TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Nachrichtentechnik

Coding for Relay Networks and Effective Secrecy for Wire-Tap Channels

Jie Hou

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor–Ingenieurs

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr.-Ing. Dr. rer. nat. Holger Boche
Prüfer der Dissertation:	1. Univ.-Prof. Dr. sc. techn. (ETH Zürich) Gerhard Kramer 2. Prof. Young-Han Kim, Ph.D. University of California, San Diego, USA

Die Dissertation wurde am 08.05.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 02.08.2014 angenommen.

Acknowledgements

I thank Professor Gerhard Kramer for his generous support and guidance during my time at LNT, TUM. Whether for a cup of coffee or for a technical meeting, Gerhard's door is always open, for he has genuine interest in his students and research. I also thank Gerhard for his effort and care in shaping my thinking and approach to research: solve problems in the simplest way and present results in the most concise way. I am surely going to benefit more from this motto in the future. I further thank Gerhard for the many opportunities he created for me, especially the trip to USC. Thank you Gerhard!

At the beginning of my Ph.D. study, I had the honor to work with the late Professor Ralf Kötter. Thank you Ralf, for giving me the opportunity to pursue a Dr.-Ing. degree and for providing valuable advises.

I thank Professor Young-Han Kim for acting as the co-referee of my thesis and for his suggestions. Special thanks go to Professor Hagenauer and Professor Utschick for their support during the difficult times at LNT. I would like to thank the colleagues at LNT for having created a pleasant atmosphere over the course of my Ph.D.. Several people contributed to it in a special way. Hassan Ghozlan, a true "Trojan", shared the moments of excitement and disappointment in research. I very much enjoyed the conversations with Hassan, both technical and non-technical, on some of the "long" days. My office-mate Tobias Lutz became a good friend and we had good times on and off work.

Finally, I am most indebted to my family who shared my ups and downs and who always stood by my side: my late father Gongwei and my mother Jianning. Without their love and support, I would not be where I am. The last word of thanks goes to Miao and Baobao for their love and for making my life a pleasant one.

München, May 2014 Jie Hou

Contents

Zusammenfassung

Diese Arbeit untersucht zwei Probleme in Netzwerk-Informationstheorie: Codierung für Relaisnetzwerke und (präzise) Approximationen der Wahrscheinlichkeitsverteilungen basierend auf nicht-normalisierter Kullback-Leibler Divergenz und deren Anwendungen zur sicheren Kommunikation in Netzen.

Im Rahmen der ersten Problemstellung wird zuerst Netzcodierung in rauschbehafteten Netzen mit kurzen Nachrichten (SNNC) untersucht. SNNC ermöglicht Rückwärtsdecodierung, die einfach zu analysieren ist und die gleichen Raten wie SNNC mit Sliding-Window Decodierung und Netzcodierung mit längen Nachrichten (LNNC) mit gemeinsamer Decodierung liefert. SNNC ermöglicht auch den Relais die frühzeitige Decodierung, wenn die Kanalqualität gut ist. Dies führt zu gemischten Strategien, die die Vorteile von SNNC und decode-forward (DF) vereinigen. Wir präsentieren einen iterativen Algorithmus, der diejenigen Nutzer findet, deren Nachrichten als Rauschen behandeltet werden sollten, um die besten Raten zu gewährleisten. Anschließend wird der Vielfachzugriff-Relaiskanal (MARC) mit Relais-Quelle Rückkopplung untersucht. Wir leiten mit einer neuen DF Codierung die Ratenregionen her, die die Kapazitätsregion des MARC ohne Rückkopplung einschließen. Dies zeigt, dass Rückkopplungen die Kapazitätsregionen in Mehrnutzernetzwerken vergrößen können.

Im Rahmen der zweiten Problemstellung zeigen wir zuerst, dass die minimale Rate, um eine Verteilung präzis zu approximieren, eine Transinformation ist. Die Genauigkeit ist mit Hilfe der nicht-normalisierten Kullback-Leibler Divergenz gemessen. Anschließend wenden wir das Ergebnis auf Kommunikationssicherheit in Netzen an und definieren ein neues effektives Sicherheitsmaß, das starke Sicherheit und Heimlichkeit beinhaltet. Dieses effektive Maß stellt sicher, dass der Lauscher nichts von der Nachricht mitbekommt und auch nicht in der Lage ist, die Präsenz der bedeutsamen Kommunikation zu detektieren.

Abstract

This thesis addresses two problems of network information theory: coding for relay networks and (accurate) approximations of distributions based on unnormalized informational divergence with applications to network security.

For the former problem, we first consider short message noisy network coding (SNNC). SNNC differs from long message noisy network coding (LNNC) in that one transmits many short messages in blocks rather than using one long message with repetitive encoding. Several properties of SNNC are developed. First, SNNC with backward decoding achieves the same rates as SNNC with offset encoding and sliding window decoding for memoryless networks where each node transmits a multicast message. The rates are the same as LNNC with joint decoding. Second, SNNC enables early decoding if the channel quality happens to be good. This leads to mixed strategies that unify the advantages of decode-forward and noisy network coding. Third, the best decoders sometimes treat other nodes' signals as noise and an iterative method is given to find the set of nodes that a given node should treat as noise sources. We next consider the multiple access relay channel (MARC) with relay-source feedback. We propose a new decode-forward (DF) coding scheme that enables the cooperation between the sources and the relay to achieve rate regions that include the capacity region of the MARC without feedback.

For the latter problem, we show that the minimum rate needed to accurately approximate a product distribution based on an unnormalized informational divergence is a mutual information. This result subsumes results of Wyner on common information and Han-Verdú on resolvability. The result also extends to cases where the source distribution is unknown but the entropy is known. We then apply this result to network security where an effective security measure is defined that includes both strong secrecy and stealth communication. Effective secrecy ensures that a message cannot be deciphered and that the presence of meaningful communication is hidden. To measure stealth we use resolvability and relate this to binary hypothesis testing. Results are developed for wire-tap channels and broadcast channels with confidential messages.

1

Introduction

Network information theory extends Shannon's seminal work [1] and seeks answers for two questions: what are the ultimate limits for

- ▷ reliable and secure data transmission
- ▷ data compression with fidelity criteria

in multi-user networks? For some special networks, the solutions are known, e.g., the capacity region of the two-user multiple access channel (MAC) [2,3] and the Slepian-Wolf problem [4] (compressing two correlated sources). However, the solutions for general multi-user networks remain elusive. For instance, the capacity for the relay channel is open for over 40 years.

In order to deepen our understanding in theory and gain insight for practical implementations, we address two topics in network information theory in this thesis, namely:

- ▷ Short message noisy network coding (SNNC)
- ▷ Resolvability with applications to network security.

SNNC is a coding scheme for relay networks where every node forwards a quantized version of its channel output along with its messages. SNNC combined with appropriate decoding methods [5–9], for instance backward decoding, achieves the same rates as its long message counterpart (LNNC) [10, 11] with joint decoding and provides a simpler analysis since *per-block* processing is possible. The rate bounds have a nice cut-set interpretation and generalize the results in [12–15] in a natural way. Also, SNNC allows the relays to switch between decode-forward (DF) and quantize-forward (QF) depending on the quality of the channels, thereby achieving a boost in performance.

On the other hand, resolvability addresses the minimal rate needed to mimic a target distribution with some distance measure. Wyner considered this problem based on *normalized* informational divergence [16] and Han-Verdú studied it based on *variational distance* [17]. The same minimal rate, which is a Shannon mutual information, was shown to be necessary and sufficient to produce good approximations for both measures in [16, 17]. We show that the same rate is also necessary and sufficient to generate good approximations based on *unnormalized* informational divergence. We then apply this result to establish a new and stronger security measure termed *effective secrecy* that includes both *strong secrecy* and *stealth*. The effective secrecy measure includes hiding the messages and hiding the presence of meaningful communication.

The thesis is organized as follows.

Chapter 2 introduces notation and useful definitions as well as inequalities that will be used throughout this thesis.

Chapter 3 discusses SNNC for networks with multiple multi-cast sessions. We show that SNNC with backward decoding achieves the same rates as SNNC with sliding window decoding and LNNC with joint decoding. Backward decoding also provides a simpler analysis since *per-block* processing is enabled. More importantly, we show that SNNC enables the relays to choose the best strategy, DF or QF, depending on the channel conditions which leads to mixed strategies that unify the advantages of both DF and QF. Numerical results show that mixed strategies provide substantial gains compared to SNNC (LNNC) and other strategies in rates and outage probabilities for networks without and with fading, respectively.

Chapter 4 deals with the multiple access relay channel (MARC) with relay-source feedback. We introduce a new DF coding scheme with feedback and establish an achiev-

able rate region that includes the capacity region of the MARC. We compare this region with the achievable SNNC rates developed in **Chapter 3**. The results show how cooperation improves rates and how network geometry affects the choice of coding strategy.

In **Chapter 5**, we consider the resolvability problem based on *unnormalized* informational divergence. Our result subsumes that in [17, 18] when restricting attention to product distributions and the proof is simpler.

In **Chapter 6**, we apply the resolvability result in **Chapter 5** to network security and establish a new and stronger security measure, *effective secrecy*, that includes strong secrecy and stealth. We also relate stealth to binary hypothesis testing. Results are developed for wire-tap channels and broadcast channels with confidential messages.

Finally, **Chapter 7** summarizes the results and discusses open research problems that are related to the work in this thesis.

2

Preliminaries

2.1. Random Variables

Random variables are written with upper case letters such as X and their realizations with the corresponding lower case letters such as x. Bold letters refer to random vectors and their realizations ($\mathbf{X}$ and $\mathbf{x}$). Subscripts on a variable/symbol denote the variable/symbol's source and the position in a sequence. For instance, X_{ki} denotes the i-th output of the k-th encoder. Superscripts denote finite-length sequences of variables/symbols, e.g., $X_k^n = (X_{k1}, \ldots, X_{kn})$. Calligraphic letters denote sets, e.g., we write $\mathcal{K} = \{1, 2, \ldots, K\}$. The size of a set $\mathcal{S}$ is denoted as $|\mathcal{S}|$ and the complement set of $\mathcal{S}$ is denoted as $\mathcal{S}^{\mathrm{c}}$. Set subscripts denote vectors of letters, e.g., $X_{\mathcal{S}} = [X_k : k \in \mathcal{S}]$.

A random variable X has distribution P_X and the support of P_X is denoted as $\mathrm{supp}(P_X)$. We write probabilities with subscripts $P_X(x)$ but we often drop the subscripts if the arguments of the distributions are lower case versions of the random variables. For example, we write $P(x) = P_X(x)$. If the X_i, $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) according to P_X, then we have $P(x^n) = \prod_{i=1}^n P_X(x_i)$

and we write $P_{X^n} = P_X^n$. We often also use Q_X^n to refer to strings (or sequences) of i.i.d. random variables. For X with finite alphabet $\mathcal{X}$ we write $P_X(\mathcal{S}) = \sum_{x\in\mathcal{S}} P_X(x)$ for any $\mathcal{S} \subseteq \mathcal{X}$.

We use $\mathcal{T}_\epsilon^n(P_X)$ to denote the set of letter-typical sequences of length n with respect to the probability distribution P_X and the non-negative number ϵ [19, Ch. 3], [20], i.e., we have

$$\mathcal{T}_\epsilon^n(P_X) = \left\{ x^n : \left| \frac{N(a|x^n)}{n} - P_X(a) \right| \leq \epsilon P_X(a),\ \forall a \in \mathcal{X} \right\} \tag{2.1}$$

where $N(a|x^n)$ is the number of occurrences of a in x^n.

2.2. Information Measures and Inequalities

The *entropy* of a discrete random variable X is defined as

$$H(X) = \sum_{x\in\mathrm{supp}(X)} \left(-P(x)\log P(x)\right). \tag{2.2}$$

Let X and Y be two discrete variables with joint distribution P_{XY}. We write the *mutual information* between X and Y as

$$I(X;Y) = \sum_{(x,y)\in\mathrm{supp}(P_{XY})} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}. \tag{2.3}$$

The *informational divergence* and *variational distance* between two distributions P_X and Q_X are respectively denoted as

$$D(P_X||Q_X) = \sum_{x\in\mathrm{supp}(P_X)} P(x)\log\frac{P(x)}{Q(x)} \tag{2.4}$$

$$||P_X - Q_X||_{\mathrm{TV}} = \sum_{x\in\mathcal{X}} |P(x) - Q(x)|. \tag{2.5}$$

Pinsker's inequality [21, Theorem 11.6.1] states that

$$D(P_X||Q_X) \geq \frac{1}{2\ln 2}||P_X - Q_X||_{\mathrm{TV}}^2. \tag{2.6}$$

3

Short Message Noisy Network Coding

Noisy Network Coding (NNC) extends network coding from noiseless to noisy networks. NNC is based on the compress-forward (CF) strategy of [21] and there are now two *encoding* variants: short message NNC (SNNC) [5–9, 22–27] and long message NNC (LNNC) [10, 11, 15]. Both variants achieve the same rates that include the results of [12–14] as special cases.

For SNNC, there are many *decoding* variants: step-by-step decoding [21–24], sliding window decoding [5, 6], backward decoding [7–9, 25, 26] and joint decoding [26]. There are also several *initialization* methods. The papers [5, 6, 24] use delayed (or offset) encoding, [7] uses many extra blocks to decode the last quantization messages and [9] uses extra blocks to transmit the last quantization messages by multihopping. We remark that the name of the relaying operation should not depend on which *decoder* (step-by-step, sliding window, joint, or backward decoding) is used at the *destination* but is a generic name for the *processing* at the *relays*, or in the case of SNNC and LNNC, the

overall encoding strategy of the network nodes.

More explicitly, SNNC has

- ▷ **Sources** transmit independent short messages in blocks.
- ▷ **Relays** perform CF but perhaps without hashing (or binning) which is called quantize-forward (QF).
- ▷ **Destinations** use one of the several decoders. For instance, SNNC with CF and step-by-step decoding was studied for relay networks in [22, Sec. 3.3.3], [23, Sec. V], and [24]. The papers [5, 6] studied SNNC with *sliding window* decoding. The papers [7–9, 25, 26] considered SNNC with *backward* decoding. SNNC with joint decoding was studied in [26].

We prefer backward decoding because it permits *per-block* processing and gives the most direct way of establishing rate bounds. However, we remark that the sliding window decoder of [5, 6] is preferable because of its lower decoding delay, and because it enables streaming.

LNNC uses three techniques from [15]:

- ▷ **Sources** use repetitive encoding with *long* messages.
- ▷ **Relays** use QF.
- ▷ **Destinations** decode all messages and all quantization bits jointly.

One important drawback of long messages is that they inhibit decode-forward (DF) even if the channel conditions are good [8]. For example, if one relay is close to the source and has a strong source-relay link, then the natural operation is DF which removes the noise at the relay. But this is generally not possible with a long message because of its high rate.

Our main goals are to simplify and extend the single source results of [7, 8, 25] by developing SNNC with backward decoding for networks with *multiple multicast* sessions [9]. We also introduce multihopping to initialize backward decoding. This method reduces overhead as compared to the joint decoder initialization used in [7]. The method

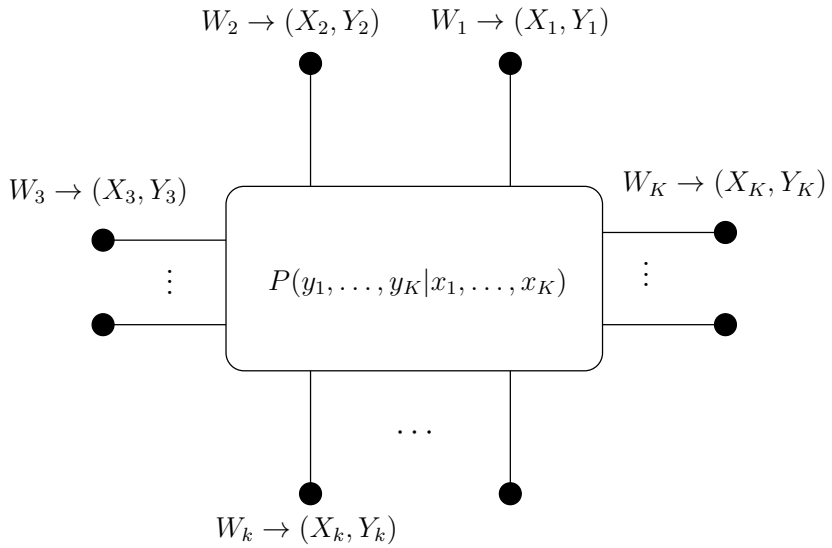


Figure 3.1.: A K-node memoryless network. The network is a DMN if the alphabets of X_k and Y_k are discrete and finite for $k = 1, \ldots, K$.

further enables *per-block* processing for all signals, i.e., all messages and quantization indices.

This chapter is organized as follows. In Section 3.1, we state the problem. In Section 3.2, we show that SNNC achieves the same rates as SNNC with sliding window decoding and LNNC for memoryless networks with multiple multicast sessions. In Section 3.3, we discuss the results and relate them to other work. In Section 3.4, we present coding schemes for mixed strategies that allow relay nodes to switch between DF and QF depending on the channel conditions. Results on Gaussian networks are discussed in Section 3.5. Finally, Section 3.6 concludes this chapter.

3.1. System Model

3.1.1. Memoryless Networks

Consider the K-node memoryless network depicted in Fig. 3.1 where each node has one message only. This model does not include broadcasting messages and was used in [11] and [28, Ch. 15] . Node k, $k \in \mathcal{K}$, has a message W_k destined for nodes in the set $\mathcal{D}_k$, $\mathcal{D}_k \subseteq \mathcal{K} \setminus \{k\}$, while acting as a relay for messages of the other nodes. We write the set

of nodes whose signals node k must decode correctly as $\widetilde{\mathcal{D}}_k = \{i \in \mathcal{K} : \ k \in \mathcal{D}_i\}$. The messages are mutually statistically independent and W_k is uniformly distributed over the set $\{1, \ldots, 2^{nR_k}\}$, where 2^{nR_k} is taken to be a non-negative integer.

The channel is described by the conditional probabilities

$$P(y^K|x^K) = P(y_1, \ldots, y_K|x_1, \ldots, x_K) \tag{3.1}$$

where $\mathcal{X}_k$ and $\mathcal{Y}_k$, $k \in \mathcal{K}$, are the respective input and output alphabets, i.e., we have

$$(x_1, \ldots, x_K) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_K$$
$$(y_1, \ldots, y_K) \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_K.$$

If all alphabets are discrete and finite sets, then the network is called a *discrete* memoryless network (DMN) [29], [30, Ch.18]. Node k transmits $x_{ki} \in \mathcal{X}_k$ at time i and receives $y_{ki} \in \mathcal{Y}_k$. The channel is *memoryless* and *time invariant* in the sense that

$$\begin{aligned} &P(y_{1i}, \ldots, y_{Ki}|w_1, \ldots, w_K, x_1^i, \ldots, x_K^i, y_1^{i-1}, \ldots, y_K^{i-1}) \\ &\quad = P_{Y^K|X^K}(y_{1i}, \ldots, y_{Ki}|x_{1i}, \ldots, x_{Ki}) \end{aligned} \tag{3.2}$$

for all i. As usual, we develop our random coding for DMNs and later extend the results to Gaussian channels.

3.1.2. Flooding

We can represent the DMN as a directed graph $\mathcal{G} = \{\mathcal{K}, \mathcal{E}\}$, where $\mathcal{E} \subset \mathcal{K} \times \mathcal{K}$ is a set of edges. Edges are denoted as $(i, j) \in \mathcal{E}, \ i, j \in \mathcal{K}, \ i \neq j$. We label edge (i, j) with the non-negative real number

$$C_{ij} = \max_{x_{\mathcal{K}\setminus i}} \max_{P_{X_i}} I(X_i; Y_j|X_{\mathcal{K}\setminus i} = x_{\mathcal{K}\setminus i}) \tag{3.3}$$

called the capacity of the link, where $I(A; B|C = c)$ is the mutual information between the random variables A and B conditioned on the event $C = c$. Let $\text{Path}_{(i,j)}$ be a path that starts from node i and ends at node j. Let $\Gamma_{(i,j)}$ to be the set of such paths. We

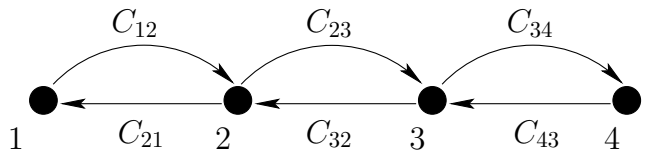


Figure 3.2.: A line network with 4 nodes. Each node can communicate reliably with any other node as long as $C_{ij} > 0$ for all i, j.

write $(k, \ell) \in \text{Path}_{(i,j)}$ if (k, ℓ) lies on the path $\text{Path}_{(i,j)}$. We may communicate reliably between nodes i and j if

$$R_{ij} = \max_{\text{Path}_{(i,j)} \in \Gamma_{(i,j)}} \min_{(k,l) \in \text{Path}_{(i,j)}} C_{kl} \tag{3.4}$$

is positive. We assume that $R_{ij} > 0$ for all nodes i with a message destined for node j. Observe that if $C_{ij} > 0$ for all i, j, then at most $K - 1$ hops are needed for node i to reliably convey its message at rate

$$\min_{j \in \mathcal{K}} R_{ij} \tag{3.5}$$

by multihopping to all other nodes in the network. Hence, for a K-node memoryless network at most $K(K-1)$ hops are needed for all nodes to "flood" their messages by multihopping through the network.

Example 3.1. A line network with 4 nodes is depicted in Fig. 3.2. Node 1 has a message for node 4 and we assume that $C_{12} > 0$, $C_{23} > 0$ and $C_{34} > 0$ so that node 1 can communicate reliably to node 4 by multihopping through nodes 2 and 3 with 3 hops.

3.1.3. Encoders and Decoders

We define two types of functions for each node k:

- ▷ n encoding functions $f_k^n = (f_{k1}, \ldots, f_{kn})$ that generate channel inputs based on the local message and past channel outputs

$$X_{ki} = f_{ki}(W_k, Y_k^{i-1}),\ i = \{1, \ldots, n\}. \tag{3.6}$$

▷ One decoding function

$$g_k(Y_k^n, W_k) = [\hat{W}_i^{(k)},\ i \in \widetilde{\mathcal{D}}_k] \tag{3.7}$$

where $\hat{W}_i^{(k)}$ is the estimate of W_i at node k.

The average error probability for the network is defined as

$$P_e^{(n)} = \Pr\left[\bigcup_{k\in\mathcal{K}} \bigcup_{i\in\widetilde{\mathcal{D}}_k} \{\hat{W}_i^{(k)} \neq W_i\}\right]. \tag{3.8}$$

A rate tuple $(R_1, \ldots, R_K)$ is achievable for the DMN if for any $\xi > 0$, there is a sufficiently large integer n and some functions $\{f_k^n\}_{k=1}^K$ and $\{g_k\}_{k=1}^K$ such that $P_e^{(n)} \leq \xi$. The capacity region is the closure of the set of achievable rate tuples. For each node k we define

$$\mathcal{K}_k = \{k\} \cup \widetilde{\mathcal{D}}_k \cup \mathcal{T}_k,\ \mathcal{T}_k \subseteq \widetilde{\mathcal{D}}_k^{\mathrm{c}} \setminus \{k\} \tag{3.9}$$

where $\mathcal{T}_k$ has the nodes whose messages node k is not interested in but whose symbol sequences are included in the typicality test in order to remove interference. We further define, for any $\mathcal{S} \subset \mathcal{L} \subseteq \mathcal{K}$, the quantities

$$I_{\mathcal{S}}^{\mathcal{L}}(k) = I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k | X_{\mathcal{S}^{\mathrm{c}}}) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{L}} \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k) \tag{3.10}$$

$$I_{\mathcal{S}}^{\mathcal{L}}(k|T) = I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k | X_{\mathcal{S}^{\mathrm{c}}} T) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{L}} \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k T) \tag{3.11}$$

where $\mathcal{S}^{\mathrm{c}}$ in (3.10) and (3.11) is the complement of $\mathcal{S}$ in $\mathcal{L}$. We write $R_{\mathcal{S}} = \sum_{k\in\mathcal{S}} R_k$.

3.2. Main Result and Proof

The following theorem is the main result of this chapter.

Theorem 3.1. For a K-node memoryless network with one multicast session per node, SNNC with backward decoding achieves the same rate tuples $(R_1, \ldots, R_K)$ as SNNC with sliding window decoding [5, 6] and LNNC with joint decoding [10, 11]. These are

the rate tuples satisfying

$$0 \leq R_{\mathcal{S}} < I_{\mathcal{S}}^{\mathcal{K}_k}(k|T) \tag{3.12}$$

for all $k \in \mathcal{K}$, all subsets $\mathcal{S} \subset \mathcal{K}_k$ with $k \in \mathcal{S}^{\mathrm{c}}$ and $\mathcal{S} \cap \widetilde{\mathcal{D}}_k \neq \emptyset$, where $\mathcal{S}^{\mathrm{c}}$ is the complement of $\mathcal{S}$ in $\mathcal{K}_k$, and for joint distributions that factor as

$$P(t)\left[\prod_{k=1}^{K} P(x_k|t)P(\hat{y}_k|y_k, x_k, t)\right] P(y^K|x^K). \tag{3.13}$$

Remark 3.1. The set $\mathcal{K}_k$ (see (3.9)) represents the set of nodes whose messages are known or decoded at node k. In other words, from node k's perspective the network has nodes $\mathcal{K}_k$ only.

Example 3.2. If $\mathcal{D} = \mathcal{D}_1 = \cdots = \mathcal{D}_K$, then the bound (3.12) is taken for all $k \in \mathcal{K}$ and all subsets $\mathcal{S} \subset \mathcal{K}_k$ with $k \in \mathcal{S}^{\mathrm{c}}$ and $\mathcal{S} \cap \mathcal{D} \neq \emptyset$, where $\mathcal{S}^{\mathrm{c}}$ is the complement of $\mathcal{S}$ in $\mathcal{K}_k$.

Example 3.3. Consider $\mathcal{K} = \{1, 2, 3, 4\}$ and suppose node 1 has a message destined for node 3, and node 2 has a message destined for node 4. We then have $\widetilde{\mathcal{D}}_3 = \{1\}$ and $\widetilde{\mathcal{D}}_4 = \{2\}$. If nodes 3 and 4 choose $\mathcal{T}_3 = \{2\}$ and $\mathcal{T}_4 = \{\emptyset\}$ respectively, then we have $\mathcal{K}_3 = \{1, 2, 3\}$ and $\mathcal{K}_4 = \{2, 4\}$. In this case the rate bounds (3.12) are:
Node 3:

$$R_1 < I(X_1; \hat{Y}_2\hat{Y}_3Y_3|X_2X_3T) \tag{3.14}$$

$$R_1 + R_2 < I(X_1X_2; \hat{Y}_3Y_3|X_3T) - I(\hat{Y}_1\hat{Y}_2; Y_1Y_2|X_1X_2X_3\hat{Y}_3Y_3T) \tag{3.15}$$

Node 4:

$$R_2 < I(X_2; \hat{Y}_4Y_4|X_4T) - I(\hat{Y}_2; Y_2|X_2X_4\hat{Y}_4Y_4T) \tag{3.16}$$

3.2.1. Encoding

To prove Theorem 3.1, we choose $\mathcal{K}_k = \mathcal{K}$ for all k for simplicity. We later discuss the case where these sets are different. For clarity, we set the time-sharing random variable T

Block	1	$\cdots$	B	$B+1 \cdots B+K\cdot(K-1)$
X_1	$\mathbf{x}_{11}(w_{11},1)$	$\cdots$	$\mathbf{x}_{1B}(w_{1B},l_{1(B-1)})$	
$\hat{Y}_1$	$\hat{\mathbf{y}}_{11}(l_{11}\|w_{11},1)$	$\cdots$	$\hat{\mathbf{y}}_{1B}(l_{1B}\|w_{1B},l_{1(B-1)})$	Multihop K messages
$\vdots$	$\vdots$	$\vdots$	$\vdots$	to $K-1$ nodes
X_K	$\mathbf{x}_{K1}(w_{K1},1)$	$\cdots$	$\mathbf{x}_{KB}(w_{KB},l_{K(B-1)})$	in $K\cdot(K-1)\cdot n'$ channel uses
$\hat{Y}_K$	$\hat{\mathbf{y}}_{K1}(l_{K1}\|w_{K1},1)$		$\hat{\mathbf{y}}_{KB}(l_{KB}\|w_{KB},l_{K(B-1)})$	

Table 3.1.: SNNC for one multicast session per node.

to be a constant. Table 3.1 shows the SNNC encoding process. We redefine R_k to be the rate of the short messages in relation to the (redefined) block length n. In other words, the message $w_k, k \in \mathcal{K}$, of nBR_k bits is split into B equally sized blocks, $w_{k1},\ldots,w_{kB}$, each of nR_k bits. Communication takes place over $B+K\cdot(K-1)$ blocks and the true rate of w_k will be

$$R_{k,\text{true}} = \frac{nBR_k}{nB+[K\cdot(K-1)\cdot n']} \tag{3.17}$$

where n' is defined in (3.20) below.

Random Code: Fix a distribution $\prod_{k=1}^{K} P(x_k)P(\hat{y}_k|y_k,x_k)$. For each block $j=1,\ldots,B$ and node $k\in\mathcal{K}$, generate $2^{n(R_k+\hat{R}_k)}$ code words $\mathbf{x}_{kj}(w_{kj},l_{k(j-1)})$, $w_{kj}=1,\ldots,2^{nR_k}$, $l_{k(j-1)}=1,\ldots,2^{n\hat{R}_k}$, according to $\prod_{i=1}^{n} P_{X_k}(x_{(kj)i})$ where $l_{k0}=1$ by convention. For each w_{kj} and $l_{k(j-1)}$, generate $2^{n\hat{R}_k}$ reconstructions $\hat{\mathbf{y}}_{kj}(l_{kj}|w_{kj},l_{k(j-1)})$, $l_{kj}=1,\ldots,2^{n\hat{R}_k}$, according to $\prod_{i=1}^{n} P_{\hat{Y}_k|X_k}(\hat{y}_{(kj)i}|x_{(kj)i}(w_{kj},l_{k(j-1)}))$. This defines the codebooks

$$\begin{aligned}\mathcal{C}_{kj} = \{&\mathbf{x}_{kj}(w_{kj},l_{k(j-1)}),\ \hat{\mathbf{y}}_{kj}(l_{kj}|w_{kj},l_{k(j-1)}),\\ &w_{kj}=1,\ldots,2^{nR_k},\ l_{k(j-1)}=1,\ldots,2^{n\hat{R}_k}, l_{kj}=1,\ldots,2^{n\hat{R}_k}\}\end{aligned} \tag{3.18}$$

for $j=1,\ldots,B$ and $k\in\mathcal{K}$.

The codebooks used in the last $K(K-1)$ blocks with $j>B$ are different. The blocks

$$j = B+(k-1)\cdot(K-1)+1,\ldots,B+k\cdot(K-1) \tag{3.19}$$

are dedicated to flooding l_{kB} through the network, and for all nodes $\tilde{k}\in\mathcal{K}$ we generate $2^{n'\hat{R}_k}$ independent and identically distributed (i.i.d.) code words $\mathbf{x}_{\tilde{k}j}(l_{kB})$, $l_{kB}=$

$1, \ldots, 2^{n'\hat{R}_k}$, according to $\prod_{i=1}^{n'} P_{X_{\tilde{k}}}(x_{(\tilde{k}j)i})$. We choose

$$n' = \max_k \frac{n\hat{R}_k}{\min\limits_{\tilde{k}\in\mathcal{K}} R_{k\tilde{k}}} \tag{3.20}$$

that is independent of k and B. The overall rate of user k is thus given by (3.17) which approaches R_k as $B \to \infty$.

Encoding: Each node k upon receiving $\mathbf{y}_{kj}$ at the end of block j, $j \leq B$, tries to find an index l_{kj} such that the following event occurs:

$$E_{0(kj)}(l_{kj}) : \left(\hat{\mathbf{y}}_{kj}(l_{kj}|w_{kj}, l_{k(j-1)}),\ \mathbf{x}_{kj}(w_{kj}, l_{k(j-1)}),\ \mathbf{y}_{kj}\right) \in \mathcal{T}_\epsilon^n \left(P_{\hat{Y}_k X_k Y_k}\right) \tag{3.21}$$

If there is no such index l_{kj}, set $l_{kj} = 1$. If there is more than one, choose one. Each node k transmits $\mathbf{x}_{kj}(w_{kj}, l_{k(j-1)})$ in block $j = 1, \ldots, B$.

In the $K-1$ blocks (3.19), node k conveys l_{kB} reliably to all other nodes by multihopping $\mathbf{x}_{kj}(l_{kB})$ through the network with blocks of length n'.

3.2.2. Backward Decoding

Let $\epsilon_1 > \epsilon$. At the end of block $B + K \cdot (K-1)$ every node $k \in \mathcal{K}$ has reliably recovered $\mathbf{l}_B = (l_{1B}, \ldots, l_{KB})$ via the multihopping of the last $K(K-1)$ blocks.

