

Is Non-Unique Decoding Necessary?

Shirin Saeedi Bidokhti, *Member, IEEE*, and Vinod M. Prabhakaran, *Member, IEEE*

Abstract—In multiterminal communication systems, signals carrying messages meant for different destinations are often observed together at any given destination receiver. Han and Kobayashi proposed a receiving strategy, which performs a joint unique decoding of messages of interest along with a subset of messages, which are not of interest. It is now well-known that this provides an achievable region, which is, in general, larger than if the receiver treats all messages not of interest as noise. Nair and El Gamal and Chong, Motani, Garg, and El Gamal independently proposed a generalization called indirect or nonunique decoding where the receiver uses the codebook structure of the messages to uniquely decode only its messages of interest. Nonunique decoding has since been used in various scenarios. The main result in this paper is to provide an interpretation and a systematic proof technique for why nonunique decoding, in all known cases where it has been employed, can be replaced by a particularly designed joint unique decoding strategy, without any penalty from a rate region viewpoint.

Index Terms—Broadcast channel, joint decoding, non-unique decoding, indirect decoding.

I. INTRODUCTION

CODING schemes for multi-terminal systems with many information sources and many destinations try to exploit the broadcast and interference nature of the communication media. A consequence of this is that in many schemes the signals received at a destination carry information, not only about messages that are expected to be decoded at the destination (*messages of interest*), but also about messages that are not of interest to that destination.

Standard methods in (random) code design (at the encoder) are rate splitting, superposition coding and Marton's coding [1], [2]. On the other hand, standard decoding techniques are successive decoding and joint decoding [1], [3]. In [3], Han and Kobayashi proposed a receiving strategy which performs a *joint decoding* of messages of interest along with a subset of messages which are not of interest. We will refer to this receiving strategy as joint *unique* decoding (and to the decoders as joint *unique* decoders) to emphasize the fact

Manuscript received October 8, 2012; revised December 17, 2013; accepted February 24, 2014. Date of publication March 18, 2014; date of current version April 17, 2014. S. Saeedi Bidokhti was supported by the Swiss National Science Foundation Fellowship under Grant 146617. V. M. Prabhakaran was supported by a Ramanujan Fellowship through the Department of Science and Technology, Government of India. This paper was presented at the 2012 IEEE International Symposium on Information Theory.

S. Saeedi Bidokhti was with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne 1015, Switzerland. She is now with the Institute for Communications Engineering, Technische Universität München, München 80333, Germany (e-mail: shirin.saeedi@tum.de).

V. M. Prabhakaran is with the School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai 400005, India (e-mail: vinodmp@tifr.res.in).

Communicated by A. B. Wagner, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2014.2312285

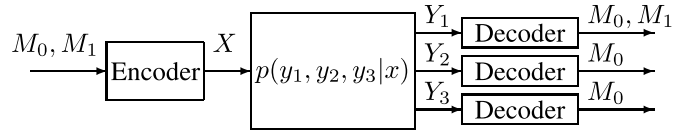


Fig. 1. The 3-receiver broadcast channel with two degraded message sets: message M_0 is destined to all receivers and message M_1 is destined to receiver Y_1 .

that it seeks a unique choice not only for the messages of interest, but also for the rest of the messages being jointly decoded. It is now well-known that employing such a joint unique decoder in the code design provides an achievable region which is, in general, larger than if the receiver decodes the messages of interest while treating all messages not of interest as noise. Recently, Nair and El Gamal [4] and Chong, Motani, Garg, and El Gamal [5] independently proposed a generalization called *indirect or non-unique decoding* where the decoder looks for the unique messages of interest while using the codebook structure of all the messages (including the ones not of interest). Unlike the joint unique decoder, such a decoder does not necessarily uniquely decode messages not of interest, though it might narrow them down to a smaller list. We refer to such a decoder as a non-unique decoder. With such a distinction, non-unique decoders perform at least as well as joint-unique decoders. Coding schemes which employ non-unique decoders have since played a role in achievability schemes in different multi-terminal problems such as [6]–[10]. It is of interest, therefore, to see if they can achieve higher reliable transmission rates compared to codes that employ joint unique decoders.

In [4], the idea of non-unique (indirect) decoding is studied in the context of broadcast channels with degraded message sets. Nair and El Gamal consider a 3-receiver general broadcast channel where a source communicates a common message M_0 to three receivers Y_1 , Y_2 , and Y_3 and a private message M_1 only to one of the receivers, Y_1 (Fig. 1). They characterize an inner-bound to the capacity region of this problem using non-unique decoding and show its tightness for some special cases. It turns out that the same inner-bound of [4] can be achieved using a joint unique decoding strategy at all receivers. The equivalence of the rate region achievable by non-unique decoding and that of joint unique decoding was observed in [4], but it was arrived at by comparing single letter expressions for the two rate regions. A similar equivalence was also noticed in [5], again by comparing single-letter expressions. For noisy network coding [6], it has been shown that the same rate region can be obtained using joint unique decoding and without the use of non-unique decoding [11]–[14]. It was also observed in [7] that non-unique decoding

is not essential to achieve the capacity region of certain state-dependent multiple access channels and joint unique decoding suffices.

In this paper, we will provide a proof technique which systematically shows an equivalence between the rate region achievable through non-unique decoders and joint unique decoders in several examples. In particular, our line of argument is applicable to all known instances where non-unique decoding has been employed in the literature as we discuss in Section III. Our technique is based on designing a special auxiliary joint unique decoder which replaces the non-unique decoder and sheds some light on why this equivalence holds. However, we would like to note that analysis using non-unique decoding is often simpler and gives a more compact representation of the rate-region—a fact observed in [4] and [5]—which still makes it a valuable tool for analysis.

Three remarks follow.

Remark 1: The reader might wonder if such an equivalence holds on the rate-regions of schemes employing joint unique decoders and non-unique decoders more generally. While our proof technique is systematic and general, it is coupled with the random nature of the codebook generation and the encoder design. Indeed, any decoding scheme is coupled with the encoding scheme and therefore asking for a more general equivalence (for any encoding scheme) seems to be a challenging problem (even to properly pose).

Remark 2: Non-unique decoders are usually easier to work with (analytically), and they capture the correct error events (conceptually). One might wonder what the advantages of joint unique decoders are. It is generally interesting to know if certain messages may be uniquely decoded at a receiver at no rate-cost. In principle, such messages may be exploited to improve the encoding schemes. We refer the interested reader to [15] where an application of using joint unique decoders is illustrated in designing a block Markov encoding scheme for the broadcast channel with degraded messages.

Remark 3: In a related line of research, [16] proves optimality of non-unique decoding for general discrete memoryless interference channels, when encoding is restricted to randomly generated codebooks, superposition coding, and time sharing. The result of this paper and the techniques we develop indicate that the same performance can be achieved by employing joint unique decoding, and that joint unique decoding is also optimal in the sense discussed in [16].

In Section II, we develop our proof technique in the context of [4]. While much of the discussion in this paper is confined to this framework, we show in Section III that the technique applies more generally.

II. WHY JOINT UNIQUE DECODING SUFFICES IN THE INNER-BOUND OF NAIR AND EL GAMAL IN [4]

We start this section by briefly reviewing the work of [4] where inner and outer bounds are derived for the capacity region of a 3-receiver broadcast channel with degraded message sets. In particular, we consider the case where a source communicates a common message (of rate R_0) to all receivers, and a private message (of rate R_1) only to one of the receivers.

A coding scheme is a sequence of $((2^{nR_0}, 2^{nR_1}), n)$ codes consisting of an encoder and a decoder and is said to achieve a rate-tuple (R_0, R_1) if the probability of error at the decoders decays to zero as n grows large.

Joint Unique Decoder vs. Non-Unique Decoder: We consider joint typical set decoding. A decoder at a certain destination may, in general, *examine* a subset of messages which includes, but is not necessarily limited to, the messages of interest to that destination. By the term *examine*, we mean that the decoder will try to make use of the structure (of the codebook) associated with the messages it examines. We say a coding scheme employs a *joint unique decoder* if the decoder tries to uniquely decode all the messages it considers (and declares an error if there is ambiguity in any of the messages, irrespective of whether such messages are of interest to the destination or not). In contrast, we say that a coding scheme employs a *non-unique decoder* if the decoder tries to decode uniquely only the messages of interest to the destination and tolerates ambiguity in messages which are not of interest.

Within this framework, [4, Proposition 5] establishes an achievable rate region for the problem of 3-receiver broadcast channel with degraded message sets. The achievability is through a coding scheme that employs a non-unique decoder. It turns out that employing a joint unique decoder, one can still achieve the same inner-bound of [4]. In this section, we develop a new proof technique to show this equivalence systematically. The same technique allows us to show the equivalence in all the examples considered in Section III.

A. Non-Unique Decoding in the Achievable Scheme of Nair and El Gamal

The main problem studied in [4] is that of sending two messages over a 3-receiver discrete memoryless broadcast channel $p(y_1, y_2, y_3|x)$. The source intends to communicate messages M_0 and M_1 to receiver 1 and message M_0 to receivers 2 and 3. Rates of messages M_0 and M_1 are denoted by R_0 and R_1 , respectively. In [4] an inner-bound to the capacity region is proved using a standard encoding scheme based on superposition coding and Marton's coding, and a non-unique decoding scheme called indirect decoding. We briefly review this scheme.

1) *Random Codebook Generation and Encoding:* To design the codebook, split the private message M_1 into four independent parts M_{10}, M_{11}, M_{12} , and M_{13} of non-negative rates S_0, S_1, S_2, S_3 , respectively. Let $R_1 = S_0 + S_1 + S_2 + S_3$, $T_2 \geq S_2$ and $T_3 \geq S_3$. Fix a joint probability distribution $p(u, v_2, v_3, x)$. Randomly and independently generate $2^{n(R_0+S_0)}$ sequences $U^n(m_0, s_0)$, $m_0 \in [1 : 2^{nR_0}]$ and $s_0 \in [1 : 2^{nS_0}]$, each distributed according to $\prod_i p_U(u_i)$. For each sequence $U^n(m_0, s_0)$, generate randomly and conditionally independently (i) 2^{nT_2} sequences $V_2^n(m_0, s_0, t_2)$, $t_2 \in [1 : 2^{nT_2}]$, each according to $\prod_i p_{V_2|U}(v_{2i}|u_i)$, and (ii) 2^{nT_3} sequences $V_3^n(m_0, s_0, t_3)$, $t_3 \in [1 : 2^{nT_3}]$, each distributed according to $\prod_i p_{V_3|U}(v_{3i}|u_i)$. Randomly partition sequences $V_2^n(m_0, s_0, t_2)$ into 2^{nS_2} bins $\mathcal{B}_2(m_0, s_0, t_2)$ and sequences $V_3^n(m_0, s_0, t_3)$ into 2^{nS_3} bins $\mathcal{B}_3(m_0, s_0, t_3)$. In each product bin $\mathcal{B}_2(m_0, s_0, t_2) \times \mathcal{B}_3(m_0, s_0, t_3)$, choose one (random)

jointly typical sequence pair $(V_2^n(m_0, s_0, t_2), V_3^n(m_0, s_0, t_3))$. If there is no such pair, declare an error whenever the message (m_0, s_0, s_2, s_3) is to be transmitted. Finally for each chosen jointly typical pair $(V_2^n(m_0, s_0, t_2), V_3^n(m_0, s_0, t_3))$ in each product bin (s_2, s_3) , randomly and conditionally independently generate 2^{nS_1} sequences $X^n(m_0, s_0, s_2, s_3, s_1)$, $s_1 \in [1 : 2^{nS_1}]$, each distributed according to $\prod_i p_{X|UV_2V_3}(x_i|u_i, v_{2i}, v_{3i})$. To send the message pair (m_0, m_1) , where m_1 is expressed as (s_0, s_1, s_2, s_3) , the encoder sends the codeword $X^n(m_0, s_0, s_2, s_3, s_1)$.

2) *Non-Unique Decoding*: Receiver Y_1 jointly uniquely decodes all messages $M_0, M_{10}, M_{11}, M_{12}$, and M_{13} . Receivers Y_2 and Y_3 , however, decode M_0 indirectly, through a non-unique decoding scheme. More precisely,

- Receiver Y_1 declares that the message tuple $(m_0, s_0, s_2, s_3, s_1)$ was sent if it is the unique quintuple such that the received signal Y_1^n is jointly typical with $(U^n(m_0, s_0), V_2^n(m_0, s_0, t_2), V_3^n(m_0, s_0, t_3), X^n(m_0, s_0, s_2, s_3, s_1))$, where s_2 is the bin index of $V_2^n(m_0, s_0, t_2)$ and s_3 is the bin index of $V_3^n(m_0, s_0, t_3)$.
- Receiver Y_2 declares that the message pair $(M_0, M_{10}) = (m_0, s_0)$ was sent if it finds a unique pair of indices (m_0, s_0) for which the received signal Y_2^n is jointly typical with $(U^n(m_0, s_0), V_2^n(m_0, s_0, t_2))$ for some index $t_2 \in [1 : 2^{nT_2}]$.
- Receiver Y_3 is similar to receiver Y_2 with V_3 and t_3 , respectively, instead of V_2 and t_2 .

The above encoding/decoding scheme achieves rate pairs (R_0, R_1) for which inequalities (1) to (12) below are satisfied for a joint distribution $p(u, v_2, v_3, x)$. The reader is referred to [4] for the analysis of the error probabilities.

Rate splitting constraints:

$$R_1 = S_0 + S_1 + S_2 + S_3 \quad (1)$$

$$T_2 \geq S_2 \quad (2)$$

$$T_3 \geq S_3 \quad (3)$$

$$S_0, S_1, S_2, S_3 \geq 0 \quad (4)$$

Encoding constraint:

$$T_2 + T_3 \geq S_2 + S_3 + I(V_2; V_3|U) \quad (5)$$

Joint unique decoding constraints at receiver Y_1 :

$$S_1 \leq I(X; Y_1|U, V_2, V_3) \quad (6)$$

$$S_1 + S_2 \leq I(X; Y_1|UV_3) \quad (7)$$

$$S_1 + S_3 \leq I(X; Y_1|UV_2) \quad (8)$$

$$S_1 + S_2 + S_3 \leq I(X; Y_1|U) \quad (9)$$

$$R_0 + S_0 + S_1 + S_2 + S_3 \leq I(X; Y_1) \quad (10)$$

Non-unique decoding constraint at receiver Y_2 :

$$R_0 + S_0 + T_2 \leq I(UV_2; Y_2) \quad (11)$$

Non-unique decoding constraint at receiver Y_3 :

$$R_0 + S_0 + T_3 \leq I(UV_3; Y_3). \quad (12)$$

B. Joint Unique Decoding Suffices in the Achievable Scheme of Nair and El Gamal in [4]

Fix the codebook generation and encoding scheme to be that of Section II-A. We will demonstrate how a joint unique decoding scheme suffices by following these steps:

- (1) We first analyze the non-unique decoder to characterize regimes where it uniquely decodes all the messages it considers and regimes where it decodes some of the messages non-uniquely.
- (2) For each of the regimes, we deduce that the non-unique decoder may be replaced by a joint unique decoder.

For the rest of this section, we only consider decoding schemes at receiver Y_2 . Similar arguments are valid for receiver Y_3 due to the symmetry of the problem. We refer to inequality (11), which is shown in [4] to ensure reliability of the non-unique decoder at receiver Y_2 , as the non-unique decoding constraint (11).

Let the rate pair (R_0, R_1) be such that the non-unique decoder of receiver Y_2 decodes message M_0 with high probability; i.e., the non-unique decoding constraint (11) is satisfied. Consider the following two regimes:

- (a) $R_0 + S_0 < I(U; Y_2)$,
- (b) $R_0 + S_0 > I(U; Y_2)$.

