*. Using bars to denote random variables generated

uniformly from their respective marginal type classes and arbitrary sequences in the respective marginal type classes, define as in [6]

$$\xi_n(V, \tilde{Q}_{Y_2|X_2})$$
$$:= \exp\left[n \Delta R_2(\tilde{Q}_{Y_2|X_2})\right] \mathbb{P}\left[(\bar{x}_2^n, \bar{\hat{Y}}_2^n, y_3^n) \in \mathcal{T}_{Q_{X_2} \tilde{Q}_{\hat{Y}_2|X_2} V_{Q_{X_1}}}\right],$$
$$\zeta_n(V, \tilde{Q}_{Y_2|X_2})$$
$$:= \exp(nR) \mathbb{P}\left[(\bar{x}_2^n, \bar{y}_2^n, \bar{X}_1^n, \bar{y}_3^n) \in \mathcal{T}_{Q_{X_2} Q_{X_1}(\tilde{Q}_{\hat{Y}_2|X_2} \times V)}\right].$$

By applying the union bound *one at a time to the two unions* that define $\mathcal{E}_V(\tilde{Q}_{Y_2|X_2})$ as was done in [6], we obtain

$$\mathbb{P}\left[\mathcal{E}_V(\tilde{Q}_{Y_2|X_2}) \,\big|\, \mathcal{T}_{P_{X_1 X_2 \hat{Y}_2 Y_3}}\right] \leq \gamma_n(V, \tilde{Q}_{Y_2|X_2})$$

where

$$\gamma_n(V, \tilde{Q}_{Y_2|X_2}) := \min\{1, \xi_n(V, \tilde{Q}_{Y_2|X_2}) \min\{1, \zeta_n(V, \tilde{Q}_{Y_2|X_2})\}\}.$$

Hence, $\gamma_n(V, \tilde{Q}_{Y_2|X_2})$ has the following exponential behavior:

$$\gamma_n(V, \tilde{Q}_{Y_2|X_2}) \doteq \exp\Big[-n\big|I(\tilde{Q}_{\hat{Y}_2|X_2}, V_{Q_{X_1}} | Q_{X_2})$$
$$- \Delta R_2(\tilde{Q}_{Y_2|X_2}) + \big|I(Q_{X_1}, \tilde{Q}_{\hat{Y}_2|X_2} \times V | Q_{X_2}) - R\big|^+\big|^+\Big].$$

Observe the exponent of $\gamma_n(V, \tilde{Q}_{Y_2|X_2})$ matches (7) upon using the definition of the excess rate $\Delta R_2(\tilde{Q}_{Y_2|X_2})$. Putting all the bounds together completes the proof. ∎

REFERENCES

[1] E. C. van der Meulen, "Three-terminal communication channels," *Adv. Appl. Prob.*, vol. 3, pp. 120–54, 1971.

[2] T. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. on Inf. Th.*, vol. 25, no. 5, pp. 572–84, 1979.

[3] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2012.

[4] V. D. Goppa, "Nonprobabilistic mutual information without memory," *Probl. of Control and Inf. Th.*, vol. 4, pp. 97–102, 1975.

[5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[6] J. Scarlett and A. Guillén i Fàbregas, "An achievable error exponent for the mismatched multiple-access channel," in *Allerton Conference*, 2012.

[7] G. J. Bradford and J. N. Laneman, "Error exponents for block Markov superposition encoding with varying decoding latency," in *IEEE Information Theory Workshop (ITW)*, 2012.

[8] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[9] E. A. Haroutunian, M. E. Haroutunian, and A. N. Harutyunyan, *Reliability Criteria in Information Theory and Statistical Hypothesis Testing*, ser. Foundations and Trends in Communications and Information Theory. Now Publishers Inc, 2008, vol. 4.

[10] B. Kelly and A. Wagner, "Reliability in source coding with side information," *IEEE Trans. on Inf. Th.*, vol. 58, no. 8, pp. 5086–5111, 2012.

[11] P. Moulin and Y. Wang, "Capacity and random-coding exponents for channel coding with side information," *IEEE Trans. on Inf. Th.*, vol. 53, no. 4, pp. 1326–47, Apr 2007.

[12] G. Dasarathy and S. C. Draper, "On reliability of content identification from databases based on noisy queries," in *Int. Symp. Inf. Th.*, St Petersburg, Russia, 2011.

[13] H. Q. Ngo, T. Q. S. Quek, and H. Shin, "Amplify-and-forward two-way relay networks: Error exponents and resource allocation,," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2653–66, Sep 2010.

[14] V. Y. F. Tan, "Error exponents for the relay channel," *arXiv:1304.3553*, Apr 2013.

[15] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. on Inf. Th.*, vol. 27, pp. 5–11, Jan 1981.