For block $j = B, \ldots, 1$, node k tries to find tuples $\hat{\mathbf{w}}_j^{(k)} = (\hat{w}_{1j}^{(k)}, \ldots, \hat{w}_{Kj}^{(k)})$ and $\hat{\mathbf{l}}_{j-1}^{(k)} = (\hat{l}_{1(j-1)}^{(k)}, \ldots, \hat{l}_{K(j-1)}^{(k)})$ such that the following event occurs:

$$\begin{aligned} E_{1(kj)}(\hat{\mathbf{w}}_j^{(k)}, \hat{\mathbf{l}}_{j-1}^{(k)}, \mathbf{l}_j) : \Big(&\mathbf{x}_{1j}(\hat{w}_{1j}^{(k)}, \hat{l}_{1(j-1)}^{(k)}), \ldots, \mathbf{x}_{Kj}(\hat{w}_{Kj}^{(k)}, \hat{l}_{K(j-1)}^{(k)}), \\ &\hat{\mathbf{y}}_{1j}(l_{1j}|\hat{w}_{1j}^{(k)}, \hat{l}_{1(j-1)}^{(k)}), \ldots, \hat{\mathbf{y}}_{Kj}(l_{Kj}|\hat{w}_{Kj}^{(k)}, \hat{l}_{K(j-1)}^{(k)}), \mathbf{y}_{kj}\Big) \in \mathcal{T}_{\epsilon_1}^n \left(P_{X_{\mathcal{K}} \hat{Y}_{\mathcal{K}} Y_k}\right) \end{aligned} \tag{3.22}$$

where $\mathbf{l}_j = (l_{1j}, \ldots, l_{Kj})$ has already been reliably recovered from the previous block $j+1$.

Error Probability: Let $\mathbf{1} = (1, \ldots, 1)$ and assume without loss of generality that $\mathbf{w}_j = \mathbf{1}$ and $\mathbf{l}_{j-1} = \mathbf{1}$. In each block j, the error events at node k are:

$$E_{(kj)0} : \cap_{l_{kj}}\ E_{0(kj)}^{\mathrm{c}}(l_{kj}) \tag{3.23}$$

$$E_{(kj)1} : E_{1(kj)}^{\mathrm{c}}(\mathbf{1}, \mathbf{1}, \mathbf{1}) \tag{3.24}$$

$$E_{(kj)2} : \cup_{(\mathbf{w}_j, \mathbf{l}_{j-1}) \neq (\mathbf{1}, \mathbf{1})} E_{1(kj)}(\mathbf{w}_j, \mathbf{l}_{j-1}, \mathbf{1}) \tag{3.25}$$

The error event $E_{kj} = \cup_{i=0}^{2} E_{(kj)i}$ at node k in block j thus satisfies

$$\Pr\Big[E_{kj}\Big] \leq \sum_{i=0}^{2} \Pr\Big[E_{(kj)i}\Big] \tag{3.26}$$

where we have used the union bound. $\Pr\Big[E_{(kj)0}\Big]$ can be made small with large n, as long as (see [20])

$$\hat{R}_k > I(\hat{Y}_k; Y_k | X_k) + \delta_\epsilon(n) \tag{3.27}$$

where $\delta_\epsilon(n) \to 0$ as $n \to \infty$. Similarly, $\Pr\Big[E_{(kj)1}\Big]$ can be made small with large n.

To bound $\Pr\Big[E_{(kj)2}\Big]$, for each $\mathbf{w}_j$ and $\mathbf{l}_{j-1}$ we define

$$\mathcal{M}(\mathbf{w}_j) = \{i \in \mathcal{K} : w_{ij} \neq 1\} \tag{3.28}$$

$$\mathcal{Q}(\mathbf{l}_{j-1}) = \{i \in \mathcal{K} : l_{i(j-1)} \neq 1\} \tag{3.29}$$

$$\mathcal{S}(\mathbf{w}_j, \mathbf{l}_{j-1}) = \mathcal{M}(\mathbf{w}_j) \cup \mathcal{Q}(\mathbf{l}_{j-1}) \tag{3.30}$$

and write $\mathcal{S} = \mathcal{S}(\mathbf{w}_j, \mathbf{l}_{j-1})$. The important observations are:

- ▷ $(\mathbf{X}_{\mathcal{S}}, \hat{\mathbf{Y}}_{\mathcal{S}})$ is independent of $(\mathbf{X}_{\mathcal{S}^{\mathrm{c}}}, \hat{\mathbf{Y}}_{\mathcal{S}^{\mathrm{c}}}, \mathbf{Y}_{kj})$ in the random coding experiment;
- ▷ The $(X_i, \hat{Y}_i)$, $i \in \mathcal{S}$, are mutually independent.

For $k \in \mathcal{S}^{\mathrm{c}}$ and $(\mathbf{w}_j, \mathbf{l}_{j-1}) \neq (\mathbf{1}, \mathbf{1})$, we thus have

$$\Pr\Big[E_{1(kj)}(\mathbf{w}_j, \mathbf{l}_{j-1}, \mathbf{l}_j)\Big] \leq 2^{-n(I_{\mathcal{S}} - \delta_{\epsilon_1}(n))} \tag{3.31}$$

where $\delta_{\epsilon_1}(n) \to 0$ as $n \to \infty$ and

$$\begin{aligned} I_{\mathcal{S}} &= \left[\sum_{i \in \mathcal{S}} H(X_i \hat{Y}_i)\right] + H(X_{\mathcal{S}^{\mathrm{c}}} \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k) - H(X_{\mathcal{S}} \hat{Y}_{\mathcal{S}} X_{\mathcal{S}^{\mathrm{c}}} \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k) \\ &= I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k | X_{\mathcal{S}^{\mathrm{c}}}) + \left[\sum_{i \in \mathcal{S}} H(\hat{Y}_i | X_i)\right] - H(\hat{Y}_{\mathcal{S}} | X_{\mathcal{K}} \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k). \end{aligned} \tag{3.32}$$

By the union bound, we have

$$\begin{aligned}
\Pr\Big[E_{(kj)2}\Big] &\leq \sum_{(\mathbf{w}_j,\mathbf{l}_{j-1})\neq(\mathbf{1},\mathbf{1})} \Pr\Big[E_{1(kj)}(\mathbf{w}_j,\mathbf{l}_{j-1},\mathbf{1})\Big] \\
&\overset{(a)}{\leq} \sum_{(\mathbf{w}_j,\mathbf{l}_{j-1})\neq(\mathbf{1},\mathbf{1})} 2^{-n(I_{\mathcal{S}(\mathbf{w}_j,\mathbf{l}_{j-1})}-\delta_{\epsilon_1}(n))} \\
&\overset{(b)}{=} \sum_{\substack{\mathcal{S}:k\in\mathcal{S}^c\\ \mathcal{S}\neq\emptyset}} \sum_{\substack{(\mathbf{w}_j,\mathbf{l}_{j-1}):\\ \mathcal{S}(\mathbf{w}_j,\mathbf{l}_{j-1})=\mathcal{S}}} 2^{-n(I_{\mathcal{S}}-\delta_{\epsilon_1}(n))} \\
&\overset{(c)}{=} \sum_{\substack{\mathcal{S}:k\in\mathcal{S}^c\\ \mathcal{S}\neq\emptyset}} \sum_{\substack{\mathcal{M}\subseteq\mathcal{S},\mathcal{Q}\subseteq\mathcal{S}\\ \mathcal{M}\cup\mathcal{Q}=\mathcal{S}}} \left(\prod_{i\in\mathcal{M}}(2^{nR_i}-1)\prod_{i\in\mathcal{Q}}(2^{n\hat{R}_i}-1)\right)\cdot 2^{-n(I_{\mathcal{S}}-\delta_{\epsilon_1}(n))} \\
&< \sum_{\substack{\mathcal{S}:k\in\mathcal{S}^c\\ \mathcal{S}\neq\emptyset}} \sum_{\substack{\mathcal{M}\subseteq\mathcal{S},\mathcal{Q}\subseteq\mathcal{S}\\ \mathcal{M}\cup\mathcal{Q}=\mathcal{S}}} 2^{nR_{\mathcal{M}}}2^{n\hat{R}_{\mathcal{Q}}}2^{-n(I_{\mathcal{S}}-\delta_{\epsilon_1}(n))} \\
&\overset{(d)}{\leq} \sum_{\substack{\mathcal{S}:k\in\mathcal{S}^c\\ \mathcal{S}\neq\emptyset}} 3^{|\mathcal{S}|}2^{n\left(R_{\mathcal{S}}+\hat{R}_{\mathcal{S}}-(I_{\mathcal{S}}-\delta_{\epsilon_1}(n))\right)} \\
&= \sum_{\substack{\mathcal{S}:k\in\mathcal{S}^c\\ \mathcal{S}\neq\emptyset}} 2^{n\left[R_{\mathcal{S}}-(I_{\mathcal{S}}-\hat{R}_{\mathcal{S}}-\frac{|\mathcal{S}|\log_2 3}{n}-\delta_{\epsilon_1}(n))\right]}
\end{aligned} \tag{3.33}$$

where

(a) follows from (3.31)

(b) follows by collecting the $(\mathbf{w}_j,\mathbf{l}_{j-1})\neq(\mathbf{1},\mathbf{1})$ into classes where $\mathcal{S}=\mathcal{S}(\mathbf{w}_j,\mathbf{l}_{j-1})$

(c) follows because there are

$$\prod_{i\in\mathcal{M}}(2^{nR_i}-1)\prod_{i\in\mathcal{Q}}(2^{n\hat{R}_i}-1) \tag{3.34}$$

different $(\mathbf{w}_j,\mathbf{l}_{j-1})\neq(\mathbf{1},\mathbf{1})$ that result in the same $\mathcal{M}$ and $\mathcal{Q}$ such that $\mathcal{M}\subseteq\mathcal{S}$, $\mathcal{Q}\subseteq\mathcal{S}$ and $\mathcal{S}=\mathcal{M}\cup\mathcal{Q}$

(d) is because for every node $i\in\mathcal{S}$, we must have one of the following three cases occur:

1) $i\in\mathcal{M}$ and $i\notin\mathcal{Q}$

2) $i\notin\mathcal{M}$ and $i\in\mathcal{Q}$

3) $i\in\mathcal{M}$ and $i\in\mathcal{Q}$

so there are $3^{|\mathcal{S}|}$ different ways of choosing $\mathcal{M}$ and $\mathcal{Q}$.

Since we require $\hat{R}_k \geq I(\hat{Y}_k; Y_k|X_k) + \delta_\epsilon(n)$, we have

$$\begin{aligned} I_{\mathcal{S}} - \hat{R}_{\mathcal{S}} &\leq I_{\mathcal{S}} - \sum_{i\in\mathcal{S}} I(\hat{Y}_i; Y_i|X_i) - \delta_\epsilon(n) \\ &= I_{\mathcal{S}}^{\mathcal{K}}(k) - \delta_\epsilon(n). \end{aligned} \tag{3.35}$$

Combining (3.26), (3.27), (3.33) and (3.35) we find that we can make $\Pr\left[E_{kj}\right] \to 0$ as $n \to \infty$ if

$$0 \leq R_{\mathcal{S}} < I_{\mathcal{S}}^{\mathcal{K}}(k) \tag{3.36}$$

for all subsets $\mathcal{S} \subset \mathcal{K}$ such that $k \in \mathcal{S}^{\mathrm{c}}$ and $\mathcal{S} \neq \emptyset$. Of course, if $I_{\mathcal{S}}^{\mathcal{K}}(k) \leq 0$, then we require that $R_{\mathcal{S}} = 0$.

We can split the bounds in (3.36) into two classes:

$$\text{Class 1} : \mathcal{S} \cap \widetilde{\mathcal{D}}_k \neq \emptyset \tag{3.37}$$

$$\text{Class 2} : \mathcal{S} \cap \widetilde{\mathcal{D}}_k = \emptyset \text{ or equivalently } \mathcal{S} \subseteq \widetilde{\mathcal{D}}_k^{\mathrm{c}} \tag{3.38}$$

LNNC requires only the Class 1 bounds. SNNC requires both the Class 1 and Class 2 bounds to guarantee reliable decoding of the quantization indices $\mathbf{l}_{j-1}$ for each backward decoding step. With the same argument as in [5, Sec. IV-C], we can show that the Class 2 bounds can be ignored when determining the best SNNC rates. For completeness, the proof is given in Appendix A.1. SNNC with backward decoding thus performs as well as SNNC with sliding window decoding and LNNC with joint decoding. Adding a time-sharing random variable T completes the proof of Theorem 3.1 for $\mathcal{K}_k = \mathcal{K}$ for all k. The proof with general $\mathcal{K}_k$ follows by having each node k treat the signals of nodes in $\mathcal{K} \setminus \mathcal{K}_k$ as noise

3.3. Discussion

3.3.1. Decoding Subsets of Messages

From Theorem 3.1 we know that if node k decodes messages from nodes in $\mathcal{K}_k$ and some of the Class 2 constraints in (3.38) are violated, then we should treat the signals from the corresponding nodes as noise. In this way, we eventually wind up with some

$$\widetilde{\mathcal{K}}_k = \{k\} \cup \widetilde{\mathcal{D}}_k \cup \mathcal{T}_k, \mathcal{T}_k \subseteq \widetilde{\mathcal{D}}_k^{\mathrm{c}} \setminus \{k\}, \tag{3.39}$$

where all Class 2 constraints are satisfied, i.e., we have

$$0 \leq R_{\mathcal{S}} < I_{\mathcal{S}}^{\widetilde{\mathcal{K}}_k}(k|T), \text{ for all } \mathcal{S} \subseteq \mathcal{T}_k, \ \mathcal{S} \neq \emptyset \tag{3.40}$$

and we achieve as good or better rates. In this sense, the sets $\widetilde{\mathcal{K}}_k$ are important even for LNNC. These sets seem difficult to find in large networks because many constraints need to be checked. However, provided that the sets $\widetilde{\mathcal{K}}_k$ are known, we have the following lemma.

Lemma 3.2. For the K-node DMN, the rate tuples $(R_1, \ldots, R_K)$ are achievable if

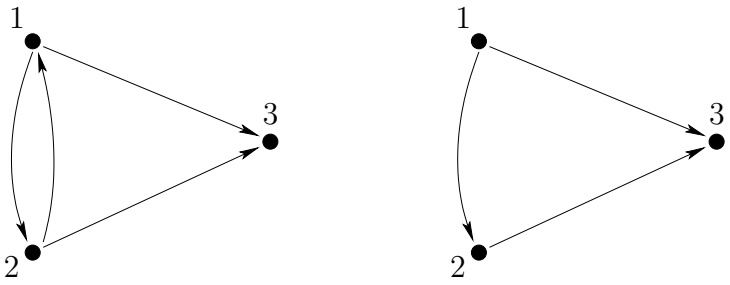
$$R_{\mathcal{S}} < I_{\mathcal{S}}^{\widetilde{\mathcal{K}}_k}(k|T)$$

for all $k \in \mathcal{K}$, all subsets $\mathcal{S} \subset \widetilde{\mathcal{K}}_k$ with $k \in \mathcal{S}^{\mathrm{c}}$ and $\mathcal{S} \cap \widetilde{\mathcal{D}}_k \neq \emptyset$, $\widetilde{\mathcal{K}}_k = \{k\} \cup \widetilde{\mathcal{D}}_k \cup \mathcal{T}_k$, $\mathcal{T}_k \subseteq \widetilde{\mathcal{D}}_k^{\mathrm{c}} \setminus \{k\}$, where $\mathcal{T}_k$ satisfies (3.40) and for any joint distribution that factors as (3.13).

Proof: The proof follows by including the messages from nodes in $\widetilde{\mathcal{K}}_k$ satisfying (3.40) in the typicality test at every destination k in Theorem 3.1. ■

3.3.2. Choice of Typicality Test

Theorem 3.1 has a subtle addition to [25] and difference to [11, Theorem 2] and [30, Theorem 18.5], namely that in (3.12) each $k \in \mathcal{K}$ may have a *different* set $\widetilde{\mathcal{K}}_k$ of nodes satisfying all Class 2 constraints whose messages and quantization indices are included



(a) Both nodes 1 and 2 are sources and relays

(b) Node 2 acts as a relay for node 1

Figure 3.3.: Examples of a three-node network with different rate pairs.

in the typicality test. But we can achieve the rates in (3.12) at node k with SNNC by using backward decoding and treating the signals from the nodes in $\mathcal{K} \setminus \widetilde{\mathcal{K}}_k$ as noise. Hence we may ignore the Class 2 constraints in (3.38) when determining the *best* SNNC rates.

The following example suggests that it may not be surprising that the SNNC and LNNC rate regions are the same. Consider the network in Fig. 3.3, where $\mathcal{K} = \{1, 2, 3\}$. Suppose both nodes 1 and 2 act as sources as well as relays for each other in transmitting information to node 3 (see Fig. 3.3(a)). Referring to Theorem 3.1, the SNNC and LNNC bounds are (see Fig. 3.4):

$$R_1 < I(X_1; \hat{Y}_2 Y_3 | X_2) - I(\hat{Y}_1; Y_1 | X_1 X_2 \hat{Y}_2 Y_3) \tag{3.41}$$

$$R_2 < I(X_2; \hat{Y}_1 Y_3 | X_1) - I(\hat{Y}_2; Y_2 | X_1 X_2 \hat{Y}_1 Y_3) \tag{3.42}$$

$$R_1 + R_2 < I(X_1 X_2; Y_3) - I(\hat{Y}_1 \hat{Y}_2; Y_1 Y_2 | X_1 X_2 Y_3) \tag{3.43}$$

However, suppose now that node 2 has no message ($R_2 = 0$) and acts as a relay node only (see Fig. 3.3(b)). Then LNNC does not have the bound (3.42) while SNNC has the bound (3.42) with $R_2 = 0$ and $\hat{Y}_1 = \emptyset$. We ask whether (3.42) reduces the SNNC rate. This is equivalent to asking whether SNNC achieves point 1 in Fig. 3.4. It would be strange if there was a discontinuity in the achievable rate region at $R_2 = 0$.

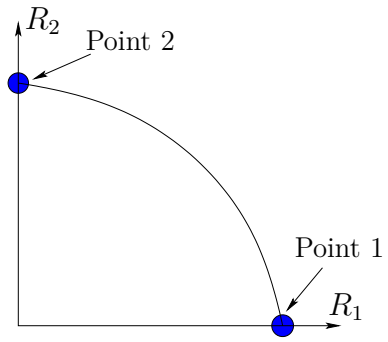


Figure 3.4.: Illustration of the achievable rates for the network of Fig. 3.3(b).

3.3.3. Optimal Decodable Sets

SNNC was studied for relay networks in [7]. For such networks there is one message at node 1 that is destined for node K. We thus have $\widetilde{\mathcal{D}}_K = \{1\}$ and $\widetilde{\mathcal{D}}_K^{\mathrm{c}} \setminus \{K\} = \{2, \ldots, K-1\}$. The authors of [7] showed that for a given random coding distribution

$$P(t)P(x_1|t) \prod_{k \in \widetilde{\mathcal{D}}_K^{\mathrm{c}}} P(x_k|t)P(\hat{y}_k|y_k, x_k, t) \tag{3.44}$$

there exists a *unique largest optimal decodable set* $\mathcal{T}^*$, $\mathcal{T}^* \subseteq \widetilde{\mathcal{D}}_K^{\mathrm{c}} \setminus \{K\}$, of the relay nodes that provides the same best achievable rates for both SNNC and LNNC [7, Theorem 2.8]. We now show that the concept of *optimal decodable set* extends naturally to *multi-source networks*.

Lemma 3.3. For a K-node memoryless network with a fixed random coding distribution

$$P(t) \prod_{k=1}^{K} P(x_k|t)P(\hat{y}_k|y_k, x_k, t) \tag{3.45}$$

there exists for each node k a *unique largest* set $\mathcal{T}_k^*$ among all subsets $\mathcal{T}_k \subseteq \widetilde{\mathcal{D}}_k^{\mathrm{c}} \setminus \{k\}$ satisfying (3.40). The messages of the nodes in $\mathcal{T}_k^*$ should be included in the typicality test to provide the best achievable rates.

Proof: We prove Lemma 3.3 without a time-sharing random variable T. The proof with T is similar. We show that $\mathcal{T}_k^*$ is unique by showing that the union of any two sets $\mathcal{T}_k^1$ and $\mathcal{T}_k^2$ satisfying all constraints also satisfies all constraints and provides as good or

better rates. Continuing taking the union, we eventually reach a unique largest set $\mathcal{T}_k^*$ that satisfies all constraints and gives the best rates.

Partition the subsets $\mathcal{T}_k \subseteq \widetilde{\mathcal{D}}_k^{\mathrm{c}} \setminus \{k\}$ into two classes:

Class 1: $R_{\mathcal{S}} < I_{\mathcal{S}}^{\mathcal{K}_k}(k)$ for all $\mathcal{S} \subseteq \mathcal{T}_k$;

Class 2: There exists one $\mathcal{S} \subseteq \mathcal{T}_k$ such that $R_{\mathcal{S}} \geq I_{\mathcal{S}}^{\mathcal{K}_k}(k)$.

We may ignore the $\mathcal{T}_k$ in Class 2 because the proof of Theorem 3.1 shows that we can treat the signals of nodes associated with violated constraints as noise and achieve as good or better rates. Hence, we focus on $\mathcal{T}_k$ in Class 1.

Suppose $\mathcal{T}_k^1$ and $\mathcal{T}_k^2$ are in Class 1 and let $\mathcal{T}_k^3 = \mathcal{T}_k^1 \cup \mathcal{T}_k^2$. We define

$$\widetilde{\mathcal{K}}_k^1 = \{k\} \cup \widetilde{\mathcal{D}}_k \cup \mathcal{T}_k^1 \tag{3.46}$$

$$\widetilde{\mathcal{K}}_k^2 = \{k\} \cup \widetilde{\mathcal{D}}_k \cup \mathcal{T}_k^2 \tag{3.47}$$

$$\widetilde{\mathcal{K}}_k^3 = \{k\} \cup \widetilde{\mathcal{D}}_k \cup \mathcal{T}_k^3. \tag{3.48}$$

Further, for every $\mathcal{S} \subseteq \widetilde{\mathcal{K}}_k^3$, define $\mathcal{S}_1 = \mathcal{S} \cap \widetilde{\mathcal{K}}_k^1$ and $\mathcal{S}_2 = \mathcal{S} \cap (\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}_1)$. We have $\mathcal{S}_1 \subseteq \widetilde{\mathcal{K}}_k^1$, $\mathcal{S}_2 \subseteq \widetilde{\mathcal{K}}_k^2$, $\mathcal{S}_1 \cup \mathcal{S}_2 = \mathcal{S}$ and $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$. We further have

$$\begin{aligned}
R_{\mathcal{S}} &\overset{(a)}{=} R_{\mathcal{S}_1} + R_{\mathcal{S}_2} \\
&\overset{(b)}{<} I_{\mathcal{S}_1}^{\widetilde{\mathcal{K}}_k^1}(k) + I_{\mathcal{S}_2}^{\widetilde{\mathcal{K}}_k^2}(k) \\
&\overset{(c)}{=} I(X_{\mathcal{S}_1}; \hat{Y}_{\widetilde{\mathcal{K}}_k^1 \setminus \mathcal{S}_1} Y_k | X_{\widetilde{\mathcal{K}}_k^1 \setminus \mathcal{S}_1}) - I(\hat{Y}_{\mathcal{S}_1}; Y_{\mathcal{S}_1} | X_{\widetilde{\mathcal{K}}_k^1} \hat{Y}_{\widetilde{\mathcal{K}}_k^1 \setminus \mathcal{S}_1} Y_k) \\
&\quad + I(X_{\mathcal{S}_2}; \hat{Y}_{\widetilde{\mathcal{K}}_k^2 \setminus \mathcal{S}_2} Y_k | X_{\widetilde{\mathcal{K}}_k^2 \setminus \mathcal{S}_2}) - I(\hat{Y}_{\mathcal{S}_2}; Y_{\mathcal{S}_2} | X_{\widetilde{\mathcal{K}}_k^2} \hat{Y}_{\widetilde{\mathcal{K}}_k^2 \setminus \mathcal{S}_2} Y_k) \\
&\overset{(d)}{\leq} I(X_{\mathcal{S}_1}; \hat{Y}_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}} Y_k | X_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}}) - I(\hat{Y}_{\mathcal{S}_1}; Y_{\mathcal{S}_1} | X_{\widetilde{\mathcal{K}}_k^1} \hat{Y}_{\widetilde{\mathcal{K}}_k^1 \setminus \mathcal{S}_1} Y_k) \\
&\quad + I(X_{\mathcal{S}_2}; \hat{Y}_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}} Y_k | X_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}_2}) - I(\hat{Y}_{\mathcal{S}_2}; Y_{\mathcal{S}_2} | X_{\widetilde{\mathcal{K}}_k^2} \hat{Y}_{\widetilde{\mathcal{K}}_k^2 \setminus \mathcal{S}_2} Y_k) \\
&\overset{(e)}{\leq} I(X_{\mathcal{S}_1}; \hat{Y}_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}} Y_k | X_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}}) - I(\hat{Y}_{\mathcal{S}_1}; Y_{\mathcal{S}_1} | X_{\widetilde{\mathcal{K}}_k^3} \hat{Y}_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}} Y_k) \\
&\quad + I(X_{\mathcal{S}_2}; \hat{Y}_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}} Y_k | X_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}_2}) - I(\hat{Y}_{\mathcal{S}_2}; Y_{\mathcal{S}_2} | X_{\widetilde{\mathcal{K}}_k^3} \hat{Y}_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}_2} Y_k) \\
&\overset{(f)}{=} I(X_{\mathcal{S}}; \hat{Y}_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}} Y_k | X_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}}) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\widetilde{\mathcal{K}}_k^3} \hat{Y}_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}} Y_k) \\
&\overset{(g)}{=} I_{\mathcal{S}}^{\widetilde{\mathcal{K}}_k^3}(k)
\end{aligned} \tag{3.49}$$

where

(a) follows from the definition of $\mathcal{S}_1$ and $\mathcal{S}_2$

(b) follows because both $\mathcal{T}_k^1$ and $\mathcal{T}_k^2$ are in Class 1

(c) follows from the definition (3.10)

(d) follows because all X_k are independent and conditioning does not increase entropy

(e) follows because conditioning does not increase entropy and by the Markov chains

$$X_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}_1} \hat{Y}_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}} Y_k - Y_{\mathcal{S}_1} X_{\mathcal{S}_1} - \hat{Y}_{\mathcal{S}_1} \tag{3.50}$$

$$X_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}_2} \hat{Y}_{\widetilde{\mathcal{K}}_k^3 \setminus \mathcal{S}_2} Y_k - Y_{\mathcal{S}_2} X_{\mathcal{S}_2} - \hat{Y}_{\mathcal{S}_2} \tag{3.51}$$

(f) follows from the chain rule for mutual information and the Markov chains (3.50) and (3.51)

(g) follows from the definition (3.10).

The bound (3.49) shows that $\mathcal{T}_k^3$ is also in Class 1. Moreover, by (3.49) if k includes the messages of nodes in $\widetilde{\mathcal{K}}_k^3$ in the typicality test, then the rates are as good or better than those achieved by including the messages of nodes in $\widetilde{\mathcal{K}}_k^1$ or $\widetilde{\mathcal{K}}_k^2$ in the typicality test. Taking the union of all $\mathcal{T}_k$ in Class 1, we obtain the *unique largest* set $\mathcal{T}_k^*$ that gives the best achievable rates. ■

Remark 3.2. There are currently no efficient algorithms for finding an optimal decodable set. Such algorithms would be useful for applications with time-varying channels.

3.3.4. Related Work

Sliding Window Decoding

SNNC with sliding window decoding was studied in [5, 6] and LNNC [11] and achieves the same rates as in [5]. SNNC has extra constraints that turn out to be redundant [5, Sec. IV-C], [6, Sec. V-B]. The sliding window decoding in [5] resembles that in [31]

where encoding is delayed (or offset) and different decoders are chosen depending on the rate point. The rates achieved by one decoder may not give the entire rate region of Theorem 3.1, but the union of achievable rates of all decoders does [6, Theorem 1]. The advantage of sliding window decoding is a small decoding delay of $K+1$ blocks as compared to backward decoding that requires $B+K(K-1)$ blocks, where $B \gg K$.

Backward Decoding

SNNC with *backward decoding* was studied in [7] for single source networks. For these networks, [7] showed that LNNC and SNNC achieve the same rates. Further, for a fixed random coding distribution there is a subset of the relay nodes whose messages should be decoded to achieve the best LNNC and SNNC rates. Several other interesting properties of the coding scheme were derived. It was also shown in [27] that SNNC with a layered network analysis [15] achieves the same LNNC rates for single source networks. In [32], SNNC with partial cooperation between the sources was considered for multi-source networks.

Joint Decoding

It turns out that SNNC with joint decoding achieves the same rates as in Theorem 3.1. Recently, the authors of [26] showed that SNNC with joint decoding fails to achieve the LNNC rates for a *specific* choice of SNNC protocol. However, by multihopping the last quantization indices and then performing joint decoding with the messages and remaining quantization bits, SNNC with joint decoding performs as well as SNNC with sliding window or backward decoding, and LNNC. This makes sense, since joint decoding should perform at least as well as backward decoding. Details are given in Appendix A.2.

Multihopping

We compare how the approaches of Theorem 3.1 and [7, Theorem 2.5] reliably convey the last quantization indices $\mathbf{l}_B$. Theorem 3.1 uses multihopping while Theorem 2.5 in [7] uses a QF-style method with M extra blocks after block B with the same block length n. In these M blocks every node transmits as before except that the messages are set to a default value. The initialization method in [7] has two disadvantages:

▷ Both B and M must go to infinity to reliably decode $\mathbf{l}_B$ [7, Sec.IV-A, Equ. (34)]. The true rate of node k's message w_k is

$$R'_{k,\text{true}} = \frac{nBR_k}{nB+nM} = \frac{B}{B+M} \cdot R_k \tag{3.52}$$

and we choose $B \gg M$ so that $R'_{k,\text{true}} \to R_k$ as $B \to \infty$.

▷ *Joint* rather than *per-block* processing is used.

Multihopping seems to be a better choice for reliably communicating $\mathbf{l}_B$, because the QF-style approach has a large decoding delay due to the large value of M and does not use per-block processing.

3.4. SNNC with a DF option

One of the main advantages of SNNC is that the relays can switch between QF (or CF) and DF depending on the channel conditions. If the channel conditions happen to be good, then the natural choice is DF which removes the noise at the relays. This not possible with LNNC due to the high rate of the long message. On the other hand, if a relay happens to experience a deep fade, then this relay should use QF (or CF).

In the following, we show how mixed strategies called SNNC-DF work for the multiple-relay channel. These mixed strategies are similar to those in [23, Theorem 4]. However, in [23] the relays use CF with a prescribed binning rate to enable *step-by-step* decoding (CF-S) instead of QF. In Section 3.5 we give numerical examples to show that SNNC-DF can outperform DF, CF-S and LNNC.

As in [23], we partition the relays $\mathcal{T} = \{2, \ldots, K-1\}$ into two sets

$$\begin{aligned} \mathcal{T}_1 &= \{k : \ 2 \le k \le K_1\} \\ \mathcal{T}_2 &= \mathcal{T} \setminus \mathcal{T}_1 \end{aligned}$$

where $1 \le K_1 \le K-1$. The relays in $\mathcal{T}_1$ use DF while the relays in $\mathcal{T}_2$ use QF. Let $\pi(\cdot)$ be a permutation on $\{1, \ldots, K\}$ with $\pi(1) = 1$ and $\pi(K) = K$ and let $\pi(j:k) = \{\pi(j), \pi(j+1), \ldots, \pi(k)\}$, $1 \le j \le k \le K$. Define $\mathcal{T}_{i(\pi)} = \{\pi(k), k \in \mathcal{T}_i\}$, $i = 1, 2$. We

have the following theorem.

Theorem 3.4. SNNC-DF achieves the rates satisfying

$$R_{\text{SNNC-DF}} < \max_{\pi(\cdot)} \max_{K_1} \min \Big\{ \min_{1\le k\le K_1-1} I\left(X_{\pi(1:k)}; Y_{\pi(k+1)}|X_{\pi(k+1:K_1)}\right), \\ I(X_1 X_{\mathcal{T}_{1(\pi)}} X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c} Y_K|X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}}|X_1 X_{\mathcal{T}} \hat{Y}_{\mathcal{S}^c} Y_K)\Big\} \quad (3.53)$$

for all $\mathcal{S} \subseteq \mathcal{T}_{2(\pi)}$, where $\mathcal{S}^c$ is the complement of $\mathcal{S}$ in $\mathcal{T}_{2(\pi)}$, and where the joint distribution factors as

$$P(x_1 x_{\mathcal{T}_{1(\pi)}}) \cdot \left[\prod_{k\in\mathcal{T}_{2(\pi)}} P(x_k) P(\hat{y}_k|y_k, x_k)\right] \cdot P(y_2, \ldots, y_K|x_1, \ldots, x_{K-1}). \quad (3.54)$$

Remark 3.3. As usual, we may add a time-sharing random variable to improve rates.