In regime (a), it is clear from the defining condition that a joint unique decoder which decodes $(M_0, M_{10}) = (m_0, s_0)$ by finding the unique sequence $U^n(m_0, s_0)$ such that $(U^n(m_0, s_0), Y_2^n)$ is jointly typical will succeed with high probability. This is the joint unique decoder we may use in place of the non-unique decoder for this regime. Notice that in this regime, while the non-unique decoder obtains (m_0, s_0) uniquely with high probability, it may not necessarily succeed in uniquely decoding t_2 . Indeed, in this regime insisting on joint unique decoding of $U^n(m_0, s_0), V_2^n(m_0, s_0, t_2)$ could, in some cases, result in a strictly smaller achievable region.

Regime (b) is the more interesting regime. Here it is clear that simply decoding for (M_0, M_{10}) and treating all other messages as noise will not work. Non-unique decoding must indeed be taking advantage of the codeword V_2^n as well. The non-unique decoder looks for a unique pair of messages (m_0, s_0) such that there exists some t_2 for which $(U^n(m_0, s_0), V_2^n(m_0, s_0, t_2), Y_2^n)$ is jointly typical. One may, in general, expect that there could be several choices of t_2 even in this regime. An important observation is that, in this regime, there is (with high probability) only one choice for t_2 . In other words, *in this regime, receiver 2 decodes t_2 uniquely along with m_0 and s_0* . To see this, notice that using inequality (11) and (b) above, we have

$$T_2 < I(V_2; Y_2|U). \quad (13)$$

Inequalities (11) and (13) together guarantee that a joint unique decoder can decode messages M_0, M_{10} , and M_{12} with high probability. Note that condition (11) makes the probability of an incorrect estimate for (M_0, M_{10}) vanish; and condition on M_0, M_{10} being correctly estimated, inequality (13) drives the probability of an incorrect estimate for M_{12} to zero. In other words, in regime (b) the non-unique decoder ends up with a unique decoding of the satellite codeword $V_2^n(m_0, s_0, t_2)$ with high probability; i.e., we may replace the non-unique decoder with a joint unique decoder for messages M_0, M_{10}, M_{12} . To summarize loosely, whenever the non-unique decoder is forced to derive information from the codeword V_2^n (i.e., when treating V_2^n as noise will not result in correct decoding), the non-unique decoder will recover this codeword also

uniquely. We make this loose intuition more concrete in Section II-C.

The same argument goes through for receiver Y_3 and this shows that insisting on jointly uniquely decoding at all receivers is not restrictive in this problem. Thus, we arrive at the following:

Theorem 1: For every rate pair (R_0, R_1) satisfying the inner-bound of (1)-(12), there exists a coding scheme employing joint unique decoders which achieves the same rate pair.

The idea behind the proof of Theorem 1 was simple and general. Consider a non-unique decoder which is decoding some messages of interest. The message of interest in our problem is M_0 . Along with this message of interest, the decoder might also decode certain other messages, M_{10} and M_{12} for example. The two main steps of the proof is then as follows.

- (1) Analyze the non-unique decoder to determine what messages it decodes uniquely. Depending on the regime of operation, the non-unique decoder ends up uniquely decoding a subset of its intended messages, and non-uniquely the rest of its intended messages. For example in regime (a) above, the non-unique decoder uniquely decodes only M_0 and M_{10} and it might not be able to settle on M_{12} . While in regime (b), the non-unique decoder ends up decoding all of its three messages M_0 , M_{10} , and M_{12} uniquely.
- (2) In each regime of operation characterized in step (1), use a joint unique decoder to only decode the messages that the non-unique decoder uniquely decodes. In the above proof, this would be a joint unique decoder that decodes M_0 and M_{10} in regime (a) and a joint unique decoder that decodes messages M_0 , M_{10} , and M_{12} in regime (b). Verify that the resulting joint unique decoder does support the corresponding part of the rate region achieved by the non-unique decoding scheme.

Though the idea is generalizable, analyzing the non-unique decoder in step (1) is a tedious task. Even for this very specific problem, it may not be entirely clear how the condition dividing cases (a) and (b) can be derived. Next, we try to resolve this using an approach which generalizes more easily.

C. An Alternative Proof to Theorem 1: An Auxiliary Decoder

We take an alternative approach in this section to prove Theorem 1. The proof technique we present here has the same spirit as the proof in Section II-B, but the task of determining which subset of messages should be decoded in what regimes will be implicit rather than explicit as before. To this end, we introduce an auxiliary decoder which serves as a tool to help us develop the proof ideas. We do not propose this more complicated auxiliary decoder as a new decoding technique, but only as a proof technique to show sufficiency of joint unique decoding in the problem of [4]. We analyze the error probability of the auxiliary decoder at receiver Y_2 and show that under the random coding experiment, it decodes correctly with high probability if the non-unique decoding constraint

(11) holds. From this auxiliary decoder and its performance, we will then be able to conclude that there exists a joint unique decoding scheme that succeeds with high probability.

We now define the auxiliary decoder. The auxiliary decoder at receiver Y_2 is a more involved decoder which has access to two component (joint unique) decoders:

- Decoder 1 is a joint unique decoder which decodes messages M_0 and M_{10} . It finds M_0 , and M_{10} by looking for the unique sequence $U^n(m_0, s_0)$ for which the pair $(U^n(m_0, s_0), Y_2^n)$ is jointly typical, and declares an error if there exists no such unique sequence.
- Decoder 2 is a joint unique decoder which decodes messages M_0 , M_{10} , M_{12} . It finds M_0 , M_{10} , M_{12} by looking for the unique sequences $U^n(m_0, s_0)$ and $V_2^n(m_0, s_0, t_2)$ such that triple $(U^n(m_0, s_0), V_2^n(m_0, s_0, t_2), Y_2^n)$ is jointly typical, and declares an error when such sequences do not exist.

The auxiliary decoder declares an error if either (a) both component decoders declare errors, or (b) if both of them decode, but their decoded (M_0, M_{10}) messages do not match. In all other cases it declares the (M_0, M_{10}) output of the component decoder which did not declare an error as the decoded message.

We analyze the error probability under the random coding experiment of such an auxiliary decoder at receiver Y_2 and prove that for any $\epsilon > 0$, there is a large enough n such that

$$\begin{aligned} & \Pr(\text{error at the auxiliary decoder}) \\ & \leq \epsilon + 2^{n(R_0 + S_0 + T_2 - I(UV_2; Y_2) + \gamma(\epsilon))}, \end{aligned} \quad (14)$$

where $\gamma(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$. Inequality (14) shows that for large enough n and under the non-unique decoding constraint (11), the auxiliary decoder has an arbitrary small probability of failure.

To prove (14), assume that $(m_0, s_0, s_1, s_2, s_3) = (1, 1, 1, 1, 1)$ is sent and indices t_1 and t_2 in the encoding procedure are $(t_2, t_3) = (1, 1)$. This assumption causes no loss of generality due to the symmetry of the codebook construction. We denote the random variables corresponding to these indices by $\mathcal{I}_{m_0}, \mathcal{I}_{s_0}, \dots, \mathcal{I}_{t_3}$ and we refer to the tuple $(\mathcal{I}_{m_0}, \mathcal{I}_{s_0}, \mathcal{I}_{s_1}, \mathcal{I}_{s_2}, \mathcal{I}_{s_3}, \mathcal{I}_{t_2}, \mathcal{I}_{t_3})$ by \mathcal{I} . In the rest of this section, we assume $\mathcal{I} = \mathbf{1}$, the all 1's vector, and analyze the probability that receiver Y_2 declares $M_0 \neq 1$. Receiver Y_2 makes an error in decoding M_0 only if at least one of the following events occur:

- \mathcal{E}_1 : *The channel and/or the encoder is atypical:* the triple $(U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), Y_2^n)$ is not jointly typical.
- \mathcal{E}_2 : *Both decoders fail to decode uniquely and declare errors:* there are at least two distinct pairs $(\tilde{m}_0, \tilde{s}_0)$ and $(\check{m}_0, \check{s}_0)$ such that both pairs $(U^n(\tilde{m}_0, \tilde{s}_0), Y_2^n)$ and $(U^n(\check{m}_0, \check{s}_0), Y_2^n)$ are jointly typical; and similarly there are at least two distinct triples $(\hat{m}_0, \hat{s}_0, \hat{t}_2)$ and $(\check{m}_0, \check{s}_0, \check{t}_2)$ such that both triples $(U^n(\hat{m}_0, \hat{s}_0), V_2^n(\hat{m}_0, \hat{s}_0, \hat{t}_2), Y_2^n)$ and $(U^n(\check{m}_0, \check{s}_0), V_2^n(\check{m}_0, \check{s}_0, \check{t}_2), Y_2^n)$ are jointly typical.

Therefore, the probability that receiver Y_2 makes an error is upper-bounded in terms of the above events.

$$\begin{aligned} & \Pr(\text{error at the auxiliary decoder} | \mathcal{I} = \mathbf{1}) \\ & \leq \Pr(\mathcal{E}_1 | \mathcal{I} = \mathbf{1}) + \Pr(\mathcal{E}_2 \cap \bar{\mathcal{E}}_1 | \mathcal{I} = \mathbf{1}) \\ & \leq \epsilon + \Pr(\mathcal{E}_2 \cap \bar{\mathcal{E}}_1 | \mathcal{I} = \mathbf{1}). \end{aligned} \quad (15)$$

where (15) follows because $\Pr(\mathcal{E}_1 | \mathcal{I} = \mathbf{1}) = \Pr((U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), Y_2^n) \notin A_\epsilon^n | \mathcal{I} = \mathbf{1}) \leq \epsilon$ (ensured by the encoding and the Asymptotic Equipartition Property). To upper-bound $\Pr(\mathcal{E}_2 \cap \bar{\mathcal{E}}_1 | \mathcal{I} = \mathbf{1})$, we write

$$\Pr(\mathcal{E}_2 \cap \bar{\mathcal{E}}_1 | \mathcal{I} = \mathbf{1}) \quad (16)$$

$$\stackrel{(a)}{\leq} \Pr \left(\begin{array}{c} (U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), Y_2^n) \in A_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0, \tilde{s}_0), Y_2^n) \in A_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{s}_0) \neq (1, 1), \\ \text{and} \\ (U^n(\hat{m}_0, \hat{s}_0), V_2^n(\hat{m}_0, \hat{s}_0, \hat{t}_2), Y_2^n) \in A_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{s}_0, \hat{t}_2) \neq (1, 1, 1) \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (17)$$

$$\begin{aligned} & \leq \Pr \left(\begin{array}{c} (U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), Y_2^n) \in A_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0, \tilde{s}_0), Y_2^n) \in A_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{s}_0) \neq (1, 1), \\ \text{and} \\ (U^n(\hat{m}_0, \hat{s}_0), V_2^n(\hat{m}_0, \hat{s}_0, \hat{t}_2), Y_2^n) \in A_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{s}_0) \neq (1, 1) \text{ and } \hat{t}_2 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\ & + \Pr \left(\begin{array}{c} (U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), Y_2^n) \in A_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0, \tilde{s}_0), Y_2^n) \in A_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{s}_0) \neq (1, 1), \\ \text{and} \\ \text{all } (U^n(\hat{m}_0, \hat{s}_0), V_2^n(\hat{m}_0, \hat{s}_0, \hat{t}_2), Y_2^n) \in A_\epsilon^n \\ \text{are s.t. } (\hat{m}_0, \hat{s}_0) = (1, 1) \\ \text{with at least one s.t. } \hat{t}_2 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \end{aligned} \quad (18)$$

In the above chain of inequalities, (a) holds because event $\mathcal{E}_2 \cap \bar{\mathcal{E}}_1$ is a subset of the event on the right hand side.

It is worthwhile to interpret inequality (18). The error event of interest, roughly speaking, is partitioned into the following two events.

- (1) The auxiliary decoder makes an error and the non-unique decoder of Section II-A also makes an error.
- (2) The auxiliary decoder makes an error but the non-unique decoder of Section II-A decodes correctly. We will show that the probability of this event is small. Note that under this error event, (a) component decoder 1 fails (i.e., it is not possible to decode (M_0, M_{10}) by treating V_2^n as noise), but still (b) non-unique decoder succeeds (i.e., the non-unique decoder must be deriving useful information by considering V_2^n). By showing that this error event has a small probability, we in effect show that whenever (a) and (b) hold, it is possible to jointly uniquely decode the V_2^n codeword as well. This makes the rough intuition from Section II-B more concrete.

To bound the error probability, we bound the two terms of inequality (18) separately. The first term of (18) is bounded by the probability of the non-unique decoder making an error:

$$\begin{aligned} & \Pr \left(\begin{array}{c} (U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), Y_2^n) \in A_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0, \tilde{s}_0), Y_2^n) \in A_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{s}_0) \neq (1, 1) \\ \text{and,} \\ (U^n(\hat{m}_0, \hat{s}_0), V_2^n(\hat{m}_0, \hat{s}_0, \hat{t}_2), Y_2^n) \in A_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{s}_0) \neq (1, 1) \text{ and } \hat{t}_2 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\ & \leq \Pr \left((U^n(\hat{m}_0, \hat{s}_0), V_2^n(\hat{m}_0, \hat{s}_0, \hat{t}_2), Y_2^n) \in A_\epsilon^n \middle| \mathcal{I} = \mathbf{1} \right) \\ & \leq \sum_{\substack{(\hat{m}_0, \hat{s}_0) \neq (1, 1) \\ \hat{t}_2}} \Pr(U^n(\hat{m}_0, \hat{s}_0), V_2^n(\hat{m}_0, \hat{s}_0, \hat{t}_2), Y_2^n) \in A_\epsilon^n | \mathcal{I} = \mathbf{1}) \\ & \leq 2^n T_2 2^{n(R_0 + S_0)} 2^{-n(I(U V_2; Y_2) - \gamma_1(\epsilon))}. [5pt] \end{aligned} \quad (19)$$

The second term of (18) is upper-bounded by the expression in (20), as we elaborate.

$$\begin{aligned} & \Pr \left(\begin{array}{c} (U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), Y_2^n) \in A_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0, \tilde{s}_0), Y_2^n) \in A_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{s}_0) \neq (1, 1), \\ \text{and} \\ \text{all } (U^n(\hat{m}_0, \hat{s}_0), V_2^n(\hat{m}_0, \hat{s}_0, \hat{t}_2), Y_2^n) \in A_\epsilon^n \\ \text{are s.t. } (\hat{m}_0, \hat{s}_0) = (1, 1) \\ \text{with at least one s.t. } \hat{t}_2 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\ & \leq 2^{n(R_0 + S_0 + T_2)} 2^{-n(I(U V_2; Y_2) - \gamma_2(\epsilon) - \delta(\epsilon))}. [5pt] \end{aligned} \quad (20)$$

We derive the bound (20) as follows. First, we write the following chain of inequalities.

$$\begin{aligned} & \Pr \left(\begin{array}{c} (U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), Y_2^n) \in A_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0, \tilde{s}_0), Y_2^n) \in A_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{s}_0) \neq (1, 1), \\ \text{and} \\ \text{all } (U^n(\hat{m}_0, \hat{s}_0), V_2^n(\hat{m}_0, \hat{s}_0, \hat{t}_2), Y_2^n) \in A_\epsilon^n \\ \text{are s.t. } (\hat{m}_0, \hat{s}_0) = (1, 1) \\ \text{with at least one s.t. } \hat{t}_2 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\ & \leq \Pr \left(\begin{array}{c} (U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), Y_2^n) \in A_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0, \tilde{s}_0), Y_2^n) \in A_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{s}_0) \neq (1, 1), \\ \text{and} \\ (U^n(1, 1), V_2^n(1, 1, \hat{t}_2), Y_2^n) \in A_\epsilon^n \\ \text{for some } \hat{t}_2 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\ & \leq \sum_{\substack{(\tilde{m}_0, \tilde{s}_0) \neq (1, 1) \\ \hat{t}_2 \neq 1}} \Pr \left(\begin{array}{c} (U^n(1, 1), V_2^n(1, 1, 1), \\ V_3^n(1, 1, 1), Y_2^n) \in A_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0, \tilde{s}_0), Y_2^n) \in A_\epsilon^n \\ \text{and} \\ (U^n(1, 1), V_2^n(1, 1, \hat{t}_2), Y_2^n) \in A_\epsilon^n, \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \end{aligned}$$

$$\leq 2^{n(R_0+S_0+T_2)} \times \Pr \left(\begin{array}{c} (U^n(1, 1), V_2^n(1, 1, 1), \\ V_3^n(1, 1, 1), Y_2^n) \in \mathcal{A}_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0, \tilde{s}_0), Y_2^n) \in \mathcal{A}_\epsilon^n, \\ \text{and} \\ (U^n(1, 1), V_2^n(1, 1, \hat{t}_2), Y_2^n) \in \mathcal{A}_\epsilon^n \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (21)$$

where we have $(\tilde{m}_0, \tilde{s}_0) \neq \mathbf{1}$ and $\hat{t}_2 \neq 1$ in the event in inequality (21).