Proof Sketch: For a given permutation $\pi(\cdot)$ and K_1, the first mutual information term in (3.53) describes the DF bounds [23, Theorem 1] (see also [33, Theorem 3.1]). The second mutual information term in (3.53) describes the SNNC bounds. Using a similar analysis as for Theorem 3.1 and by treating $(X_1 X_{\mathcal{T}_{1(\pi)}})$ as the "new" source signal at the destination, we have the SNNC bounds

$$R_{\text{SNNC-DF}} < I(X_1 X_{\mathcal{T}_{1(\pi)}} X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c} Y_K|X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}}|X_1 X_{\mathcal{T}} \hat{Y}_{\mathcal{S}^c} Y_K) \quad (3.55)$$

$$0 \le I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c} Y_K|X_1 X_{\mathcal{T}_{1(\pi)}} X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}}|X_1 X_{\mathcal{T}} \hat{Y}_{\mathcal{S}^c} Y_K) \quad (3.56)$$

for all $\mathcal{S} \subseteq \mathcal{T}_{2(\pi)}$. The same argument used to prove Theorem 3.1 shows that if any of the constraints (3.56) is violated, then we get rate bounds that can be achieved with SNNC-DF by treating the signals from the corresponding relay nodes as noise. Thus we may ignore the constraints (3.56). ■

Example 3.4. Consider $K = 4$ and $K_1 = 2$. There are two possible permutations $\pi_1(1:4) = \{1, 2, 3, 4\}$ and $\pi_2(1:4) = \{1, 3, 2, 4\}$. For $\pi_1(1:4) = \{1, 2, 3, 4\}$, Theorem 3.4 states that SNNC-DF achieves any rate up to

$$R_{\text{SNNC-DF}} = \min \Big\{ I(X_1; Y_2|X_2), I(X_1 X_2; \hat{Y}_3 Y_4|X_3),$$

Block	1	2	$\cdots$	B	$B+1$	$B+2 \cdots B+4$
X_1	$\mathbf{x}_{11}(w_1, 1)$	$\mathbf{x}_{12}(w_2, w_1)$	$\cdots$	$\mathbf{x}_{1B}(w_B, w_{B-1})$	$\mathbf{x}_{1(B+1)}(1, w_B)$	
X_2	$\mathbf{x}_{21}(1)$	$\mathbf{x}_{22}(w_1)$	$\cdots$	$\mathbf{x}_{2B}(w_{(B-1)})$	$\mathbf{x}_{2(B+1)}(w_B)$	Multihop l_{B+1}
X_3	$\mathbf{x}_{31}(1)$	$\mathbf{x}_{32}(l_1)$	$\cdots$	$\mathbf{x}_{3B}(l_{B-1})$	$\mathbf{x}_{3(B+1)}(l_B)$	to node 4 in $3n'$
$\hat{Y}_3$	$\hat{\mathbf{y}}_{31}(l_1\|1)$	$\hat{\mathbf{y}}_{32}(l_2\|l_1)$	$\cdots$	$\hat{\mathbf{y}}_{3B}(l_B\|l_{B-1})$	$\hat{\mathbf{y}}_{3(B+1)}(l_{B+1}\|l_B)$	channel uses

Table 3.2.: Coding scheme for the two-relay channel with block Markov coding at the source.

$$I(X_1X_2X_3; Y_4) - I(\hat{Y}_3; Y_3|X_1X_2X_3Y_4)\Big\} \tag{3.57}$$

where the joint distribution factors as

$$P(x_1, x_2)P(x_3)P(\hat{y}_3|y_3, x_3) \cdot P(y_2, y_3, y_4|x_1, x_2, x_3). \tag{3.58}$$

The corresponding coding scheme is given in Table 3.2.

If relay node 2 uses DF while relay node 3 uses CF-S, then by [23, Theorem 4] with $U_2 = 0$, any rate up to

$$R_{\text{[CF-S]-DF}} < \min\ \Big\{I(X_1; Y_2|X_2), I(X_1X_2; \hat{Y}_3Y_4|X_3)\Big\} \tag{3.59}$$

can be achieved, subject to

$$I(\hat{Y}_3; Y_3|X_3Y_4) \le I(X_3; Y_4) \tag{3.60}$$

and the joint distribution factors as (3.58). It turns out that $R_{\text{[CF-S]-DF}}$ in (3.59)-(3.60) is the same as $R_{\text{SNNC-DF}}$ (3.57), since LNNC and SNNC do not improve the CF-S rate for one relay [8]. But $R_{\text{SNNC-DF}}$ is better than $R_{\text{[CF-S]-DF}}$ in general.

Remark 3.4. For rapidly changing channels it is advantageous to use independent inputs so all nodes can use the same encoder for all channel states. If X_1 and X_2 in the above example are independent, there is no need to use block Markov coding (BMC). However, we need to use two backward (or sliding window) decoders to recover the rates (3.57). See Appendix A.3.

Remark 3.5. How to perform DF for multiple sources is not obvious. Consider again the three node network in Fig. 3.3, but now every node wishes to send a message to the other two nodes. How should one set up cooperation if all nodes may use DF? Such questions are worth addressing, since their answers will give insight on how to incorporate mixed strategies to boost system performance.

3.5. Gaussian Networks

We next consider additive white Gaussian noise (AWGN) networks. We use $X \sim \mathcal{CN}(\mu, \sigma^2)$ to denote a circularly symmetric complex Gaussian random variable X with mean μ and variance σ^2. Let $Z^K = Z_1 Z_2 \dots Z_K$ be a noise string whose symbols are i.i.d. and $Z_k \sim \mathcal{CN}(0,1)$ for all k. The channel output at node k is

$$Y_k = \left[\sum_{\substack{j=1 \\ j \neq k}}^{K} G_{jk} X_j \right] + Z_k \tag{3.61}$$

where the channel gain is

$$G_{jk} = \frac{H_{jk}}{\sqrt{d_{jk}^{\alpha}}} \tag{3.62}$$

and d_{jk} is the distance between nodes j and k, α is a path-loss exponent and H_{jk} is a complex fading random variable.

We consider two kinds of fading:

- ▷ No fading: H_{jk} is a constant and known at all nodes. We set $H_{jk} = 1$ for all $(j,k) \in \mathcal{K} \times \mathcal{K}$.
- ▷ Rayleigh fading: we have $H_{jk} \sim \mathcal{CN}(0,1)$. We assume that each destination node k knows G_{jk} for all $(j,k) \in \mathcal{K} \times \mathcal{K}$ and each relay node k knows G_{jk} for all $j \in \mathcal{K}$ and knows the statistics of all other $G_{jl}, (j,l) \in \mathcal{K} \times \mathcal{K}$. We focus on slow fading, i.e., all G_{jk} remain unchanged once chosen.

We avoid issues of power control by imposing a per-symbol power constraint $\mathrm{E}[|X_k|^2] \leq$

P_k. We choose the inputs to be Gaussian, i.e., $X_k \sim \mathcal{CN}(0, P_k)$, $k \in \mathcal{K}$.

In the following we give numerical examples for four different channels

- ▷ the relay channel;
- ▷ the two-relay channel;
- ▷ the multiple access relay channel (MARC);
- ▷ the two-way relay channel (TWRC).

We evaluate performance for no fading in terms of achievable rates (in bits per channel use) and for Rayleigh fading in terms of outage probability [34] for a target rate R_{tar}.

Relay node k chooses

$$\hat{Y}_k = Y_k + \hat{Z}_k \tag{3.63}$$

where $\hat{Z}_k \sim \mathcal{CN}(0, \hat{\sigma}_k^2)$. For the no fading case, relay node k numerically calculates the optimal $\hat{\sigma}_k^2$ for CF-S and SNNC, and the optimal binning rate $R_{k(\text{bin})}$ for CF-S, in order to maximize the rates. For DF, the source and relay nodes numerically calculate the power allocation for superposition coding that maximizes the rates. For the Rayleigh fading case, relay node k knows only the G_{jk}, $j \in \mathcal{K}$, but it can calculate the optimal $\hat{\sigma}_k^2$ and $R_{k(\text{bin})}$ based on the statistics of G_{jl}, for all $(j, l) \in \mathcal{K} \times \mathcal{K}$ so as to minimize the outage probability. For DF, the fraction of power that the source and relay nodes allocate for cooperation is calculated numerically based on the statistics of G_{jk}, for all $(j, k) \in \mathcal{K} \times \mathcal{K}$, to minimize the outage probability. Details of the derivations are given in Appendix A.4.

3.5.1. Relay Channels

The Gaussian relay channel (Fig. 3.5) has

$$\begin{aligned} Y_2 &= G_{12}X_1 + Z_2 \\ Y_3 &= G_{13}X_1 + G_{23}X_2 + Z_3 \end{aligned} \tag{3.64}$$

and source node 1 has a message destined for node 3.

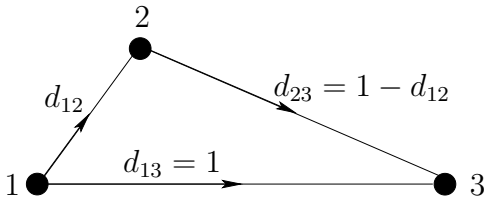


Figure 3.5.: A relay channel.

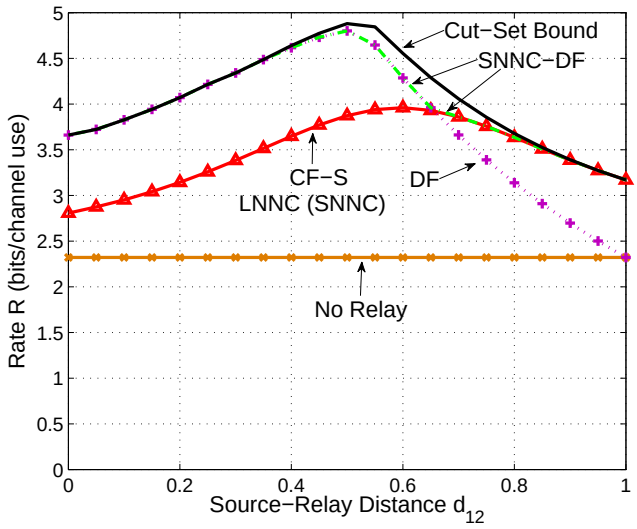


Figure 3.6.: Achievable rates R (in bits per channel use) for a relay channel with no fading.

No Fading

Fig. 3.5 depicts the geometry and Fig. 3.6 depicts the achievable rates as a function of d_{12} for $P_1 = 4, P_2 = 2$ and $\alpha = 3$. DF achieves rates close to capacity when the relay is close to the source while CF-S dominates as the relay moves towards the destination. For the relay channel, CF-S performs as well as SNNC (LNNC). SNNC-DF unifies the advantages of both SNNC and DF and achieves the best rates for all relay positions.

Slow Rayleigh Fading

Fig. 3.7 depicts the outage probabilities with $R_{\text{tar}} = 2$, $P_1 = 2P, P_2 = P$, $d_{12} = 0.3$, $d_{23} = 0.7$, $d_{13} = 1$ and $\alpha = 3$.

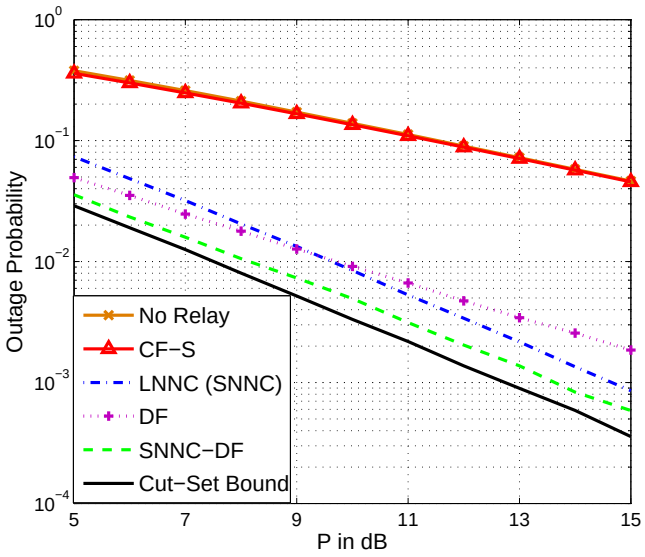


Figure 3.7.: Outage probabilities f or a relay channel with Rayleigh fading.

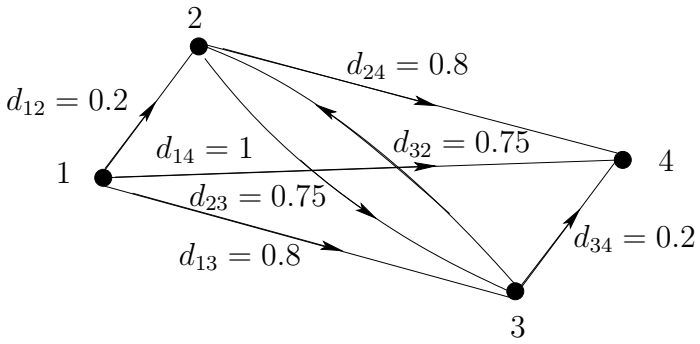


Figure 3.8.: A two-relay channel.

Over the entire power range CF-S gives the worst outage probability. This is because CF-S requires a reliable relay-destination link so that both the bin and quantization indices can be recovered. Both DF and SNNC improve on CF-S. DF performs better at low power while SNNC is better at high power. SNNC-DF has the relay decode if possible and perform QF otherwise, and gains 1 dB over SNNC and DF.

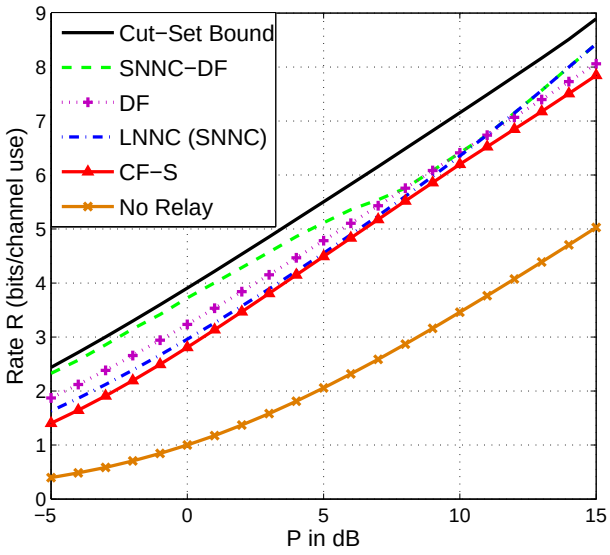


Figure 3.9.: Achievable rates R (in bits per channel use) for a TRC with no fading.

3.5.2. Two-Relay Channels

The Gaussian two-relay channel (Fig. 3.8) has

$$\begin{aligned} Y_2 &= G_{12}X_1 + G_{32}X_3 + Z_2 \\ Y_3 &= G_{13}X_1 + G_{23}X_2 + Z_3 \\ Y_4 &= G_{14}X_1 + G_{24}X_2 + G_{34}X_3 + Z_4 \end{aligned} \tag{3.65}$$

where the relay nodes 2 and 3 help node 1 transmit a message to node 4.

No Fading

Fig. 3.8 depicts the geometry and Fig. 3.9 depicts the achievable rates for $P_1 = P_2 = P_3 = P$ and $\alpha = 3$. The CF-S rates are the lowest over the entire power range. As expected, SNNC improves on CF-S. DF performs better than SNNC at low power but worse at high power. SNNC-DF achieves the best rates and exhibits reasonable rate and power gains over SNNC and DF for $P = -5$ dB to 5 dB. The gains are because in this power range SNNC-DF has relay 2 perform DF and relay 3 perform QF.

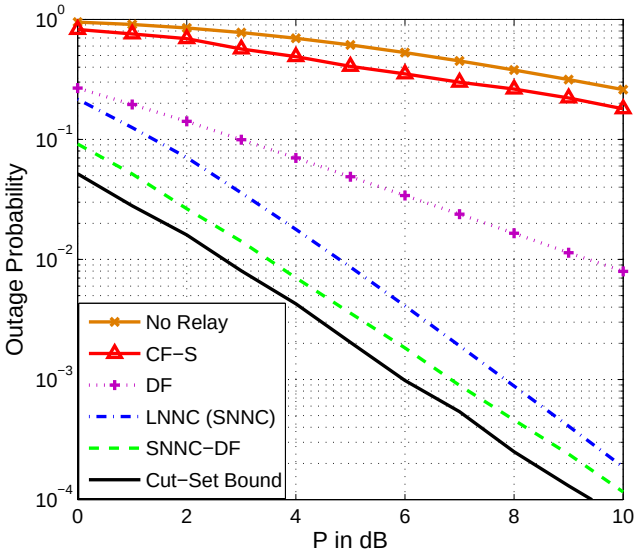


Figure 3.10.: Outage probabilities for a TRC with Rayleigh fading.

Slow Rayleigh Fading

Fig. 3.10 depicts the outage probabilities with $R_{\text{tar}} = 2$, $P_1 = P_2 = P_3 = P$, the geometry of Fig. 3.8 and $\alpha = 3$. CF-S gives the worst performance over the entire power range. This is because CF-S requires a reliable relay-destination link for *both* relays so that the bin and quantization indices for both relays can be decoded. DF provides better outage probabilities than CF-S but is worse than SNNC or LNNC, since it requires reliable decoding at both relays. SNNC-DF has the two relays decode if possible and perform QF otherwise and gains about 1 dB over LNNC (SNNC). In general, we expect larger gains of SNNC-DF over LNNC for networks with more relays.

3.5.3. Multiple Access Relay Channels

The Gaussian MARC (Fig. 3.11) has

$$\begin{aligned} Y_3 &= G_{13}X_1 + G_{23}X_2 + Z_3 \\ Y_4 &= G_{14}X_1 + G_{24}X_2 + G_{34}X_3 + Z_4 \end{aligned} \tag{3.66}$$

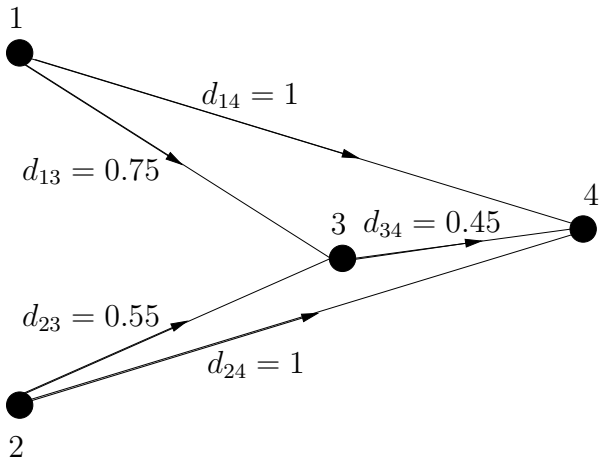


Figure 3.11.: A MARC.

and nodes 1 and 2 have messages destined for node 4.

No Fading

Fig. 3.11 depicts the geometry and Fig. 3.12 depicts the achievable rate regions for $P_1 = P_2 = P_3 = P$, $P = 15$ dB and $\alpha = 3$. The SNNC rate region includes the CF-S rate region. Through time-sharing, the SNNC-DF region is the convex hull of the union of DF and SNNC regions. SNNC-DF again improves on LNNC (or SNNC) and DF.

Slow Rayleigh Fading

Fig. 3.13 depicts the outage probabilities with $R_{\text{tar1}} = R_{\text{tar2}} = 1$, $P_1 = P_2 = P_3 = P$, $d_{13} = 0.3$, $d_{23} = 0.4$, $d_{14} = d_{24} = 1$, $d_{34} = 0.6$ and $\alpha = 3$. CF-S has the worst outage probability because it requires a reliable relay-destination link to decode the bin and quantization indices. DF has better outage probability than CF-S, while LNNC (or SNNC) improves on DF over the entire power range. SNNC-DF has the relay perform DF or QF depending on channel quality and gains 1 dB at low power and 0.5 dB at high power over SNNC.

Remark 3.6. The gain of SNNC-DF over SNNC is not very large at high power. This is because the MARC has one relay only. For networks with more relays we expect larger gains from SNNC-DF.

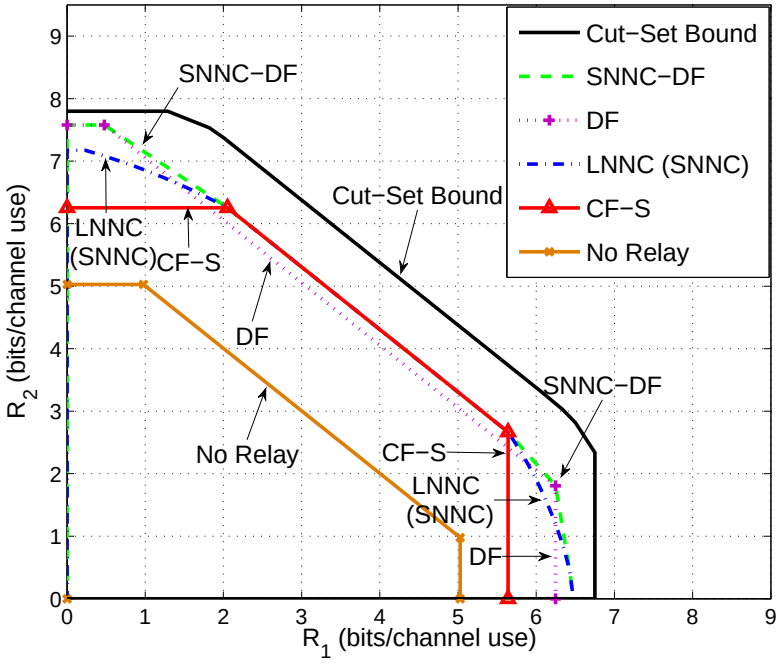


Figure 3.12.: Achievable rate regions for a MARC with no fading.

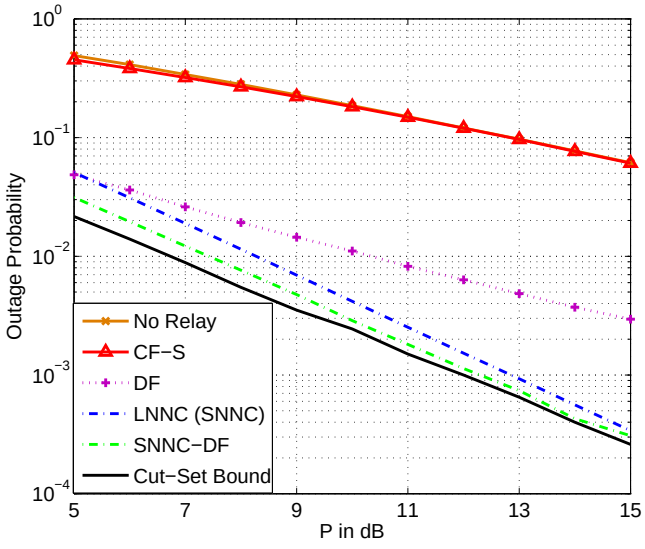


Figure 3.13.: Outage probabilities for a MARC with Rayleigh fading.

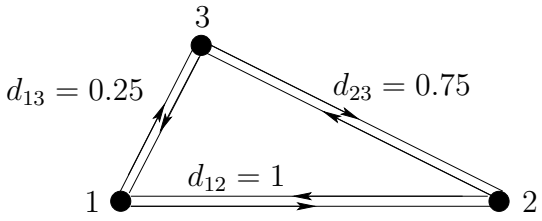


Figure 3.14.: A TWRC.

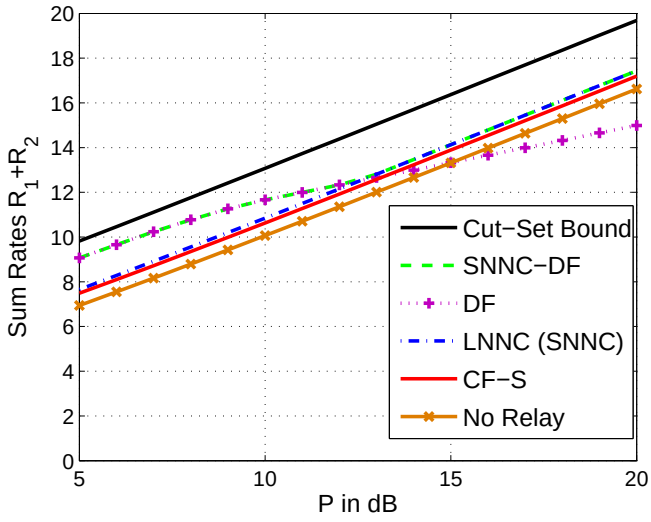


Figure 3.15.: Achievable sum rates (in bits per channel use) for a TWRC with no fading.

3.5.4. Two-Way Relay Channels

The Gaussian TWRC (Fig. 3.14) has

$$\begin{aligned} Y_1 &= G_{21}X_2 + G_{31}X_3 + Z_1 \\ Y_2 &= G_{12}X_1 + G_{32}X_3 + Z_2 \\ Y_3 &= G_{13}X_1 + G_{23}X_2 + Z_3 \end{aligned} \tag{3.67}$$

where nodes 1 and 2 exchange messages with the help of relay node 3.

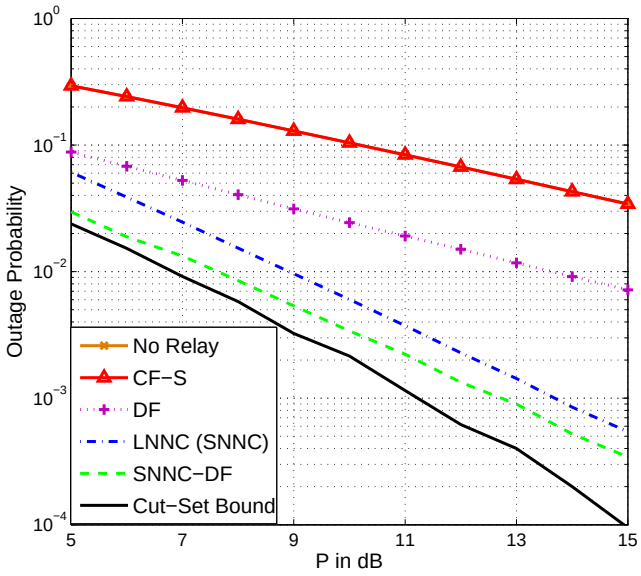


Figure 3.16.: Outage probabilities for a TWRC with Rayleigh fading.

No Fading

Fig. 3.14 depicts the geometry and Fig. 3.15 depicts the achievable sum rates for $P_1 = 5P, P_2 = 2P, P_3 = P$ and $\alpha = 3$. DF gives the best rates at low power while SNNC provides better rates at high power. The CF-S rates are slightly lower than the SNNC rates over the entire power range. SNNC-DF combines the advantages of SNNC and DF and achieves the best rates throughout.

Slow Rayleigh Fading

Fig. 3.16 depicts the outage probabilities with $R_{\text{tar1}} = 2$, $R_{\text{tar2}} = 1$, $P_1 = 5P$, $P_2 = 2P$, $P_3 = P$, the geometry of Fig. 3.14 and $\alpha = 3$. CF-S has the worst outage probability since it requires that both relay-destination links ($3-1$ and $3-2$) are reliable so that the bin and quantization indices can be recovered at both destinations 1 and 2. DF is better than CF-S, while LNNC (or SNNC) improves on DF. SNNC-DF lets the relay use DF or QF depending on the channel conditions and gains over about 2 dB at low power and 1 dB at high power over LNNC (or SNNC).

3.6. Concluding Remarks

SNNC with joint or backward decoding was shown to achieve the same rates as LNNC for multicasting multiple messages in memoryless networks. Although SNNC has extra constraints on the rates, these constraints give insight on the best decoding procedure. SNNC enables early decoding at nodes, and this enables the use of SNNC-DF. Numerical examples demonstrate that SNNC-DF shows reasonable gains as compared to DF, CF-S and LNNC in terms of rates and outage probabilities.

4

Multiple Access Relay Channel with Relay-Source Feedback

The multiple access relay channel (MARC) [35] has multiple sources communicate with one destination with the help of a relay node (Fig. 3.11). Results on coding strategies for the MARC were discussed in [23, 31, 35–38]. In all previous work, the information flow from the sources to the relay and the destination was considered. However, no information flow in the opposite direction (feedback from the relay or the destination to the sources) was considered. It turns out [39–41] that feedback can increase capacity of the multiple-access channel (MAC) [2, 3].

In this chapter, we incorporate feedback from the relay (Fig. 4.1) and establish an achievable rate region that includes the capacity region of the MARC without feedback [35]. We use a DF coding scheme where the sources cooperate with one another and with the relay due to the feedback. As a result, the relay can serve the sources simultaneously rather than separately [35–37] and higher rates are achieved. We also compare these results with the achievable SNNC rates developed in Chapter 3 with and without the

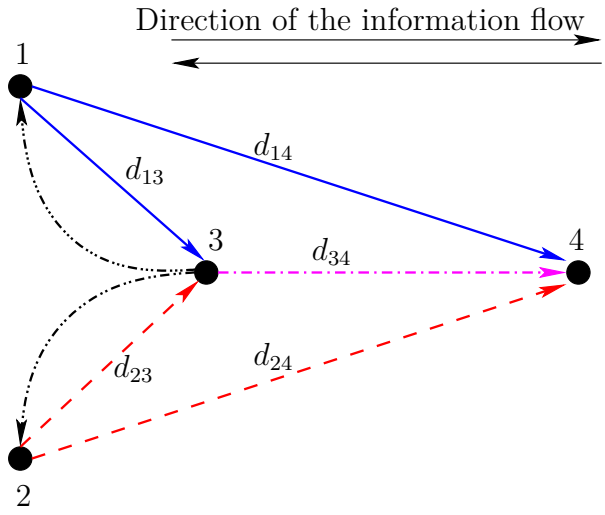


Figure 4.1.: A two-source MARC with relay-source feedback.

relay feedback and show that cooperation improves rates, and that network geometry influences the choice of coding strategies.

This chapter is organized as follows. In Section 4.1 we state the problem. In Section 4.2 we derive an achievable rate region for the MARC with relay-source feedback. Section 4.3 discusses Gaussian cases and shows that feedback can increase the capacity region of the MARC.

4.1. System Model

We use the notation developed in Sec. 3.1. We study the two-user *discrete memoryless* MARC with feedback (Fig. 4.1); the results extend naturally to multi-user and Gaussian cases. The node k, $k \in \{1, 2\}$, has a message W_k destined for node 4. The messages are statistically independent and W_k is uniformly distributed over $1, \ldots, 2^{nR_k}$, $k = 1, 2$, where 2^{nR_k} is taken to be a non-negative integer. The relay node 3 assists the communication by forwarding information to node 4 and feeding back its channel outputs to nodes 1 and 2. The channel is described by

$$P(y_3, y_4|x_1, x_2, x_3) \tag{4.1}$$

and the feedback link is assumed to be *noiseless*.

We define the following functions:

- ▷ n encoding functions $f_k^n = (f_{k1}, \ldots, f_{kn})$, $k = 1, 2, 3$, that generate channel inputs based on the local messages and past relay channel outputs

$$X_{ki} = f_{ki}(W_k, Y_3^{i-1}),\ i = \{1, \ldots, n\} \tag{4.2}$$

  where $W_3 = \emptyset$.

- ▷ One decoding function that puts out estimates of the messages

$$g_4(Y_4^n) = [\hat{W}_1, \hat{W}_2]. \tag{4.3}$$

The average error probability is

$$P_e^{(n)} = \Pr\left[(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\right]. \tag{4.4}$$

The rate pair (R_1, R_2) is said to be *achievable* for the MARC with feedback if for any ξ there is a sufficiently large n and some functions $\{f_k^n\}_{k=1}^3$ and g_4 such that $P_e^{(n)} \leq \xi$. The capacity region is the closure of the set of all achievable rate pairs (R_1, R_2).

4.2. Main Result and Proof

We have the following theorems.

Theorem 4.1. The capacity region of the MARC with feedback is contained in the set

$$\bigcup \left\{ (R_1, R_2) : \begin{array}{r} 0 \leq R_1 \leq I(X_1; Y_3 Y_4 | X_2 X_3) \\ 0 \leq R_2 \leq I(X_2; Y_3 Y_4 | X_1 X_3) \\ R_1 + R_2 \leq \min\{I(X_1 X_2; Y_3 Y_4 | X_3), I(X_1 X_2 X_3; Y_4)\} \end{array} \right\} \tag{4.5}$$

where the union is taken over all joint distributions $P_{X_1 X_2 X_3 Y_3 Y_4}$.

Proof: Theorem 1 is a special case of the cut-set bound [28, Theorem 15.10.1]. ■

Theorem 4.2. An achievable rate region of the MARC with relay-source feedback is

$$\bigcup \left\{ (R_1, R_2): \begin{array}{l} 0 \le R_1 \le I(X_1; Y_3|UX_2X_3) \\ 0 \le R_2 \le I(X_2; Y_3|UX_1X_3) \\ R_1 + R_2 \le \min\{I(X_1X_2; Y_3|UX_3), I(X_1X_2X_3; Y_4)\} \end{array} \right\} \tag{4.6}$$

where the union is taken over joint distributions that factor as

$$P(u) \prod_{k=1}^{3} P(x_k|u) P(y_3, y_4|x_1, x_2, x_3). \tag{4.7}$$

Proof: In every block, the sources and the relay cooperate through the feedback to help the receiver resolve the remaining uncertainty from the previous block. At the same time, the sources send fresh information to the relay and the receiver.

Random Code: Fix a distribution $P(u)\prod_{k=1}^{3} P(x_k|u)P(y_3, y_4|x_1, x_2, x_3)$. For each block $j = 1, \ldots, B$, generate $2^{n(R_1+R_2)}$ code words $\mathbf{u}_j(w_{j-1})$, $w_{j-1} = (w_{1(j-1)}, w_{2(j-1)})$, $w_{1(j-1)} = 1, \ldots, 2^{nR_1}$, $w_{2(j-1)} = 1, \ldots, 2^{nR_2}$, using $\prod_{i=1}^{n} P(u_{ji})$. For each w_{j-1}, generate 2^{nR_k} $\mathbf{x}_{kj}(w_{kj}|w_{j-1})$, $w_{kj} = 1, \ldots, 2^{nR_k}$, $k = 1, 2$, using $\prod_{i=1}^{n} P(x_{(kj)i}|u_{ji}(w_{j-1}))$ and generate an $\mathbf{x}_{3j}(w_{j-1})$ using $\prod_{i=1}^{n} P(x_{(3j)i}|u_{ji}(w_{j-1}))$. This defines the codebooks

$$\begin{aligned} \mathcal{C}_j = \{&\mathbf{u}_j(w_{j-1}), \mathbf{x}_{1j}(w_{1j}|w_{j-1}), \mathbf{x}_{2j}(w_{2j}|w_{j-1}), \mathbf{x}_{3j}(w_{j-1}), \\ &w_{j-1} = (w_{1(j-1)}, w_{2(j-1)}), w_{1(j-1)} = 1, \ldots, 2^{nR_1}, w_{2(j-1)} = 1, \ldots, 2^{nR_2}, \\ &w_{1j} = 1, \ldots, 2^{nR_1}, w_{2j} = 1, \ldots, 2^{nR_2}\Big\} \end{aligned} \tag{4.8}$$

for $j = 1, \ldots, B$.

Encoding: In blocks $j = 1, \ldots, B$, nodes 1 and 2 send $\mathbf{x}_{1j}(w_{1j}|w_{j-1})$ and $\mathbf{x}_{2j}(w_{2j}|w_{j-1})$, and node 3 sends $\mathbf{x}_{3j}(w_{j-1})$ where $w_0 = w_{1B} = w_{2B} = 1$. The coding scheme is depicted in Table 4.1.

Decoding at the relay: For block $j = 1, \ldots, B$, knowing w_{j-1} the relay node 3 puts out $(\hat{w}_{1j}, \hat{w}_{2j})$ if there is a unique pair $(\hat{w}_{1j}, \hat{w}_{2j})$ satisfying the typicality check

$$(\mathbf{x}_1(\hat{w}_{1j}|w_{j-1}), \mathbf{x}_2(\hat{w}_{2j}|w_{j-1}), \mathbf{u}(w_{j-1}), \mathbf{x}_{3j}(w_{j-1}), \mathbf{y}_{3j}) \in \mathcal{T}_\epsilon^n(P_{UX_1X_2X_3Y_3}). \tag{4.9}$$

Block	1	2	3	$\cdots$	B
U	$\mathbf{u}_1(1)$	$\mathbf{u}_2(w_1)$	$\mathbf{u}_3(w_2)$	$\cdots$	$\mathbf{u}_B(w_{B-1})$
X_1	$\mathbf{x}_{11}(w_{11}\|1)$	$\mathbf{x}_{12}(w_{12}\|w_1)$	$\mathbf{x}_{13}(w_{13}\|w_2)$	$\cdots$	$\mathbf{x}_{1B}(1\|w_{B-1})$
X_2	$\mathbf{x}_{21}(w_{21}\|1)$	$\mathbf{x}_{22}(w_{22}\|w_1)$	$\mathbf{x}_{23}(w_{23}\|w_2)$	$\cdots$	$\mathbf{x}_{2B}(1\|w_{B-1})$
X_3	$\mathbf{x}_{31}(1)$	$\mathbf{x}_{32}(w_1)$	$\mathbf{x}_{33}(w_2)$	$\cdots$	$\mathbf{x}_{3B}(w_{B-1})$

Table 4.1.: Coding scheme for MARC with feedback.

Otherwise it puts out $(\hat{w}_{1j}, \hat{w}_{2j}) = (1, 1)$. Node 3 can decode reliably as $n \to \infty$ if (see [20])

$$\begin{aligned} R_1 &< I(X_1; Y_3|UX_2X_3) \\ R_2 &< I(X_2; Y_3|UX_1X_3) \\ R_1 + R_2 &< I(X_1X_2; Y_3|UX_3). \end{aligned} \tag{4.10}$$

Decoding at the sources: For block $j = 1, \ldots, B-1$, assuming knowledge of w_{j-1}, source 1 puts out $\hat{w}_{2j}$ if there is a unique $\hat{w}_{2j}$ satisfying the typicality check

$$(\mathbf{x}_1(w_{1j}|w_{j-1}), \mathbf{x}_2(\hat{w}_{2j}|w_{j-1}), \mathbf{u}(w_{j-1}), \mathbf{x}_3(w_{j-1}), \mathbf{y}_{3j}) \in \mathcal{T}_\epsilon^n(P_{UX_1X_2X_3Y_3}). \tag{4.11}$$

Otherwise it puts out $\hat{w}_{2j} = 1$. Node 1 can reliably decode w_{2j} if (see [20])

$$R_2 < I(X_2; Y_3|UX_1X_3) \tag{4.12}$$

and n is sufficiently large. Similarly, source 2 can reliably decode w_{1j} if

$$R_1 < I(X_1; Y_3|UX_2X_3) \tag{4.13}$$

and n is sufficiently large. Both sources then calculate $w_j = (w_{1j}, w_{2j})$ for the cooperation in block $j+1$. The constraints (4.12)-(4.13) are already included in (4.10) and do not further constrain the rates. This is because the sources observe the relay's channel outputs and have knowledge about their own messages.

Backward decoding at the receiver: For block $j = B, \ldots, 1$, assuming correct decoding of (w_{1j}, w_{2j}), the receiver puts out $\hat{w}_{j-1}$ if there is a unique $\hat{w}_{j-1}$ satisfying the typicality

check

$$(\mathbf{x}_1(w_{1j}|\hat{w}_{j-1}), \mathbf{x}_2(w_{2j}|\hat{w}_{j-1}), \mathbf{x}_3(\hat{w}_{j-1}), \mathbf{y}_{4j}) \in \mathcal{T}_\epsilon^n(P_{X_1X_2X_3Y_4}). \tag{4.14}$$

Otherwise it puts out $\hat{w}_{j-1} = 1$. The receiver can decode reliably as $n \to \infty$ if (see [20])

$$R_1 + R_2 < I(X_1X_2X_3; Y_4) \tag{4.15}$$

which yields the reliable estimate $w_{j-1} = (w_{1(j-1)}, w_{2(j-1)})$. Continuing in this way, the receiver successively finds all (w_{1j}, w_{2j}). This completes the proof. ■

Remark 4.1. In (4.2), instead of three destination bounds, we have only one since the receiver needs only one joint decoding to recover both sources' messages. Without feedback, a common approach is to use successive interference cancellation, i.e., first decode one source's message while treat the other source's message as noise. The resulting rate region is smaller and time-sharing between different decoding orders is useful.

Remark 4.2. The rate region (4.6) is achieved with *backward* decoding which incurs a substantial decoding delay. With offset encoding [31] we may enable sliding window decoding to achieve the same region as in (4.6) and enjoy a reduced delay.

Corollary 4.3. An achievable SNNC rate region of the MARC is

$$\bigcup \left\{ (R_1, R_2) : \begin{array}{l} 0 \le R_1 \le I(X_1; \hat{Y}_3Y_4|X_2X_3) \\ 0 \le R_1 \le I(X_1X_3; Y_4|X_2) - I(\hat{Y}_3; Y_3|X_1X_2X_3Y_4) \\ 0 \le R_2 \le I(X_2; \hat{Y}_3Y_4|X_1X_3) \\ 0 \le R_2 \le I(X_2X_3; Y_4|X_1) - I(\hat{Y}_3; Y_3|X_1X_2X_3Y_4) \\ R_1 + R_2 \le I(X_1X_2; \hat{Y}_3Y_4|X_3) \\ R_1 + R_2 \le I(X_1X_2X_3; Y_4) - I(\hat{Y}_3; Y_3|X_1X_2X_3Y_4) \end{array} \right\} \tag{4.16}$$

Proof: Apply Theorem 3.1. ■

Corollary 4.4. An achievable SNNC rate region of the MARC where the sources use the feedback from the relay is

$$\bigcup \left\{ (R_1, R_2) : \begin{array}{l} 0 \le R_1 \le I(X_1; \hat{Y}_2\hat{Y}_3Y_4|X_2X_3) - I(\hat{Y}_1; Y_3|X_1X_2X_3\hat{Y}_2\hat{Y}_3Y_4) \\ 0 \le R_1 \le I(X_1X_3; \hat{Y}_2Y_4|X_2) - I(\hat{Y}_1\hat{Y}_3; Y_3|X_1X_2X_3\hat{Y}_2Y_4) \\ 0 \le R_2 \le I(X_2; \hat{Y}_1\hat{Y}_3Y_4|X_1X_3) - I(\hat{Y}_2; Y_3|X_1X_2X_3\hat{Y}_1\hat{Y}_3Y_4) \\ 0 \le R_2 \le I(X_2X_3; \hat{Y}_1Y_4|X_1) - I(\hat{Y}_2\hat{Y}_3; Y_3|X_1X_2X_3\hat{Y}_1Y_4) \\ R_1 + R_2 \le I(X_1X_2; \hat{Y}_3Y_4|X_3) - I(\hat{Y}_1\hat{Y}_2; Y_3|X_1X_2X_3\hat{Y}_3Y_4) \\ R_1 + R_2 \le I(X_1X_2X_3; Y_4) - I(\hat{Y}_1\hat{Y}_2\hat{Y}_3; Y_3|X_1X_2X_3Y_4) \end{array} \right\} \tag{4.17}$$

Proof: Apply Theorem 3.1. ■

4.3. The Gaussian Case

The Gaussian MARC with feedback from the relay has

$$\begin{aligned} Y_3 &= G_{13}X_1 + G_{23}X_2 + Z_3 \\ Y_4 &= G_{14}X_1 + G_{24}X_2 + G_{34}X_3 + Z_4 \end{aligned} \tag{4.18}$$

where $Z_3 \sim \mathcal{CN}(0,1)$, $Z_4 \sim \mathcal{CN}(0,1)$ and Z_3 and Z_4 are statistically independent, and the channel gain is

$$G_{jk} = \frac{1}{\sqrt{d_{jk}^{\alpha}}} \tag{4.19}$$

where d_{jk} is the distance between nodes j and k, and α is a path-loss exponent. We impose a per-symbol power constraint $\mathrm{E}[|X_k|^2] \le P_k$ and choose the inputs to be

$$\begin{aligned} X_1 &= \sqrt{\alpha_1 P_1} \cdot U + \sqrt{\bar{\alpha}_1 P_1} \cdot X_1' \\ X_2 &= \sqrt{\alpha_2 P_2} \cdot U + \sqrt{\bar{\alpha}_2 P_2} \cdot X_2' \\ X_3 &= \sqrt{P_3} \cdot U \end{aligned} \tag{4.20}$$

where U, X_1', X_2' are independent Gaussian $\mathcal{CN}(0,1)$, $0 \le \alpha_k \le 1$ and $\bar{\alpha}_k = 1 - \alpha_k$, $k = 1, 2$. In blocks j, $j = 1 \ldots, B$, the two sources devote a fraction α_k of the power, $k = 1, 2$, to resolving the residual uncertainty from block $j-1$ and the remaining fraction $\bar{\alpha}_k$ to sending fresh information. The residual uncertainty can be resolved with an effective power

$$P_{\text{eff}} = \left(G_{14}\sqrt{\alpha_1 P_1} + G_{24}\sqrt{\alpha_2 P_2} + G_{34}\sqrt{P_3} \right)^2 \tag{4.21}$$

while it in [35] can be resolved only with an effective power

$$P'_{\text{eff}} = \left(G_{14}\sqrt{\alpha_1 P_1} + G_{34}\sqrt{\alpha_3 P_3} \right)^2 + \left(G_{24}\sqrt{\alpha_2 P_2} + G_{34}\sqrt{\bar{\alpha}_3 P_3} \right)^2 \tag{4.22}$$

where $0 \le \alpha_3 \le 1$, $\bar{\alpha}_3 = 1 - \alpha_3$, because the relay splits its power to serve the sources separately rather than simultaneously. Let $C(x) = \log_2(1+x)$, $x \ge 0$. Referring to Theorem 4.2, an achievable rate region of the Gaussian MARC with feedback is

$$\bigcup \left\{ (R_1, R_2) : \begin{array}{l} 0 \le R_1 \le C\left(\bar{\alpha}_1 G_{13}^2 P_1\right) \\ 0 \le R_2 \le C\left(\bar{\alpha}_2 G_{23}^2 P_2\right) \\ R_1 + R_2 \le C\left(\bar{\alpha}_1 G_{13}^2 P_1 + \bar{\alpha}_2 G_{23}^2 P_2\right) \\ R_1 + R_2 \le C\left(\bar{\alpha}_1 G_{14}^2 P_1 + \bar{\alpha}_2 G_{24}^2 P_2 + P_{\text{eff}}\right) \end{array} \right\} \tag{4.23}$$

where the union is over α_k satisfying $0 \le \alpha_k \le 1, k = 1, 2$. Referring to Corollary 4.3, an achievable SNNC rate region of the Gaussian MARC is

$$\bigcup \left\{ (R_1, R_2) : \begin{array}{l} 0 \le R_1 \le C\left(\frac{G_{13}^2 P_1}{1+\hat{\sigma}_3^2} + G_{14}^2 P_1\right) \\ 0 \le R_1 \le C\left(G_{14}^2 P_1 + G_{34}^2 P_3\right) - C\left(\frac{1}{\hat{\sigma}_3^2}\right) \\ 0 \le R_2 \le C\left(\frac{G_{23}^2 P_2}{1+\hat{\sigma}_3^2} + G_{24}^2 P_2\right) \\ 0 \le R_2 \le C\left(G_{24}^2 P_2 + G_{34}^2 P_3\right) - C\left(\frac{1}{\hat{\sigma}_3^2}\right) \\ R_1 + R_2 \le C\left(G_{14}^2 P_1 + G_{24}^2 P_2 + \frac{G_{13}^2 P_1 + G_{23}^2 P_2 + P_1 P_2 (G_{13} G_{24} - G_{14} G_{23})^2}{1+\hat{\sigma}_3^2}\right) \\ R_1 + R_2 \le C\left(G_{14}^2 P_1 + G_{24}^2 P_2 + G_{34}^2 P_3\right) - C\left(\frac{1}{\hat{\sigma}_3^2}\right) \end{array} \right\} \tag{4.24}$$

where the union is over $\hat{\sigma}_3^2$ satisfying $\hat{\sigma}_3^2 > \frac{1}{G_{34}^2 P_3}$. Referring to Corollary 4.4, an achievable SNNC rate region of the Gaussian MARC with feedback is

$$\bigcup \left\{ (R_1, R_2): \begin{array}{l} 0 \le R_1 \le C\left(\frac{G_{13}^2 P_1(\hat{\sigma}_2^2+\hat{\sigma}_3^2)}{\hat{\sigma}_2^2+\hat{\sigma}_3^2+\hat{\sigma}_2^2\hat{\sigma}_3^2} + G_{14}^2 P_1\right) - C\left(\frac{\hat{\sigma}_2^2\hat{\sigma}_3^2}{\hat{\sigma}_1^2(\hat{\sigma}_2^2+\hat{\sigma}_3^2)+\hat{\sigma}_1^2\hat{\sigma}_2^2\hat{\sigma}_3^2}\right) \\ 0 \le R_1 \le C\left(\frac{G_{13}^2 P_1(1+G_{34}^2 P_3)}{1+\hat{\sigma}_2^2} + G_{14}^2 P_1 + G_{34}^2 P_3\right) - C\left(\frac{\hat{\sigma}_2^2(\hat{\sigma}_1^2+\hat{\sigma}_3^2)}{\hat{\sigma}_1^2\hat{\sigma}_3^2(1+\hat{\sigma}_2^2)}\right) \\ 0 \le R_2 \le C\left(\frac{G_{23}^2 P_2(\hat{\sigma}_1^2+\hat{\sigma}_3^2)}{\hat{\sigma}_1^2+\hat{\sigma}_3^2+\hat{\sigma}_1^2\hat{\sigma}_3^2} + G_{24}^2 P_2\right) - C\left(\frac{\hat{\sigma}_1^2\hat{\sigma}_3^2}{\hat{\sigma}_2^2(\hat{\sigma}_1^2+\hat{\sigma}_3^2)+\hat{\sigma}_1^2\hat{\sigma}_2^2\hat{\sigma}_3^2}\right) \\ 0 \le R_2 \le C\left(\frac{G_{23}^2 P_2(1+G_{34}^2 P_3)}{1+\hat{\sigma}_1^2} + G_{24}^2 P_2 + G_{34}^2 P_3\right) - C\left(\frac{\hat{\sigma}_1^2(\hat{\sigma}_2^2+\hat{\sigma}_3^2)}{\hat{\sigma}_2^2\hat{\sigma}_3^2(1+\hat{\sigma}_1^2)}\right) \\ R_1 + R_2 \le C\left(\frac{G_{13}^2 P_1+G_{23}^2 P_2+P_1P_2(G_{13}G_{24}-G_{14}G_{23})^2}{1+\hat{\sigma}_3^2} + G_{14}^2 P_1 + G_{24}^2 P_2\right) \\ \qquad\qquad -C\left(\frac{\hat{\sigma}_3^2(\hat{\sigma}_1^2+\hat{\sigma}_2^2)}{\hat{\sigma}_1^2\hat{\sigma}_2^2(1+\hat{\sigma}_3^2)}\right) \\ R_1 + R_2 \le C\left(G_{14}^2 P_1 + G_{24}^2 P_2 + G_{34}^2 P_3\right) - C\left(\frac{1}{\hat{\sigma}_1^2} + \frac{1}{\hat{\sigma}_2^2} + \frac{1}{\hat{\sigma}_3^2}\right) \end{array} \right\} \tag{4.25}$$

where the union is over all $(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2)$ such that (R_1, R_2) is non-negative.

Remark 4.3. The rate region in (4.25) recovers that in (4.24) when both $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ go to infinity, i.e., both nodes 1 and 2 quantize so coarsely that $\hat{Y}_1$ and $\hat{Y}_2$ become statistically independent of X_1, X_2 and Y_3. This implies that better rates can be achieved if all nodes in the network cooperate by quantizing the feedback (channel outputs) and forwarding extra information to the other nodes. The cooperation incurs no additional cost in terms of power but requires more sophisticated encoders and decoders. More importantly, it relies on the non-selfish behavior of the nodes. For example, in the MARC case, nodes 1 and 2 could just ignore the feedback Y_3, because they are not obliged to help each other. But when they do, they both gain.

Fig. 4.2 depicts the rate regions of the MARC with and without feedback for $P_1 = P_2 = P_3 = 15$ dB, $d_{13} = d_{23} = 0.2$, $d_{14} = d_{24} = 1$, $d_{34} = 0.8$ and $\alpha = 3$. The DF region with feedback is larger than that without feedback [35]. Also, the DF region with feedback includes the cut-set region without feedback which confirms that feedback can indeed increase the capacity of the MARC. Further, the SNNC region with feedback includes that without feedback. This shows that cooperation helps. Note that the DF region with feedback is larger than the SNNC regions with and without feedback suggesting that DF is preferable for a network geometry where the relay is close to the sources. Fig. 4.3 depicts the rate regions for $P_1 = P_2 = P_3 = 15$ dB, $d_{13} = 0.7$,

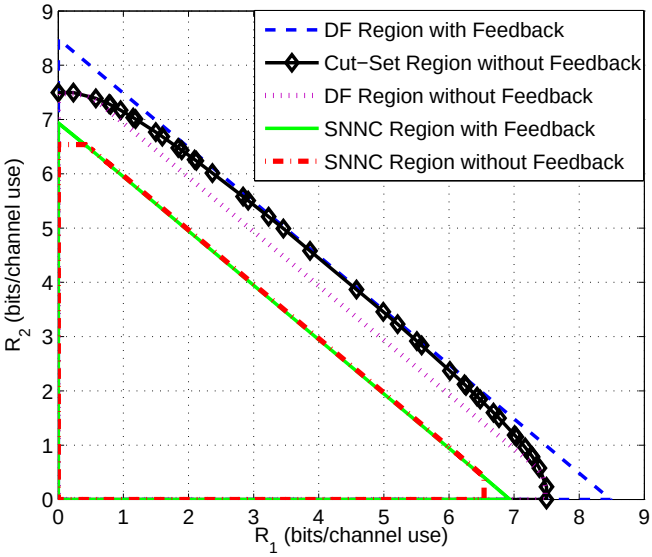


Figure 4.2.: Rate regions for MARC with and without feedback for $P_1 = P_2 = P_3 = 15$ dB, $d_{13} = d_{23} = 0.2$, $d_{14} = d_{24} = 1$, $d_{34} = 0.8$ and $\alpha = 3$.

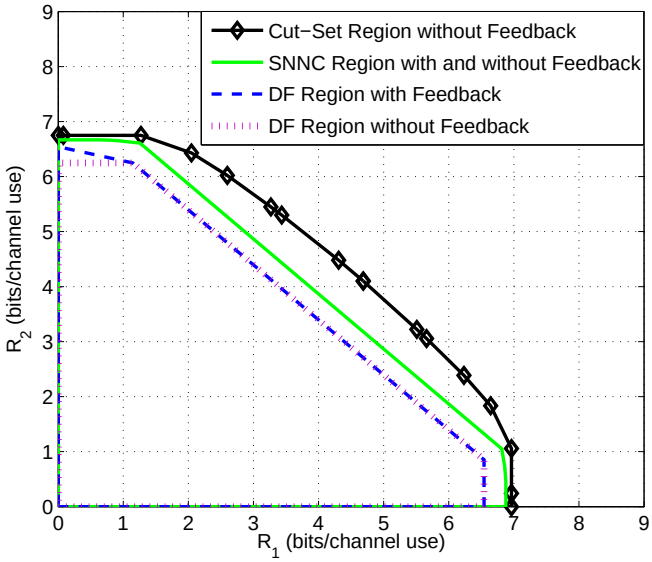


Figure 4.3.: Rate regions for MARC with and without feedback for $P_1 = P_2 = P_3 = 15$ dB, $d_{13} = 0.7$, $d_{23} = 0.75$, $d_{14} = d_{24} = 1$, $d_{34} = 0.25$ and $\alpha = 3$.

$d_{23} = 0.75$, $d_{14} = d_{24} = 1$, $d_{34} = 0.25$ and $\alpha = 3$. The cut-set region without feedback is the largest. The DF region with feedback is larger than the DF region without feedback, but is included by the SNNC regions with and without feedback which overlap in this case. Thus, for a geometry where the relay is close to the destination, SNNC provides better rates. In any case, DF with feedback performs at least as well as DF without feedback.

5

Resolvability

What is the minimal rate needed to generate a good approximation of a target distribution with respect to some distance measure? Wyner considered such a problem and characterized the smallest rate needed to approximate a *product* distribution accurately when using the *normalized* informational divergence as the distance measure between two distributions. The smallest rate is a Shannon mutual information [16]. Han-Verdú [17] showed that the same rate is necessary and sufficient to generate distributions arbitrarily close to an *information stable* distribution in terms of *variational distance*. Note that normalized informational divergence and variational distance are not necessarily larger or smaller than the other.

The main contributions of this chapter are to develop a simple and direct proof to show that the minimal rate needed to make the *unnormalized* informational divergence between a target product distribution and the approximating distribution arbitrarily small is the same Shannon mutual information as in [16, 17] and we extend the proof to cases where the encoder has a non-uniform input distribution. Our result implies results in [16] and [17] when restricting attention to product distributions (in particular

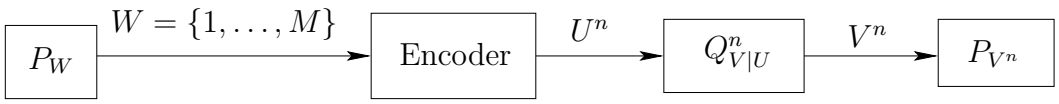


Figure 5.1.: Coding problem with the goal of making $P_{V^n} \approx Q_V^n$.

Theorem 6.3 in [16] and Theorem 4 in [17]). We remark that Hayashi developed closely related theory via Gallager's error exponent in [42] and Bloch and Kliewer considered non-uniform distributions for secrecy in [43]. We also refer to results by Csiszar [44, p. 44, bottom] who treats strong secrecy by showing that a variational distance exhibits an exponential behavior with block length n [44, Prop. 2]. This result implies that an unnormalized mutual information expression can be made small with growing n via [44, Lemma 1]. Finally, Winter states such a result in [45] but provides no proof.

This chapter is organized as follows. In Section 5.1, we state the problem. In Section 5.2 we state and prove the main result. Section 5.3 discusses related work and extensions.

5.1. System Model

Consider the system depicted in Fig. 5.1. The random variable W is *uniformly* distributed over $\{1, \ldots, M\}$, $M = 2^{nR}$, and is encoded to the sequence

$$U^n = f(W). \tag{5.1}$$

The sequence V^n is generated from U^n through a memoryless channel $Q^n_{V|U}$ and has distribution P_{V^n}. A rate R is *achievable* if for any $\xi > 0$ there is a sufficiently large n and an encoder such that the informational divergence

$$D(P_{V^n}||Q_V^n) = \sum_{v^n \in \mathrm{supp}(P_{V^n})} P(v^n) \log \frac{P(v^n)}{Q_V^n(v^n)} \tag{5.2}$$

is less than ξ. We wish to determine the smallest achievable rate.

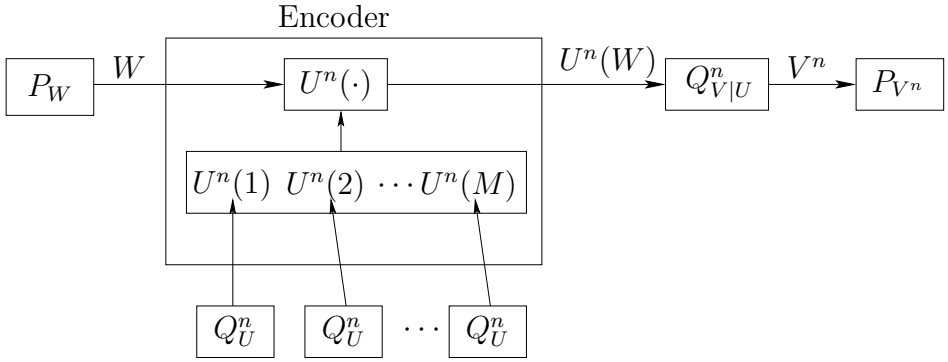


Figure 5.2.: The random coding experiment.

5.2. Main Result and Proof

Theorem 5.1. For a given target distribution Q_V, the rate R is achievable if $R > I(V;U)$, where $I(V;U)$ is calculated with some joint distribution Q_{UV} that has marginal Q_V and $|\text{supp}(Q_U)| \leq |\mathcal{V}|$. The rate R is not achievable if $R < I(V;U)$ for all Q_{UV} with $|\text{supp}(Q_U)| \leq |\mathcal{V}|$.

We provide two proofs, one with Shannon's typicality argument and the other with Gallager's error exponent [46] where we extend results in [42]. Consider the random coding experiment in Fig. 5.2. Suppose U and V have finite alphabets $\mathcal{U}$ and $\mathcal{V}$, respectively. Let Q_{UV} be a probability distribution with marginals Q_U and Q_V. Let $U^nV^n \sim Q^n_{UV}$, i.e., for any $u^n \in \mathcal{U}^n$, $v^n \in \mathcal{V}^n$ we have

$$Q(u^n, v^n) = \prod_{i=1}^{n} Q_{UV}(u_i, v_i) = Q^n_{UV}(u^n, v^n) \tag{5.3}$$

$$Q(v^n|u^n) = \prod_{i=1}^{n} Q_{V|U}(v_i|u_i) = Q^n_{V|U}(v^n|u^n). \tag{5.4}$$

Let $\widetilde{\mathcal{C}} = \{U^n(1), U^n(2), \ldots, U^n(M)\}$ be a random codebook, where the $U^n(w), w = 1, \ldots, M$, are generated in an i.i.d. manner using Q^n_U and occur with probability $\frac{1}{M}$. V^n is generated from $U^n(W)$ through the channel $Q^n_{V|U}$ (see Fig. 5.2) and we have

$$P(v^n|\widetilde{\mathcal{C}}) = \sum_{w=1}^{M} \frac{1}{M} \cdot Q^n_{V|U}(v^n|U^n(w)). \tag{5.5}$$

Note that if for a v^n we have

$$Q_V^n(v^n) = \sum_{u^n \in \mathrm{supp}(Q_U^n)} Q_U^n(u^n) Q_{V|U}^n(v^n|u^n) = 0 \tag{5.6}$$

then we have

$$Q_{V|U}^n(v^n|u^n) = 0, \text{ for all } u^n \in \mathrm{supp}(Q_U^n). \tag{5.7}$$

This means $P(v^n|\widetilde{\mathcal{C}}) = 0$ and $\mathrm{supp}(P_{V^n|\widetilde{\mathcal{C}}}) \subseteq \mathrm{supp}(Q_V^n)$ so that $D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n) < \infty$. We further have

$$\mathrm{E}\left[\frac{Q_{V|U}^n(v^n|U^n)}{Q_V^n(v^n)}\right] = \sum_{u^n} Q_U^n(u^n) \cdot \frac{Q_{V|U}^n(v^n|u^n)}{Q_V^n(v^n)} = 1. \tag{5.8}$$

5.2.1. Typicality Argument

The informational divergence averaged over W, $\widetilde{\mathcal{C}}$ and V^n is (recall that $P(w) = \frac{1}{M}, w = 1, \ldots, M$):

$$\begin{aligned}
\mathrm{E}[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)] &\overset{(a)}{=} \mathrm{E}\left[\log \frac{\sum_{j=1}^M \frac{1}{M} \cdot Q_{V^n|U^n}(V^n|U^n(j))}{Q_V^n(V^n)}\right] \\
&= \sum_{w=1}^M \frac{1}{M} \cdot \mathrm{E}\left[\log \frac{\sum_{j=1}^M Q_{V|U}^n(V^n|U^n(j))}{M Q_V^n(V^n)} \middle| W = w\right] \\
&\overset{(b)}{\leq} \sum_{w=1}^M \frac{1}{M} \cdot \mathrm{E}\left[\log \left(\frac{Q_{V|U}^n(V^n|U^n(w))}{M Q_V^n(V^n)} + \frac{M-1}{M}\right) \middle| W = w\right] \\
&\leq \sum_{w=1}^M \frac{1}{M} \cdot \mathrm{E}\left[\log \left(\frac{Q_{V|U}^n(V^n|U^n(w))}{M Q_V^n(V^n)} + 1\right) \middle| W = w\right] \\
&\overset{(c)}{=} \mathrm{E}\left[\log \left(\frac{Q_{V|U}^n(V^n|U^n)}{M \cdot Q_V^n(V^n)} + 1\right)\right]
\end{aligned} \tag{5.9}$$

where

(a) follows by taking the expectation over W, V^n and $U^n(1), \ldots, U^n(M)$;

(b) follows by the concavity of the logarithm and Jensen's inequality applied to the expectation over the $U^n(j), j \neq w$, and by using (5.8);

(c) follows by choosing $U^n V^n \sim Q_{UV}^n$.

Alternatively, we can make the steps (5.9) more explicit:

$$
\begin{aligned}
&\mathrm{E}[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)] \\
&\overset{(a)}{=} \sum_{u^n(1)} \cdots \sum_{u^n(M)} \prod_{k=1}^{M} Q_U^n(u^n(k)) \sum_{v^n} \sum_{w=1}^{M} \frac{1}{M} \cdot Q_{V|U}^n(v^n|u^n(w)) \left[\log \frac{\sum_{j=1}^{M} Q_{V|U}^n(v^n|u^n(j))}{M \cdot Q_V^n(v^n)} \right] \\
&= \sum_{w=1}^{M} \frac{1}{M} \sum_{v^n} \sum_{u^n(w)} Q_{UV}^n(u^n(w), v^n) \sum_{k \neq w}^{M} \sum_{u^n(k)} \prod_{l \neq w}^{M} Q_U^n(u^n(l)) \left[\log \frac{\sum_{j=1}^{M} Q_{V|U}^n(v^n|u^n(j))}{M \cdot Q_V^n(v^n)} \right] \\
&\overset{(b)}{\leq} \sum_{w=1}^{M} \frac{1}{M} \sum_{v^n} \sum_{u^n(w)} Q_{UV}^n(u^n(w), v^n) \left[\log \left(\frac{Q_{V|U}^n(v^n|u^n(w))}{M \cdot Q_V^n(v^n)} + \sum_{j \neq w}^{M} \sum_{u^n(j)} \left[\frac{Q_{UV}^n(u^n(j), v^n)}{M \cdot Q_V^n(v^n)} \right] \right) \right] \\
&= \sum_{w=1}^{M} \frac{1}{M} \sum_{v^n} \sum_{u^n(w)} Q_{UV}^n(u^n(w), v^n) \left[\log \left(\frac{Q_{V|U}^n(v^n|u^n(w))}{M \cdot Q_V^n(v^n)} + \frac{M-1}{M} \right) \right] \\
&\leq \sum_{w=1}^{M} \frac{1}{M} \sum_{v^n} \sum_{u^n(w)} Q_{UV}^n(u^n(w), v^n) \left[\log \left(\frac{Q_{V|U}^n(v^n|u^n(w))}{M \cdot Q_V^n(v^n)} + 1 \right) \right] \\
&\overset{(c)}{=} \mathrm{E} \left[\log \left(\frac{Q_{V|U}^n(V^n|U^n)}{M \cdot Q_V^n(V^n)} + 1 \right) \right]. \qquad (5.10)
\end{aligned}
$$

We remark that the identity after (a) is valid for $M = 1$ by interpreting the empty sum followed by an empty product to be 1. We may write (5.9) or (5.10) as

$$
\mathrm{E} \left[\log \left(\frac{Q_{V|U}^n(V^n|U^n)}{M \cdot Q_V^n(V^n)} + 1 \right) \right] = d_1 + d_2 \qquad (5.11)
$$

where

$$
\begin{aligned}
d_1 &= \sum_{(u^n, v^n) \in \mathcal{T}_\epsilon^n(Q_{UV})} Q(u^n, v^n) \log \left(\frac{Q(v^n|u^n)}{M \cdot Q(v^n)} + 1 \right) \\
d_2 &= \sum_{\substack{(u^n, v^n) \notin \mathcal{T}_\epsilon^n(Q_{UV}) \\ (u^n, v^n) \in \mathrm{supp}(Q_{UV}^n)}} Q(u^n, v^n) \log \left(\frac{Q(v^n|u^n)}{M \cdot Q(v^n)} + 1 \right).
\end{aligned}
$$

Using standard inequalities (see [20]) we have

$$\begin{aligned} d_1 &\leq \sum_{(u^n,v^n)\in\mathcal{T}_\epsilon^n(Q_{UV})} Q(u^n,v^n)\log\left(\frac{2^{-n(1-\epsilon)H(V|U)}}{M\cdot 2^{-n(1+\epsilon)H(V)}}+1\right) \\ &\leq \log\left(\frac{2^{-n(1-\epsilon)H(V|U)}}{M\cdot 2^{-n(1+\epsilon)H(V)}}+1\right) \\ &= \log\left(2^{-n(R-I(V;U)-\epsilon(H(V|U)+H(V)))}+1\right) \\ &\leq \log(e)\cdot 2^{-n(R-I(V;U)-2\epsilon H(V))} \end{aligned} \tag{5.12}$$

and $d_1 \to 0$ if $R > I(V;U) + 2\epsilon H(V)$ and $n \to \infty$. We further have

$$\begin{aligned} d_2 &\leq \sum_{\substack{(u^n,v^n)\notin\mathcal{T}_\epsilon^n(Q_{UV}) \\ (u^n,v^n)\in\mathrm{supp}(Q_{UV}^n)}} Q(u^n,v^n)\log\left(\left(\frac{1}{\mu_V}\right)^n+1\right) \\ &\leq 2|\mathcal{V}|\cdot|\mathcal{U}|\cdot e^{-2n\epsilon^2\mu_{UV}^2}\log\left(\left(\frac{1}{\mu_V}\right)^n+1\right) \\ &\leq 2|\mathcal{V}|\cdot|\mathcal{U}|\cdot e^{-2n\epsilon^2\mu_{UV}^2}\cdot n\cdot\log\left(\frac{1}{\mu_V}+1\right) \end{aligned} \tag{5.13}$$

and $d_2 \to 0$ as $n \to \infty$, where

$$\begin{aligned} \mu_V &= \min_{v\in\mathrm{supp}(Q_V)} Q(v) \\ \mu_{UV} &= \min_{(v,u)\in\mathrm{supp}(Q_{UV})} Q(u,v). \end{aligned} \tag{5.14}$$

Combining the above we have

$$\mathrm{E}[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)] \to 0 \tag{5.15}$$

if $R > I(V;U) + 2\epsilon H(V)$ and $n \to \infty$. As usual, this means that there must exist a good code $\mathcal{C}^*$ with rate $R > I(V;U)$ that achieves a divergence $D(P_{V^n|\mathcal{C}^*}||Q_V^n)$ smaller than or equal to the average in (5.15) for sufficiently large n. This proves the coding theorem.