Next, we bound the probability term in (21). In what follows, $U^n, V_2^n, V_3^n, \tilde{U}^n, \hat{V}_2^n$ denote $U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), U^n(\tilde{m}_0, \tilde{s}_0), V_2^n(1, 1, \hat{t}_2)$, respectively. Also, $p_{U^n|\mathcal{I}}(u^n|\mathbf{1})$ denotes $\Pr(U^n = u^n|\mathcal{I} = \mathbf{1})$. We sometimes drop the subscripts of probabilities if there is no ambiguity; e.g., $p(u^n|\mathbf{1})$ is just $p_{U^n|\mathcal{I}}(u^n|\mathbf{1})$.

In order to bound the probability term in (21), one should treat $p_{U^n V_2^n V_3^n Y_2^n \tilde{U}^n \hat{V}_2^n|\mathcal{I}}(u^n, v_2^n, v_3^n, y_2^n, \tilde{u}^n, \hat{v}_2^n|\mathbf{1})$. This would have been a straightforward task if the generated codebook was independent of indices \mathcal{I} . Nonetheless, it is an important observation that this is not the case.¹ For example, given U^n (and under the conditioning $\mathcal{I} = \mathbf{1}$), Y_2^n may not be independent of \hat{V}_2^n . Interestingly however, almost the same result holds. We address this in the following. We follow the proof idea in [17] to address this technicality.

$$\begin{aligned} & \Pr \left(\begin{array}{c} (U^n, V_2^n, V_3^n, Y_2^n) \in \mathcal{A}_\epsilon^n \text{ and} \\ (\tilde{U}^n, Y_2^n) \in \mathcal{A}_\epsilon^n \text{ and } (U^n, \hat{V}_2^n, Y_2^n) \in \mathcal{A}_\epsilon^n \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\ &= \sum_{(u^n, v_2^n, v_3^n, y_2^n) \in \mathcal{A}_\epsilon^n} \sum_{\substack{\tilde{u}^n: \\ (\tilde{u}^n, y_2^n) \in \mathcal{A}_\epsilon^n}} \sum_{\substack{\hat{v}_2^n: \\ (\hat{v}_2^n, y_2^n) \in \mathcal{A}_\epsilon^n}} p(u^n, v_2^n, v_3^n, y_2^n, \tilde{u}^n, \hat{v}_2^n|\mathbf{1}) \\ &= \sum_{(u^n, v_2^n, v_3^n, y_2^n) \in \mathcal{A}_\epsilon^n} \sum_{\substack{\tilde{u}^n: \\ (\tilde{u}^n, y_2^n) \in \mathcal{A}_\epsilon^n}} \sum_{\substack{\hat{v}_2^n: \\ (\hat{v}_2^n, y_2^n) \in \mathcal{A}_\epsilon^n}} [p(u^n, v_2^n, v_3^n, y_2^n|\mathbf{1}) \\ & \quad \times p(\tilde{u}^n|u^n, v_2^n, v_3^n, y_2^n, \mathbf{1}) \\ & \quad \times p(\hat{v}_2^n|u^n, v_2^n, v_3^n, y_2^n, \tilde{u}^n, \mathbf{1})] \\ &= \sum_{(u^n, v_2^n, v_3^n, y_2^n) \in \mathcal{A}_\epsilon^n} \sum_{\substack{\tilde{u}^n: \\ (\tilde{u}^n, y_2^n) \in \mathcal{A}_\epsilon^n}} \sum_{\substack{\hat{v}_2^n: \\ (\hat{v}_2^n, y_2^n) \in \mathcal{A}_\epsilon^n}} [p(u^n, v_2^n, v_3^n, y_2^n|\mathbf{1}) \\ & \quad \times p_{U^n}(\tilde{u}^n) \\ & \quad \times p(\hat{v}_2^n|u^n, v_2^n, v_3^n, y_2^n, \mathbf{1})] \\ &\stackrel{(a)}{=} \sum_{(u^n, v_2^n, v_3^n, y_2^n) \in \mathcal{A}_\epsilon^n} \sum_{\substack{\tilde{u}^n: \\ (\tilde{u}^n, y_2^n) \in \mathcal{A}_\epsilon^n}} \sum_{\substack{\hat{v}_2^n: \\ (\hat{v}_2^n, y_2^n) \in \mathcal{A}_\epsilon^n}} [p(u^n, v_2^n, v_3^n, y_2^n|\mathbf{1}) \\ & \quad \times p_{U^n}(\tilde{u}^n) \\ & \quad \times p(\hat{v}_2^n|u^n, v_2^n, v_3^n, \mathbf{1})] \\ &\stackrel{(b)}{\leq} \sum_{(u^n, v_2^n, v_3^n, y_2^n) \in \mathcal{A}_\epsilon^n} \sum_{\substack{\tilde{u}^n: \\ (\tilde{u}^n, y_2^n) \in \mathcal{A}_\epsilon^n}} \sum_{\substack{\hat{v}_2^n: \\ (\hat{v}_2^n, y_2^n) \in \mathcal{A}_\epsilon^n}} [p(u^n, v_2^n, v_3^n, y_2^n|\mathbf{1}) \\ & \quad \times p_{U^n}(\tilde{u}^n) \\ & \quad \times 2^{n\delta(\epsilon)} p_{\hat{V}_2^n|U^n}(\hat{v}_2^n|u^n)] \\ &\leq 2^{n\delta(\epsilon)} 2^{-n(I(U; Y_2) - \gamma'(\epsilon))} 2^{-n(I(V_2; Y_2|U) - \gamma'(\epsilon))} \\ & \quad \times \sum_{(u^n, v_2^n, v_3^n, y_2^n) \in \mathcal{A}_\epsilon^n} p(u^n, v_2^n, v_3^n, y_2^n|\mathbf{1}) \\ &\leq 2^{-n(I(UV_2; Y_2) - \gamma_2(\epsilon) - \delta(\epsilon))} \end{aligned}$$

¹This was pointed out to us by anonymous reviewers, to whom we are grateful. Similar observations are made in [17] and [18] where proof techniques were developed to handle such technicalities.

Step (a) follows from the fact that $\hat{V}_2^n - U^n, V_2^n, V_3^n, \mathcal{I} - Y_2^n$ forms a Markov chain. In order to prove step (b), we show that conditioned on U^n , \hat{V}_2^n is “almost” independent of $V_2^n, V_3^n, \mathcal{I}$. More precisely, we use similar steps as in [17, Lemma 1] and show in Appendix A that for any jointly typical tuple $(u^n, v_2^n, v_3^n) \in \mathcal{A}_\epsilon^n$ and any $\epsilon > 0$, there is a large enough n such that $p(\hat{v}_2^n|u^n, v_2^n, v_3^n, \mathbf{1}) \leq 2^{n\delta(\epsilon)} p(\hat{v}_2^n|u^n)$, where $\delta(\epsilon)$ tends to zero as $\epsilon \rightarrow 0$.

We conclude the error probability analysis by putting together inequalities (15), (18), (19), and (20) to obtain that the error probability at the auxiliary decoder is bounded as in inequality (14). So for large enough n , the auxiliary decoder succeeds with high probability if the non-unique decoding constraint (11) is satisfied; i.e., when the non-unique decoder succeeds with high probability.

One can now argue that if the auxiliary decoder succeeds with high probability for an operating point, then there also exists a joint unique decoding scheme that succeeds with high probability. The idea is that for all operating points (except in a subset of the rate region of measure zero), each of the two component (joint unique) decoders 1 and 2 have either a high or a low probability of success. So, if the operating point is such that the auxiliary decoder decodes correctly with high probability, then at least one of the component decoders should also decode correctly with high probability, giving us the joint unique decoding scheme we were looking for. This is summarized in Lemma 1, and the reader is referred to Appendix B for the proof.

Lemma 1: Given any operating point (except in a subset of the rate region of measure zero), if the auxiliary decoder succeeds with high probability under the random coding experiment, then there exists a joint unique decoding scheme that also succeeds with high probability.

A similar argument goes through for receiver Y_3 . The random coding argument for the joint unique decoding scheme can now be completed as usual.

D. Discussion

Remark 4: In Sections II-B and II-C, we did not consider cases where $R_0 + S_0 = I(U; Y_2)$ or $R_0 + S_0 = I(U; Y_3)$ (i.e., a subset of measure zero). This is enough since we may get arbitrarily close to such points.

Remark 5: In Sections II-B and II-C, we fixed the encoding scheme to be that of [4]. The message splitting and the structure of the codebook is therefore a priori assumed to be that of [4], even when $R_0 + S_0 < I(U; Y_2)$ and message M_{12} is not jointly decoded at Y_2 . However, in such cases this extra message structure is not required and one can consider message M_{12} as a part of message M_{11} .

III. MORE EXAMPLES

We saw that joint unique decoding was sufficient to achieve the inner-bound of [4]. This is not coincidental and the same phenomenon can be observed for example in the work of Chong, Motani, Garg and El Gamal [5] where the region obtained by non-unique decoding turned out to be equivalent to that of Han and Kobayashi in [3]. Similarly for noisy

network coding [6], it has been shown that the same rate region can be obtained employing joint unique decoding [11]–[14]. It was also observed in [7] that non-unique decoding is not essential to achieve the capacity region of certain state-dependent multiple access channels and joint unique decoding suffices. Non-unique decoding schemes have appeared also in [8]–[10]. We consider these three problems next and show that employing joint unique decoders, one can achieve the same proposed inner-bounds. To show such equivalence, we use the proof technique that we developed in Section II-C.

A. Two-Receiver Compound Channel With State Noncausally Available at the Encoder

An inner-bound to the common message capacity region of a 2-receiver compound channel with discrete memoryless state noncausally available at the encoder is derived in [8]. The inner-bound is established using superposition coding, Marton's coding, and non-unique decoding schemes. More precisely, the achievable scheme is as follows:

1) *Codebook Generation*: Fix $p_{WUV}(w, u, v)$ and $f(w, u, v, s)$. For each message m , generate randomly and independently 2^{nT_0} sequences $W^n(m, l_0)$ according to $\prod_i p_W(w_i)$. For each (m, l_0) , generate randomly and conditionally independently (i) 2^{nT_1} sequences $U^n(m, l_0, l_1)$ according to $\prod_i p_{U|W}(u_i|w_i)$ and (ii) 2^{nT_2} sequences $V^n(m, l_0, l_2)$ according to $\prod_i p_{V|W}(v_i|w_i)$.

2) *Encoding*: Given message m and state s^n , the encoder finds l_0 such that $(W^n(m, l_0), s^n) \in \mathcal{A}_\epsilon^n$. If there is more than one such index, one is chosen uniformly at random.² If there is no such index, a random index is chosen among $\{1, \dots, 2^{nT_0}\}$. Next, the encoder finds l_1 and l_2 such that $(W^n(m, l_0), s^n, U^n(m, l_0, l_1), V^n(m, l_0, l_2)) \in \mathcal{A}_\epsilon^n$. If there is more than one such index pair, one pair is chosen uniformly at random. If there is none, a random index pair (l_1, l_2) is chosen among $\{1, \dots, 2^{nT_1}\} \times \{1, \dots, 2^{nT_2}\}$. The encoder transmits x^n , $x_i = f(w_i, u_i, v_i, s_i)$, where $w^n = W^n(m, l_0)$, $u^n = U^n(m, l_0, l_1)$, and $v^n = V^n(m, l_0, l_2)$.

3) *Decoding*: Receiver Y_1 declares message M to be the unique index m for which $(W^n(m, l_0), U^n(m, l_0, l_1), Y_1^n)$ is jointly typical for some $l_0 \in \{1, \dots, 2^{nT_0}\}$ and $l_1 \in \{1, \dots, 2^{nT_1}\}$. Receiver Y_2 follows a similar scheme.

In this problem, we show that employing joint unique decoders lets us achieve the same inner-bound of Theorem 1 of [8]. We outline the proof which is built on the proof technique of Subsection II-C. Define the auxiliary decoder (at receiver Y_1) to have access to two component (joint unique) decoders: one jointly uniquely decoding indices m_0, l_0 , and one jointly uniquely decoding indices m_0, l_0, l_1 . The auxiliary decoder declares an error if either (a) both component decoders declare an error or (b) neither of them declare an error but they do not agree on their decoded m_0 and l_0 indices.

We now analyze the error probability. Assume, without any loss of generality, that the originally sent indices were $(m, l_0, l_1, l_2) = (1, 1, 1, 1)$. We denote this event by $\mathcal{I} = \mathbf{1}$. Proceeding as in Section II-C, the error probability of the

auxiliary decoder is bounded by the following probability term.

$$\Pr(\text{error}|\mathcal{I} = \mathbf{1}) \leq \epsilon + \Pr \left(\begin{array}{l} (W^n(1, 1), S^n, U^n(1, 1, 1), \\ V^n(1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n, \\ \text{and} \\ (W^n(\tilde{m}, \tilde{l}_0), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}, \tilde{l}_0) \neq (1, 1), \\ \text{and} \\ (W^n(\hat{m}, \hat{l}_0), U^n(\hat{m}, \hat{l}_0, \hat{l}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}, \hat{l}_0, \hat{l}_1) \neq (1, 1, 1) \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (22)$$

The probability term on the right hand side of inequality (22) is very similar to what we obtained in inequality (17) and is analyzed in the same manner (with the subtle difference that $W^n(m, l_0)$ is indexed not only by the message but also by the state, which asks for a more careful treatment). See Appendix C. We follow similar steps to conclude that the auxiliary decoder performs reliably under the *non-unique decoding constraints* of [8]. So, there exists a joint unique decoding scheme that performs reliably under those decoding constraints. More explicitly, the proposed joint unique decoding scheme would be joint unique decoding of m and l_0 , if $R_0 + T_0 < I(W; Y_1)$; and joint unique decoding of m, l_0 and l_1 , otherwise.

B. Three-User Deterministic Interference Channel

In [9], an inner-bound to the capacity region of a class of deterministic interference channels with three user pairs is derived. The key idea is to simultaneously decode the combined interference signal and the intended message at each receiver and this is done by a non-unique decoding scheme. We focus on [9, Th. 1] and to have the paper self contained we briefly mention the encoding and decoding scheme. The deterministic interference channels that are considered here are described by the following deterministic relations between the inputs and the outputs³: $Y_k = f_k(X_{kk}, S_k)$ where $S_1 = h_1(X_{21}, X_{31})$, $S_2 = h_2(X_{12}, X_{32})$, an $S_3 = h_3(X_{23}, X_{13})$ and $X_{lk} = g_{lk}(X_l)$ for every $l, k \in \{1, 2, 3\}$. It is assumed that functions h_k and f_k are one-to-one mappings when either of their arguments is fixed.

Codebook Generation: Fix the probability mass function (pmf) $p(q)p(x_1|q)p(x_2|q)p(x_3|q)$. Sequence Q^n is generated according to $\prod_i p_Q(q_i)$. For each $k = 1, 2, 3$, sequences $X_k^n(m_k)$, $m_k \in \{1, \dots, 2^{nR_k}\}$, are generated randomly and conditionally independently according to $\prod_i p_{X_k|Q}(x_{k,i}|q_i)$.

Encoding: To send message m_k , transmitter k transmits $X_k^n(m_k)$.

Decoding: Upon receiving Y_1^n , decoder 1 declares that m_1 is sent if it is the unique message such that

$$(Q^n, X_1^n(m_1), S_1^n(m_2, m_3), X_{21}^n(m_2), X_{31}^n(m_3), Y_1^n) \in \mathcal{A}_\epsilon^n$$

for some $m_2 \in [1 : 2^{nR_2}]$ and $m_3 \in [1 : 2^{nR_3}]$. Decoders 2 and 3 work similarly.

²We allow a small modification to [8] in randomly choosing the index l_0 whenever there is not a unique choice.

³All results easily generalize to interference channels with noisy observations (e.g., [9, Th. 4]).

Here, we use the proof technique of Section II-C to prove that a code design that employs joint unique decoders achieves the same inner-bound.

Define the auxiliary decoder (at receiver Y_1) to have access to four component (joint unique) decoders: one jointly uniquely decoding $X^n(m_1)$, one jointly uniquely decoding $X_1^n(m_1)$ and $X_{21}^n(m_2)$, one jointly uniquely decoding $X_1^n(m_1)$ and $X_{31}^n(m_3)$ and finally one jointly uniquely decoding all sequences $X^n(m_1)$, $X_{21}^n(m_2)$, $X_{31}^n(m_3)$, and $S_1^n(m_2, m_3)$. The auxiliary decoder declares an error if either (a) all component decoders declare error, or (b) not all of the decoders that decode without declaring an error agree on the decoded index m_0 (i.e., among those component decoders that do not declare an error, there is not a common agreement on the decoded index m_0).

We now analyze the error probability of the auxiliary decoder. We assume without any loss of generality that $(m_1, m_2, m_3) = (1, 1, 1)$ was sent. Proceeding as in Section II-C, the error probability of the auxiliary decoder is bounded by inequality (23) as follows.

$$\begin{aligned}
 & \Pr(\text{error}) \\
 & \leq \epsilon + \Pr \left(\begin{array}{l} (Q^n, X_1^n(\bar{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \bar{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\check{m}_1), X_{21}^n(\check{m}_2), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\check{m}_1, \check{m}_2) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_1, \hat{m}_3) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_1, \hat{m}_2, \hat{m}_3) \neq (1, 1, 1) \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\
 & \leq \epsilon + \Pr \left(\begin{array}{l} (Q^n, X_1^n(\bar{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \bar{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\check{m}_1), X_{21}^n(\check{m}_2), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\check{m}_1, \check{m}_2) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_1, \hat{m}_3) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_2, \hat{m}_3), \hat{m}_1 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\
 & + \Pr \left(\begin{array}{l} (Q^n, X_1^n(\bar{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \bar{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\check{m}_1), X_{21}^n(\check{m}_2), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\check{m}_1, \check{m}_2) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_1, \hat{m}_3) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2)X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_2, \hat{m}_3) \neq (1, 1), \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (23)
 \end{aligned}$$

As before, the first probability term of inequality (23) is upper-bounded by the probability of an indirect decoder making an error; i.e., by the expression below.

$$\Pr \left(\begin{array}{l} (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_2, \hat{m}_3) \text{ and } \hat{m}_1 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (24)$$

In [9], constraints on rates have been derived under which this error probability approaches 0 as n grows large and an achievable rate region has been characterized. We refer to these constraints as the *non-unique decoding constraints* of [9]. One can show that under these decoding constraints, the second probability term can also be made arbitrarily small by choosing a sufficiently large n (Appendix D). It then becomes clear that the auxiliary decoder succeeds with high probability if the non-unique decoding constraints of [9] are satisfied. So, analogous to Section II-C, we conclude that there exists a joint unique decoding scheme that achieves the same inner-bound of [9, Th. 1].

C. Three-Receiver Broadcast Channel With Common and Confidential Messages

In [10] a general 3-receiver broadcast channel with one common and one confidential message set is studied. Inner-bounds and outer-bounds are derived for the capacity regions under two setups of this problem: when the confidential message is intended for one of the receivers and when the confidential message is intended for two of the receivers. We only address the first setup here, and in particular [10, Th. 2]. The other inner-bounds can be similarly dealt with. In Theorem 2, the authors establish an inner-bound to the secrecy capacity region using the ideas of superposition coding, Wyner wiretap channel coding, and non-unique decoding. We briefly explain the achievable scheme.

Codebook Construction: Fix $p_{UV_0V_1V_2X}(u, v_0, v_1, v_2, x)$. Choose $R_r \geq 0$ such that $R_1 - R_e + R_r \geq I(V_0; Z|U) + \delta(\epsilon)$. Randomly and independently generate 2^{nR_0} sequences $u^n(m_0)$, each according to $\prod_i P_U(u_i)$. For each m_0 , randomly and conditionally independently generate sequences $v_0^n(m_0, m_1, m_r)$, $(m_1, m_r) \in [1 : 2^{n(R_1+R_r)}]$, each according to $\prod_i P_{V_0|U}(v_{0i}|u_i)$. For each (m_0, m_1, m_r) : (i) generate sequences $v_1^n(m_0, m_1, m_r, t_1)$, $t_1 \in \{1, \dots, 2^{nT_1}\}$, each according to $\prod_i P_{V_1|V_0}(v_{1i}|v_{0i})$, and partition the set $\{1, \dots, 2^{nT_1}\}$ into $2^{n\bar{R}_1}$ equal size bins $\mathcal{B}(m_0, m_1, m_r, l_1)$, (ii) generate sequences $v_2^n(m_0, m_1, m_r, t_2)$, $t_2 \in \{1, \dots, 2^{nT_2}\}$, each according to the product distribution $\prod_i P_{V_2|V_0}(v_{2i}|v_{0i})$ and partition the set $\{1, \dots, 2^{nT_2}\}$ into $2^{n\bar{R}_2}$ equal size bins $\mathcal{B}(m_0, m_1, m_r, l_2)$. For each product bin $\mathcal{B}(l_1) \times \mathcal{B}(l_2)$, find a jointly typical sequence pair $(v_1^n(m_0, m_1, m_r, t_1(l_1)), v_2^n(m_0, m_1, m_r, t_2(l_2)))$, and associate it to the product bin. If there is more than one pair, one of the jointly typical pairs is picked uniformly at random. If there is no such pair, one pair is picked uniformly at random from the set of all possible pairs. Finally, for all (m_0, m_1, m_r) and all their associated sequence pairs $(v_1^n(m_0, m_1, m_r, t_1(l_1)), v_2^n(m_0, m_1, m_r, t_2(l_2)))$ a codeword $X^n(m_0, m_1, m_r, t_1(l_1), t_2(l_2))$ is generated according to $\prod_i P_{X|V_0V_1V_2}(x_i|v_{0i}, v_{1i}, v_{2i})$.

Encoding: To send the message pair (m_0, m_1) , the encoder chooses a random index $m_r \in \{1, \dots, 2^{nR_r}\}$ and thus the sequence pair $(u^n(m_0), v_0^n(m_0, m_1, m_r))$. It then chooses a product bin index (L_1, L_2) at random and selects the corresponding jointly typical pair $(v_1^n(m_0, m_1, m_r, t_1(L_1)), v_2^n(m_0, m_1, m_r, t_2(L_2)))$ in it. Finally the corresponding codeword $X^n(m_0, m_1, m_r, t_1(L_1), t_2(L_2))$ is sent.

Decoding: Both legitimate receivers Y_1 and Y_2 decode their messages of interest, M_0 and M_1 , by non-unique decoding schemes. More precisely, receiver Y_1 looks for the unique triple (m_0, m_1, m_r) such that the tuple $(U^n(m_0), V_0^n(m_0, m_1, m_r), V_1^n(m_0, m_1, m_r, t_1), Y_1^n)$ is jointly typical for some $t_1 \in [1 : 2^{nT_1}]$. Receiver Y_2 follows a similar scheme. Receiver Z decodes m_0 directly by finding the jointly typical pair $(U^n(m_0), Z^n)$.

We use the proof technique of Subsection II-C to show that a code design that employs joint unique decoders achieves the same inner-bound. To do so, we first present an auxiliary decoder which succeeds with high probability under the decoding constraints of [10], and then conclude that there exists a joint unique decoding scheme that succeeds with high probability.

Define the auxiliary decoder (at receiver Y_1) to have access to two component (joint unique) decoders, one jointly uniquely decoding indices m_0, m_1, m_r and the other jointly uniquely decoding indices m_0, m_1, m_r, t_1 . The auxiliary decoder declares an error if either (a) both component decoders declare errors, or (b) if both of them decode and their declared (m_0, m_1, m_r) indices do not match. In all other cases it declares the index triple (m_0, m_1, m_r) according to the output of the component decoder which did not declare an error. Proceeding as in Section II-C, the error probability of the auxiliary decoder can be bounded by (25) as follows. As before, we assume without any loss of generality that the all-1-indices are chosen at the encoding stage, and we denote this event by $\mathcal{I} = \mathbf{1}$.

$\Pr(\text{error} \mid \mathcal{I} = \mathbf{1})$

$$\begin{aligned} &\leq \epsilon + \Pr \left(\begin{array}{c} (U^n(1), V_0^n(1, 1, 1), V_1^n(1, 1, 1, 1), \\ V_2^n(1, 1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{and} \\ (U^n(\tilde{m}_0), V_0^n(\tilde{m}_0, \tilde{m}_1, \tilde{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{m}_1, \tilde{m}_r) \neq (1, 1, 1) \\ \text{and} \\ (U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), \\ V_1^n(\hat{m}_0, \hat{m}_1, \hat{m}_r, \hat{t}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{m}_1, \hat{m}_r, \hat{t}_1) \neq (1, 1, 1, 1) \end{array} \mid \mathcal{I} = \mathbf{1} \right) \\ &\leq \epsilon + \Pr \left(\begin{array}{c} (U^n(1), V_0^n(1, 1, 1), V_1^n(1, 1, 1, 1), \\ V_2^n(1, 1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{and} \\ (U^n(\tilde{m}_0), V_0^n(\tilde{m}_0, \tilde{m}_1, \tilde{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{m}_1, \tilde{m}_r) \neq (1, 1, 1) \\ \text{and} \\ (U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), \\ V_1^n(\hat{m}_0, \hat{m}_1, \hat{m}_r, \hat{t}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{m}_1, \hat{m}_r) \neq (1, 1, 1), \hat{t}_1 \neq 1 \end{array} \mid \mathcal{I} = \mathbf{1} \right) \\ &+ \Pr \left(\begin{array}{c} (U^n(1), V_0^n(1, 1, 1), V_1^n(1, 1, 1, 1), \\ V_2^n(1, 1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{and} \\ (U^n(\tilde{m}_0), V_0^n(\tilde{m}_0, \tilde{m}_1, \tilde{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{m}_1, \tilde{m}_r) \neq (1, 1, 1) \\ \text{and} \\ (U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), \\ V_1^n(\hat{m}_0, \hat{m}_1, \hat{m}_r, \hat{t}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for } (\hat{m}_0, \hat{m}_1, \hat{m}_r) = (1, 1, 1), \hat{t}_1 \neq 1 \end{array} \mid \mathcal{I} = \mathbf{1} \right) \end{aligned}$$

$$\begin{aligned} &\stackrel{(a)}{\leq} \epsilon + 2^{n(R_0+R_1+T_1+R_r-I(UV_0V_1;Y_1)+\gamma_1(\epsilon))} \\ &+ 2^{n(R_1+T_1+R_r-I(V_0V_1;Y_1|U)+\gamma'_1(\epsilon))} \\ &+ 2^{n(R_0+R_1+T_1+R_r-I(UV_0V_1;Y_1)+\gamma_2(\epsilon)+\delta(\epsilon))} \\ &+ 2^{n(R_1+T_1+R_r-I(V_0V_1;Y_1|U)+\gamma'_2(\epsilon)+\delta(\epsilon))} \end{aligned} \quad (25)$$

Here, $\gamma_1(\epsilon)$, $\gamma'_1(\epsilon)$, $\gamma_2(\epsilon)$, $\gamma'_2(\epsilon)$, $\delta(\epsilon)$ all go to zero as $\epsilon \rightarrow 0$. To prove the inequality in step (a), we bound each probability term separately.

The first probability term above is upper-bounded by the probability of a non-unique decoder making an error (i.e., $2^{n(R_0+R_1+T_1+R_r-I(UV_0V_1;Y_1)+\gamma_1(\epsilon))} + 2^{n(R_1+T_1+R_r-I(V_0V_1;Y_1|U)+\gamma'_1(\epsilon))}$). This non-unique decoder is analyzed in [10] and shown to be reliable under the following two constraints to which we refer as the *non-unique decoding constraints* of [10].

$$R_0 + R_1 + T_1 + R_r < I(UV_0V_1; Y_1) - \gamma_1(\epsilon) \quad (26)$$

$$R_1 + T_1 + R_r < I(V_0V_1; Y_1|U) - \gamma'_1(\epsilon) \quad (27)$$

The second term is upper-bounded by further splitting the event and following steps similar to that of Subsection II-C.

$$\begin{aligned} &\Pr \left(\begin{array}{c} (U^n(1), V_0^n(1, 1, 1), V_1^n(1, 1, 1, 1), \\ V_2^n(1, 1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0), V_0^n(\tilde{m}_0, \tilde{m}_1, \tilde{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{m}_1, \tilde{m}_r) \neq (1, 1, 1) \\ \text{and} \\ (U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), \\ V_1^n(\hat{m}_0, \hat{m}_1, \hat{m}_r, \hat{t}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{m}_1, \hat{m}_r) = (1, 1, 1), \hat{t}_1 \neq 1 \end{array} \mid \mathcal{I} = \mathbf{1} \right) \\ &\leq \Pr \left(\begin{array}{c} (U^n(1), V_0^n(1, 1, 1), V_1^n(1, 1, 1, 1), \\ V_2^n(1, 1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0), V_0^n(\tilde{m}_0, \tilde{m}_1, \tilde{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{m}_1, \tilde{m}_r) \neq (1, 1, 1), \tilde{m}_0 \neq 1 \\ \text{and} \\ (U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), \\ V_1^n(\hat{m}_0, \hat{m}_1, \hat{m}_r, \hat{t}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{m}_1, \hat{m}_r) = (1, 1, 1), \hat{t}_1 \neq 1 \end{array} \mid \mathcal{I} = \mathbf{1} \right) \\ &+ \Pr \left(\begin{array}{c} (U^n(1), V_0^n(1, 1, 1), V_1^n(1, 1, 1, 1), \\ V_2^n(1, 1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n, \\ \text{and} \\ (U^n(\tilde{m}_0), V_0^n(\tilde{m}_0, \tilde{m}_1, \tilde{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{m}_1, \tilde{m}_r) \neq (1, 1, 1), \tilde{m}_0 = 1 \\ \text{and} \\ (U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), \\ V_1^n(\hat{m}_0, \hat{m}_1, \hat{m}_r, \hat{t}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{m}_1, \hat{m}_r) = (1, 1, 1), \hat{t}_1 \neq 1 \end{array} \mid \mathcal{I} = \mathbf{1} \right) \\ &\leq 2^{n(R_0+R_1+T_1+R_r-I(UV_0V_1;Y_1)+\gamma_2(\epsilon)+\delta(\epsilon))} \\ &+ 2^{n(R_1+T_1+R_r-I(V_0V_1;Y_1|U)+\gamma'_2(\epsilon)+\delta(\epsilon))} \end{aligned} \quad (28)$$

In the last step, the first probability term is bounded by $2^{n(R_0+R_1+T_1+R_r-I(UV_0V_1;Y_1)+\gamma_2(\epsilon)+\delta(\epsilon))}$ based on the derivation in Section II-C, and the second probability term is bounded by $2^{n(R_1+T_1+R_r-I(V_0V_1;Y_1|U)+\gamma'_2(\epsilon)+\delta(\epsilon))}$ for similar reasons

(in the conditional form), the details of which are presented in Appendix E.