Remark 5.1. The cardinality bound on $\mathrm{supp}(Q_U)$ can be derived using techniques from [47, Ch. 15].

Remark 5.2. If $V = U$, then we have $R > H(V)$.

Theorem 5.1 is proved using a uniform W which represents strings of uniform bits. If we use a non-uniform W for the coding scheme, can we still drive the unnormalized informational divergence to zero? We give the answer in the following lemma.

Lemma 5.2. Let $W = B^{nR}$ be a bit stream with nR bits that are generated i.i.d. with a binary distribution P_X with $P_X(0) = p$, $0 < p \leq \frac{1}{2}$. The rate R is achievable if

$$R > \frac{I(V;U)}{H_2(p)} \tag{5.16}$$

where $H_2(\cdot)$ is the binary entropy function.

Proof: The proof is given in Appendix B.1. ■

Remark 5.3. Lemma 5.2 states that even if W is not uniformly distributed, the informational divergence can be made small. This is useful because if the distribution of W is not known exactly, then we can choose R large enough to guarantee the desired resolvability result. A similar result was developed in [43] for secrecy.

5.2.2. Error Exponents

We provide a second proof using Gallager's error exponent [46] by extending [42, Lemma 2] to asymptotic cases. Consider $-\frac{1}{2} \leq \rho \leq 0$ and define

$$E_0^n(\rho, Q_{UV}^n) = \ln \sum_{v^n} \left\{ \mathrm{E}[P_{V^n|\widetilde{\mathcal{C}}}(v^n|\widetilde{\mathcal{C}})^{\frac{1}{1+\rho}}] \right\}^{1+\rho} \tag{5.17}$$

$$E_0(\rho, Q_{UV}) = \ln \sum_{v} \left\{ \sum_{u} Q(u) Q(v|u)^{\frac{1}{1+\rho}} \right\}^{1+\rho} \tag{5.18}$$

$$E_G(R, Q_{UV}) = \inf_{-\frac{1}{2} \leq \rho < 0} \left\{ E_0(\rho, Q_{UV}) + \rho R \right\}. \tag{5.19}$$

Due to [42, Lemma 2], we have the following properties concerning $E_0^n(\rho, Q_{UV}^n)$ and $E_0(\rho, Q_{UV})$:

Property 1:

$$E_0^n(0, Q_{UV}^n) = E_0(0, Q_{UV}) = 0 \tag{5.20}$$

Property 2:

$$\left.\frac{\partial E_0^n(\rho, Q_{UV}^n)}{\partial \rho}\right|_{\rho=0} = -\mathrm{E}[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)]$$
$$\left.\frac{\partial E_0(\rho, Q_{UV})}{\partial \rho}\right|_{\rho=0} = -I(V;U) \tag{5.21}$$

Property 3:

$$\frac{\partial^2 E_0^n(\rho, Q_{UV}^n)}{\partial \rho^2} \geq 0$$
$$\frac{\partial^2 E_0(\rho, Q_{UV})}{\partial \rho^2} \geq 0 \tag{5.22}$$

Due to [46, Theorem 5.6.3], we have

$$\begin{cases} E_G(R, Q_{UV}) < 0 & \text{if } R > I(V;U) \\ E_G(R, Q_{UV}) = 0 & \text{if } R \leq I(V;U) \end{cases} \tag{5.23}$$

By extending [42, Sec. III, Inequality (15)] to asymptotic cases, we have the following lemma.

Lemma 5.3. We have

$$E_0^n(\rho, Q_{UV}^n) \leq e^{nE_G(R, Q_{UV})}. \tag{5.24}$$

Proof: The proof is given in Appendix B.2. ■

Combining Properties 1-3, we have $E_0^n(\rho, Q_{UV}^n)$ and $E_0(\rho, Q_{UV})$ are convex in ρ, for $-\frac{1}{2} \leq \rho \leq 0$ and (see Fig. 5.3)

$$\rho \cdot (-\mathrm{E}[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)]) \leq E_0^n(\rho, Q_{UV}^n) \tag{5.25}$$

which means

$$\mathrm{E}[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)] \leq \frac{E_0^n(\rho, Q_{UV}^n)}{-\rho}$$

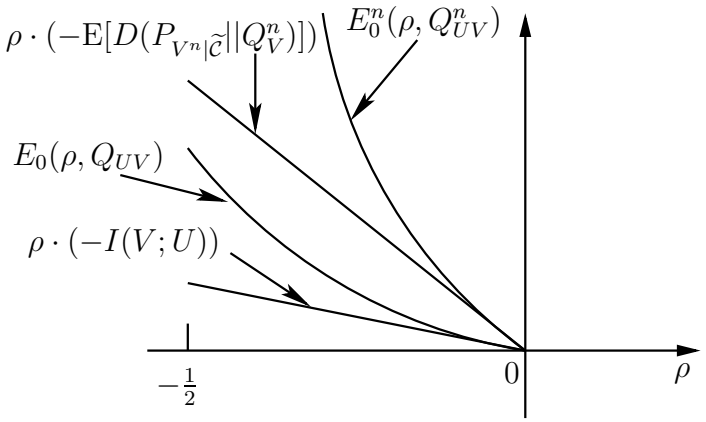


Figure 5.3.: An example of $E_0^n(\rho, Q_{UV}^n)$ and $E_0(\rho, Q_{UV})$.

$$\overset{(a)}{\leq} \frac{e^{nE_G(R,Q_{UV})}}{-\rho} \tag{5.26}$$

where (a) follows from Lemma 5.3. The right hand side of (5.26) goes to 0 as $n \to \infty$ as long as (see (5.23))

$$R > I(V;U). \tag{5.27}$$

Remark 5.4. This proof applies to *continuous* random variables by replacing the sums in the proof of Lemma 5.3 with integrals.

Remark 5.5. The average divergence $\mathrm{E}[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)]$ can be viewed as the mutual information $I(\widetilde{\mathcal{C}}; V^n)$ from the random codebook $\widetilde{\mathcal{C}}$ to the output V^n [42, Sec. III]. To show this, denote $\mathcal{C}$ as a realization of $\widetilde{\mathcal{C}}$ and we have (see (5.10))

$$\begin{aligned}
I(\widetilde{\mathcal{C}}; V^n) &= \sum_{\mathcal{C}} P(\mathcal{C}) \sum_{v^n} P(v^n|\mathcal{C}) \log \frac{P(v^n|\mathcal{C})}{Q_V^n(v^n)} \\
&= \sum_{u^n(1)} \cdots \sum_{u^n(M)} \prod_{k=1}^{M} Q_U^n(u^n(k)) \sum_{v^n} \sum_{w=1}^{M} \frac{1}{M} \cdot Q_{V|U}^n(v^n|u^n(w)) \log \frac{\sum_{j=1}^{M} \frac{1}{M} Q_{V|U}^n(v^n|u^n(j))}{Q_V^n(v^n)} \\
&= \mathrm{E}\left[\log \frac{\sum_{j=1}^{M} \frac{1}{M} Q_{V|U}^n(V^n|U^n(j))}{Q_V^n(V^n)}\right] \\
&= \mathrm{E}[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)].
\end{aligned} \tag{5.28}$$

Thus, as $\mathrm{E}[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)] \to 0$ we have $I(\widetilde{\mathcal{C}}; V^n) \to 0$ which means that $\widetilde{\mathcal{C}}$ and V^n are (almost) independent. This makes sense, since as $P_{V^n|\widetilde{\mathcal{C}}} \to Q_V^n$ one is not able to distinguish which codebook is used to generate the output.

5.2.3. Converse

The converse follows by removing the normalization factor $\frac{1}{n}$ in [16, Theorem 5.2] . We here provide a direct and simpler proof. Consider a given target distribution Q_V and any code $\mathcal{C}$ with rate R and code words of length n. Using the requirement $D(P_{V^n}||Q_V^n) < n\xi$, for some $\xi > 0$, we have

$$\begin{aligned}
n\xi &> D(P_{V^n}||Q_V^n) \\
&= \left[\sum_{v^n} P(v^n) \sum_{i=1}^{n} \log \frac{1}{Q_V(v_i)}\right] - H(V^n) \qquad (5.29) \\
&= \left[\sum_{v^n} P(v^n) \sum_{i=1}^{n} \log \frac{1}{Q_V(v_i)}\right] - \sum_{i=1}^{n} H(V_i|V^{i-1}) \\
&\overset{(a)}{\geq} \left[\sum_{i=1}^{n} \sum_{v} P_{V_i}(v) \log \frac{1}{Q_V(v)}\right] - \sum_{i=1}^{n} H(V_i) \\
&= \sum_{i=1}^{n} D(P_{V_i}||Q_V) \\
&\overset{(b)}{\geq} nD(P_{\overline{V}}||Q_V) \qquad (5.30)
\end{aligned}$$

where $P_{\overline{V}} = \frac{1}{n}\sum_{i=1}^{n} P_{V_i}$, (a) follows because conditioning does not increase entropy, and (b) follows by the convexity of $D(P_X||Q_X)$ in P_X. We also have

$$\begin{aligned}
nR &\geq I(V^n; U^n) \\
&\geq I(V^n; U^n) + D(P_{V^n}||Q_V^n) - n\xi \\
&\overset{(a)}{=} H(V^n) - H(V^n|U^n) + \left[\sum_{i=1}^{n} \sum_{v} P_{V_i}(v) \log \frac{1}{Q_V(v)}\right] - H(V^n) - n\xi \\
&\overset{(b)}{\geq} \left[\sum_{i=1}^{n} \sum_{v} P_{V_i}(v) \log \frac{1}{Q(v)}\right] - \left[\sum_{i=1}^{n} H(V_i|U_i)\right] - n\xi \\
&= \left[\sum_{i=1}^{n} \sum_{v} P_{V_i}(v) \log \frac{1}{Q_V(v)}\right] + \left[\sum_{i=1}^{n} \sum_{(u,v)} P_{U_iV_i}(u,v) \log Q_{V|U}(v|u)\right] - n\xi
\end{aligned}$$

$$
\begin{aligned}
&= n\left[\sum_{(u,v)} P_{\overline{U}}(u)Q(v|u)\log\frac{Q(v|u)}{Q(v)}\right] - n\xi \\
&\overset{(c)}{\geq} n\left[\sum_{(u,v)} P_{\overline{U}}(u)Q(v|u)\log\frac{Q(v|u)}{Q(v)}\right] - n\xi - nD(P_{\overline{V}}||Q_V) \\
&= n\left[\sum_{(u,v)} P_{\overline{U}}(u)Q(v|u)\log\frac{Q(v|u)}{Q(v)}\right] - n\left[\sum_{(u,v)} P_{\overline{U}}(u)Q(v|u)\log\frac{P_{\overline{V}}(v)}{Q(v)}\right] - n\xi \\
&= nI(\overline{V};\overline{U}) - n\xi
\end{aligned} \tag{5.31}
$$

where $P_{\overline{U}} = \frac{1}{n}\sum_{i=1}^{n} P_{U_i}$, (a) follows by (5.29), (b) follows because conditioning does not increase entropy, and (c) follows by $D(P_{\overline{V}}||Q_V) \geq 0$. We further require that $D(P_{\overline{V}}||Q_V) \to 0$ as $\xi \to 0$. Hence, we have

$$
R \geq \min_{P_U : P_V = Q_V} I(V;U) \tag{5.32}
$$

where $P(v) = \sum_u P(u)Q(v|u)$.

5.3. Discussion

Hayashi studied the resolvability problem using unnormalized divergence and he derived bounds for nonasymptotic cases [42, Lemma 2]. We have outlined his proof steps in Sec. 5.2.2. Theorem 5.1 can be derived by extending [42, Lemma 2] to asymptotic cases (see 5.2.2) and it seems that such a result was the underlying motivation for [42, Lemma 2]. Unfortunately, Theorem 5.1 is not stated explicitly in [42] and the ensuing asymptotic analysis was done for *normalized* informational divergence. Hayashi's proofs (he developed two approaches) were based on Shannon random coding.

Theorem 5.1 implies [16, Theorem 6.3] which states that for $R > I(V;U)$ the normalized divergence $\frac{1}{n}D(P_{V^n}||Q_V^n)$ can be made small. Theorem 5.1 implies [17, Theorem 4] for product distributions through Pinsker's inequality (Equ. (2.6)).Moreover, the speed of decay in (5.26) is exponential with n. We can thus make

$$
\alpha(n) \cdot \mathrm{E}\left[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)\right] \tag{5.33}
$$

vanishingly small as $n \to \infty$, where $\alpha(n)$ represents a *sub-exponential* function of n that satisfies,

$$\lim_{n\to\infty} \frac{\alpha(n)}{e^{\beta n}} = 0 \tag{5.34}$$

where β is positive and independent of n (see also [42]). For example, we may choose $\alpha(n) = n^m$ for any integer m. We may also choose $\alpha(n) = e^{\gamma n}$ where $\gamma < \beta$.

Since all achievability results in [48] are based on [17, Theorem 4], Theorem 5.1 extends the results in [48] as well. Theorem 5.1 is closely related to *strong* secrecy [49, 50] and provides a simple proof that Shannon random coding suffices to drive an *unnormalized* mutual information between messages and eavesdropper observations to zero (see Chapter 6 below).

Theorem 5.1 is valid for approximating product distributions only. However extensions to a broader class of distributions, e.g., *information stable* distributions [17], are clearly possible.

Finally, an example code is as follows (courtesy of F. Kschischang). Consider a channel with input and output alphabet the 2^7 binary 7-tuples. Suppose the channel maps each input uniformly to a 7-tuple that is distance 0 or 1 away, i.e., there are 8 channel transitions for every input and each transition has probability $\frac{1}{8}$. A simple "modulation" code for this channel is the $(7, 4)$ Hamming code. The code is perfect and if we choose each code word with probability $\frac{1}{16}$, then the output V^7 of the channel is uniformly distributed over all 2^7 values. Hence $I(V; U) = 4$ bits suffice to "approximate" the product distribution (here there is no approximation).

6

Effective Secrecy: Reliability, Confusion and Stealth

Wyner [51] derived the *secrecy capacity* for *degraded* wire-tap channels (see Fig. 6.1). Csiszár and Körner [52] extended the results to broadcast channels with confidential messages. In both [51] and [52], secrecy was measured by a *normalized* mutual information between the message M and the eavesdropper's output Z^n under a secrecy constraint

$$\frac{1}{n} I(M; Z^n) \leq S \tag{6.1}$$

which is referred to as *weak secrecy*. Weak secrecy has the advantage that one can trade off S for rate. The drawback is that even $S \approx 0$ is usually considered too weak because the eavesdropper can decipher nS bits of M, which grows with n. Therefore, [49] (see also [44]) advocated using *strong secrecy* where secrecy is measured by the *unnormalized*

mutual information $I(M; Z^n)$ and requires

$$I(M; Z^n) \leq \xi \tag{6.2}$$

for any $\xi > 0$ and sufficiently large n.

In related work, Han and Verdú [17] studied *resolvability* based on *variational distance* that addresses the number of bits needed to mimic a marginal distribution of a prescribed joint distribution. Bloch and Laneman [43] used the resolvability approach of [17] and extended the results in [52] to continuous random variables and channels with memory.

The main contribution of this chapter is to define and justify the usefulness of a new and stronger security measure for wire-tap channels that includes not only reliability and (wiretapper) confusion but also *stealth*. The measure is satisfied by random codes and by using the simplified proof of resolvability based on *unnormalized* informational divergence developed in Chapter 5 (see also [45, Lemma 11]). In particular, we measure secrecy by the informational divergence

$$D(P_{MZ^n}||P_M Q_{Z^n}) \tag{6.3}$$

where P_{MZ^n} is the joint distribution of MZ^n, P_M is the distribution of M, P_{Z^n} is the distribution of Z^n, and Q_{Z^n} is the distribution that the eavesdropper expects to observe when the source is *not* communicating useful messages. We call this security measure *effective secrecy*.

One can easily check that (see (6.7) below)

$$D(P_{MZ^n}||P_M Q_{Z^n}) = \underbrace{I(M; Z^n)}_{\text{Non-Confusion}} + \underbrace{D(P_{Z^n}||Q_{Z^n})}_{\text{Non-Stealth}} \tag{6.4}$$

where we interpret $I(M; Z^n)$ as a measure of "non-confusion" and $D(P_{Z^n}||Q_{Z^n})$ as a measure of "non-stealth". We justify the former interpretation by using error probability in Sec. 6.2 and the latter by using binary hypothesis testing in Sec. 6.3. Thus, by making $D(P_{MZ^n}||P_M Q_{Z^n}) \to 0$ we not only keep the message secret from the eavesdropper but also hide the presence of meaningful communication.

We remark that the choice of default behavior Q_{Z^n} in (6.3) will depend on the appli-

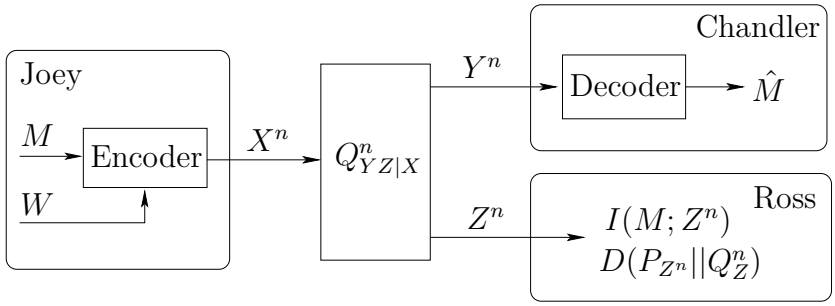


Figure 6.1.: A wire-tap channel.

cation. For example, if the default behavior is to send a code word, then $Q_{Z^n} = P_{Z^n}$ and one achieves stealth for "free". On the other hand, if the default behavior is of the type $Q_{Z^n} = Q_Z^n$, then we must be more careful. We mostly focus on the case $Q_{Z^n} = Q_Z^n$. We also remark that one need not consider stealth as being combined with confusion as in (6.4), see Sec. 6.2.4 below. We combine these concepts mainly for convenience of the proofs.

This chapter is organized as follows. In Section 6.1, we state the problem. In Section 6.2 we state and prove the main result. Section 6.3 relates the result to hypothesis testing. Section 6.4 discusses related works.

6.1. Wire-Tap Channel

Consider the wire-tap channel depicted in Fig. 6.1. Joey has a message M which is destined for Chandler but should be kept secret from Ross. The message M is uniformly distributed over $\{1, \ldots, L\}$, $L = 2^{nR}$, and an encoder $f(\cdot)$ maps M to the sequence

$$X^n = f(M, W) \tag{6.5}$$

with help of a randomizer variable W that is independent of M and uniformly distributed over $\{1, \ldots, L_1\}$, $L_1 = 2^{nR_1}$. The purpose of W is to confuse Ross so that he learns little about M. X^n is transmitted through a memoryless channel $Q^n_{YZ|X}$. Chandler observes the channel output Y^n while Ross observes Z^n. The pair MZ^n has the joint distribution

P_{MZ^n}. Chandler estimates $\hat{M}$ from Y^n and the average error probability is

$$P_e^{(n)} = \Pr\left[\hat{M} \neq M\right]. \tag{6.6}$$

Ross tries to learn M from Z^n and secrecy is measured by

$$\begin{aligned}
&D(P_{MZ^n}||P_M Q_Z^n) \\
&= \sum_{\substack{(m,z^n) \\ \in \mathrm{supp}(P_{MZ^n})}} P(m,z^n) \log\left(\frac{P(m,z^n)}{P(m)\cdot Q_Z^n(z^n)} \cdot \frac{P(z^n)}{P(z^n)}\right) \\
&= \sum_{\substack{(m,z^n) \\ \in \mathrm{supp}(P_{MZ^n})}} P(m,z^n) \left(\log\frac{P(z^n|m)}{P(z^n)} + \log\frac{P(z^n)}{Q_Z^n(z^n)}\right) \\
&= \underbrace{I(M;Z^n)}_{\text{Non-Confusion}} + \underbrace{D(P_{Z^n}||Q_Z^n)}_{\text{Non-Stealth}}
\end{aligned} \tag{6.7}$$

where P_{Z^n} is the distribution Ross observes at his channel output and Q_Z^n is the *default* distribution Ross expects to observe if Joey is *not* sending useful information. For example, if Joey transmits X^n with probability $Q_X^n(X^n)$ through the channel, then we have

$$Q_Z^n(z^n) = \sum_{x^n \in \mathrm{supp}(Q_X^n)} Q_X^n(x^n) Q_{Z|X}^n(z^n|x^n) = P_{Z^n}(z^n). \tag{6.8}$$

When Joey sends useful messages, then P_{Z^n} and Q_Z^n are different. But a small $D(P_{MZ^n}||P_M Q_Z^n)$ implies that both $I(M;Z^n)$ and $D(P_{Z^n}||Q_Z^n)$ are small which in turn implies that Ross learns little about M and cannot recognize whether Joey is communicating anything meaningful. A rate R is *achievable* if for any $\xi_1, \xi_2 > 0$ there is a sufficiently large n and an encoder and a decoder such that

$$P_e^{(n)} \leq \xi_1 \tag{6.9}$$

$$D(P_{MZ^n}||P_M Q_Z^n) \leq \xi_2. \tag{6.10}$$

The *effective secrecy capacity* C_S is the supremum of the set of achievable R. We wish to determine C_S.

6.2. Main result and Proof

We prove the following result.

Theorem 6.1. The effective secrecy capacity of the wire-tap channel is the same as the weak and strong secrecy capacity, namely

$$C_S = \sup_{Q_{VX}} \left[I(V;Y) - I(V;Z) \right] \tag{6.11}$$

where the supremum is taken over all joint distributions Q_{VX} satisfying

$$Q_Z(z) = \sum_{v,x} Q_{VX}(v,x) Q_{Z|X}(z|x) \tag{6.12}$$

and the Markov chain

$$V - X - YZ \tag{6.13}$$

where Q_Z is the default distribution at the eavesdropper's channel output.

One may restrict the cardinality of V to $|\mathcal{V}| \leq |\mathcal{X}|$.

6.2.1. Achievability

We use random coding and the proof technique of Sec. 5.2.1.

Random Code: Fix a distribution Q_X and generate $L \cdot L_1$ code words $x^n(m,w)$, $m = 1,\ldots,L$, $w = 1,\ldots,L_1$ using $\prod_{i=1}^n Q_X(x_i(m,w))$. This defines the codebook

$$\mathcal{C} = \{x^n(m,w), m = 1,\ldots,L,\ w = 1,\ldots,L_1\} \tag{6.14}$$

and we denote the random codebook by

$$\widetilde{\mathcal{C}} = \{X^n(m,w)\}_{(m,w)=(1,1)}^{(L,L_1)}. \tag{6.15}$$

Encoding: To send a message m, Joey chooses w uniformly from $\{1,\ldots,L_1\}$ and transmits $x^n(m,w)$. Hence, for a fixed codebook $\widetilde{\mathcal{C}} = \mathcal{C}$ every $x^n(m,w)$ occurs with

probability

$$P_{X^n}(x^n(m,w)) = \frac{1}{L \cdot L_1} \tag{6.16}$$

rather than $Q_X^n(x^n(m,w))$ (see (6.8)) and Q_Z^n may not be the same as P_{Z^n}. Further, for every pair (m, z^n) we have

$$P(z^n|m) = \sum_{w=1}^{L_1} \frac{1}{L_1} \cdot Q_{Z|X}^n(z^n|x^n(m,w)) \tag{6.17}$$

$$P(z^n) = \sum_{m=1}^{L} \sum_{w=1}^{L_1} \frac{1}{L \cdot L_1} \cdot Q_{Z|X}^n(z^n|x^n(m,w)). \tag{6.18}$$

Chandler: Chandler puts out $(\hat{m}, \hat{w})$ if there is a unique pair $(\hat{m}, \hat{w})$ satisfying the typicality check

$$(x^n(\hat{m}, \hat{w}), y^n) \in \mathcal{T}_\epsilon^n(Q_{XY}). \tag{6.19}$$

Otherwise he puts out $(\hat{m}, \hat{w}) = (1, 1)$.

Analysis: Define the events

$$\begin{aligned} E_1 &: \{(\hat{M}, \hat{W}) \neq (M, W)\} \\ E_2 &: D(P_{MZ^n}||P_M Q_Z^n) > \xi_2. \end{aligned} \tag{6.20}$$

Let $E = E_1 \cup E_2$ so that we have

$$\Pr[E] \leq \Pr[E_1] + \Pr[E_2] \tag{6.21}$$

where we have used the union bound. $\Pr[E_1]$ can be made small with large n as long as

$$R + R_1 < I(X;Y) - \delta_\epsilon(n) \tag{6.22}$$

where $\delta_\epsilon(n) \to 0$ as $n \to \infty$ (see [20]) which implies that $P_e^{(n)}$ is small.

$\Pr[E_2]$ can be made small with large n as long as (Theorem 5.1)

$$R_1 > I(X;Z) + \delta'_\epsilon(n) \tag{6.23}$$

where $\delta'_\epsilon(n) \to 0$ as $n \to \infty$. This is because the divergence averaged over M, W, $\widetilde{\mathcal{C}}$ and Z^n satisfies (see Equ. (5.9))

$$\begin{aligned}
&\mathrm{E}[D(P_{MZ^n|\widetilde{\mathcal{C}}}||P_M Q_Z^n)] \\
&\overset{(a)}{=} \mathrm{E}[D(P_M||P_M) + D(P_{Z^n|M\widetilde{\mathcal{C}}}||Q_Z^n|P_M)] \\
&\overset{(b)}{=} \mathrm{E}\left[\log \frac{\sum_{j=1}^{L_1} Q_{Z|X}^n(Z^n|X^n(M,j))}{L_1 \cdot Q_Z^n(Z^n)}\right] \\
&= \sum_{m=1}^{L}\sum_{w=1}^{L_1} \frac{1}{L\cdot L_1} \mathrm{E}\left[\log \frac{\sum_{j=1}^{L_1} Q_{Z|X}^n(Z^n|X^n(m,j))}{L_1 \cdot Q_Z^n(Z^n)} \middle| M=m, W=w\right] \\
&\overset{(c)}{\leq} \sum_{m=1}^{L}\sum_{w=1}^{L_1} \frac{1}{L\cdot L_1} \mathrm{E}\left[\log\left(\frac{Q_{Z|X}^n(Z^n|X^n(m,w))}{L_1 \cdot Q_Z^n(Z^n)} + 1\right) \middle| M=m, W=w\right] \\
&\overset{(d)}{=} \mathrm{E}\left[\log\left(\frac{Q_{Z|X}^n(Z^n|X^n)}{L_1 \cdot Q_Z^n(Z^n)} + 1\right)\right]
\end{aligned} \tag{6.24}$$

where

(a) follows by the chain rule for informational divergence;

(b) follows by (6.17) and by taking the expectation over $M, W, X^n(1,1), \ldots, X^n(L,L_1), Z^n$;

(c) follows by the concavity of the logarithm and Jensen's inequality applied to the expectation over the $X^n(m,j), j \neq w$ for a fixed m;

(d) follows by choosing $X^n Z^n \sim Q_{XZ}^n$.

Next we can show that the right hand side (RHS) of (6.24) is small if (6.23) is valid by splitting the expectation in (6.24) into sums of typical and atypical pairs (see Equ. (5.11)-(5.14)]). But if the RHS of (6.24) approaches 0, then using (6.7) we have

$$\mathrm{E}\left[I(M;Z^n|\widetilde{\mathcal{C}}) + D(P_{Z^n|\widetilde{\mathcal{C}}}||Q_Z^n)\right] \to 0. \tag{6.25}$$

Combining (6.21), (6.22) and (6.23) we can make $\Pr[E] \to 0$ as $n \to \infty$ as long as

$$R + R_1 < I(X;Y) \tag{6.26}$$

$$R_1 > I(X;Z). \tag{6.27}$$

which means that there must exist one good code $\mathcal{C}^*$ satisfying the above conditions and achieves an error probability smaller than or equal to the average error probability (see (6.22) and (6.23)). We hence have the achievability of any R satisfying

$$0 \leq R < \sup_{Q_X} [I(X;Y) - I(X;Z)]. \tag{6.28}$$

where the supremum is taken over all Q_X such that

$$Q_Z(z) = \sum_x Q_X(x) Q_{Z|X}(z|x). \tag{6.29}$$

Of course, if the RHS of (6.28) is non-positive, then we require $R = 0$. Now we prefix a channel $Q^n_{X|V}$ to the original channel $Q^n_{YZ|X}$ and obtain a new channel $Q^n_{YZ|V}$ where

$$Q^n_{YZ|V}(y^n, z^n|v^n) = \sum_{x^n \in \mathrm{supp}(Q^n_{X|V}(\cdot|v^n))} Q^n_{X|V}(x^n|v^n) Q^n_{YZ|X}(y^n, z^n|x^n). \tag{6.30}$$

Using a similar analysis as above, we have the achievability of any R satisfying

$$0 \leq R < \sup_{Q_{VX}} [I(V;Y) - I(V;Z)] \tag{6.31}$$

where the supremum is taken over all Q_{VX} satisfying (6.12) and (6.13). Again, if the RHS of (6.31) is non-positive, then we require $R = 0$. As usual, the purpose of adding the auxiliary variable V is to potentially increase R. Note that $V = X$ recovers (6.28). Hence, the RHS of (6.28) is always smaller than or equal to the RHS of (6.31).

Remark 6.1. The steps (6.24) imply that secrecy and stealth are attained for every message m and not just for the average over all messages. This gives a guarantee for a good worst case performance.

Remark 6.2. The average divergence $\mathrm{E}[D(P_{MZ^n|\widetilde{\mathcal{C}}}||P_M Q^n_Z)]$ is the sum of $I(M\widetilde{\mathcal{C}}; Z^n)$

and $D(P_{Z^n}||Q_Z^n)$ [42, Sec. III] (see also Remark 5.5). To see this, consider

$$\begin{aligned}
&\mathrm{E}[D(P_{MZ^n|\widetilde{\mathcal{C}}}||P_M Q_Z^n)] \\
&= D(P_{MZ^n|\widetilde{\mathcal{C}}}||P_M Q_Z^n|P_{\widetilde{\mathcal{C}}}) \\
&\overset{(a)}{=} D(P_{Z^n|M\widetilde{\mathcal{C}}}||Q_Z^n|P_M P_{\widetilde{\mathcal{C}}}) \\
&= D(P_{Z^n|M\widetilde{\mathcal{C}}}||P_{Z^n}|P_M P_{\widetilde{\mathcal{C}}}) + D(P_{Z^n}||Q_Z^n) \\
&= I(M\widetilde{\mathcal{C}}; Z^n) + D(P_{Z^n}||Q_Z^n) \qquad (6.32)
\end{aligned}$$

where (a) follows by the independence of M and the code words. Thus, as $\mathrm{E}[D(P_{MZ^n|\widetilde{\mathcal{C}}}||P_M Q_Z^n)] \to 0$ we have $I(M\widetilde{\mathcal{C}}; Z^n) \to 0$ which means that $M\widetilde{\mathcal{C}}$ and Z^n are (almost) independent. This makes sense, since for effective secrecy the adversary learns little about M and the presence of meaningful transmission.