It becomes clear from (25), that the auxiliary decoder also succeeds with high probability under the non-unique decoding constraints of [10]. Similar to Subsection II-C, one can conclude that if for an operating point the non-unique decoder succeeds with high probability, then there also exists a joint unique decoding scheme that succeeds with high probability.

One can also use the auxiliary decoder to (explicitly) devise the joint unique decoding scheme. Analogous to Subsection II-C, the decoding scheme could be joint unique decoding of m_0, m_1, m_r in the regime where it succeeds (with high probability) and joint unique decoding of m_0, m_1, m_r, t_1 otherwise. To express the two regimes, we analyze the error probability of the component (joint unique) decoder that decodes m_0, m_1 and m_r .

$$\Pr(\text{error}) \leq \epsilon + 2^{n(R_0+R_1+R_r-I(UV_0;Y_1)+\sigma(\epsilon))} + 2^{n(R_1+R_r-I(V_0;Y_1|U)+\sigma(\epsilon))}$$

where $\sigma(\epsilon) \rightarrow 0$ if $\epsilon \rightarrow 0$. Therefore, joint unique decoding of m_0, m_1 and m_r succeeds with high probability if the following two inequalities hold in addition to the indirect decoding constraints of [10].

$$R_0 + R_1 + R_r < I(UV_0; Y_1) \quad (30)$$

$$R_1 + R_r < I(V_0; Y_1|U) \quad (31)$$

If either of the above inequalities does not hold, then joint unique decoding of m_0, m_1, m_r fails with high probability (see Appendix F). Nonetheless, while the non-unique decoding constraint of [10] is satisfied, since the auxiliary decoder succeeds with high probability, we conclude that joint unique decoding of m_0, m_1, m_r, t_1 succeeds with high probability. So the following joint unique decoding scheme achieves the inner-bound of [10]: If inequalities (30) and (31) hold, jointly uniquely decode indices m_0, m_1 , and m_r , and otherwise, jointly uniquely decode all four indices m_0, m_1, m_r, t_1 .

IV. CONCLUSION

We examined the non-unique decoding strategy of [4] where messages of interest are decoded jointly with other messages even when the decoder is unable to disambiguate uniquely some of the messages which are not of interest to it. We showed that in all known cases where it has been employed, non-unique decoding can be replaced by the classic joint unique decoding strategy without any penalty from a rate region viewpoint. We believe that this technique may be applicable more generally to show the equivalence of rate regions achievable using random coding employing non-unique decoders and joint unique decoders.

APPENDIX A

FOR ANY $\delta > 0$ AND JOINTLY TYPICAL TRIPLES

$$(u^n, v_2^n, v_3^n) \quad p_{\hat{V}_2^n|U^n V_2^n V_3^n \mathcal{I}}(\hat{v}_2^n|u^n, v_2^n, v_3^n, \mathbf{1}) \leq 2^{n\delta} p_{\hat{V}_2^n|U^n}(\hat{v}_2^n|u^n) \text{ FOR } n \text{ LARGE ENOUGH}$$

We proceed along the lines of [17, Lemma 1]. Recall the codebook structure, where (i) $V_2^n(m_0, s_0, t_2)$

and $V_3^n(m_0, s_0, t_3)$ are superposed on $U^n(m_0, s_0)$, (ii) $V_2^n(m_0, s_0, t_2)$ and $V_3(m_0, s_0, t_3)$ are distributed into bins $\mathcal{B}_2(m_0, s_0, s_2)$ and $\mathcal{B}_3(m_0, s_0, s_3)$ and (iii) that a jointly typical pair $(V_2^n(m_0, s_0, t_2), V_3^n(m_0, s_0, t_3))$ is chosen randomly in each product bin. In the error analysis of Section II-C, we assumed all sent indices to be 1, and we considered the event of decoding a wrong index \hat{t}_2 (and thus an incorrect sequence $V_2^n(m_0, s_0, \hat{t}_2)$). We denote $U^n(1, 1), V_2^n(1, 1, 1), V_3^n(1, 1, 1), V_2^n(1, 1, \hat{t}_2)$ by $U^n, V_2^n, V_3^n, \hat{V}_2^n$, respectively. If $V_2^n(1, 1, 1)$ and $V_2^n(1, 1, \hat{t}_2)$ belong to two different bins $\mathcal{B}_1(1, 1, s_2)$ and $\mathcal{B}_1(1, 1, s'_2)$, $s_2 \neq s'_2$, then it is easy to see that the relation $p_{\hat{V}_2^n|U^n V_2^n V_3^n \mathcal{I}}(\hat{v}_2^n|u^n, v_2^n, v_3^n, \mathbf{1}) = p_{\hat{V}_2^n|U^n}(\hat{v}_2^n|u^n)$ holds. Here we only need to consider the case where \hat{t}_2 is such that $V_2^n(1, 1, 1)$ and $V_2^n(1, 1, \hat{t}_2)$ belong to the same bin, i.e., $\mathcal{B}_2(1, 1, 1)$. We assume without any loss of generality that $\hat{t}_2 = 2$.

Define the random ensemble $\mathcal{C}' \in \mathcal{C}'$ as the overall collection of all sequences $(V_2^n(1, 1, t_2))$ and $(V_3^n(1, 1, t_3))$, where $t_2 \in \{3, \dots, 2^{n(T_2-S_2)}\}$ and $t_3 \in \{2, \dots, 2^{n(T_3-S_3)}\}$. For a given \mathbf{c}' , define $N_1(v_2^n, v_3^n, \mathbf{c}')$ to be the number of jointly typical pairs $(v_2^n(1, 1, t_2), v_3^n(1, 1, t_3))$ for all $t_2 \neq 2, t_3$. Similarly, given \mathbf{c}' and $v_2^n(1, 1, 2)$, let $N_2(v_2^n(1, 1, 2), v_3^n, \mathbf{c}')$ be the number of jointly typical pairs $(v_2^n(1, 1, 2), v_3^n(1, 1, t_3))$ for all t_3 .

We now write

$$\begin{aligned} & p_{\hat{V}_2^n|U^n V_2^n V_3^n \mathcal{I}}(\hat{v}_2^n|u^n, v_2^n, v_3^n, \mathbf{1}) \\ &= p_{\hat{V}_2^n|U^n V_2^n V_3^n \mathcal{I}_{t_2} \mathcal{I}_{t_3}}(\hat{v}_2^n|u^n, v_2^n, v_3^n, 1, 1) \\ &= \sum_{\mathbf{c}' \in \mathcal{C}'} p_{\hat{V}_2^n \mathbf{C}'|U^n V_2^n V_3^n \mathcal{I}_{t_2} \mathcal{I}_{t_3}}(\hat{v}_2^n, \mathbf{c}'|u^n, v_2^n, v_3^n, 1, 1) \\ &= \sum_{\mathbf{c}' \in \mathcal{C}'} \left[p(\mathbf{c}'|u^n, v_2^n, v_3^n, 1, 1) p_{\hat{V}_2^n|U^n}(\hat{v}_2^n|u^n, v_2^n, v_3^n, \mathbf{c}') \right. \\ & \quad \left. \times \frac{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1|u^n, v_2^n, v_3^n, \hat{v}_2^n, \mathbf{c}')}{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1|u^n, v_2^n, v_3^n, \mathbf{c}')} \right] \\ &= p_{\hat{V}_2^n|U^n}(\hat{v}_2^n|u^n) \sum_{\mathbf{c}' \in \mathcal{C}'} \left[p(\mathbf{c}'|u^n, v_2^n, v_3^n, 1, 1) \right. \\ & \quad \left. \times \frac{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1|u^n, v_2^n, v_3^n, \hat{v}_2^n, \mathbf{c}')}{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1|u^n, v_2^n, v_3^n, \mathbf{c}')} \right] \end{aligned} \quad (32)$$

To continue bounding (32), we consider two cases.

1) $T_3 - S_3 - I(V_2; V_3|U) < 0$: We bound the fraction in (32). The numerator is bounded from above by disregarding \hat{v}_2^n .

$$\begin{aligned} & p(\mathcal{I}_{t_2} = 1, \mathcal{I}_{t_3} = 1|u^n, v_2^n, v_3^n, \hat{v}_2^n, \mathbf{c}') \\ &= \frac{1}{N_1(v_2^n, v_3^n, \mathbf{c}') + N_2(\hat{v}_2^n, v_3^n, \mathbf{c}')} \\ &\leq \frac{1}{N_1(v_2^n, v_3^n, \mathbf{c}')}. \end{aligned} \quad (33)$$

The denominator is bounded from below by the expression in (34).

$$\begin{aligned}
& p(\mathcal{I}_{t_2} = 1, \mathcal{I}_{t_3} = 1 | u^n, v_2^n, v_3^n, \mathbf{c}') \\
& \geq p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1, N_2(\hat{V}_2^n, v_3^n, \mathbf{c}')=0 | u^n, v_2^n, v_3^n, \mathbf{c}') \\
& \geq p(N_2(\hat{V}_2^n, v_3^n, \mathbf{c}') = 0 | u^n, v_2^n, v_3^n, \mathbf{c}') \\
& \quad \times p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1 | N_2(\hat{V}_2^n, v_3^n, \mathbf{c}')=0, u^n, v_2^n, v_3^n, \mathbf{c}') \\
& = p(N_2(\hat{V}_2^n, v_3^n, \mathbf{c}') = 0 | u^n, v_2^n, v_3^n, \mathbf{c}') \\
& \quad \times \frac{1}{N_1(v_2^n, v_3^n, \mathbf{c}')} \\
& \geq \left(1 - 2^{n(T_3-S_3)} 2^{-n(I(V_2; V_3|U) - \sigma(\epsilon))}\right) \\
& \quad \times \frac{1}{N_1(v_2^n, v_3^n, \mathbf{c}')} \tag{34}
\end{aligned}$$

Comparing (33) and (34) (under the assumption that $T_3 - S_3 - I(V_2; V_3|U) < 0$), it becomes clear that we have $p(\hat{\delta}_2^n | u^n, v_2^n, v_3^n) \leq 2^{n\delta} p(\delta_2^n | u^n)$ for every $\delta > 0$ and n large enough.

- 2) $T_3 - S_3 - I(V_2; V_3|U) > 0$: in this case, we first re-write expression (32) as follows.

$$\begin{aligned}
& p_{\hat{V}_2^n | U^n}(\hat{\delta}_2^n | u^n) \sum_{\mathbf{c}' \in \mathcal{C}'} \left[p(\mathbf{c}' | u^n, v_2^n, v_3^n, 1, 1) \right. \\
& \quad \left. \times \frac{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1 | u^n, v_2^n, v_3^n, \hat{\delta}_2^n, \mathbf{c}')}{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1 | u^n, v_2^n, v_3^n, \mathbf{c}')} \right] \\
& = p(\hat{\delta}_2^n | u^n) \sum_{\mathbf{c}' \in \mathcal{C}'} \left[\frac{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1, \mathbf{c}' | u^n, v_2^n, v_3^n)}{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1 | u^n, v_2^n, v_3^n)} \right. \\
& \quad \left. \times \frac{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1 | u^n, v_2^n, v_3^n, \hat{\delta}_2^n, \mathbf{c}')}{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1 | u^n, v_2^n, v_3^n, \mathbf{c}')} \right] \\
& = p(\hat{\delta}_2^n | u^n) \sum_{\mathbf{c}' \in \mathcal{C}'} \left[\frac{p(\mathbf{c}' | u^n, v_2^n, v_3^n)}{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1 | u^n, v_2^n, v_3^n)} \right. \\
& \quad \left. \times p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1 | u^n, v_2^n, v_3^n, \hat{\delta}_2^n, \mathbf{c}') \right] \\
& = p(\hat{\delta}_2^n | u^n) \sum_{\mathbf{c}' \in \mathcal{C}'} \left[\frac{p(\mathbf{c}' | u^n, v_2^n, v_3^n, \hat{\delta}_2^n)}{p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1 | u^n, v_2^n, v_3^n)} \right. \\
& \quad \left. \times p(\mathcal{I}_{t_2}=1, \mathcal{I}_{t_3}=1 | u^n, v_2^n, v_3^n, \hat{\delta}_2^n, \mathbf{c}') \right] \\
& = p(\hat{\delta}_2^n | u^n) \frac{p(\mathcal{I}_{t_2} = 1, \mathcal{I}_{t_3} = 1 | u^n, v_2^n, v_3^n, \hat{\delta}_2^n)}{p(\mathcal{I}_{t_2} = 1, \mathcal{I}_{t_3} = 1 | u^n, v_2^n, v_3^n)} \tag{35}
\end{aligned}$$

The following claim will be the key in bounding the fraction in (35).

Claim 1: Let all sequences $V_2^n(1, 1, t_2)$, $t_2 \neq 1$, and $V_3^n(1, 1, t_3)$, $t_3 \neq 1$, be picked randomly in the product bin of interest. The event where the number of jointly typical pairs in a row is much larger than the total remaining number of jointly typical pairs in the bin has a probability which decays to zero double exponentially fast with n (under the assumption $T_3 - S_3 > I(V_2; V_3|U)$); i.e., for some constant $\alpha, \beta > 0$,

$$\begin{aligned}
& \Pr(N_2(\hat{V}_2^n, v_3^n, \mathbf{C}') > 2^{2+2n\delta(\epsilon)} N_1(v_2^n, v_3^n, \mathbf{C}') | U^n = u^n) \\
& \leq \beta \exp\left(-\alpha 2^{n(T_3-S_3-I(V_2; V_3|U)-\delta(\epsilon))}\right) \tag{36}
\end{aligned}$$

Proof: Let $N_3(v_3^n, \mathbf{C}')$ be the number of jointly typical pairs $(V_2^n(1, 1, 3), V_3^n(1, 1, t_3))$, where $t_3 = 1, \dots, 2^{n(T_3-S_3)}$. Obviously, $N_1(v_2^n, v_3^n, \mathbf{C}') \geq N_3(v_3^n, \mathbf{C}')$. To prove the claim, it is sufficient to show that

$$\begin{aligned}
& \Pr\left(N_2(\hat{V}_2^n, v_3^n, \mathbf{C}') > 2^{1+n\delta(\epsilon)} 2^{n(T_3-S_3-I(V_2; V_3|U))} | U^n = u^n\right) \\
& \leq \beta_1 \exp\left(-\alpha_1 2^{n(T_3-S_3-I(V_2; V_3|U)-\delta(\epsilon))}\right) \tag{37}
\end{aligned}$$

for some $\alpha_1, \beta_1 > 0$, and that

$$\begin{aligned}
& \Pr\left(N_3(v_3^n, \mathbf{C}') < 2^{-1-n\delta(\epsilon)} 2^{n(T_3-S_3-I(V_2; V_3|U))} | U^n = u^n\right) \\
& \leq \beta_2 \exp\left(-\alpha_2 2^{n(T_3-S_3-I(V_2; V_3|U)-\delta(\epsilon))}\right) \tag{38}
\end{aligned}$$

for some $\alpha_2, \beta_2 > 0$. Both of the above inequalities can be shown using standard Chernoff bounding techniques. We defer the interested reader to Appendix G. ■ The numerator of (35) is bounded from above by disregarding $\hat{\delta}_2^n$.

$$\begin{aligned}
& \Pr(\mathcal{I}_{t_2} = 1, \mathcal{I}_{t_3} = 1 | u^n, v_2^n, \hat{\delta}_2^n, v_3^n) \\
& \leq \sum_{\mathbf{c}'} p(\mathbf{c}' | u^n) \frac{1}{N_1(v_2^n, v_3^n, \mathbf{c}')} \\
& = \mathbb{E}\left[\frac{1}{N_1(v_2^n, v_3^n, \mathbf{C}')} \mid U^n = u^n\right] \tag{39}
\end{aligned}$$

Similarly for the denominator we have the lower bound in (40).

$$\begin{aligned}
& \Pr(\mathcal{I}_{t_2} = 1, \mathcal{I}_{t_3} = 1 | u^n, v_2^n, v_3^n) \\
& = \sum_{\mathbf{c}', \tilde{v}_2^n} \left[p_{\mathbf{C}' \hat{V}_2^n | U^n}(\mathbf{c}', \tilde{v}_2^n | u^n) \right. \\
& \quad \left. \times p_{\mathcal{I}_{t_2} \mathcal{I}_{t_3} | U^n v_2^n v_3^n \hat{V}_2^n \mathbf{C}'}(1, 1 | u^n, v_2^n, v_3^n, \tilde{v}_2^n, \mathbf{c}') \right] \\
& = \sum_{\mathbf{c}', \tilde{v}_2^n} \left[p_{\mathbf{C}' \hat{V}_2^n | U^n}(\mathbf{c}', \tilde{v}_2^n | u^n) \right. \\
& \quad \left. \times \frac{1}{N_1(v_2^n, v_3^n, \mathbf{c}') + N_2(\tilde{v}_2^n, v_3^n, \mathbf{c}')} \right] \\
& \geq \sum_{\mathbf{c}', \tilde{v}_2^n: N_2(\tilde{v}_2^n, v_3^n, \mathbf{c}') \leq 2^{2+2n\delta(\epsilon)} N_1(v_2^n, v_3^n, \mathbf{c}')} \left[p_{\mathbf{C}' \hat{V}_2^n | U^n}(\mathbf{c}', \tilde{v}_2^n | u^n) \right. \\
& \quad \left. \times \frac{1}{N_1(v_2^n, v_3^n, \mathbf{c}') + N_2(\tilde{v}_2^n, v_3^n, \mathbf{c}')} \right] \\
& \geq \sum_{\mathbf{c}', \tilde{v}_2^n: N_2(\tilde{v}_2^n, v_3^n, \mathbf{c}') \leq 2^{2+2n\delta(\epsilon)} N_1(v_2^n, v_3^n, \mathbf{c}')} \left[p_{\mathbf{C}' \hat{V}_2^n | U^n}(\mathbf{c}', \tilde{v}_2^n | u^n) 2^{-3-2n\delta(\epsilon)} \frac{1}{N_1(v_2^n, v_3^n, \mathbf{c}')} \right] \\
& = 2^{-3-2n\delta(\epsilon)} \left[\sum_{\mathbf{c}', \tilde{v}_2^n} p_{\mathbf{C}' \hat{V}_2^n | U^n}(\mathbf{c}', \tilde{v}_2^n | u^n) \frac{1}{N_1(v_2^n, v_3^n, \mathbf{c}')} \right. \\
& \quad \left. - \sum_{\mathbf{c}', \tilde{v}_2^n: N_2(\tilde{v}_2^n, v_3^n, \mathbf{c}') > 2^{2+2n\delta(\epsilon)} N_1(v_2^n, v_3^n, \mathbf{c}')} p_{\mathbf{C}' \hat{V}_2^n | U^n}(\mathbf{c}', \tilde{v}_2^n | u^n) \frac{1}{N_1(v_2^n, v_3^n, \mathbf{c}')} \right]
\end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{\geq} 2^{-3-2n\delta(\epsilon)} \mathbb{E} \left[\frac{1}{N_1(v_2^n, v_3^n, \mathbf{C}')} \middle| U^n = u^n \right] \\
&\quad - 2^{-3-2n\delta(\epsilon)} \Pr \left(\frac{N_2(\hat{V}_2^n, v_3^n, \mathbf{C}')}{2^{2+2n\delta(\epsilon)} N_1(v_2^n, v_3^n, \mathbf{C}')} > 1 \middle| U^n = u^n \right) \\
&\stackrel{(b)}{\geq} 2^{-3-2n\delta(\epsilon)} (1-c) \mathbb{E} \left[\frac{1}{N_1(v_2^n, v_3^n, \mathbf{C}')} \middle| U^n = u^n \right], \quad c > 0
\end{aligned} \tag{40}$$

In the above, (a) holds because $N_1(v_2^n, v_3^n, \mathbf{C}') \geq 1$ (ensured by the assumption that $(v_2^n, v_3^n) \in \mathcal{A}_\epsilon^n$). Step (b) holds, for any constant $c > 0$ and large enough n , by Claim 1 as follows.

$$\begin{aligned}
&\Pr(N_2(\hat{V}_2^n, v_3^n, \mathbf{C}') > 2^{2+2n\delta(\epsilon)} N_1(v_2^n, v_3^n, \mathbf{C}') | U^n = u^n) \\
&\leq \beta \exp \left(-\alpha 2^n (T_3 - S_3 - I(V_2; V_3 | U) - \delta(\epsilon)) \right) \\
&\stackrel{(a)}{\leq} \frac{c}{2} 2^{-n(T_2 - S_2 + T_3 - S_3 - I(V_2; V_3 | U) + \delta(\epsilon))} \\
&\leq c \frac{1}{\mathbb{E} [N_1(v_2^n, v_3^n, \mathbf{C}') | U^n = u^n]} \\
&\leq c \mathbb{E} \left[\frac{1}{N_1(v_2^n, v_3^n, \mathbf{C}')} \middle| U^n = u^n \right]
\end{aligned}$$

In step (a) above, we have used the fact that $T_3 - S_3 > I(V_2; V_3 | U) - \delta(\epsilon)$, $T_2 \geq S_2$, and that n is large enough. Finally, upper bounding the numerator of (35) by (39) and lower bounding its denominator by (40), we reach to a factor with an exponent of order $n\delta(\epsilon)$. Inserting this back into (35), we conclude that for every $\delta > 0$ and n large enough $p(\hat{v}_2^n | u^n, v_2^n, v_3^n) \leq 2^{n\delta} p(\hat{v}_2^n | u^n)$.

APPENDIX B

PROOF TO LEMMA 1

We start by proving the following claim.

Claim 2: Component decoder 1 succeeds with high probability (averaged over codebooks) if $R_0 + S_0 < I(U; Y_2)$, and fails with high probability, if $R_0 + S_0 > I(U; Y_2)$.

Proof of Claim 2: Component decoder 1 makes an error in decoding only if one of the following events occur:

- (i) $(U^n(1, 1), Y_2^n)$ is not jointly typical. The probability of this event can be made arbitrarily small by choosing a large enough n .
- (ii) There exists a pair of indices $(\hat{m}_0, \hat{s}_0) \neq (1, 1)$ such that $(U^n(\hat{m}_0, \hat{s}_0), Y_2^n)$ is jointly typical.

To analyze the error probability, we assume without any loss of generality that the originally sent indices are $m_0 = 1$ and $s_0 = 1$. The error probability is thus upper-bounded by

$$\begin{aligned}
&\Pr(\text{error at component decoder 1}) \\
&\leq \epsilon + \Pr \left(\begin{array}{l} (U^n(\hat{m}_0, \hat{s}_0), Y_2^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{s}_0) \neq (1, 1) \end{array} \middle| m_0 = 1, s_0 = 1 \right) \\
&\leq \epsilon + 2^{n(R_0 + S_0 - I(U; Y_2) + \delta(\epsilon))},
\end{aligned}$$

where $\delta(\epsilon) \rightarrow 0$ if $\epsilon \rightarrow 0$. This proves that for large enough n , the error probability of component decoder 1 could be made arbitrarily small if $R_0 + S_0 < I(U; Y_2)$.

On the other hand, decoder 1 makes an error if there exists an index pair $(\hat{m}_0, \hat{s}_0) \neq (1, 1)$ such that $(U^n(\hat{m}_0, \hat{s}_0), Y_2^n)$

is jointly typical. The probability of error at decoder 1 is, therefore, lower-bounded by

$$\Pr \left(\begin{array}{l} (U^n(\hat{m}_0, \hat{s}_0), Y_2^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{s}_0) \neq (1, 1) \end{array} \middle| m_0 = 1, s_0 = 1 \right), \tag{41}$$

and we want to show that it is arbitrarily close to 1 if $R_0 + S_0 > I(U; Y_2)$. We instead look at the complementary event, $\{(U^n(\hat{m}_0, \hat{s}_0), Y_2^n) \notin \mathcal{A}_\epsilon^n \text{ for all } (\hat{m}_0, \hat{s}_0) \neq (1, 1)\}$, and show that its probability can be made arbitrarily small.

$$\begin{aligned}
&\Pr \left(\begin{array}{l} (U^n(\hat{m}_0, \hat{s}_0), Y_2^n) \notin \mathcal{A}_\epsilon^n \\ \text{for all } (\hat{m}_0, \hat{s}_0) \neq (1, 1) \end{array} \middle| m_0 = 1, s_0 = 1 \right) \\
&= \sum_{y_2^n} \left[\Pr(Y_2^n = y_2^n | m_0 = 1, s_0 = 1) \right. \\
&\quad \left. \times \Pr \left(\begin{array}{l} (U^n(\hat{m}_0, \hat{s}_0), y_2^n) \notin \mathcal{A}_\epsilon^n \\ \text{for all } (\hat{m}_0, \hat{s}_0) \neq (1, 1) \end{array} \middle| \begin{array}{l} Y_2^n = y_2^n, \\ m_0 = 1, \\ s_0 = 1 \end{array} \right) \right] \\
&\leq \epsilon + \sum_{y_2^n \in \mathcal{A}_\epsilon^n} \left[\Pr(Y_2^n = y_2^n | m_0 = 1, s_0 = 1) \right. \\
&\quad \left. \times \Pr \left(\begin{array}{l} (U^n(\hat{m}_0, \hat{s}_0), y_2^n) \notin \mathcal{A}_\epsilon^n \\ \text{for all } (\hat{m}_0, \hat{s}_0) \neq (1, 1) \end{array} \middle| \begin{array}{l} Y_2^n = y_2^n, \\ m_0 = 1, \\ s_0 = 1 \end{array} \right) \right] \\
&= \epsilon + \sum_{y_2^n \in \mathcal{A}_\epsilon^n} \left[\Pr(Y_2^n = y_2^n | m_0 = 1, s_0 = 1) \right. \\
&\quad \left. \times \prod_{(\hat{m}_0, \hat{s}_0) \neq (1, 1)} \Pr \left(\begin{array}{l} (U^n(\hat{m}_0, \hat{s}_0), y_2^n) \notin \mathcal{A}_\epsilon^n \\ \begin{array}{l} Y_2^n = y_2^n, \\ m_0 = 1, \\ s_0 = 1 \end{array} \end{array} \right) \right] \\
&\leq \epsilon + \sum_{y_2^n \in \mathcal{A}_\epsilon^n} \left[\Pr(Y_2^n = y_2^n | m_0 = 1, s_0 = 1) \right. \\
&\quad \left. \times \left(1 - (1 - \epsilon) 2^{-n(I(U; Y_2) + 2\epsilon)} \right)^{(2^{n(R_0 + S_0)} - 1)} \right] \\
&\leq \epsilon + \left(1 - (1 - \epsilon) 2^{-n(I(U; Y_2) + 2\epsilon)} \right)^{(2^{n(R_0 + S_0)} - 1)}.
\end{aligned}$$

In the limit of $n \rightarrow \infty$, we have

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \left(1 - (1 - \epsilon) 2^{-n(I(U; Y_2) + 2\epsilon)} \right)^{(2^{n(R_0 + S_0)} - 1)} \\
&= \lim_{n \rightarrow \infty} \exp \left\{ - \left(2^{n(R_0 + S_0)} (1 - \epsilon) 2^{-n(I(U; Y_2) + 2\epsilon)} \right) \right\},
\end{aligned}$$

which (for any $0 < \epsilon < 1$) goes to 0 as n grows large, if $R_0 + S_0 > I(U; Y_2) + 2\epsilon$. ■

From Claim 2, it becomes clear that for each operating point, averaged over codebooks, component decoder 1 either succeeds with high probability if $R_0 + S_0 < I(U; Y_2)$ or fails with high probability if $R_0 + S_0 > I(U; Y_2)$. In the former case, we let the joint unique decoding scheme be that of decoder 1, and in the latter, we let the joint unique decoding scheme be that of decoder 2. We prove in the following that this joint unique decoding scheme is reliable (averaged over the codebooks) since the auxiliary decoder is reliable.

Consider an operating point for which decoder 1 fails with high probability. In such cases, we assumed the decoding scheme to be joint unique decoding of messages M_0 , M_{10} , and M_{12} . For this operating point, the probability of error of our joint unique decoder is

$$\begin{aligned} & \Pr(\text{error at component decoder 2}) \\ & \leq \Pr \left(\begin{array}{l} \text{error at component decoder 2} \\ \text{and component decoder 1 succeeds} \end{array} \right) \\ & \quad + \Pr \left(\begin{array}{l} \text{error at component decoder 2} \\ \text{and component decoder 1 fails} \end{array} \right) \\ & \stackrel{(a)}{\leq} \delta + \Pr \left(\begin{array}{l} \text{error at component decoder 2} \\ \text{and component decoder 1 fails} \end{array} \right) \\ & \leq \delta + \Pr(\text{error at the auxiliary decoder}). \end{aligned}$$

In the above chain of inequalities, (a) follows from the assumption on the operating point. Also, δ and ϵ can both be taken arbitrarily close to 0 for large enough n . It is now easy to see that given an operating point for which component decoder 1 fails, component decoder 2 succeeds with high probability if the auxiliary decoder succeeds with high probability.