6.2.2. Converse

The converse follows as in [52, Theorem 1]. We provide an alternative proof using the *telescoping identity* [53, Sec. G]. Consider any code $\mathcal{C}$ with rate R and code words of length n satisfying (6.9) and (6.10) for some ξ_1, $\xi_2 > 0$. We have

$$\begin{aligned}
&\log_2 L = nR \\
&= H(M) \\
&= I(M; Y^n) + H(M|Y^n) \\
&\overset{(a)}{\leq} I(M; Y^n) + (1 + \xi_1 \cdot nR) \\
&\overset{(b)}{\leq} I(M; Y^n) - I(M; Z^n) + \xi_2 \cdot n + (1 + \xi_1 \cdot nR) \qquad (6.33)
\end{aligned}$$

where (a) follows from Fano's inequality and (b) follows from (6.7) and (6.10). Using the telescoping identity [53, Equ. (9) and (11)] we have

$$\begin{aligned}
&\frac{1}{n}\left[I(M; Y^n) - I(M; Z^n)\right] \\
&= \sum_{i=1}^{n} [I(M; Z_{i+1}^n Y^i) - I(M; Z_i^n Y^{i-1})]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}[I(M;Y_i|Y^{i-1}Z_{i+1}^n) - I(M;Z_i;|Y^{i-1}Z_{i+1}^n)] \\
&\overset{(a)}{=} I(M;Y_T|Y^{T-1}Z_{T+1}^n T) - I(M;Z_T|Y^{T-1}Z_{T+1}^n T) \\
&\overset{(b)}{=} I(V;Y|U) - I(V;Z|U) \\
&\overset{(c)}{\leq} \max_{Q_{UVX}:Q_Z(z)=\sum_{u,v,x}Q_{UVX}(u,v,x)Q_{Z|X}(z|x)} [I(V;Y|U) - I(V;Z|U)] \\
&\leq \max_{u} \max_{Q_{VX|U=u}:Q_Z(z)=\sum_{v,x}Q_{VX|U}(v,x|u)Q_{Z|X}(z|x)} [I(V;Y|U=u) - I(V;Z|U=u)] \quad (6.34) \\
&\overset{(d)}{=} \max_{Q_{VX}:Q_Z(z)=\sum_{u,v,x}Q_{VX}(v,x)Q_{Z|X}(z|x)} [I(V;Y) - I(V;Z)] \quad (6.35)
\end{aligned}
$$

where

(a) follows by letting T be independent of all other random variables and uniformly distributed over $\{1,\ldots,n\}$;

(b) follows by defining

$$
\begin{aligned}
&U = Y^{T-1}Z_{T+1}^n T,\ V = MU, \\
&X = X_T,\ Y = Y_T,\ Z = Z_T;
\end{aligned} \quad (6.36)
$$

(c) follows from (6.10) (see also Sec. 5.2.3)

(d) follows because if the maximum in (6.34) is achieved for $U = u^*$ and $Q_{VX|U=u^*}$, then the same can be achieved in (6.35) by choosing a $Q_{VX} = Q_{VX|U=u^*}$.

Combining (6.33) and (6.35) we have

$$
R \leq \frac{\sup_{Q_{VX}} [I(V;Y) - I(V;Z)]}{1-\xi_1} + \frac{\xi_2 \cdot n + 1}{(1-\xi_1)n}. \quad (6.37)
$$

where the supremum is taken over all Q_{VX} satisfying (6.12) and (6.13). Letting $n \to \infty$, $\xi_1 \to 0$, and $\xi_2 \to 0$, we have

$$
R \leq \sup_{Q_{VX}} [I(V;Y) - I(V;Z)] \quad (6.38)
$$

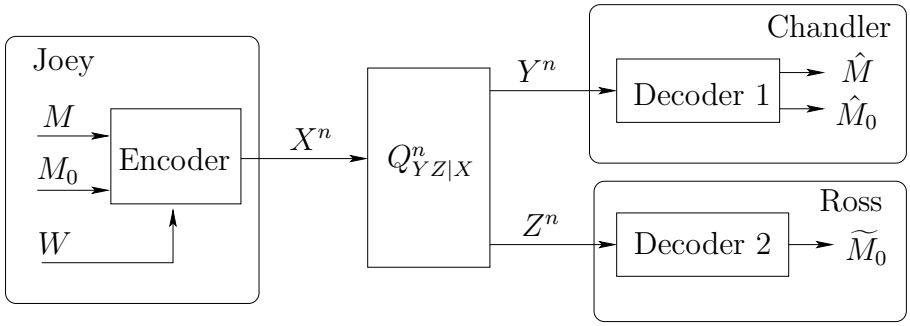


Figure 6.2.: A broadcast channel with a confidential message.

where the supremum is taken over all Q_{VX} satisfying (6.12) and (6.13). The cardinality bound in Theorem 1 was derived in [30, Theorem 22.1].

6.2.3. Broadcast Channels with Confidential Messages

Broadcast channels with confidential messages (BCC) [52] are wire-tap channels with common messages. For the BCC (Fig. 6.2), Joey has a common message M_0 destined for both Chandler and Ross which is independent of M and uniformly distributed over $\{1, \ldots, L_0\}$, $L_0 = 2^{nR_0}$. An encoder maps M_0 and M to

$$X^n = f(M_0, M, W) \tag{6.39}$$

which is sent through the channel $Q^n_{YZ|X}$. Chandler estimates $(\hat{M}_0, \hat{M})$ from Y^n while Ross estimates $\widetilde{M}_0$ from Z^n. The average error probability is

$$P_e^{*(n)} = \Pr\left[\left\{(\hat{M}_0, \hat{M}) \neq (M_0, M)\right\} \cup \left\{\widetilde{M}_0 \neq M_0\right\}\right] \tag{6.40}$$

and non-secrecy is measured by $D(P_{MZ^n}||P_M Q_Z^n)$. A rate pair (R_0, R) is achievable if, for any $\xi_1, \xi_2 > 0$, there is a sufficiently large n, an encoder and two decoders such that

$$P_e^{*(n)} \leq \xi_1 \tag{6.41}$$

$$D(P_{MZ^n}||P_M Q_Z^n) \leq \xi_2. \tag{6.42}$$

The effective secrecy capacity region C_{BCC} is the closure of the set of achievable (R_0, R). We have the following theorem.

Theorem 6.2. C_{BCC} is the same as the weak and strong secrecy capacity region

$$C_{\text{BCC}} = \bigcup \left\{ \begin{array}{l} (R_0, R): \\ 0 \leq R_0 \leq \min\left\{I(U;Y), I(U;Z)\right\} \\ 0 \leq R \leq I(V;Y|U) - I(V;Z|U) \end{array} \right\} \tag{6.43}$$

where the union is over all distributions Q_{UVX} satisfying

$$Q_Z(z) = \sum_{u,v,x} Q_{UVX}(u,v,x) Q_{Z|X}(z|x) \tag{6.44}$$

and the Markov chain

$$U - V - X - YZ. \tag{6.45}$$

One may restrict the alphabet sizes to

$$|\mathcal{U}| \leq |\mathcal{X}| + 3; \quad |\mathcal{V}| \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3. \tag{6.46}$$

Proof: The proof is omitted due to the similarity to the proof of Theorem 6.1. ■

6.2.4. Choice of Security Measures

Effective secrecy includes both strong secrecy and stealth communication. One may argue that using only $I(M; Z^n)$ or $D(P_{Z^n}||Q_Z^n)$ would suffice to measure secrecy. However, we consider two examples where secrecy is achieved but not stealth, and where stealth is achieved but not secrecy.

Example 6.1. $I(M; Z^n) \to 0$, $D(P_{Z^n}||Q_Z^n) = D > 0$. Suppose that Joey inadvertently uses $\widetilde{Q}_X$ rather than Q_X for codebook generation, where (6.23) is still satisfied. For example, $\widetilde{Q}_X$ might represent no energy while Q_X must have positive energy. The new

$\widetilde{Q}_X$ could result in a different expected $\widetilde{Q}_Z^n \neq Q_Z^n$. Hence, as n grows large we have

$$D(P_{MZ^n}||P_M Q_Z^n) = I(M;Z^n) + D(\widetilde{Q}_Z^n||Q_Z^n) \tag{6.47}$$

where $I(M;Z^n) \to 0$ but we have

$$D(\widetilde{Q}_Z^n||Q_Z^n) = D, \text{ for some } D > 0. \tag{6.48}$$

Ross thus recognizes that Joey is transmitting useful information even though he cannot decode.

Example 6.2. $I(M;Z^n) = I > 0,\ D(P_{Z^n}||Q_Z^n) \to 0$. Note that $\mathrm{E}[D(P_{Z^n}||Q_Z^n)] \to 0$ as $n \to \infty$ as long as (see [54, Theorem 1])

$$R + R_1 > I(X;Z). \tag{6.49}$$

If Joey is not careful and chooses R_1 such that (6.23) is violated and (6.49) is satisfied, then $D(P_{Z^n}||Q_Z^n)$ can be made small but we have

$$I(M;Z^n) = I \text{ for some } I > 0. \tag{6.50}$$

For example, Joey might choose $R_1 = 0$. Thus, although the communication makes $D(P_{Z^n}||Q_Z^n)$ small, Ross can learn

$$I(M;Z^n) \approx n[I(X;Z) - R_1] \tag{6.51}$$

bits about M if he is willing to pay a price and always tries to decode (see Sec. 6.3).

6.3. Hypothesis Testing

The reader may wonder how $D(P_{Z^n}||Q_Z^n)$ relates to stealth. We consider a hypothesis testing framework and show that as long as (6.49) is satisfied, the best Ross can do to detect Joey's action is to guess.

For every channel output z^n, Ross considers two hypotheses

$$H_0 = Q_Z^n \tag{6.52}$$

$$H_1 = P_{Z^n}. \tag{6.53}$$

If H_0 is accepted, then Ross decides that Joey's transmission is not meaningful, whereas if H_1 is accepted, then Ross decides that Joey is sending useful messages. We define two kinds of error probabilities

$$\alpha = \Pr\{H_1 \text{ is accepted} \mid H_0 \text{ is true}\} \tag{6.54}$$

$$\beta = \Pr\{H_0 \text{ is accepted} \mid H_1 \text{ is true}\}. \tag{6.55}$$

The value α is referred to as *the level of significance* [55] and corresponds to the probability of raising a false alarm, while β corresponds the probability of mis-detection. In practice, raising a false alarm can be expensive. Therefore, Ross would like to minimize β for a given tolerance level of α. To this end, Ross performs for every z^n a ratio test

$$\frac{Q_Z^n(z^n)}{P_{Z^n}(z^n)} = r \tag{6.56}$$

and makes a decision depending on a threshold F, $F \geq 0$, namely

$$\begin{cases} H_0 \text{ is accepted} & \text{if } r > F \\ H_1 \text{ is accepted} & \text{if } r \leq F \end{cases}. \tag{6.57}$$

Define the set of z^n for which H_0 is accepted as

$$\mathcal{A}_F^n = \left\{ z^n : \frac{Q_Z^n(z^n)}{P_{Z^n}(z^n)} > F \right\} \tag{6.58}$$

and $(\mathcal{A}_F^n)^{\mathrm{c}}$ is the set of z^n for which H_1 is accepted (see Fig. 6.3). Ross chooses the threshold F and we have

$$\alpha = Q_Z^n((\mathcal{A}_F^n)^{\mathrm{c}}) = 1 - Q_Z^n\left(\mathcal{A}_F^n\right)$$

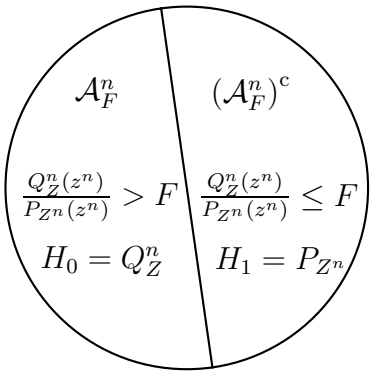


Figure 6.3.: Example of the decision regions $\mathcal{A}_F^n$ and $(\mathcal{A}_F^n)^{\mathrm{c}}$.

$$\beta = P_{Z^n}(\mathcal{A}_F^n). \tag{6.59}$$

The ratio test in (6.56) is the *Neyman-Pearson test* which is *optimal* [55, Theorem 3.2.1] in the sense that it minimizes β for a given α. We have the following lemma.

Lemma 6.3. If $D(P_{Z^n}||Q_Z^n) \leq \xi_2$, $\xi_2 > 0$, then with the Neyman-Pearson test we have

$$1 - g(\xi_2) \leq \alpha + \beta \leq 1 + g(\xi_2) \tag{6.60}$$

where

$$g(\xi_2) = \sqrt{\xi_2 \cdot 2 \ln 2} \tag{6.61}$$

which goes to 0 as $\xi_2 \to 0$.

Proof: Since $D(P_{Z^n}||Q_Z^n) \leq \xi_2$, we have

$$||P_{Z^n} - Q_Z^n||_{\mathrm{TV}} \leq \sqrt{\xi_2 \cdot 2 \ln 2} = g(\xi_2) \tag{6.62}$$

where the inequality follows by (2.6). We further have

$$\begin{aligned} &||P_{Z^n} - Q_Z^n||_{\mathrm{TV}} \\ &= \sum_{z^n \in \mathcal{A}_F^n} |P_{Z^n}(z^n) - Q_Z^n(z^n)| + \sum_{z^n \in \left(\mathcal{A}_F^n\right)^{\mathrm{c}}} |P_{Z^n}(z^n) - Q_Z^n(z^n)| \end{aligned}$$

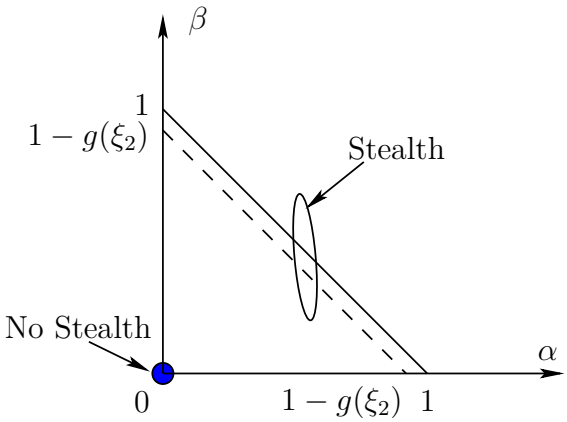


Figure 6.4.: Optimal tradeoff between α and β.

$$
\begin{aligned}
&\geq \sum_{z^n \in \mathcal{A}_F^n} |P_{Z^n}(z^n) - Q_Z^n(z^n)| \\
&\overset{(a)}{\geq} \left| \sum_{z^n \in \mathcal{A}_F^n} [P_{Z^n}(z^n) - Q_Z^n(z^n)] \right| \\
&= |P_{Z^n}(\mathcal{A}_F^n) - Q_Z^n(\mathcal{A}_F^n)| \\
&= |\beta - (1-\alpha)| \qquad (6.63)
\end{aligned}
$$

where (a) follows by the triangle inequality. Combining (6.62) and (6.63), we have the bounds (6.60). ■

Fig. 6.4 illustrates the optimal tradeoff between α and β for stealth communication, i.e., when (6.49) is satisfied. As $n \to \infty$ and $\xi_2 \to 0$, we have

$$D(P_{Z^n} || Q_Z^n) \to 0 \qquad (6.64)$$

$$(\alpha + \beta) \to 1. \qquad (6.65)$$

If Ross allows no false alarm ($\alpha = 0$), then he always ends up with mis-detection ($\beta = 1$). If Ross tolerates no mis-detection ($\beta = 0$), he pays a high price ($\alpha = 1$). Further, for any given α, the optimal mis-detection probability is

$$\beta_{\text{opt}} = 1 - \alpha. \qquad (6.66)$$

But Ross does not need to see Z^n or perform an optimal test to achieve β_{opt}. He may randomly choose some $\mathcal{A}'$ such that

$$Q_Z^n((\mathcal{A}')^{\text{c}}) = \alpha \tag{6.67}$$

and achieves $\beta'_{\text{opt}} = 1 - \alpha$. The best strategy is thus to guess. On the other hand, if

$$\lim_{n\to\infty} D(P_{Z^n}||Q_Z^n) > 0 \tag{6.68}$$

then Ross detects Joey's action and we can have

$$\alpha + \beta = 0. \tag{6.69}$$

We thus operate in one of two regimes in Fig. 6.4, either near $(\alpha, \beta) = (0, 0)$ or near the line $\alpha + \beta = 1$.

6.4. Discussion

Our resolvability proof differs from that in [43] in that we rely on *unnormalized* informational divergence [54] instead of variational distance [17]. Our proof is simpler and the result is stronger than that in [43] when restricting attention to product distributions and memoryless channels because a small $D(P_{MZ^n}||P_M Q_Z^n)$ implies small $I(M; Z^n)$ and $D(P_{Z^n}||Q_Z^n)$ while a small $||P_{X^n} - Q_X^n||_{\text{TV}}$ implies only a small $I(M; Z^n)$ [44, Lemma 1].

Hayashi studied strong secrecy for wire-tap channels using resolvability based on unnormalized divergence and he derived bounds for nonasymptotic cases [42, Theorem 3]. We remark that Theorem 1 can be derived by extending [42, Lemma 2] to asymptotic cases. However, Hayashi did not consider stealth but focused on strong secrecy, although he too noticed a formal connection to (6.7) [42, p. 1568].

7

Conclusion

We have addressed two problems in network information theory: short message noisy network coding (SNNC) and resolvability based on unnormalized informational divergence with applications to network security. SNNC with backward decoding simplifies the analysis and enables mixed strategies of DF and QF that provide better rates and outage probabilities. Resolvability based on informational divergence gives a stronger result with a simpler proof that also applies to establish a new and stronger *effective secrecy* measure. This measure includes strong secrecy and the hiding of the presence of meaningful communication.

There are several open research problems worth addressing:

1. In [11], the authors characterized the gap between the cut-set bound and the achievable LNNC (SNNC) rates under the assumption that *all* network nodes quantize at the noise level. The resulting gap increases *linearly* with the number of nodes K in the network. Recently, the papers [56–58] showed that if the quantization noise level is proportional to the number of nodes K at *all* nodes, then the gap increases *logarithmically* with K.

However, we may improve the gap by letting the nodes quantize at *individual* noise levels depending on the network geometry rather than quantizing at the *same* noise level. Suppose we have a network with several sources, destinations and two classes of relay nodes:

Class 1: close to the sources but far from the destinations

Class 2: close to the destinations but far from sources

Then, the Class 1 relays should quantize coarsely, because their relay-destination links are weak and may not support fine details. For the Class 2 relays the situation is the other way around: they should quantize finely to exploit the strong relay-destination links. In this way, we may get a better gap.

2. The investigation of wire-tap channels with effective secrecy was done information-theoretically. It would be interesting to find explicit codes with rate close to the secrecy capacity that achieve strong secrecy and stealth at the same time. Recently, the paper [59] showed that polar coding achieves strong secrecy for degraded binary symmetric wire-tap channels. It's worth checking whether polar coding can also achieve effective secrecy for a broader class of wire-tap channels.

A

Proofs for Chapter 3

A.1. Treating Class 2 Nodes as Noise

If all Class 2 bounds in (3.38) are satisfied, then SNNC achieves the same rates as NNC. In this case the decoder recovers the signals from the nodes in $\widetilde{\mathcal{D}}_k^{\mathrm{c}}$ and thereby removes interference from these nodes.

Now suppose that an $\mathcal{S}$ in Class 2 has $R_{\mathcal{S}} \geq I_{\mathcal{S}}^{\mathcal{K}}(k)$. We use the argument presented in [25] (see also [5]): for any $\mathcal{J}$ satisfying $\mathcal{S} \subset \mathcal{J} \subset \mathcal{K}$ such that $\mathcal{J} \cap \widetilde{\mathcal{D}}_k \neq \emptyset$ and $k \in \mathcal{J}^{\mathrm{c}}$ we have

$$
\begin{aligned}
R_{\mathcal{J}\setminus\mathcal{S}} &< I_{\mathcal{J}}^{\mathcal{K}}(k) - R_{\mathcal{S}} \\
&\overset{(a)}{\leq} I_{\mathcal{J}}^{\mathcal{K}}(k) - I_{\mathcal{S}}^{\mathcal{K}}(k) \\
&\overset{(b)}{=} I(X_{\mathcal{J}}; \hat{Y}_{\mathcal{J}^{\mathrm{c}}} Y_k | X_{\mathcal{J}^{\mathrm{c}}}) - I(\hat{Y}_{\mathcal{J}}; Y_{\mathcal{J}} | X_{\mathcal{K}} \hat{Y}_{\mathcal{J}^{\mathrm{c}}} Y_k) - I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k | X_{\mathcal{S}^{\mathrm{c}}}) + I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{K}} \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k) \\
&\overset{(c)}{=} I(X_{\mathcal{J}}; \hat{Y}_{\mathcal{J}^{\mathrm{c}}} Y_k | X_{\mathcal{J}^{\mathrm{c}}}) - I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k | X_{\mathcal{S}^{\mathrm{c}}}) - I(\hat{Y}_{\mathcal{J}\setminus\mathcal{S}}; Y_{\mathcal{S}} Y_{\mathcal{J}\setminus\mathcal{S}} | X_{\mathcal{K}} \hat{Y}_{\mathcal{J}^{\mathrm{c}}} Y_k) \\
&\quad - I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} Y_{\mathcal{J}\setminus\mathcal{S}} | X_{\mathcal{K}} \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k) + I(\hat{Y}_{\mathcal{S}}; Y_{\mathcal{S}} | X_{\mathcal{K}} \hat{Y}_{\mathcal{S}^{\mathrm{c}}} Y_k)
\end{aligned}
$$

$$\begin{aligned}
&\overset{(d)}{=} I(X_{\mathcal{J}}; \hat{Y}_{\mathcal{J}^c} Y_k | X_{\mathcal{J}^c}) - I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c} Y_k | X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{J}\setminus\mathcal{S}}; Y_{\mathcal{S}} Y_{\mathcal{J}\setminus\mathcal{S}} | X_{\mathcal{K}} \hat{Y}_{\mathcal{J}^c} Y_k) \\
&\overset{(e)}{=} I(X_{\mathcal{J}\setminus\mathcal{S}}; \hat{Y}_{\mathcal{J}^c} Y_k | X_{\mathcal{J}^c}) + I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{J}^c} Y_k | X_{\mathcal{S}^c}) - I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{S}^c} Y_k | X_{\mathcal{S}^c}) - I(\hat{Y}_{\mathcal{J}\setminus\mathcal{S}}; Y_{\mathcal{J}\setminus\mathcal{S}} | X_{\mathcal{K}} \hat{Y}_{\mathcal{J}^c} Y_k) \\
&\overset{(f)}{=} I(X_{\mathcal{J}\setminus\mathcal{S}}; \hat{Y}_{\mathcal{J}^c} Y_k | X_{\mathcal{J}^c}) - I(X_{\mathcal{S}}; \hat{Y}_{\mathcal{J}\setminus\mathcal{S}} | X_{\mathcal{S}^c} \hat{Y}_{\mathcal{J}^c} Y_k) - I(\hat{Y}_{\mathcal{J}\setminus\mathcal{S}}; Y_{\mathcal{J}\setminus\mathcal{S}} | X_{\mathcal{K}} \hat{Y}_{\mathcal{J}^c} Y_k) \\
&\overset{(g)}{=} I(X_{\mathcal{J}\setminus\mathcal{S}}; \hat{Y}_{\mathcal{J}^c} Y_k | X_{\mathcal{J}^c}) - I(\hat{Y}_{\mathcal{J}\setminus\mathcal{S}}; X_{\mathcal{S}} Y_{\mathcal{J}\setminus\mathcal{S}} | X_{\mathcal{S}^c} \hat{Y}_{\mathcal{J}^c} Y_k) \\
&\overset{(h)}{=} I(X_{\mathcal{J}\setminus\mathcal{S}}; \hat{Y}_{\mathcal{J}^c} Y_k | X_{\mathcal{J}^c}) - I(\hat{Y}_{\mathcal{J}\setminus\mathcal{S}}; Y_{\mathcal{J}\setminus\mathcal{S}} | X_{\mathcal{S}^c} \hat{Y}_{\mathcal{J}^c} Y_k) \\
&\overset{(i)}{=} I^{\mathcal{K}\setminus\mathcal{S}}_{\mathcal{J}\setminus\mathcal{S}}(k) \qquad \text{(A.1)}
\end{aligned}$$

where

(a) follows because $R_{\mathcal{S}} \geq I^{\mathcal{K}}_{\mathcal{S}}(k)$ by assumption

(b) follows from the definition (3.10)

(c) follows from the chain rule for mutual information

(d) follows from the Markov chain

$$X_{\mathcal{S}^c} Y_{\mathcal{J}\setminus\mathcal{S}} \hat{Y}_{\mathcal{S}^c} Y_k - Y_{\mathcal{S}} X_{\mathcal{S}} - \hat{Y}_{\mathcal{S}} \qquad \text{(A.2)}$$

(e)-(h) follow from the chain rule for mutual information and the Markov chain

$$X_{\mathcal{K}\setminus\mathcal{J}} X_{\mathcal{S}} Y_{\mathcal{S}} \hat{Y}_{\mathcal{J}^c} Y_k - Y_{\mathcal{J}\setminus\mathcal{S}} X_{\mathcal{J}\setminus\mathcal{S}} - \hat{Y}_{\mathcal{J}\setminus\mathcal{S}} \qquad \text{(A.3)}$$

(i) follows from the definition (3.10).

The rates satisfying (A.1) are the NNC rates for the nodes in $\mathcal{K} \setminus \mathcal{S}$ while treating the signals from the nodes in $\mathcal{S}$ as noise. This shows that if any of the constraints in Class 2 is violated for NNC or SNNC, then the destination node should treat the signals from the corresponding nodes as noise rather than decoding them.

Now repeat the above argument for the nodes in $\mathcal{K} \setminus \mathcal{S}$, until we reach a set $\widetilde{\mathcal{K}}_k \subset \mathcal{K}$, for which

$$0 \leq R_{\mathcal{S}} < I^{\widetilde{\mathcal{K}}_k}_{\mathcal{S}}(k) \qquad \text{(A.4)}$$

for all subsets $\mathcal{S} \subset \widetilde{\mathcal{K}}_k$ such that $k \in \mathcal{S}^{\mathrm{c}}$ and $\mathcal{S} \neq \emptyset$. By the union bound, the error probability for all destinations tends to zero as $n \to \infty$ if the rate tuple $(R_1, \ldots, R_K)$ satisfies (A.4) for all subsets $\mathcal{S} \subset \widetilde{\mathcal{K}}_k$ with $k \in \mathcal{S}^{\mathrm{c}}$ and $\mathcal{S} \cap \widetilde{\mathcal{D}}_k \neq \emptyset$, where $\mathcal{S}^{\mathrm{c}}$ is the complement of $\mathcal{S}$ in $\widetilde{\mathcal{K}}_k$, and for joint distributions that factor as

$$\left[\prod_{k=1}^{K} P(x_k)P(\hat{y}_k|y_k, x_k)\right] P(y^K|x^K). \tag{A.5}$$

A.2. SNNC with joint Decoding

After block $B + K \cdot (K-1)$ every node $k \in \mathcal{K}$ can reliably recover $\mathbf{l}_B = (l_{1B}, \ldots, l_{KB})$ via the multihopping of the last $K(K-1)$ blocks.

Let $\epsilon_1 > \epsilon$. Node k tries to find a $(\hat{\mathbf{w}}_1^{(k)}, \ldots, \hat{\mathbf{w}}_B^{(k)})$ and $(\hat{\mathbf{l}}_1^{(k)}, \ldots, \hat{\mathbf{l}}_B^{(k)})$ such that the event (3.22) occurs for *all* $j = 1, \ldots, B$, where $\mathbf{l}_B$ is already known. The difference between *joint decoding* and *backward decoding* is that the typicality test is performed jointly over all blocks (see (A.6)-(A.8) below) while it is performed in only one block in (3.23)-(3.25).

Error Probability: Let $\mathbf{1} = (1, \ldots, 1)$. Assume without loss of generality that $\mathbf{w}_j = \mathbf{1}$ and $\mathbf{l}_j = \mathbf{1}$ for $j = 1, \ldots, B$. For any $\mathcal{S} \subset \mathcal{K}$, define

$$\mathbf{w}_{(\mathcal{S})j} = [w_{ij} : \ i \in \mathcal{S}].$$

The error events at decoder k are:

$$E_{k0} : \cup_{j=1}^{B} \cap_{l_{kj}} \ E_{0(kj)}^{\mathrm{c}}(l_{kj}) \tag{A.6}$$

$$E_{k1} : (\cap_{j=1}^{B} E_{1(kj)}(\mathbf{1}, \mathbf{1}, \mathbf{1}))^{\mathrm{c}} \tag{A.7}$$

$$E_{k2} : \cup_{(\mathbf{w}^B_{\widetilde{\mathcal{D}}_k} \neq \mathbf{1}, \mathbf{w}^B_{\widetilde{\mathcal{D}}_k^{\mathrm{c}}})} \cup_{\mathbf{l}^B} \cap_{j=1}^{B} \ E_{1(kj)}(\mathbf{w}_j, \mathbf{l}_{j-1}, \mathbf{l}_j) \tag{A.8}$$

The error event $E_k = \cup_{i=0}^{2} E_{ki}$ at node k thus satisfies

$$\Pr[E_k] \leq \Pr[E_{k0}] + \Pr[E_{k1}] + \Pr[E_{k2}] \tag{A.9}$$

where we have used the union bound. $\Pr[E_{k0}]$ can be made small with large n as long

as (see (3.27))

$$\hat{R}_k > I(\hat{Y}_k; Y_k|X_k) + \delta_\epsilon(n). \tag{A.10}$$

Also, we have

$$\begin{aligned}
\Pr[E_{k1}] &= \Pr[(\cap_{j=1}^{B} E_{1(kj)}(\mathbf{1},\mathbf{1},\mathbf{1}))^{\mathrm{c}}] \\
&= \Pr[\cup_{j=1}^{B} E_{1(kj)}^{\mathrm{c}}(\mathbf{1},\mathbf{1},\mathbf{1})] \\
&\leq \sum_{j=1}^{B} \Pr[E_{1(kj)}^{\mathrm{c}}(\mathbf{1},\mathbf{1},\mathbf{1})] \\
&\overset{(a)}{\leq} B \cdot \delta_{\epsilon_1}(n) \\
&= \delta_{\epsilon_1}(n, B)
\end{aligned} \tag{A.11}$$

where (a) follows because $\Pr[E_{1(kj)}^{\mathrm{c}}(\mathbf{1},\mathbf{1},\mathbf{1})] \leq \delta_{\epsilon_1}(n)$, which goes to zero as $n \to \infty$, for $j = 1, \ldots, B$ [20].

To bound $\Pr\Big[E_{k2}\Big]$, for each $(\mathbf{w}_j, \mathbf{l}_{j-1})$, we define

$$\mathcal{S}_j(\mathbf{w}_j, \mathbf{l}_{j-1}) = \{i \in \mathcal{K} : w_{ij} \neq 1 \text{ or } l_{i(j-1)} \neq 1\} \tag{A.12}$$

and write $\mathcal{S}_j = \mathcal{S}_j(\mathbf{w}_j, \mathbf{l}_{j-1})$. Observe that for $j = 1, \ldots, B$:

- ▷ $(\mathbf{X}_{\mathcal{S}_j}, \hat{\mathbf{Y}}_{\mathcal{S}_j})$ is independent of $(\mathbf{X}_{\mathcal{S}_j^{\mathrm{c}}}, \hat{\mathbf{Y}}_{\mathcal{S}_j^{\mathrm{c}}}, \mathbf{Y}_{kj})$ in the random coding experiment;
- ▷ the $(X_{ij}, \hat{Y}_{ij})$, $i \in \mathcal{S}_j$, are mutually independent.

We have (see (3.31) and (3.32)):

$$\Pr\Big[E_{1(kj)}(\mathbf{w}_j, \mathbf{l}_{j-1}, \mathbf{l}_j)\Big] \leq P_{(kj)}(\mathcal{S}_j) \tag{A.13}$$

where

$$P_{(kj)}(\mathcal{S}_j) = \begin{cases} 2^{-n(I_{\mathcal{S}_j} - \delta_{\epsilon_1}(n))} & \text{if } \mathcal{S}_j \neq \emptyset \\ 1 & \text{otherwise} \end{cases} \tag{A.14}$$

and $\delta_{\epsilon_1}(n) \to 0$ as $n \to \infty$.

By the union bound, we have

$$\begin{aligned}
\Pr[E_{k2}] &\leq \sum_{\left(\mathbf{w}^B_{\widetilde{\mathcal{D}}_k} \neq \mathbf{1}, \mathbf{w}^B_{\widetilde{\mathcal{D}}^{\mathrm{c}}_k}\right)} \sum_{\mathbf{l}^{B-1}} \Pr[\cap_{j=1}^{B} E_{1(kj)}\left(\mathbf{w}_j, \mathbf{l}_{j-1}, \mathbf{l}_j\right)] \\
&\overset{(a)}{=} \sum_{\left(\mathbf{w}^B_{\widetilde{\mathcal{D}}_k} \neq \mathbf{1}, \mathbf{w}^B_{\widetilde{\mathcal{D}}^{\mathrm{c}}_k}\right)} \sum_{\mathbf{l}^{B-1}} \prod_{j=1}^{B} \Pr[E_{1(kj)}\left(\mathbf{w}_j, \mathbf{l}_{j-1}, \mathbf{l}_j\right)] \\
&\overset{(b)}{\leq} \left[\sum_{\mathbf{w}^B, \mathbf{l}^{B-1}} \prod_{j=1}^{B} \Pr[E_{1(kj)}\left(\mathbf{w}_j, \mathbf{l}_{j-1}, \mathbf{l}_j\right)]\right] - \prod_{j=1}^{B} \Pr[E_{1(kj)}\left(\mathbf{1}, \mathbf{1}, \mathbf{1}\right)] \\
&\overset{(c)}{\leq} \left[\sum_{\mathbf{w}^B, \mathbf{l}^{B-1}} \prod_{j=1}^{B} P_{(kj)}(\mathcal{S}_j)\right] - (1 - \delta_{\epsilon_1}(n, B)) \\
&\overset{(d)}{=} \left[\prod_{j=1}^{B} \sum_{\mathbf{w}_j, \mathbf{l}_{j-1}} P_{(kj)}(\mathcal{S}_j)\right] - (1 - \delta_{\epsilon_1}(n, B)) \\
&\overset{(e)}{<} \prod_{j=1}^{B} \left(1 + \sum_{\substack{\mathcal{S}: k \in \mathcal{S}^{\mathrm{c}} \\ \mathcal{S} \neq \emptyset}} \sum_{\substack{(\mathbf{w}_j, \mathbf{l}_{j-1}) \neq (\mathbf{1}, \mathbf{1}): \\ \mathcal{S}_j(\mathbf{w}_j, \mathbf{l}_{j-1}) = \mathcal{S}}} 2^{-n(I_{\mathcal{S}} - \delta_{\epsilon_1}(n))}\right) - (1 - \delta_{\epsilon_1}(n, B)) \\
&\overset{(f)}{<} \left(1 + \sum_{\substack{\mathcal{S}: k \in \mathcal{S}^{\mathrm{c}} \\ \mathcal{S} \neq \emptyset}} 3^{|\mathcal{S}|} 2^{n\left(R_{\mathcal{S}} + \hat{R}_{\mathcal{S}}) - (I_{\mathcal{S}} - \delta_{\epsilon_1}(n))\right)}\right)^{B} - (1 - \delta_{\epsilon_1}(n, B)) \qquad \text{(A.15)}
\end{aligned}$$

where

(a) follows because the codebooks are independent and the channel is memoryless

(b) follows by adding $\left(\mathbf{w}^B_{\widetilde{\mathcal{D}}_k} = \mathbf{1}, \mathbf{w}^B_{\widetilde{\mathcal{D}}^{\mathrm{c}}_k}\right)$ to the sum

(c) follows from (A.13) and because (see (A.11))

$$\begin{aligned}
&\Pr[\cap_{j=1}^{B} E_{1(kj)}(\mathbf{1}, \mathbf{1}, \mathbf{1})] \\
&= 1 - \Pr[(\cap_{j=1}^{B} E_{1(kj)}(\mathbf{1}, \mathbf{1}, \mathbf{1}))^{\mathrm{c}}] \\
&\geq 1 - \delta_{\epsilon_1}(n, B) \qquad \text{(A.16)}
\end{aligned}$$

(d) follows because $P_{(kj)}(\mathcal{S}_j)$ depends only on $\mathcal{S}_j$ which in turn depends only on $(\mathbf{w}_j, \mathbf{l}_{j-1})$

(e) follows from (A.14)

(f) follows from (3.33).