APPENDIX C

THE ERROR PROBABILITY ANALYSIS OF (22) (SECTION III-A)

We proceed as in Section II-C. We start by splitting the error event into two events:

$$\begin{aligned} & \Pr \left(\begin{array}{l} (W^n(1, 1), S^n, U^n(1, 1, 1), \\ V^n(1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n, \\ \text{and} \\ (W^n(\tilde{m}, \tilde{l}_0), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}, \tilde{l}_0) \neq (1, 1), \\ \text{and} \\ (W^n(\hat{m}, \hat{l}_0), U^n(\hat{m}, \hat{l}_0, \hat{l}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}, \hat{l}_0, \hat{l}_1) \neq (1, 1, 1) \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\ & \leq \epsilon + \Pr \left(\begin{array}{l} (W^n(1, 1), S^n, U^n(1, 1, 1), \\ V^n(1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n, \\ \text{and} \\ (W^n(\tilde{m}, \tilde{l}_0), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}, \tilde{l}_0) \neq (1, 1), \\ \text{and} \\ (W^n(\hat{m}, \hat{l}_0), U^n(\hat{m}, \hat{l}_0, \hat{l}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}, \hat{l}_0) \neq (1, 1) \text{ and } \hat{l}_1 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\ & \quad + \Pr \left(\begin{array}{l} (W^n(1, 1), S^n, U^n(1, 1, 1), \\ V^n(1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n, \\ \text{and} \\ (W^n(\tilde{m}, \tilde{l}_0), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}, \tilde{l}_0) \neq (1, 1), \\ \text{and} \\ (W^n(1, 1), U^n(1, 1, \hat{l}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{l}_1 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (42) \end{aligned}$$

The first term in (42) is bounded by

$$\Pr \left(\begin{array}{l} (W^n(\tilde{m}, \tilde{l}_0), U^n(\tilde{m}, \tilde{l}_0, \tilde{l}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}, \tilde{l}_0) \neq (1, 1) \text{ and } \tilde{l}_1 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right). \quad (43)$$

If $W^n(\tilde{m}, \tilde{l}_0)$ was independent of Y_1^n (for $(\tilde{m}, \tilde{l}_0) \neq (1, 1)$), this would have been the non-unique decoding error probability. However, the conditioning on $\mathcal{I} = \mathbf{1}$ makes this not exactly true. Nonetheless, this probability term is still “almost” the non-unique decoding error probability. We make this statement more precise. Let W^n , U^n , V^n , \tilde{W}^n , and \tilde{U}^n denote respectively $W^n(1, 1)$, $U^n(1, 1, 1)$, $V^n(1, 1, 1)$, $W^n(\tilde{m}, \tilde{l}_0)$, and $U^n(\tilde{m}, \tilde{l}_0, \tilde{l}_1)$. The above probability term is upper-bounded by

$$2^{n(R+T_0+T_1)} \sum_{(\tilde{w}^n, \tilde{u}^n, y_1^n) \in \mathcal{A}_\epsilon^n} p_{Y_1^n \tilde{W}^n \tilde{U}^n | \mathcal{I}}(y_1^n, \tilde{w}^n, \tilde{u}^n | \mathbf{1}),$$

and the inner pmf may be written as follows.

$$\begin{aligned} & p_{Y_1^n \tilde{W}^n \tilde{U}^n | \mathcal{I}}(y_1^n, \tilde{w}^n, \tilde{u}^n | \mathbf{1}) \\ & = p_{Y_1^n | \mathcal{I}}(y_1^n | \mathbf{1}) p_{\tilde{W}^n | Y_1^n \mathcal{I}}(\tilde{w}^n | y_1^n, \mathbf{1}) p_{\tilde{U}^n | \tilde{W}^n Y_1^n \mathcal{I}}(\tilde{u}^n | \tilde{w}^n, y_1^n, \mathbf{1}) \\ & \stackrel{(a)}{\leq} p_{W^n}(\tilde{w}^n) (1 + c(\epsilon)) p_{Y_1^n | \mathcal{I}}(y_1^n | \mathbf{1}) p_{U^n | W^n}(\tilde{u}^n | \tilde{w}^n) \end{aligned}$$

In the above inequality, (a) follows by $p_{\tilde{W}^n | Y_1^n \mathcal{I}}(\tilde{w}^n | y_1^n, \mathbf{1}) \leq (1 + c(\epsilon)) p_{W^n}(\tilde{w}^n)$ (see [17, Lemma 1]), and $c(\epsilon) \rightarrow 0$ as n grows large. Standard typicality arguments then bound this term (for every $\epsilon > 0$ and for some large enough n) by $(1 + c(\epsilon)) 2^{n(R+T_0+T_1)} 2^{-n(I(WU; Y) - \gamma_1(\epsilon))}$.

To analyze the second probability term, let \tilde{W}^n and \tilde{U}^n denote $W^n(\tilde{m}, \tilde{l}_0)$ and $U^n(1, 1, \hat{l}_1)$ respectively. The second term of (42) is bounded from above by

$$\begin{aligned} & 2^{n(R+T_0+T_1)} \\ & \times \sum_{(w^n, s^n, u^n, v^n, y_1^n) \in \mathcal{A}_\epsilon^n} \sum_{\substack{\tilde{w}^n: \\ (\tilde{w}^n, y_1^n) \in \mathcal{A}_\epsilon^n}} \sum_{\substack{\hat{u}^n: \\ (\hat{u}^n, y_1^n) \in \mathcal{A}^n}} p(w^n, s^n, u^n, v^n, y_1^n, \tilde{w}^n, \hat{u}^n | \mathbf{1}), \end{aligned}$$

and we treat the inner pmf in a similar way as in Section II-C.

$$\begin{aligned} & p_{W^n S^n U^n V^n Y_1^n \tilde{W}^n \hat{U}^n | \mathcal{I}}(w^n, s^n, u^n, v^n, y_1^n, \tilde{w}^n, \hat{u}^n | \mathbf{1}) \\ & = p(w^n, s^n, u^n, v^n, y_1^n | \mathbf{1}) p(\tilde{w}^n | w^n, s^n, u^n, v^n, y_1^n, \mathbf{1}) \\ & \quad \times p(\hat{u}^n | \tilde{w}^n, w^n, s^n, u^n, v^n, y_1^n, \mathbf{1}) \end{aligned}$$

It is now easy to see that (e.g., see [17])

$$\begin{aligned} p(\tilde{w}^n | w^n, s^n, u^n, v^n, y_1^n, \mathbf{1}) & = p(\tilde{w}^n | w^n, s^n, \mathbf{1}) \\ & \leq (1 + c(\epsilon)) p_{\tilde{W}^n}(\tilde{w}^n). \end{aligned}$$

Similarly, it turns out that (see Appendix A and follow a similar line of argument)

$$\begin{aligned} p(\hat{u}^n | \tilde{w}^n, w^n, s^n, u^n, v^n, y_1^n, \mathbf{1}) & = p(\hat{u}^n | w^n, s^n, u^n, v^n, \mathbf{1}) \\ & \leq 2^{n\delta(\epsilon)} p_{\hat{U}^n | W^n}(\hat{u}^n | w^n). \end{aligned}$$

Therefore, the second term of (42) is bounded by $2^{n(R+T_0+T_1)} 2^{-n(I(WU; Y) - \gamma_2(\epsilon) - \delta(\epsilon))}$.

One sees that the non-unique decoding constraints are sufficient to drive both terms of (42) to zero, as n goes large.

APPENDIX D

THE SECOND PROBABILITY TERM OF INEQUALITY (23) CAN BE MADE ARBITRARILY SMALL BY CHOOSING SUFFICIENTLY LARGE n UNDER THE NON-UNIQUE DECODING CONSTRAINTS IN [9]

To upper-bound the second probability term of inequality (23), we use union bound and inclusion of events to obtain

the expression in (44). We then show that each probability term of inequality (44) can be made arbitrarily small by choosing a sufficiently large n , if the non-unique decoding constraints of [9] hold.

$$\begin{aligned}
 & \Pr \left(\begin{array}{c} (Q^n, X_1^n(\bar{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \bar{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\check{m}_1), X_{21}^n(\check{m}_2), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\check{m}_1, \check{m}_2) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_1, \hat{m}_3) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_2, \hat{m}_3) \neq (1, 1), \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\
 & \leq \Pr \left(\begin{array}{c} (Q^n, X_1^n(\bar{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \bar{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_1, \hat{m}_3) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1, \hat{m}_3 = 1, \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\
 & + \Pr \left(\begin{array}{c} (Q^n, X_1^n(\bar{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \bar{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\check{m}_1), X_{21}^n(\check{m}_2), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\check{m}_1, \check{m}_2) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 = 1, \hat{m}_3 \neq 1, \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\
 & + \Pr \left(\begin{array}{c} (Q^n, X_1^n(\bar{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \bar{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1, \hat{m}_3 \neq 1, \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (44)
 \end{aligned}$$

The first probability term of (44) (and similiary the second term) is analyzed below.

$$\Pr \left(\begin{array}{c} (Q^n, X_1^n(\bar{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \bar{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_1, \hat{m}_3) \neq (1, 1), \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1, \hat{m}_3 = 1, \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (45)$$

$$\leq \Pr \left(\begin{array}{c} (Q^n, X_1^n(\hat{m}_1), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_1 \neq 1, \hat{m}_3 = 1, \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1, \hat{m}_3 = 1, \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (46)$$

$$+ \Pr \left(\begin{array}{c} (Q^n, X_1^n(\hat{m}_1), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_1 \neq 1, \hat{m}_3 \neq 1, \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1, \hat{m}_3 = 1, \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (47)$$

$$+ \Pr \left(\begin{array}{c} (Q^n, X_1^n(\bar{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \bar{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_1 = 1, \hat{m}_3 \neq 1, \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1, \hat{m}_3 = 1, \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (48)$$

$$\leq 2^{nR_1} 2^{n \min\{R_2, H(X_{21}|Q)\}} 2^{-nI(X_1 X_{21}; Y_1 | Q X_{31}) + 2n\delta(\epsilon)} \quad (49)$$

$$+ 2^{nR_1} 2^{n \min\{R_3, H(X_{31}|Q)\}} 2^{n \min\{R_2, H(X_{21}|Q)\}} \\ \times 2^{-nI(X_1 X_{21} X_{31}; Y_1 | Q) + 2n\delta(\epsilon)} \quad (50)$$

$$+ 2^{nR_1} 2^{n \min\{R_3, H(X_{31}|Q)\}} 2^{n \min\{R_2, H(X_{21}|Q), H(S_1|X_{31} Q)\}} \\ \times 2^{-nI(X_1 X_{21} X_{31}; Y_1 | Q) + 2n\delta(\epsilon)} \quad (51)$$

where $\delta(\epsilon)$ vanishes to zero as $\epsilon \rightarrow 0$. The first inequality above is obtained by considering the different cases of (\hat{m}_1, \hat{m}_2) , and using inclusion of events. In the second inequality, the probability terms in (46), (47), and (48) are bounded by (49), (50), and (51), respectively. Essentially, the derivation follows from an analysis similar to that of (20) in Section II-C, together with the bounding techniques of [9] (where the key is in that depending on the input pmfs and the message rates, the number of possible combined interference sequences can be equal to the number of interfering message pairs, the number of typical combined interference sequences, or some combination of the two— see [9, Lemmas 2 and 3]). Here, we briefly outline how (46) is bounded by (49), and we leave the derivation of the other two terms to the interested reader.

We start with the following bound.

$$\Pr \left(\begin{array}{c} (Q^n, X_1^n(\hat{m}_1), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_1 \neq 1, \hat{m}_3 = 1, \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1, \hat{m}_3 = 1, \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\
 \leq 2^{nR_1} \Pr \left(\begin{array}{c} (Q^n, \hat{X}_1^n, X_{31}^n, Y_1^n) \in \mathcal{A}_\epsilon^n, \text{ and} \\ (Q^n, X_1^n, X_{21}^n(\hat{m}_2), X_{31}^n, Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (52)$$

Using the bounding technique of [9], the above probability term can be upper bounded in two different manners. By counting the number of different messages \hat{m}_2 , we find

$$\Pr \left(\begin{array}{c} (Q^n, \hat{X}_1^n, X_{31}^n, Y_1^n) \in \mathcal{A}_\epsilon^n, \text{ and} \\ (Q^n, X_1^n, X_{21}^n(\hat{m}_2), X_{31}^n, Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\
 \leq 2^{nR_2} \Pr \left(\begin{array}{c} (Q^n, \hat{X}_1^n, X_{31}^n, Y_1^n) \in \mathcal{A}_\epsilon^n, \text{ and} \\ (Q^n, X_1^n, \hat{X}_{21}^n, X_{31}^n, Y_1^n) \in \mathcal{A}_\epsilon^n \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\
 \leq 2^{nR_2} 2^{-nI(X_1 X_{21}; Y_1 | Q X_{31}) + 2n\delta(\epsilon)}. \quad (53)$$

Furthermore, by counting the number of typical sequences X_{21}^n , we find

$$\Pr \left(\begin{array}{c} (Q^n, \hat{X}_1^n, X_{31}^n, Y_1^n) \in \mathcal{A}_\epsilon^n, \text{ and} \\ (Q^n, X_1^n, X_{21}^n(\hat{m}_2), X_{31}^n, Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right)$$

$$\begin{aligned}
&\leq \sum_{(q^n, x_{31}^n) \in \mathcal{A}_\epsilon^n} \left[p(q^n, x_{31}^n) \right. \\
&\quad \times \Pr \left(\begin{array}{c} (q^n, \hat{X}_1^n, x_{31}^n, Y_1^n) \in \mathcal{A}_\epsilon^n \\ (q^n, X_1^n, \hat{X}_{21}^n(\hat{m}_2), \\ x_{31}^n, Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1 \end{array} \middle| \begin{array}{c} Q^n = q^n, \\ X_3^n = x_3^n, \\ \mathcal{I} = \mathbf{1} \end{array} \right) \Big] \\
&\leq 2^{nH(X_{21}|Q)} 2^{-nI(X_1 X_{21}; Y_1 | Q X_{31}) + 2n\delta(\epsilon)}. \quad (54)
\end{aligned}$$

Putting together (52), (53), and (54) results in the bound (49).

Finally, the third probability term of (44) is bounded as follows.

$$\begin{aligned}
&\Pr \left(\begin{array}{c} (Q^n, X_1^n(\tilde{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \tilde{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), \\ X_{21}^n(\hat{m}_2), X_{31}^n(\hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1, \hat{m}_3 \neq 1, \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (55) \\
&\leq \Pr \left(\begin{array}{c} (Q^n, X_1^n(\tilde{m}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \tilde{m}_1 \neq 1, \text{ and} \\ (Q^n, X_1^n(\hat{m}_1), S_1^n(\hat{m}_2, \hat{m}_3), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } \hat{m}_2 \neq 1, \hat{m}_3 \neq 1, \hat{m}_1 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \\
&\leq 2^{nR_1} 2^{n \min\{R_2+R_3, R_2+H(X_{31}|Q), H(X_{21}|Q)+R_3, H(S_1|Q)\}} \\
&\quad \times 2^{-nI(X_1 S_1; Y_1 | Q) + \delta(\epsilon)}, \quad (57)
\end{aligned}$$

where $\delta(\epsilon) \rightarrow 0$ when $\epsilon \rightarrow 0$.

APPENDIX E

THE SECOND PROBABILITY TERM IN (28) IS BOUNDED BY $2^{n(R_1+T_1+R_r-I(V_0 V_1; Y_1|U)+\gamma_2(\epsilon)+\delta(\epsilon))}$

We now show that for any $\epsilon > 0$, the second probability term in (28) is bounded (for a large enough n)

by $2^{n(R_1+T_1+R_r-I(V_0 V_1; Y_1|U)+\gamma_2(\epsilon)+\delta(\epsilon))}$. Let us denote $U^n(1)$, $V_0^n(1, 1, 1)$, $V_0^n(1, \tilde{m}_1, \tilde{m}_r)$, and $V_1^n(1, 1, 1, \hat{t}_1)$ by U^n , V_0^n , \tilde{V}_0^n , and \hat{V}_1^n respectively. Proceeding as in Section II-C, we bound (28) in (60)–(66) at the bottom of this page. Note that inequality (65) follows for the same reasons as in Appendix IV.

APPENDIX F

THE PROBABILITY THAT THE JOINT UNIQUE DECODER OF m_0, m_1 AND m_r IN SUBSECTION III-C FAILS

We analyze the probability that a joint unique decoder fails to uniquely decode indices m_0, m_1, m_r and show that it fails with high probability if either (30) or (31) is violated. Note that

$$\begin{aligned}
&\Pr \left((U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \middle| \mathcal{I} = \mathbf{1} \right) \\
&\geq \Pr \left((U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \middle| \begin{array}{c} \text{for some } (\hat{m}_0, \hat{m}_1, \hat{m}_r) \neq (1, 1, 1), \\ \hat{m}_0 = 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right), \quad (58)
\end{aligned}$$

and

$$\begin{aligned}
&\Pr \left((U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \middle| \mathcal{I} = \mathbf{1} \right) \\
&\geq \Pr \left((U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \middle| \begin{array}{c} \text{for some } (\hat{m}_0, \hat{m}_1, \hat{m}_r) \neq (1, 1, 1), \\ \hat{m}_0 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right). \quad (59)
\end{aligned}$$

It is now not hard to see that the probability term on the right hand side of inequality (58) is arbitrarily close to 1 if $R_1 + R_r > I(V_0; Y_1|U)$ and the probability term on the right hand side of inequality (59) is arbitrarily close to 1 if $R_1 + R_r > I(U V_0; Y_1)$.

$$\Pr \left(\begin{array}{c} (U^n(1), V_0^n(1, 1, 1), V_1^n(1, 1, 1, 1), V_2^n(1, 1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n \text{ and} \\ (U^n(\tilde{m}_0), V_0^n(\tilde{m}_0, \tilde{m}_1, \tilde{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\tilde{m}_0, \tilde{m}_1, \tilde{m}_r) \neq (1, 1, 1), \tilde{m}_0 = 1 \text{ and} \\ (U^n(\hat{m}_0), V_0^n(\hat{m}_0, \hat{m}_1, \hat{m}_r), V_1^n(\hat{m}_0, \hat{m}_1, \hat{m}_r, \hat{t}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \\ \text{for some } (\hat{m}_0, \hat{m}_1, \hat{m}_r) = (1, 1, 1), \hat{t}_1 \neq 1 \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (60)$$

$$\leq 2^{n(R_1+R_r+T_1)} \Pr \left(\begin{array}{c} (U^n(1), V_0^n(1, 1, 1), V_1^n(1, 1, 1, 1), V_2^n(1, 1, 1, 1), Y_1^n) \in \mathcal{A}_\epsilon^n \text{ and} \\ (U^n(1), V_0^n(1, \tilde{m}_1, \tilde{m}_r), Y_1^n) \in \mathcal{A}_\epsilon^n \text{ and} \\ (U^n(1), V_0^n(1, 1, 1), V_1^n(1, 1, 1, \hat{t}_1), Y_1^n) \in \mathcal{A}_\epsilon^n \end{array} \middle| \mathcal{I} = \mathbf{1} \right) \quad (61)$$

$$\leq 2^{n(R_1+R_r+T_1)} \sum_{(u^n, v_0^n, v_1^n, v_2^n, y_1^n) \in \mathcal{A}_\epsilon^n} \sum_{\substack{\tilde{v}_0^n: \\ (u^n, \tilde{v}_0^n, y_1^n) \in \mathcal{A}_\epsilon^n}} \sum_{\substack{\hat{v}_1^n: \\ (u^n, v_0^n, \hat{v}_1^n, y_1^n) \in \mathcal{A}_\epsilon^n}} P_{U^n V_0^n V_1^n V_2^n Y_1^n \tilde{V}_0^n \hat{V}_1^n | \mathcal{I}}(u^n, v_0^n, v_1^n, v_2^n, y_1^n, \tilde{v}_0^n, \hat{v}_1^n | \mathbf{1}) \quad (62)$$

$$\leq 2^{n(R_1+R_r+T_1)} \sum_{(u^n, v_0^n, v_1^n, v_2^n, y_1^n) \in \mathcal{A}_\epsilon^n} \sum_{\substack{\tilde{v}_0^n: \\ (u^n, \tilde{v}_0^n, y_1^n) \in \mathcal{A}_\epsilon^n}} \sum_{\substack{\hat{v}_1^n: \\ (u^n, v_0^n, \hat{v}_1^n, y_1^n) \in \mathcal{A}_\epsilon^n}} p(u^n, v_0^n, v_1^n, v_2^n, y_1^n | \mathbf{1}) p(\tilde{v}_0^n | u^n, v_0^n, v_1^n, v_2^n, y_1^n, \mathbf{1}) \\ \times p(\hat{v}_1^n | u^n, v_0^n, v_1^n, v_2^n, y_1^n, \tilde{v}_0^n, \mathbf{1}) \quad (63)$$

$$= 2^{n(R_1+R_r+T_1)} \sum_{(u^n, v_0^n, v_1^n, v_2^n, y_1^n) \in \mathcal{A}_\epsilon^n} \sum_{\substack{\tilde{v}_0^n: \\ (u^n, \tilde{v}_0^n, y_1^n) \in \mathcal{A}_\epsilon^n}} \sum_{\substack{\hat{v}_1^n: \\ (u^n, v_0^n, \hat{v}_1^n, y_1^n) \in \mathcal{A}_\epsilon^n}} p(u^n, v_0^n, v_1^n, v_2^n, y_1^n | \mathbf{1}) p(\tilde{v}_0^n | u^n) p(\hat{v}_1^n | u^n, v_0^n, v_1^n, v_2^n, \mathbf{1}) \quad (64)$$

$$\leq 2^{n(R_1+R_r+T_1)} \sum_{(u^n, v_0^n, v_1^n, v_2^n, y_1^n) \in \mathcal{A}_\epsilon^n} \sum_{\substack{\tilde{v}_0^n: \\ (u^n, \tilde{v}_0^n, y_1^n) \in \mathcal{A}_\epsilon^n}} \sum_{\substack{\hat{v}_1^n: \\ (u^n, v_0^n, \hat{v}_1^n, y_1^n) \in \mathcal{A}_\epsilon^n}} 2^{n\delta(\epsilon)} p(u^n, v_0^n, v_1^n, v_2^n, y_1^n | \mathbf{1}) p(\tilde{v}_0^n | u^n) p(\hat{v}_1^n | u^n, v_0^n) \quad (65)$$

$$\leq 2^{n(R_1+R_r+T_1)} 2^{-n(I(V_0 V_1; Y|U) - \gamma_2(\epsilon) - \delta(\epsilon))} \quad (66)$$

APPENDIX G

CHERNOFF BOUNDS AND INEQUALITIES (37) AND (38)

Let $N = 2^{n(T_3 - S_3)}$, $M = 2^{n(T_2 - S_2)}$. To simplify notation, we define $X_{i,j}$ to be a binary random variable which takes value 0 when $(V_2^n(1, 1, i), V_3^n(1, 1, j)) \in \mathcal{A}_\epsilon^n$. For example, $X_{1,1} = 1$ by the assumption that $(v_2^n, v_3^n) \in \mathcal{A}_\epsilon^n$. Also, $N_2(\hat{V}_2^n, v_2^n, v_3^n, \mathbf{C}') = \sum_{t_3=1}^N X_{2,t_3}$ and $N_3(v_3^n, \mathbf{C}') = \sum_{t_3=1}^N X_{3,t_3}$. Furthermore, we define $p_{\hat{v}_2^n} = \Pr((\hat{v}_2^n, V_3^n(1, 1, 2)) \in \mathcal{A}_\epsilon^n | V_2^n(1, 1, 2) = \hat{v}_2^n, U^n = u^n)$. For ϵ_1 -typical sequences \hat{v}_2^n (where $\epsilon_1 < \epsilon$), we have $2^{-n(I(V_2; V_3|U) + \delta(\epsilon))} \leq p_{\hat{v}_2^n} \leq 2^{-n(I(V_2; V_3|U) - \delta(\epsilon))}$. We let $p_l = 2^{-n(I(V_2; V_3|U) + \delta(\epsilon))}$ and $p_u = 2^{-n(I(V_2; V_3|U) - \delta(\epsilon))}$. To prove Claim 1, we show that

$$\Pr(N_2(\hat{V}_2^n, v_2^n, v_3^n, \mathbf{C}') > 2Np_u | U^n = u^n) \leq \beta_1 \exp(-\alpha_1 Np_l) \quad (67)$$

for some $\alpha_1, \beta_1 > 0$, and

$$\Pr(N_3(v_3^n, \mathbf{C}') < \frac{1}{2}Np_l | U^n = u^n) \leq \beta_2 \exp(-\alpha_2 Np_l) \quad (68)$$

for some $\alpha_2, \beta_2 > 0$.

We start with (67).

$$\begin{aligned} & \Pr(N_2(\hat{V}_2^n, v_2^n, v_3^n, \mathbf{C}') > 2Np_u | U^n = u^n) \\ &= \Pr\left(\sum_{t_3=1}^N X_{2,t_3} > 2Np_u \mid U^n = u^n\right) \\ &\leq \Pr\left(\sum_{t_3=2}^N X_{2,t_3} > 2Np_u - 1 \mid U^n = u^n\right) \\ &\leq \frac{\mathbb{E}\left[e^{t \sum_{t_3=2}^N X_{2,t_3}} \mid U^n = u^n\right]}{e^{t2Np_u - t}}, \quad t > 0 \\ &= \frac{\mathbb{E}\left[\mathbb{E}\left[e^{t \sum_{t_3=2}^N X_{2,t_3}} \mid \hat{V}_2^n, U^n = u^n\right] \mid U^n = u^n\right]}{e^{t2Np_u - t}} \\ &= \frac{\mathbb{E}\left[\prod_{t_3=2}^N \mathbb{E}\left[e^{tX_{2,t_3}} \mid \hat{V}_2^n, U^n = u^n\right] \mid U^n = u^n\right]}{e^{t2Np_u - t}} \\ &= \frac{\mathbb{E}\left[\prod_{t_3=2}^N (p_{\hat{V}_2^n} e^t + (1 - p_{\hat{V}_2^n})) \mid U^n = u^n\right]}{e^{t2Np_u - t}} \\ &\leq \frac{(1 + p_u(e^t - 1))^N}{e^{t2Np_u - t}} \end{aligned}$$

Set $t = \frac{1}{2}$. Then

$$\begin{aligned} & \Pr(N_2(\hat{V}_2^n, v_2^n, v_3^n, \mathbf{C}') > 2Np_u | U^n = u^n) \\ &\leq \frac{(1 + p_u(e^{\frac{1}{2}} - 1))^N}{e^{-\frac{1}{2} + 2Np_u}} \\ &= e^{\frac{1}{2}} \left(\frac{1 + p_u(e^{\frac{1}{2}} - 1)}{e^{p_u}} \right)^N \\ &\leq e^{\frac{1}{2}} e^{-Np_u(2 - e^{\frac{1}{2}})} \\ &\leq \beta_1 e^{-\alpha_1 Np_l}, \quad \text{for } \alpha_1 = 2 - e^{\frac{1}{2}}, \beta_1 = e^{\frac{1}{2}}. \end{aligned}$$

Similarly, to show (68) we proceed as follows.

$$\begin{aligned} & \Pr\left(N_3(v_3^n, \mathbf{C}') < \frac{1}{2}Np_l \mid U^n = u^n\right) \\ &= \Pr\left(\sum_{t_3=1}^N X_{3,t_3} < \frac{1}{2}Np_l \mid U^n = u^n\right) \\ &\leq \Pr\left(\sum_{t_3=2}^N X_{3,t_3} < \frac{1}{2}Np_l \mid U^n = u^n\right) \\ &\leq \frac{\mathbb{E}\left[e^{-t \sum_{t_3=2}^N X_{3,t_3}} \mid U^n = u^n\right]}{e^{-t \frac{1}{2} Np_l}}, \quad t > 0 \\ &= \frac{\mathbb{E}\left[\mathbb{E}\left[e^{-t \sum_{t_3=2}^N X_{3,t_3}} \mid V_2^n(1, 1, 3), U^n = u^n\right] \mid U^n = u^n\right]}{e^{-t \frac{1}{2} Np_l}} \\ &= \frac{\mathbb{E}\left[\prod_{t_3=2}^N \mathbb{E}\left[e^{-tX_{3,t_3}} \mid V_2^n(1, 1, 3), U^n = u^n\right] \mid U^n = u^n\right]}{e^{-t \frac{1}{2} Np_l}} \\ &= \frac{\mathbb{E}\left[\prod_{t_3=2}^N (1 - p_{V_2^n(1,1,3)}(1 - e^{-t})) \mid U^n = u^n\right]}{e^{-t \frac{1}{2} Np_l}} \\ &\leq \frac{(1 - p_l(1 - e^{-t}))^N}{e^{-t \frac{1}{2} Np_l}} \end{aligned}$$

Set $t = 1$. Then

$$\begin{aligned} & \Pr\left(N_3(v_3^n, \mathbf{C}') < \frac{1}{2}Np_l \mid U^n = u^n\right) \\ &\leq \left(\frac{1 - p_l(1 - e^{-1})}{e^{-\frac{1}{2} p_l}}\right)^N \\ &\leq e^{-p_l(\frac{1}{2} - e^{-1})N} \\ &\leq \beta_2 e^{-\alpha_2 Np_l}, \quad \text{for } \alpha_2 = \frac{1}{2} - e^{-1}, \beta_2 = 1. \end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank Suhas Diggavi for helpful comments and discussions. The authors would also like to thank Aaron Wagner and the anonymous reviewers for their comments that helped improve the manuscript.

REFERENCES

- [1] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [2] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.
- [3] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.
- [4] C. Nair and A. El Gamal, "The capacity region of a class of 3-receiver broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4479–4493, Oct. 2009.
- [5] H. F. Chong, M. Motani, H. K. Garg, and H. El Gamal, "On the Han-Kobayashi region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 3188–3195, Jul. 2008.
- [6] S. Lim, Y.-H. Kim, A. El-Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

- [7] A. Zaidi, P. Piantanida, and S. Shamai. (2012, Jan.). *Capacity Region of Multiple Access Channel With States Known Noncausally at One Encoder and Only Strictly Causally at the Other Encoder* [Online]. Available: <http://adsabs.harvard.edu/abs/2012arXiv1201.3278Z>
- [8] C. Nair, A. El Gamal, and Y. K. Chia. (2010, Apr.). *An Achievability Scheme for the Compound Channel with State Noncausally Available at the Encoder* [Online]. Available: <http://arxiv.org/abs/1004.3427>
- [9] B. Bandemer and A. El Gamal, "Interference decoding for deterministic channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2966–2975, May 2011.
- [10] Y. K. Chia and A. El Gamal. (2011, Jun.). *3-Receiver Broadcast Channels with Common and Confidential Messages* [Online]. Available: <http://arxiv.org/abs/0910.1407>
- [11] X. Wu and L.-L. Xie. (2010, Sep.). *On the Optimal Compressions in the Compress-and-Forward Relay Schemes* [Online]. Available: <http://adsabs.harvard.edu/abs/2010arXiv1009.5959W>
- [12] X. Wu and L.-L. Xie, "On the optimality of successive decoding in compress-and-forward relay schemes," in *Proc. Allerton Conf. Commun., Control Comput.*, Oct. 2010, pp. 534–541.
- [13] G. Kramer and J. Hou, "On message lengths for noisy network coding," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2011, pp. 430–431.
- [14] J. Hou and G. Kramer, "Short message noisy network coding for multiple sources," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 1677–1681.
- [15] S. Saeedi Bidokhti, V. Prabhakaran, and S. Diggavi, "A block Markov encoding scheme for broadcasting nested message sets," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2975–2979.
- [16] B. Bandemer, A. El Gamal, and Y.-H. Kim, "Simultaneous nonunique decoding is rate-optimal," in *Proc. Allerton Conf. Commun., Control Comput.*, Oct. 2012, pp. 9–16.
- [17] P. Minero, S. H. Lim, and Y.-H. Kim. (2013, Jun.). *Hybrid Coding: An Interface for Joint Source-Channel Coding and Network Communication* [Online]. Available: <http://arxiv.org/abs/1306.0530>
- [18] P. Grover, A. B. Wagner, and A. Sahai. (2013, Jun.). *Information Embedding and the Triple Role of Control* [Online]. Available: <http://arxiv.org/abs/1306.5018>

Shirin Saeedi Bidokhti (S'09–M'12) received the B.Sc. degree in electrical engineering from the University of Tehran, Iran, in 2005, and the M.Sc and Ph.D. degrees in Communication Systems from the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2007 and 2012, respectively. In 2013, she was awarded a Swiss National Science Foundation Prospective Researcher Fellowship.

Since 2012, she has been a postdoctoral researcher at the Institute for Communication Engineering, Technische Universität München (TUM). Her research interests include information theory and coding, multi-user communications systems, and network coding.

Vinod M. Prabhakaran (S'97–M'98) received his Ph.D. in 2007 from the EECS Department, University of California, Berkeley. He was a Postdoctoral Researcher at the Coordinated Science Laboratory, University of Illinois, Urbana-Champaign from 2008 to 2010 and at Ecole Polytechnique Fédérale de Lausanne, Switzerland in 2011. In Fall 2011, he joined the Tata Institute of Fundamental Research, Mumbai, where he currently holds the position of a Reader. His research interests are in information theory, wireless communication, cryptography, and signal processing.

He has received the Tong Leong Lim Pre-Doctoral Prize and the Demetri Angelakos Memorial Achievement Award from the EECS Department, University of California, Berkeley, and the Ramanujan Fellowship from the Department of Science and Technology, Government of India.