Performing the same steps as in (3.35) and (3.36), we require

$$R_{\mathcal{S}} < I_{\mathcal{S}}^{\mathcal{K}}(k) \tag{A.17}$$

for all subsets $\mathcal{S} \subset \mathcal{K}$ such that $k \in \mathcal{S}^{\mathrm{c}}$ and $\mathcal{S} \neq \emptyset$. We can again split the bounds in (A.17) into two classes:

$$\text{Class 1} : \mathcal{S} \cap \widetilde{\mathcal{D}}_k \neq \emptyset \tag{A.18}$$

$$\text{Class 2} : \mathcal{S} \cap \widetilde{\mathcal{D}}_k = \emptyset \text{ or equivalently } \mathcal{S} \subseteq \widetilde{\mathcal{D}}_k^{\mathrm{c}} \tag{A.19}$$

and show that the constraints in (A.19) at node k are redundant with the same argument used for backward decoding. By the union bound, the error probability for all destinations tends to zero as $n \to \infty$ if the rate tuple $(R_1, \ldots, R_K)$ satisfies (3.12) for all subsets $\mathcal{S} \subset \mathcal{K}$ such that $k \in \mathcal{S}^{\mathrm{c}}$ and $\mathcal{S} \neq \emptyset$, and for any joint distribution that factors as (3.13).

A.3. Backward Decoding for the Two-Relay Channel without Block Markov Coding

The coding scheme is the same as in Example 3.4, except that no BMC is used (see Table A.1). We show how to recover the rate (3.57) with independent inputs and with 2 different backward decoders.

Decoding at Relays:

1) Node 2. For block $j = 1, \ldots, B$, node 2 tries to find a $\hat{w}_j$ that satisfies

$$\left(\mathbf{x}_1\left(\hat{w}_j\right), \mathbf{x}_2(w_{j-1}), \mathbf{y}_{2j}\right) \in \mathcal{T}_\epsilon^n\left(P_{X_1X_2Y_2}\right). \tag{A.20}$$

Node 2 can reliably decode w_j as $n \to \infty$ if (see [20])

$$R < I(X_1; Y_2|X_2). \tag{A.21}$$

Block	1	2	$\cdots$	B	$B+1$
X_1	$\mathbf{x}_{11}(w_1)$	$\mathbf{x}_{12}(w_2)$	$\cdots$	$\mathbf{x}_{1B}(w_B)$	$\mathbf{x}_{1(B+1)}(1)$
X_2	$\mathbf{x}_{21}(1)$	$\mathbf{x}_{22}(w_1)$	$\cdots$	$\mathbf{x}_{2B}(w_{(B-1)})$	$\mathbf{x}_{2(B+1)}(w_B)$
X_3	$\mathbf{x}_{31}(1)$	$\mathbf{x}_{32}(l_1)$	$\cdots$	$\mathbf{x}_{3B}(l_{B-1})$	$\mathbf{x}_{3(B+1)}(l_B)$
$\hat{Y}_3$	$\hat{\mathbf{y}}_{31}(l_1\|1)$	$\hat{\mathbf{y}}_{32}(l_2\|l_1)$	$\cdots$	$\hat{\mathbf{y}}_{3B}(l_B\|l_{B-1})$	$\hat{\mathbf{y}}_{3B+1}(l_{B+1}\|l_B)$

Table A.1.: Coding scheme for the two-relay channel without block Markov coding at the source.

2) Node 3. For block $j = 1, \ldots, B+1$, node 3 finds an l_j such that

$$(\hat{\mathbf{y}}_{3j}(l_j|l_{j-1}),\mathbf{x}_{3j}(l_{j-1}), \mathbf{y}_{3j}) \in \mathcal{T}_\epsilon^n(P_{\hat{Y}_3X_3Y_3}) \tag{A.22}$$

as $n \to \infty$ if (see [20])

$$\hat{R} > I(\hat{Y}_3; Y_3|X_3). \tag{A.23}$$

Backward Decoding at the destination: Let $\epsilon_1 > \epsilon$.

Decoder 1: 1) Multihop l_{B+1} to node 4 in blocks $B+2$ to $B+3$.

2) For block $j = B, \ldots, 1$, node 4 puts out $(\hat{w}_j, \hat{l}_j)$, if there is a unique pair $(\hat{w}_j, \hat{l}_j)$ satisfying the following typicality checks in both blocks $j+1$ and j:

$$\left(\mathbf{x}_{1(j+1)}(w_{j+1}), \mathbf{x}_{2(j+1)}(\hat{w}_j), \mathbf{x}_{3(j+1)}(\hat{l}_j), \hat{\mathbf{y}}_{3(j+1)}(l_{j+1}|\hat{l}_j), \mathbf{y}_{4(j+1)}\right) \in \mathcal{T}_{\epsilon_1}^n\left(P_{X_1X_2X_3\hat{Y}_3Y_4}\right) \tag{A.24}$$

and

$$(\mathbf{x}_{1j}(\hat{w}_j), \mathbf{y}_{4j}) \in \mathcal{T}_{\epsilon_1}^n\left(P_{X_1Y_4}\right) \tag{A.25}$$

where w_{j+1} and l_{j+1} have already been reliably decoded from the previous block $j+1$. Otherwise it puts out $(\hat{w}_j, \hat{l}_j) = (1, 1)$.

Similar analysis as in Theorem 3.1 shows that node 4 can reliably recover (w_j, l_j) if

$$R < I(X_1; Y_4) + I(X_2; \hat{Y}_3Y_4|X_1X_3) \tag{A.26}$$

$$R < I(X_1X_2X_3; Y_4) - I(\hat{Y}_3; Y_3|X_1X_2X_3Y_4) \tag{A.27}$$

$$0 \leq I(X_3; Y_4|X_1X_2) - I(\hat{Y}_3; Y_3|X_1X_2X_3Y_4) \tag{A.28}$$

If the constraint (A.28) is violated, then the rate bound (A.27) becomes

$$R < I(X_1X_2; Y_4) \tag{A.29}$$

which is a stronger bound than (A.26) and can be achieved with SNNC-DF by treating X_3 as noise. Thus, we may ignore (A.28).

Decoder 2:

1) Multihop l_{B+1} and l_B to node 4 in blocks $B+2$ to $B+5$.

2) For block $j = B, \ldots, 1$, node 4 puts out $(\hat{w}_j, \hat{l}_{j-1})$ if there is a unique pair $(\hat{w}_j, \hat{l}_{j-1})$ satisfying the following typicality checks in both blocks $j+1$ and j:

$$(\mathbf{x}_{1(j+1)}(w_{j+1}), \mathbf{x}_{2(j+1)}(\hat{w}_j), \mathbf{x}_{3(j+1)}(l_j), \hat{\mathbf{y}}_{3(j)}(l_{j+1}|l_j), \mathbf{y}_{4(j+1)}) \in \mathcal{T}_{\epsilon_1}^n \left(P_{X_1X_2X_3\hat{Y}_3Y_4}\right) \tag{A.30}$$

and

$$\left(\mathbf{x}_{1j}(\hat{w}_j), \mathbf{x}_{3j}(\hat{l}_{j-1}), \hat{\mathbf{y}}_{3j}(l_j|\hat{l}_{j-1}), \mathbf{y}_{4j}\right) \in \mathcal{T}_{\epsilon_1}^n \left(P_{X_1X_3\hat{Y}_3Y_4}\right) \tag{A.31}$$

where w_{j+1}, l_j and l_{j+1} have already been reliably decoded from the previous block $j+1$. Otherwise it puts out $(\hat{w}_j, \hat{l}_{j-1}) = (1, 1)$.

Node 4 can reliably recover (w_j, l_{j-1}) if

$$R < I(X_1X_2; \hat{Y}_3Y_4|X_3) \tag{A.32}$$

$$R < I(X_1X_2X_3; Y_4) - I(\hat{Y}_3; Y_3|X_1X_2X_3Y_4) \tag{A.33}$$

$$0 \leq I(X_3; Y_4|X_1) - I(\hat{Y}_3; Y_3|X_1X_3Y_4) \tag{A.34}$$

If the constraint (A.34) is violated, then the rate bound (A.33) becomes

$$R < I(X_1; Y_4) + I(X_2; \hat{Y}_3Y_4|X_1X_3) \tag{A.35}$$

and the resulting R can be achieved by using decoder 1 (see (A.26)). Thus, with the

combination of both decoders, we may ignore (A.34) and achieve the rate (3.57).

Remark A.1. Sliding window decoding with 2 different decoders also recovers the rate (3.57) for independent X_1 and X_2 and enjoys a smaller decoding delay.

A.4. Rates and Outage for Gaussian Networks

In the following, let $C(x) = \log_2(1+x)$, $x \geq 0$.

A.4.1. Relay Channels

No Fading

The achievable rates R with DF and CF-S are given in [23]. The SNNC and LNNC rates are simply the CF-S rate. The SNNC-DF rate is the larger of the SNNC and DF rates.

Slow Rayleigh Fading

Define the events

$$\begin{aligned} D_{\text{DF}} &= \left\{ R_{\text{tar}} < C\left(|G_{12}|^2 P_1 (1-|\beta|^2) \right) \right\} \\ D_{\text{CF-S1}} &= \left\{ R_{2(\text{bin})} < C\left(\frac{|G_{23}|^2 P_2}{1+|G_{13}|^2 P_1} \right) \right\} \\ D_{\text{CF-S2}} &= \left\{ R_{2(\text{bin})} \geq C\left(\frac{1}{\hat{\sigma}_2^2} + \frac{|G_{12}|^2 P_1}{\hat{\sigma}_2^2 (1+|G_{13}|^2 P_1)} \right) \right\} \\ D_{\text{SNNC}} &= \left\{ \hat{\sigma}_2^2 \geq \frac{1}{|G_{23}|^2 P_2} \right\} \end{aligned} \tag{A.36}$$

where $|\beta|^2$ is the fraction of power allocated by source 1 to sending new messages. The optimal β, $R_{2(\text{bin})}$ and $\hat{\sigma}_2^2$ are calculated numerically.

The DF, CF-S, SNNC and SNNC-DF rates are

$$\begin{aligned} R_{\text{DF}} &= a_1 \\ R_{\text{CF-S}} &= b_1 \end{aligned}$$

$$
\begin{aligned}
R_{\text{SNNC}} &= c_1 \\
R_{\text{SNNC-DF}} &= \begin{cases} R_{\text{DF}} & \text{if } D_{\text{DF}} \text{ occurs} \\ R_{\text{SNNC}} & \text{otherwise} \end{cases}
\end{aligned} \tag{A.37}
$$

where

$$
\begin{aligned}
a_1 &= \min\Big\{C\left(|G_{12}|^2 P_1(1-|\beta|^2)\right), \\
&\qquad C\left(|G_{13}|^2 P_1 + |G_{23}|^2 P_2 + 2\Re\{\beta G_{13} G_{23}^*\}\sqrt{P_1 P_2}\right)\Big\} \\
b_1 &= \begin{cases} C\left(\frac{|G_{12}|^2 P_1}{1+\hat{\sigma}_2^2} + |G_{13}|^2 P_1\right) & \text{if } D_{\text{CF-S1}} \cap D_{\text{CF-S2}} \\ C\left(|G_{13}|^2 P_1\right) & \text{if } D_{\text{CF-S1}} \cap D_{\text{CF-S2}}^{\text{c}} \\ C\left(\frac{|G_{13}|^2 P_1}{1+|G_{23}|^2 P_2}\right) & \text{otherwise} \end{cases} \\
c_1 &= \begin{cases} \min\Big\{C\left(|G_{13}|^2 P_1 + |G_{23}|^2 P_2\right) - C(\frac{1}{\hat{\sigma}_2^2}), \\ C\left(\frac{|G_{12}|^2 P_1}{1+\hat{\sigma}_2^2} + |G_{13}|^2 P_1\right)\Big\} & \text{if } D_{\text{SNNC}} \\ C\left(\frac{|G_{13}|^2 P_1}{1+|G_{23}|^2 P_2}\right) & \text{otherwise} \end{cases}
\end{aligned} \tag{A.38}
$$

and $\Re\{x\}$ is the real part of x and x^* is the complex conjugate of x.

Remark A.2. For SNNC, event D_{SNNC} means that

$$
I(X_2; Y_3|X_1) - I(\hat{Y}_2; Y_2|X_1 X_2 Y_3) \geq 0 \tag{A.39}
$$

and the destination can reliably recover X_2 and $\hat{Y}_2$ *jointly* which helps to decode X_1. Otherwise the destination should treat X_2 as noise to get a better rate (see Theorem 3.1). Similarly, for CF-S the events $D_{\text{CF-S1}}$ and $D_{\text{CF-S2}}$ mean that both X_2 and $\hat{Y}_2$ can be decoded in a *step-by-step* fashion [21]. If $D_{\text{CF-S1}}$ and $D_{\text{CF-S2}}^{\text{c}}$ occur, then X_2 can be recovered which removes interference at the receiver. Otherwise the relay signal should be treated as noise.

As recognized in [60], one drawback of DF is that if the source-relay link happens to be weak and the relay tries to decode, then the rate suffers. Hence the relay should decode only if the source-relay link is strong enough to support R_{tar}, i.e., if event D_{DF} occurs. Otherwise, the relay should perform CF-S or QF. Different choices of relay operations

depending on the channel conditions lead to the achievable rates with SNNC-DF.

The outage probabilities are as follows:

$$\begin{aligned} P_{\text{DF}}^{\text{out}} &= \Pr[R_{\text{DF}} < R_{\text{tar}}] \\ P_{\text{CF-S}}^{\text{out}} &= \Pr[R_{\text{CF-S}} < R_{\text{tar}}] \\ P_{\text{SNNC}}^{\text{out}} &= \Pr[R_{\text{SNNC}} < R_{\text{tar}}] \\ P_{\text{SNNC-DF}}^{\text{out}} &= \Pr[R_{\text{SNNC-DF}} < R_{\text{tar}}] \end{aligned} \tag{A.40}$$

A.4.2. Two-Relay Channels

No Fading

The achievable DF rates are [23, Theorem 1]

$$R_{\text{DF}} < \max\{R_{\text{DF1}}, R_{\text{DF2}}\} \tag{A.41}$$

where

$$\begin{aligned} R_{\text{DF1}} &= \min\ \{a_{21}, a_{22}, a_{23}\} \\ R_{\text{DF2}} &= \min\ \{b_{21}, b_{22}, b_{23}\} \end{aligned} \tag{A.42}$$

with

$$\begin{aligned} a_{21} &= C\left(|\beta_1|^2|G_{12}|^2P_1\right) \\ a_{22} &= C\left((1-|\beta_3|^2)|G_{13}|^2P_1 + |\gamma_1|^2|G_{23}|^2P_2 + 2\Re\{\beta_2 G_{13}(\gamma_1 G_{23})^*\}\sqrt{P_1P_2}\right) \\ a_{23} &= C\Big(|G_{14}|^2P_1 + |G_{24}|^2P_2 + |G_{34}|^2P_3 \\ &\quad + \left(2\Re\{\beta_2 G_{14}(\gamma_1 G_{24})^*\} + 2\Re\{\beta_3 G_{14}(\gamma_2 G_{24})^*\}\right)\sqrt{P_1P_2} \\ &\quad + 2\Re\{\beta_3 G_{14}G_{34}^*\}\sqrt{P_1P_3} + 2\Re\{\gamma_2 G_{24}G_{34}^*\}\sqrt{P_2P_3}\Big) \\ b_{21} &= C\left(|\beta_1|^2|G_{13}|^2P_1\right) \\ b_{22} &= C\left((1-|\beta_3|^2)|G_{12}|^2P_1 + |\gamma_1|^2|G_{32}|^2P_3 + 2\Re\{\beta_2 G_{12}(\gamma_1 G_{32})^*\}\sqrt{P_1P_3}\right) \end{aligned}$$

$$
\begin{aligned}
b_{23} = C\Big(&|G_{14}|^2P_1 + |G_{24}|^2P_2 + |G_{34}|^2P_3 \\
&+ \left(2\Re\{\beta_2 G_{14}(\gamma_1 G_{34})^*\} + 2\Re\{\beta_3 G_{14}(\gamma_2 G_{34})^*\}\right)\sqrt{P_1P_3} \\
&+2\Re\{\beta_2 G_{14}G_{24}^*\}\sqrt{P_1P_2} + 2\Re\{\gamma_1 G_{24}G_{34}^*\}\sqrt{P_2P_3}\Big)
\end{aligned} \tag{A.43}
$$

where $\sum_{i=1}^{3}|\beta_i|^2 = 1$ and $\sum_{i=1}^{2}|\gamma_i|^2 = 1$ and the optimal power allocation parameters are calculated numerically.

The CF-S rates are (see [23, Theorem 2] with $U_i = 0$, $i = 2, 3$)

$$
R_{\text{CF-S}} < c_{21} \tag{A.44}
$$

subject to

$$
g_2 \le d_2,\ h_2 \le e_2,\ i_2 \le f_2
$$

where

$$
\begin{aligned}
c_{21} &= C\left(\frac{|G_{12}|^2P_1}{1+\hat{\sigma}_2^2} + \frac{|G_{13}|^2P_1}{1+\hat{\sigma}_3^2} + |G_{14}|^2P_1\right) \\
d_2 &= C\left(\frac{|G_{24}|^2P_2}{1+|G_{14}|^2P_1}\right) \\
e_2 &= C\left(\frac{|G_{34}|^2P_3}{1+|G_{14}|^2P_1}\right) \\
f_2 &= C\left(\frac{|G_{24}|^2P_2 + |G_{34}|^2P_3}{1+|G_{14}|^2P_1}\right) \\
g_2 &= C\left(\frac{1}{\hat{\sigma}_2^2} + \frac{|G_{12}|^2P_1}{\hat{\sigma}_2^2(1+\frac{|G_{13}|^2P_1}{(1+\hat{\sigma}_3^2)} + |G_{14}|^2P_1)}\right) \\
h_2 &= C\left(\frac{1}{\hat{\sigma}_3^2} + \frac{|G_{13}|^2P_1}{\hat{\sigma}_3^2(1+\frac{|G_{12}|^2P_1}{1+\hat{\sigma}_2^2} + |G_{14}|^2P_1)}\right) \\
i_2 &= C\left(\frac{1+\hat{\sigma}_2^2+\hat{\sigma}_3^2}{\hat{\sigma}_2^2\hat{\sigma}_3^2} + \frac{|G_{12}|^2P_1(1+\hat{\sigma}_3^2) + |G_{13}|^2P_1(1+\hat{\sigma}_2^2)}{\hat{\sigma}_2^2\hat{\sigma}_3^2(1+|G_{14}|^2P_1)}\right).
\end{aligned} \tag{A.45}
$$

The optimal $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ are calculated numerically.

Referring to Theorem 3.1, the achievable SNNC rates are

$$R_{\text{SNNC}} < \min\ \{c_{21}, j_{21}, j_{22}, j_{23}\} \tag{A.46}$$

where

$$\begin{aligned}
j_{21} = C\Bigg(&|G_{14}|^2P_1 + |G_{24}|^2P_2 + \frac{|G_{13}|^2P_1 + |G_{23}|^2P_2}{1+\hat{\sigma}_3^2} \\
&+\frac{P_1P_2(|G_{13}|^2|G_{24}|^2 + |G_{14}|^2|G_{23}|^2)}{1+\hat{\sigma}_3^2} \\
&-\frac{2\Re\{G_{13}G_{24}G_{14}^*G_{23}^*\}P_1P_2}{1+\hat{\sigma}_3^2}\Bigg) - C\left(\frac{1}{\hat{\sigma}_2^2}\right) \\
j_{22} = C\Bigg(&|G_{14}|^2P_1 + |G_{34}|^2P_3 + \frac{|G_{12}|^2P_1 + |G_{32}|^2P_3}{1+\hat{\sigma}_3^2} \\
&+\frac{P_1P_3(|G_{12}|^2|G_{34}|^2 + |G_{14}|^2|G_{32}|^2)}{(1+\hat{\sigma}_2^2)} \\
&-\frac{2\Re\left\{G_{12}G_{34}G_{14}^*G_{32}^*\right\}P_1P_3}{(1+\hat{\sigma}_2^2)}\Bigg) - C\left(\frac{1}{\hat{\sigma}_3^2}\right) \\
j_{23} = C\Big(&|G_{14}|^2P_1 + |G_{24}|^2P_2 + |G_{34}|^2P_3\Big) - C\left(\frac{1+\hat{\sigma}_2^2+\hat{\sigma}_3^2}{\hat{\sigma}_2^2\hat{\sigma}_3^2}\right).
\end{aligned} \tag{A.47}$$

where c_{21} is defined in (A.45). The optimal $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ are calculated numerically.

If one relay uses DF and the other uses QF, rates satisfying

$$R_{\text{DQF}} < \max\ \{R_{\text{DQF1}}, R_{\text{DQF2}}\} \tag{A.48}$$

can be achieved, where

$$\begin{aligned}
R_{\text{DQF1}} &= \min\ \{k_{21}, k_{22}, k_{23}\} \\
R_{\text{DQF2}} &= \min\ \{l_{21}, l_{22}, l_{23}\}
\end{aligned}$$

with

$$k_{21} = C\left(\frac{|G_{12}|^2P_1(1-|\theta|^2)}{1+|G_{32}|^2P_3}\right)$$

$$
\begin{aligned}
k_{22} &= C\Bigg(|G_{14}|^2P_1 + |G_{24}|^2P_2 + 2\Re\{\theta G_{14}G_{24}^*\}\sqrt{P_1P_2} \\
&\quad + \frac{|G_{13}|^2P_1 + |G_{23}|^2P_2 + 2\Re\{\theta G_{13}G_{23}^*\}\sqrt{P_1P_2}}{1+\hat{\sigma}_3^2} \\
&\quad + \frac{(1-|\theta|^2)P_1P_2(|G_{13}|^2|G_{24}|^2 + |G_{14}|^2|G_{23}|^2 - 2\Re\{G_{13}G_{24}G_{14}^*G_{23}^*\})}{1+\hat{\sigma}_3^2}\Bigg) \\
k_{23} &= C\left(|G_{14}|^2P_1 + |G_{24}|^2P_2 + |G_{34}|^2P_3 + 2\Re\{\theta G_{14}G_{24}^*\}\sqrt{P_1P_2}\right) - C\left(\frac{1}{\hat{\sigma}_3^2}\right) \qquad \text{(A.49)}
\end{aligned}
$$

and

$$
\begin{aligned}
l_{21} &= C\left(\frac{|G_{13}|^2P_1(1-|\theta|^2)}{1+|G_{23}|^2P_2}\right) \\
l_{22} &= C\Bigg(|G_{14}|^2P_1 + |G_{34}|^2P_3 + 2\Re\{\theta G_{14}G_{34}^*\}\sqrt{P_1P_3} \\
&\quad + \frac{|G_{12}|^2P_1 + |G_{32}|^2P_3 + 2\Re\{\theta G_{12}G_{32}^*\}\sqrt{P_1P_3}}{1+\hat{\sigma}_2^2} \\
&\quad + \frac{(1-|\theta|^2)P_1P_3(|G_{12}|^2|G_{34}|^2 + |G_{14}|^2|G_{32}|^2 - 2\Re\{G_{12}G_{34}G_{14}^*G_{32}^*\})}{1+\hat{\sigma}_2^2}\Bigg) \\
l_{23} &= C\left(|G_{14}|^2P_1 + |G_{24}|^2P_2 + |G_{34}|^2P_3 + 2\Re\{\theta G_{14}G_{34}^*\}\sqrt{P_1P_3}\right) - C\left(\frac{1}{\hat{\sigma}_2^2}\right) \qquad \text{(A.50)}
\end{aligned}
$$

where $0 \leq |\theta|^2 \leq 1$ and the optimal θ, $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ for R_{DQF1} and R_{DQF2} are calculated numerically.

Referring to Theorem 3.4, SNNC-DF achieves rates satisfying

$$
R_{\text{SNNC-DF}} < \max\{R_{\text{DF}}, R_{\text{DQF}}, R_{\text{SNNC}}\}. \qquad \text{(A.51)}
$$

Slow Rayleigh Fading

Define the events

$$
\begin{aligned}
D_{\text{DFV}} &= \left\{ \begin{array}{l} R_{\text{tar}} < V_{21} \\ R_{\text{tar}} < V_{22} \end{array} \right\} \\
D_{\text{DF1}} &= \{\ R_{\text{tar}} < k_{21}\} \\
D_{\text{DF2}} &= \{\ R_{\text{tar}} < l_{21}\}
\end{aligned}
$$

$$
\begin{aligned}
D_{\text{CF-S1}} &= \left\{ \begin{array}{l} R_{2(\text{bin})} < d_2 \\ R_{3(\text{bin})} < e_2 \\ R_{2(\text{bin})} + R_{3(\text{bin})} < f_2 \end{array} \right\} \\
D_{\text{CF-S2}} &= \left\{ \begin{array}{l} R_{2(\text{bin})} \geq g_2 \\ R_{3(\text{bin})} \geq h_2 \\ R_{2(\text{bin})} + R_{3(\text{bin})} \geq i_2 \end{array} \right\} \\
D_{\text{SNNC1}} &= \left\{ \begin{array}{l} |G_{24}|^2 P_2 + \frac{|G_{23}|^2 P_2}{1+\hat{\sigma}_3^2} \geq \frac{1}{\hat{\sigma}_2^2} \\ |G_{34}|^2 P_3 + \frac{|G_{32}|^2 P_3}{1+\hat{\sigma}_2^2} \geq \frac{1}{\hat{\sigma}_3^2} \\ |G_{24}|^2 P_2 + |G_{34}|^2 P_3 \geq \frac{1}{\hat{\sigma}_2^2} + \frac{1}{\hat{\sigma}_3^2} + \frac{1}{\hat{\sigma}_2^2 \hat{\sigma}_3^2} \end{array} \right\} \\
D_{\text{SNNC2}} &= \left\{ \hat{\sigma}_2^2 \geq \frac{1 + |G_{32}|^2 P_3 + |G_{34}|^2 P_3}{|G_{24}|^2 P_2} \right\} \\
D_{\text{SNNC3}} &= \left\{ \hat{\sigma}_3^2 \geq \frac{1 + |G_{23}|^2 P_2 + |G_{24}|^2 P_2}{|G_{34}|^2 P_3} \right\}
\end{aligned} \tag{A.52}
$$

where $\{V_{21}, V_{22}, V_{23}\}$ takes on the value $\{a_{21}, a_{22}, a_{23}\}$ or $\{b_{21}, b_{22}, b_{23}\}$ (see (A.43)) and the choice depends on the statistics of the fading coefficients such that the DF outage probability is minimized.

The DF rates are

$$
R_{\text{DF}} = \min \, \{V_{21}, V_{22}, V_{23}\} \, . \tag{A.53}
$$

The CF-S rates are

$$
R_{\text{CF-S}} = \left\{ \begin{array}{ll} c_{21} & \text{if } D_{\text{CF-S1}} \cap D_{\text{CF-S2}} \\ c_{22} & \text{if } D_{\text{CF-S1}} \cap D_{\text{CF-S2}}^{\text{c}} \\ c_{23} & \text{otherwise} \end{array} \right. \tag{A.54}
$$

where c_{21} is defined in (A.45) and

$$
\begin{aligned}
c_{22} &= C\left(|G_{14}|^2 P_1\right) \\
c_{23} &= C\left(\frac{|G_{14}|^2 P_1}{1 + |G_{24}|^2 P_2 + |G_{34}|^2 P_3}\right) .
\end{aligned} \tag{A.55}
$$

Observe that if both $D_{\text{CF-S1}}$ and $D_{\text{CF-S2}}$ occur, then both the bin and quantization indices can be decoded. If only $D_{\text{CF-S1}}$ occurs, then only the bin index can be recovered.

Referring to Theorem 3.1 the SNNC rates are

$$R_{\text{SNNC}} = \begin{cases} \min\ \{c_{21}, j_{21}, j_{22}, j_{23}\} & \text{if } D_{\text{SNNC1}} \\ \min\ \{m_{21}, m_{22}\} & \text{if } D^{\text{c}}_{\text{SNNC1}} \cap D_{\text{SNNC2}} \\ \min\ \{q_{21}, q_{22}\} & \text{if } D^{\text{c}}_{\text{SNNC1}} \cap D_{\text{SNNC3}} \\ c_{23} & \text{otherwise} \end{cases} \tag{A.56}$$

where

$$\begin{aligned} m_{21} &= C\left(\frac{P_1(|G_{12}|^2 + (1+\hat{\sigma}_2^2)|G_{14}|^2) + P_1P_3|G_{14}|^2|G_{32}|^2}{|G_{32}|^2P_3 + (1+\hat{\sigma}_2^2)\,(1+|G_{34}|^2P_3)}\right. \\ &\qquad \left. + \frac{P_1P_3(|G_{12}|^2|G_{34}|^2 - 2\Re\{G_{12}G_{34}G_{14}^*G_{32}^*\})}{|G_{32}|^2P_3 + (1+\hat{\sigma}_2^2)\,(1+|G_{34}|^2P_3)}\right) \\ m_{22} &= C\left(\frac{|G_{14}|^2P_1 + |G_{24}|^2P_2}{1+|G_{34}|^2P_3}\right) - C\left(\frac{1}{\hat{\sigma}_2^2} + \frac{|G_{32}|^2P_3}{\hat{\sigma}_2^2\,(1+|G_{34}|^2P_3)}\right) \\ q_{21} &= C\left(\frac{P_1(|G_{13}|^2 + (1+\hat{\sigma}_3^2)|G_{14}|^2) + P_1P_2|G_{14}|^2|G_{23}|^2}{|G_{23}|^2P_2 + (1+\hat{\sigma}_3^2)\,(1+|G_{24}|^2P_2)}\right. \\ &\qquad \left. + \frac{P_1P_2(|G_{13}|^2|G_{24}|^2 - 2\Re\{G_{13}G_{24}G_{14}^*G_{23}^*\})}{|G_{23}|^2P_2 + (1+\hat{\sigma}_3^2)\,(1+|G_{24}|^2P_2)}\right) \\ q_{22} &= C\left(\frac{|G_{14}|^2P_1 + |G_{34}|^2P_3}{1+|G_{24}|^2P_2}\right) - C\left(\frac{1}{\hat{\sigma}_3^2} + \frac{|G_{23}|^2P_2}{\hat{\sigma}_3^2\,(1+|G_{24}|^2P_2)}\right). \end{aligned} \tag{A.57}$$

The event D_{SNNC1} means that both quantization indices can be recovered. The events D_{SNNC2} and D_{SNNC3} mean that only one of the two quantization indices can be decoded. The SNNC-DF rates are

$$R_{\text{SNNC-DF}} = \begin{cases} R_{\text{DF}} & \text{if } D_{\text{DFV}} \\ R_{\text{DQF1}} & \text{if } D^{\text{c}}_{\text{DFV}} \cap D_{\text{DF1}} \\ R_{\text{DQF2}} & \text{if } D^{\text{c}}_{\text{DFV}} \cap D_{\text{DF2}} \\ R_{\text{SNNC}} & \text{otherwise} \end{cases} \tag{A.58}$$

where (see (A.49) and (A.50))

$$\begin{aligned} R_{\mathrm{DQF1}} &= \min\ \{k_{21}, k_{22}, k_{23}\} \\ R_{\mathrm{DQF2}} &= \min\ \{l_{21}, l_{22}, l_{23}\}\,. \end{aligned}$$

The outage probabilities are as in (A.40).

A.4.3. Multiple Access Relay Channels

No Fading

The DF region of the Gaussian MARC is the union of all (R_1, R_2) satisfying [38, Sec. 3]

$$\begin{aligned} R_1 < R_{\mathrm{DF1}} &= \min\ \{a_{31}, a_{32}\} \\ R_2 < R_{\mathrm{DF2}} &= \min\ \{b_{31}, b_{32}\} \\ R_1 + R_2 < R_{\mathrm{DF3}} &= \min\ \{c_{31}, c_{32}\} \end{aligned} \tag{A.59}$$

where

$$\begin{aligned} a_{31} &= C\left(|G_{13}|^2 P_1 (1-|\beta|^2)\right) \\ a_{32} &= C\left(|G_{14}|^2 P_1 + |G_{34}|^2 P_3 + 2\Re\{\beta G_{14}(\theta_1 G_{34})^*\}\sqrt{P_1 P_3}\right) \\ b_{31} &= C\left(|G_{23}|^2 P_2 (1-|\gamma|^2)\right) \\ b_{32} &= C\left(|G_{24}|^2 P_2 + |G_{34}|^2 P_3 + 2\Re\{\gamma G_{24}(\theta_2 G_{34})^*\}\sqrt{P_2 P_3}\right) \\ c_{31} &= C\left(|G_{13}|^2 P_1 (1-|\beta|^2) + |G_{23}|^2 P_2 (1-|\gamma|^2)\right) \\ c_{32} &= C\Big(|G_{14}|^2 P_1 + |G_{24}|^2 P_2 + |G_{34}|^2 P_3 + 2\Re\{\beta G_{14}(\theta_1 G_{34})^*\}\sqrt{P_1 P_3} \\ &\qquad + 2\Re\{\gamma G_{24}(\theta_2 G_{34})^*\}\sqrt{P_2 P_3}\Big) \end{aligned} \tag{A.60}$$

where $0 \le |\beta|^2, |\gamma|^2 \le 1$ and $\sum_{i=1}^{2} |\theta_i|^2 = 1$. The optimal power allocation parameters are calculated numerically.

The achievable CF-S rate region is the union of all (R_1, R_2) satisfying [38, Sec. 3]

$$\begin{aligned} R_1 &< d_{31} \\ R_2 &< e_{31} \\ R_1 + R_2 &< f_{31} \end{aligned} \tag{A.61}$$

where

$$\begin{aligned} d_{31} &= C\left(\frac{|G_{13}|^2 P_1}{1+\hat{\sigma}_3^2} + |G_{14}|^2 P_1\right) \\ e_{31} &= C\left(\frac{|G_{23}|^2 P_2}{1+\hat{\sigma}_3^2} + |G_{24}|^2 P_2\right) \\ f_{31} &= C\Bigg(|G_{14}|^2 P_1 + |G_{24}|^2 P_2 + \frac{|G_{13}|^2 P_1 + |G_{23}|^2 P_2}{1+\hat{\sigma}_3^2} \\ &\quad + \frac{P_1 P_2(|G_{13}|^2|G_{24}|^2 + |G_{14}|^2|G_{23}|^2 - 2\Re\{G_{13}G_{24}G_{14}^* G_{23}^*\})}{1+\hat{\sigma}_3^2}\Bigg) \end{aligned} \tag{A.62}$$

for some

$$\begin{aligned} \hat{\sigma}_3^2 \geq{}& \frac{1 + (|G_{13}|^2 + |G_{14}|^2)P_1 + (|G_{23}|^2 + |G_{24}|^2)P_2}{|G_{34}|^2 P_3} \\ &+ \frac{P_1 P_2(|G_{13}|^2|G_{24}|^2 + |G_{14}|^2|G_{23}|^2 - 2\Re\{G_{13}G_{24}G_{14}^* G_{23}^*\})}{|G_{34}|^2 P_3}. \end{aligned}$$

Referring to Theorem 3.1, the SNNC rate region is the union of pairs (R_1, R_2) satisfying

$$\begin{aligned} R_1 &< \min\ \{d_{31}, g_{31}\} \\ R_2 &< \min\ \{e_{31}, h_{31}\} \\ R_1 + R_2 &< \min\ \{f_{31}, i_{31}\} \end{aligned} \tag{A.63}$$

where d_{31}, e_{31} and f_{31} are defined in (A.62) and

$$\begin{aligned} g_{31} &= C\left(|G_{14}|^2 P_1 + |G_{34}|^2 P_3\right) - C\left(\frac{1}{\hat{\sigma}_3^2}\right) \\ h_{31} &= C\left(|G_{24}|^2 P_2 + |G_{34}|^2 P_3\right) - C\left(\frac{1}{\hat{\sigma}_3^2}\right) \end{aligned}$$

$$i_{31} = C\left(|G_{14}|^2 P_1 + |G_{24}|^2 P_2 + |G_{34}|^2 P_3\right) - C\left(\frac{1}{\hat{\sigma}_3^2}\right)$$

for some $\hat{\sigma}_3^2 > \frac{1}{|G_{34}|^2 P_3}$. The SNNC-DF rate region is the union of the SNNC and DF rate regions.

Slow Rayleigh Fading

Define the events

$$\begin{aligned}
D_{\text{DF}} &= \left\{ \begin{array}{l} R_{\text{tar1}} < a_{31} \\ R_{\text{tar2}} < b_{31} \\ R_{\text{tar1}} + R_{\text{tar2}} < c_{31} \end{array} \right\} \\
D_{\text{CF-S1}} &= \left\{ R_{3(\text{bin})} < C\left(\frac{|G_{34}|^2 P_3}{1 + |G_{14}|^2 P_1 + |G_{24}|^2 P_2}\right) \right\} \\
D_{\text{CF-S2}} &= \left\{ R_{3(\text{bin})} \geq C\left(\frac{1}{\hat{\sigma}_3^2} + \frac{|G_{13}|^2 P_1 + |G_{23}|^2 P_2}{\hat{\sigma}_3^2\left(1 + |G_{14}|^2 P_1 + |G_{24}|^2 P_2\right)} \right.\right. \\
&\qquad \left.\left. + \frac{P_1 P_2 (|G_{13}|^2 |G_{24}|^2 + |G_{14}|^2 |G_{23}|^2 - 2\Re\{G_{13} G_{24} G_{14}^* G_{23}^*\})}{\hat{\sigma}_3^2\left(1 + |G_{14}|^2 P_1 + |G_{24}|^2 P_2\right)} \right) \right\} \\
D_{\text{SNNC}} &= \left\{ \hat{\sigma}_3^2 \geq \frac{1}{|G_{34}|^2 P_3} \right\}.
\end{aligned} \tag{A.64}$$

The DF rate region of the Gaussian MARC is the union of all (R_1, R_2) satisfying (A.59). The CF-S rate region is the union of all (R_1, R_2) satisfying [38]

$$R_1 < R_{\text{CF1}} = \left\{ \begin{array}{ll} d_{31} & \text{if } D_{\text{CF-S1}} \cap D_{\text{CF-S2}} \\ d_{32} & \text{if } D_{\text{CF-S1}} \cap D_{\text{CF-S2}}^{\text{c}} \\ d_{33} & \text{otherwise} \end{array} \right. \tag{A.65}$$

$$R_2 < R_{\text{CF2}} = \left\{ \begin{array}{ll} e_{31} & \text{if } D_{\text{CF-S1}} \cap D_{\text{CF-S2}} \\ e_{32} & \text{if } D_{\text{CF-S1}} \cap D_{\text{CF-S2}}^{\text{c}} \\ e_{33} & \text{otherwise} \end{array} \right. \tag{A.66}$$

$$R_1 + R_2 < R_{\text{CF3}} = \begin{cases} f_{31} & \text{if } D_{\text{CF-S}_1} \cap D_{\text{CF-S2}} \\ f_{32} & \text{if } D_{\text{CF-S}_1} \cap D^{\text{c}}_{\text{CF-S2}} \\ f_{33} & \text{otherwise} \end{cases} \tag{A.67}$$

where

$$\begin{aligned} d_{32} &= C\left(|G_{14}|^2 P_1\right) \\ d_{33} &= C\left(\frac{|G_{14}|^2 P_1}{1+|G_{34}|^2 P_3}\right) \\ e_{32} &= C\left(|G_{24}|^2 P_2\right) \\ e_{33} &= C\left(\frac{|G_{24}|^2 P_2}{1+|G_{34}|^2 P_3}\right) \\ f_{32} &= C\left(|G_{14}|^2 P_1 + |G_{24}|^2 P_2\right) \\ f_{33} &= C\left(\frac{|G_{14}|^2 P_1 + |G_{24}|^2 P_2}{1+|G_{34}|^2 P_3}\right). \end{aligned} \tag{A.68}$$

If both $D_{\text{CF-S1}}$ and $D_{\text{CF-S2}}$ occur, then the relay bin and quantization indices can be decoded. If only $D_{\text{CF-S1}}$ occurs, then only the bin index can be recovered.
Referring to Theorem 3.1, the SNNC rate region is the union of all (R_1, R_2) satisfying

$$\begin{aligned} R_1 < R_{\text{SNNC1}} &= \begin{cases} \min\{d_{31}, g_{31}\} & \text{if } D_{\text{SNNC}} \\ d_{33} & \text{otherwise} \end{cases} \\ R_2 < R_{\text{SNNC2}} &= \begin{cases} \min\{e_{31}, h_{31}\} & \text{if } D_{\text{SNNC}} \\ e_{33} & \text{otherwise} \end{cases} \\ R_1 + R_2 < R_{\text{SNNC3}} &= \begin{cases} \min\{f_{31}, i_{31}\} & \text{if } D_{\text{SNNC}} \\ f_{33} & \text{otherwise.} \end{cases} \end{aligned} \tag{A.69}$$

The event D_{SNNC} means that the destination should decode the relay signal to achieve better performance.

The SNNC-DF rate region is the union of all (R_1, R_2) satisfying

$$\begin{aligned} R_1 < R_{\text{SNNC-DF1}} &= \begin{cases} R_{\text{DF1}} & \text{if} \quad D_{\text{DF}} \\ R_{\text{SNNC1}} & \text{otherwise} \end{cases} \\ R_2 < R_{\text{SNNC-DF2}} &= \begin{cases} R_{\text{DF2}} & \text{if} \quad D_{\text{DF}} \\ R_{\text{SNNC2}} & \text{otherwise} \end{cases} \\ R_1 + R_2 < R_{\text{SNNC-DF3}} &= \begin{cases} R_{\text{DF3}} & \text{if} \quad D_{\text{DF}} \\ R_{\text{SNNC3}} & \text{otherwise.} \end{cases} \end{aligned} \tag{A.70}$$

If D_{DF} occurs, then the relay should decode which will remove interference at the relay. Otherwise, the relay should perform QF to avoid unnecessarily lowering the rates.

Let $R_{\text{tar3}} = R_{\text{tar1}} + R_{\text{tar2}}$. The outage probabilities are:

$$\begin{aligned} P_{\text{DF}}^{\text{out}} &= \Pr[\{R_{\text{DF1}} < R_{\text{tar1}}\} \cup \{R_{\text{DF}_2} < R_{\text{tar2}}\} \cup \{R_{\text{DF}_3} < R_{\text{tar3}}\}] \\ P_{\text{CF-S}}^{\text{out}} &= \Pr[\{R_{\text{CF-S1}} < R_{\text{tar1}}\} \cup \{R_{\text{CF-S2}} < R_{\text{tar2}}\} \cup \{R_{\text{CF-S}} < R_{\text{tar3}}\}] \\ P_{\text{SNNC}}^{\text{out}} &= \Pr[\{R_{\text{SNNC1}} < R_{\text{tar1}}\} \cup \{R_{\text{SNNC2}} < R_{\text{tar2}}\} \cup \{R_{\text{SNNC3}} < R_{\text{tar3}}\}] \\ P_{\text{SNNC-DF}}^{\text{out}} &= \Pr[\{R_{\text{SNNC-DF1}} < R_{\text{tar1}}\} \cup \{R_{\text{SNNC-DF2}} < R_{\text{tar2}}\} \cup \{R_{\text{SNNC-DF3}} < R_{\text{tar3}}\}] \end{aligned} \tag{A.71}$$

A.4.4. Two-Way Relay Channels

No Fading

The DF rate region for the Gaussian TWRC is the union of all (R_1, R_2) satisfying

$$\begin{aligned} R_1 < R_{\text{DF1}} &= \min\ \{a_{41}, a_{42}\} \\ R_2 < R_{\text{DF2}} &= \min\ \{b_{41}, b_{42}\} \\ R_1 + R_2 < R_{\text{DF3}} &= c_{41} \end{aligned} \tag{A.72}$$

where

$$a_{41} = C\left(|G_{13}|^2 P_1 (1 - |\beta|^2)\right)$$

$$
\begin{aligned}
a_{42} &= C\left(|G_{12}|^2P_1 + |G_{32}|^2P_3(1-|\theta_1|^2) + 2\Re\{\beta G_{12}(\theta_1 G_{32})^*\}\sqrt{P_1P_3}\right)\\
b_{41} &= C\left(|G_{23}|^2P_2(1-|\gamma|^2)\right)\\
b_{42} &= C\left(|G_{21}|^2P_2 + |G_{31}|^2P_3(1-|\theta_1|^2) + 2\Re\{\gamma G_{21}(\theta_2 G_{31})^*\}\sqrt{P_2P_3}\right)\\
c_{41} &= C\left(|G_{13}|^2P_1(1-|\beta|^2) + |G_{23}|^2P_2(1-|\gamma|^2)\right)
\end{aligned}
\tag{A.73}
$$

where $0 \leq |\beta|^2, |\gamma|^2 \leq 1$ and $\sum_{i=1}^{2}|\theta_i|^2 = 1$. The optimal power allocation parameters are calculated numerically.

The CF-S rate region [61, Proposition 4] is the union of all (R_1, R_2) satisfying

$$
\begin{aligned}
R_1 &< d_{41}\\
R_2 &< e_{41}
\end{aligned}
\tag{A.74}
$$

where

$$
\begin{aligned}
d_{41} &= C\left(|G_{12}|^2P_1 + \frac{|G_{13}|^2P_1}{1+\hat{\sigma}_3^2}\right)\\
e_{41} &= C\left(|G_{21}|^2P_2 + \frac{|G_{23}|^2P_2}{1+\hat{\sigma}_3^2}\right)
\end{aligned}
$$

for some

$$
\hat{\sigma}_3^2 \geq \max\ \{f_{41}, f_{42}, f_{43}, f_{44}\}
$$

where

$$
\begin{aligned}
f_{41} &= \frac{1+|G_{12}|^2P_1+|G_{13}|^2P_1}{|G_{32}|^2P_3}\\
f_{42} &= \frac{|G_{21}|^2P_2+1}{|G_{31}|^2P_3} + \frac{|G_{13}|^2P_1(|G_{21}|^2P_2+1)}{|G_{31}|^2P_3(|G_{12}|^2P_1+1)}\\
f_{43} &= \frac{1+|G_{21}|^2P_2+|G_{23}|^2P_2}{|G_{31}|^2P_3},\\
f_{44} &= \frac{|G_{12}|^2P_1+1}{|G_{32}|^2P_3} + \frac{|G_{23}|^2P_2(|G_{12}|^2P_1+1)}{|G_{32}|^2P_3(|G_{21}|^2P_2+1)}.
\end{aligned}
$$

Referring to Theorem 3.1, the SNNC rate region is the union of all (R_1, R_2) satisfying

$$\begin{aligned} R_1 &< \min\{d_{41}, g_{41}\} \\ R_2 &< \min\{e_{41}, h_{41}\} \end{aligned} \tag{A.75}$$

where

$$\begin{aligned} g_{41} &= C\left(|G_{12}|^2 P_1 + |G_{32}|^2 P_3\right) - C\left(\frac{1}{\hat{\sigma}_3^2}\right) \\ h_{41} &= C\left(|G_{21}|^2 P_2 + |G_{31}|^2 P_3\right) - C\left(\frac{1}{\hat{\sigma}_3^2}\right) \end{aligned}$$

for some $\hat{\sigma}_3^2 > 0$. The SNNC-DF rate region is the union of the DF and SNNC rate regions.

Slow Rayleigh Fading

Define the events

$$\begin{aligned} D_{\text{DF}} &= \left\{ \begin{array}{l} R_{\text{tar}_1} < a_{41} \\ R_{\text{tar}_2} < b_{41} \\ R_{\text{tar}_1} + R_{\text{tar}_2} < c_{41} \end{array} \right\} \\ D_{\text{CF-S11}} &= \left\{ R_{3(\text{bin})} < C\left(\frac{|G_{31}|^2 P_3}{1 + |G_{21}|^2 P_2}\right) \right\} \\ D_{\text{CF-S12}} &= \left\{ R_{3(\text{bin})} \geq C\left(\frac{1}{\hat{\sigma}_3^2} + \frac{|G_{23}|^2 P_2}{\hat{\sigma}_3^2 (1 + |G_{21}|^2 P_2)}\right) \right\} \\ D_{\text{CF-S21}} &= \left\{ R_{3(\text{bin})} < C\left(\frac{|G_{32}|^2 P_3}{1 + |G_{12}|^2 P_1}\right) \right\} \\ D_{\text{CF-S22}} &= \left\{ R_{3(\text{bin})} \geq C\left(\frac{1}{\hat{\sigma}_3^2} + \frac{|G_{13}|^2 P_1}{\hat{\sigma}_3^2 (1 + |G_{12}|^2 P_1)}\right) \right\} \\ D_{\text{SNNC1}} &= \left\{ \hat{\sigma}_3^2 \geq \frac{1}{|G_{32}|^2 P_3} \right\} \\ D_{\text{SNNC2}} &= \left\{ \hat{\sigma}_3^2 \geq \frac{1}{|G_{31}|^2 P_3} \right\}. \end{aligned}$$

The DF region is the union of all (R_1, R_2) satisfying (A.72). The CF-S region is the union of all (R_1, R_2) satisfying

$$R_1 < R_{\text{CF-S1}} = \begin{cases} d_{41} & \text{if } D_{\text{CF-S21}} \cap D_{\text{CF-S22}} \\ d_{42} & \text{if } D_{\text{CF-S21}} \cap D^{\text{c}}_{\text{CF-S22}} \\ d_{43} & \text{otherwise} \end{cases} \tag{A.76}$$

$$R_2 < R_{\text{CF-S2}} = \begin{cases} e_{41} & \text{if } D_{\text{CF-S11}} \cap D_{\text{CF-S12}} \\ e_{42} & \text{if } D_{\text{CF-S11}} \cap D^{\text{c}}_{\text{CF-S12}} \\ e_{43} & \text{otherwise} \end{cases} \tag{A.77}$$

where

$$\begin{aligned} d_{42} &= C\left(|G_{12}|^2 P_1\right) \\ d_{43} &= C\left(\frac{|G_{12}|^2 P_1}{1+|G_{32}|^2 P_3}\right) \\ e_{42} &= C\left(|G_{21}|^2 P_2\right) \\ e_{43} &= C\left(\frac{|G_{21}|^2 P_2}{1+|G_{31}|^2 P_3}\right). \end{aligned}$$

The optimal $R_{3(\text{bin})}$ and $\hat{\sigma}_3^2$ are calculated numerically.
Referring to Theorem 3.1, SNNC achieves all pairs (R_1, R_2) satisfying

$$R_1 < R_{\text{SNNC1}} = \begin{cases} \min\ \{d_{41}, g_{41}\} & \text{if } D_{\text{SNNC1}} \\ d_{43} & \text{otherwise} \end{cases} \tag{A.78}$$

$$R_2 < R_{\text{SNNC2}} = \begin{cases} \min\ \{e_{41}, h_{41}\} & \text{if } D_{\text{SNNC2}} \\ e_{43} & \text{otherwise.} \end{cases} \tag{A.79}$$

The SNNC-DF rate region is the union of the (R_1, R_2) satisfying

$$R_1 < R_{\text{SNNC-DF1}} = \begin{cases} R_{\text{DF1}} & \text{if } D_{\text{DF}} \\ R_{\text{SNNC1}} & \text{otherwise} \end{cases} \tag{A.80}$$

$$R_2 < R_{\text{SNNC-DF2}} = \begin{cases} R_{\text{DF2}} & \text{if } D_{\text{DF}} \\ R_{\text{SNNC2}} & \text{otherwise.} \end{cases} \tag{A.81}$$

The outage probabilities are:

$$\begin{aligned} P_{\text{DF}}^{\text{out}} &= \Pr[\{R_{\text{DF1}} < R_{\text{tar1}}\} \cup \{R_{\text{DF}_2} < R_{\text{tar2}}\}] \\ P_{\text{CF-S}}^{\text{out}} &= \Pr[\{R_{\text{CF-S1}} < R_{\text{tar1}}\} \cup \{R_{\text{CF-S2}} < R_{\text{tar2}}\}] \\ P_{\text{SNNC}}^{\text{out}} &= \Pr[\{R_{\text{SNNC1}} < R_{\text{tar1}}\} \cup \{R_{\text{SNNC2}} < R_{\text{tar2}}\}] \\ P_{\text{SNNC-DF}}^{\text{out}} &= \Pr[\{R_{\text{SNNC-DF1}} < R_{\text{tar1}}\} \cup \{R_{\text{SNNC-DF2}} < R_{\text{tar2}}\}] \end{aligned} \tag{A.82}$$

B

Proofs for Chapter 5

B.1. Proof of Lemma 5.2: Non-Uniform W

Observe that $H(W) = H(B^{nR}) = nR \cdot H_2(p)$. Following the same steps as in (5.9) we have

$$\begin{aligned}
&\mathrm{E}[D(P_{V^n|\tilde{\mathcal{C}}}||Q_V^n)] = \mathrm{E}\left[\log \frac{\sum_{j=1}^M P(j) Q_{V^n|U^n}(V^n|U^n(j))}{Q_V^n(V^n)}\right] \\
&= \sum_w P(w) \cdot \mathrm{E}\left[\log \frac{\sum_{j=1}^M P(j) Q_{V|U}^n(V^n|U^n(j))}{Q_V^n(V^n)} \middle| W = w\right] \\
&\leq \sum_w P(w) \cdot \mathrm{E}\left[\log\left(\frac{P(w) Q_{V|U}^n(V^n|U^n(w))}{Q_V^n(V^n)} + 1 - P(w)\right)\right] \\
&\leq \sum_w P(w) \cdot \mathrm{E}\left[\log\left(\frac{P(w) Q_{V|U}^n(V^n|U^n(w))}{Q_V^n(V^n)} + 1\right)\right] \\
&= d_1 + d_2 + d_3
\end{aligned} \tag{B.1}$$

where

$$
\begin{aligned}
d_1 &= \sum_{w\in\mathcal{T}_\epsilon^n(P_X^n)} P(w) \sum_{(u^n(w),v^n)\in\mathcal{T}_\epsilon^n(Q_{UV}^n)} Q_{UV}^n(u^n(w),v^n)\left[\log\left(\frac{P(w)Q_{V|U}^n(v^n|u^n(w))}{Q_V^n(v^n)}+1\right)\right]\\
d_2 &= \sum_{w\in\mathcal{T}_\epsilon^n(P_X^n)} P(w) \sum_{\substack{(u^n(w),v^n)\notin\mathcal{T}_\epsilon^n(Q_{UV})\\(u^n(w),v^n)\in\mathrm{supp}(Q_{UV}^n)}} Q_{UV}^n(u^n(w),v^n)\left[\log\left(\frac{P(w)Q_{V|U}^n(v^n|u^n(w))}{Q_V^n(v^n)}+1\right)\right]\\
d_3 &= \sum_{\substack{w\notin\mathcal{T}_\epsilon^n(P_X^n)\\w\in\mathrm{supp}(P_X^n)}} P(w) \sum_{(u^n(w),v^n)\in\mathrm{supp}(Q_{UV}^n)} Q_{UV}^n(u^n(w),v^n)\left[\log\left(\frac{P(w)Q_{V|U}^n(v^n|u^n(w))}{Q_V^n(v^n)}+1\right)\right].
\end{aligned}
\tag{B.2}
$$

We can bound d_1 as follows (see (5.12))

$$
\begin{aligned}
d_1 &\le \sum_{w\in\mathcal{T}_\epsilon^n(P_X^n)} P(w)\left[\log\left(\frac{2^{n(I(V;U)+2\epsilon H(V))}}{2^{n(1-\epsilon)R\cdot H_2(p)}}+1\right)\right]\\
&\le \log\left(2^{-n(R\cdot H_2(p)-I(V;U)-\epsilon(2H(V)+R\cdot H_2(p)))}+1\right)\\
&\le \log(e)\cdot 2^{-n(R\cdot H_2(p)-I(V;U)-\delta_\epsilon(n))}
\end{aligned}
\tag{B.3}
$$

which goes to zero if $R > \frac{I(V;U)+\delta_\epsilon(n)}{H_2(p)}$ and $n\to\infty$, where $\delta_\epsilon(n)=\epsilon(2H(V)+R\cdot H_2(p))$. We also have

$$
\begin{aligned}
d_2 &\le \sum_{w\in\mathcal{T}_\epsilon^n(P_X^n)} P(w) \sum_{\substack{(u^n(w),v^n)\notin\mathcal{T}_\epsilon^n(Q_{UV})\\(u^n(w),v^n)\in\mathrm{supp}(Q_{UV}^n)}} Q_{UV}^n(u^n(w),v^n)\left[\log\left(\left(\frac{1}{\mu_V}\right)^n+1\right)\right]\\
&\le 2|\mathcal{V}|\cdot|\mathcal{U}|\cdot e^{-2n\epsilon^2\mu_{UV}^2}\log\left(\left(\frac{1}{\mu_V}\right)^n+1\right)
\end{aligned}
\tag{B.4}
$$

which goes to zero as $n\to\infty$ (see (5.13)). We further have

$$
\begin{aligned}
d_3 &\le \sum_{\substack{w\notin\mathcal{T}_\epsilon^n(P_X^n)\\w\in\mathrm{supp}(P_X^n)}} P(w) \sum_{(u^n(w),v^n)\in\mathrm{supp}(Q_{UV}^n)} Q_{UV}^n(u^n(w),v^n)\left[\log\left(\left(\frac{1}{\mu_V}\right)^n+1\right)\right]\\
&\le \sum_{\substack{w\notin\mathcal{T}_\epsilon^n(P_X^n)\\w\in\mathrm{supp}(P_X^n)}} P(w)\left[\log\left(\left(\frac{1}{\mu_V}\right)^n+1\right)\right]
\end{aligned}
$$

$$\leq 4 \cdot e^{-2n\epsilon^2 p^2} \log\left(\left(\frac{1}{\mu_V}\right)^n + 1\right) \tag{B.5}$$

which goes to zero as $n \to \infty$ (see (5.13)).

Combining the above for non-uniform W we have

$$\mathrm{E}[D(P_{V^n|\widetilde{\mathcal{C}}}||Q_V^n)] \to 0 \tag{B.6}$$

if $R > \frac{I(V;U)+\delta_n(\epsilon)}{H_2(p)}$ and $n \to \infty$.

B.2. Proof of Lemma 5.3

We extend the proof of [42, Sec. III, Inequality (15)] to asymptotic cases to establish Lemma 5.3. Recall that $-\frac{1}{2} \leq \rho \leq 0$. Let $s = \frac{-\rho}{1+\rho}$ so we have

$$\begin{aligned} 0 \leq s \leq 1 \\ 1 + s = \frac{1}{1+\rho} \end{aligned} \tag{B.7}$$

We also have for any $a, b \geq 0$ and $0 \leq x \leq 1$

$$(a+b)^x \leq a^x + b^x. \tag{B.8}$$

Observe that for any v^n we have

$$\begin{aligned} \mathrm{E}[P(v^n|\widetilde{\mathcal{C}})] &= \mathrm{E}\left[\sum_{w=1}^{M} \frac{1}{M} \cdot Q_{V|U}^n(v^n|U^n(w))\right] \\ &= \mathrm{E}\left[Q_{V|U}^n(v^n|U^n(1))\right] \\ &= \mathrm{E}\left[\prod_{i=1}^{n} Q_{V|U}(v_i|U_i(1))\right] \\ &= \prod_{i=1}^{n} \mathrm{E}\left[Q_{V|U}(v_i|U_i(1))\right] \\ &= \prod_{i=1}^{n} \left[\sum_u Q(u) Q_{V|U}(v_i|u)\right] \end{aligned}$$

$$= \prod_{i=1}^{n} Q_V(v_i) = Q_V^n(v^n). \tag{B.9}$$

We further have

$$\begin{aligned}
e^{E_0^n(\rho, Q_{UV}^n)} &= \sum_{v^n} \left\{ \mathrm{E}[P(v^n|\widetilde{\mathcal{C}})^{\frac{1}{1+\rho}}] \right\}^{1+\rho} \\
&\overset{(a)}{=} \sum_{v^n} \left\{ \mathrm{E}[P(v^n|\widetilde{\mathcal{C}})^{1+s}] \right\}^{\frac{1}{1+s}} \\
&= \sum_{v^n} \left\{ \mathrm{E}\left[\left(\sum_{w=1}^{M} \frac{1}{M} \cdot Q_{V|U}^n(v^n|U^n(w)) \right)^{1+s} \right] \right\}^{\frac{1}{1+s}} \\
&= \frac{1}{M} \sum_{v^n} \left\{ \mathrm{E}\left[\sum_{w=1}^{M} Q_{V|U}^n(v^n|U^n(w)) \left(Q_{V|U}^n(v^n|U^n(w)) + \sum_{j \neq w}^{M} Q_{V|U}^n(v^n|U^n(j)) \right)^{s} \right] \right\}^{\frac{1}{1+s}}
\end{aligned} \tag{B.10}$$

where (a) follows from (B.7). Applying (B.8) to (B.10) we have

$$\begin{aligned}
e^{E_0^n(\rho, Q_{UV}^n)} &\leq \frac{1}{M} \sum_{v^n} \left\{ \mathrm{E}\left[\sum_{w=1}^{M} Q_{V|U}^n(v^n|U^n(w)) \right. \right. \\
&\left. \left. \left(\left(Q_{V|U}^n(v^n|U^n(w)) \right)^{s} + \left(\sum_{j \neq w}^{M} Q_{V|U}^n(v^n|U^n(j)) \right)^{s} \right) \right] \right\}^{\frac{1}{1+s}} \\
&\overset{(a)}{=} \frac{1}{M} \sum_{v^n} \left\{ \mathrm{E}\left[\sum_{w=1}^{M} \left(Q_{V|U}^n(v^n|U^n(w)) \right)^{1+s} \right] \right. \\
&\qquad \left. + \sum_{w=1}^{M} \left(\mathrm{E}\left[Q_{V|U}^n(v^n|U^n(w)) \right] \right) \cdot \mathrm{E}\left[\left(\sum_{j \neq w}^{M} Q_{V|U}^n(v^n|U^n(j)) \right)^{s} \right] \right\}^{\frac{1}{1+s}} \\
&\overset{(b)}{\leq} \frac{1}{M} \sum_{v^n} \left\{ M\mathrm{E}\left[\left(Q_{V|U}^n(v^n|U^n) \right)^{1+s} \right] + MQ_V^n(v^n) \cdot \left(\mathrm{E}\left[\sum_{j \neq w}^{M} Q_{V|U}^n(v^n|U^n(j)) \right] \right)^{s} \right\}^{\frac{1}{1+s}} \\
&\overset{(c)}{=} \frac{1}{M} \sum_{v^n} \left\{ M\mathrm{E}\left[\left(Q_{V|U}^n(v^n|U^n) \right)^{1+s} \right] + MQ_V^n(v^n) \left((M-1) Q_V^n(v^n) \right)^{s} \right\}^{\frac{1}{1+s}} \\
&\leq \frac{1}{M} \sum_{v^n} \left\{ M\mathrm{E}\left[\left(Q_{V|U}^n(v^n|U^n) \right)^{1+s} \right] + \left(MQ_V^n(v^n) \right)^{1+s} \right\}^{\frac{1}{1+s}}
\end{aligned} \tag{B.11}$$

where

(a) follows because $U^n(w)$ is independent of $U^n(j)$, $j \neq w$

(b) follows by choosing $U^n V^n \sim Q_{UV}^n$, by the concavity of x^a for $0 \leq a \leq 1$ and by (B.9)

(c) follows by (B.9)

Applying (B.8) again to (B.11) we have

$$\begin{aligned}
e^{E_0^n(\rho, Q_{UV}^n)} &\leq \frac{1}{M} \sum_{v^n} \left\{ \left(M \mathrm{E}\left[\left(Q_{V|U}^n(v^n|U^n) \right)^{1+s} \right] \right)^{\frac{1}{1+s}} + M Q_V^n(v^n) \right\} \\
&\overset{(a)}{=} 1 + M^\rho \sum_{v^n} \left(\mathrm{E}\left[\left(Q_{V|U}^n(v^n|U^n) \right)^{\frac{1}{1+\rho}} \right] \right)^{1+\rho} \\
&= 1 + M^\rho \sum_{v^n} \left(\sum_{u^n} Q_U^n(u^n) \left(Q_{V|U}^n(v^n|u^n) \right)^{\frac{1}{1+\rho}} \right)^{1+\rho} \\
&\overset{(b)}{=} 1 + e^{n\rho R} \sum_{v} \left(\sum_{u} Q(u) \left(Q(v|u) \right)^{\frac{1}{1+\rho}} \right)^{n(1+\rho)} \\
&= 1 + e^{n(E_0(\rho, Q_{UV}) + \rho R)}
\end{aligned} \tag{B.12}$$

where

(a) follows from (B.7)

(b) follows because the $U_i V_i$ are i.i.d., $i = 1, \ldots, n$

Optimizing over ρ, we have

$$\begin{aligned}
E_0^n(\rho, Q_{UV}^n) &\leq \ln\left(1 + e^{nE_G(R, Q_{UV})}\right) \\
&\leq e^{nE_G(R, Q_{UV})}.
\end{aligned} \tag{B.13}$$

C

Abbreviations

List of Abbreviations

AWGN	additive white Gaussian noise
BCC	broadcast channel with confidential messages
BMC	block Markov coding
CF	compress-forward
CF-S	compress-forward with step-by-step decoding
DF	decode-forward
DMN	discrete memoryless network
i.i.d.	independent and identically distributed
LNNC	long message noisy network coding
MAC	multiple access channel
MARC	multiple access relay channel
NNC	noisy network coding
QF	quantize-forward

RHS	right-hand side
SNNC	short message noisy network coding
TWRC	two-way relay channel

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.

[2] H. Liao, "Multiple-access channels," Ph.D. dissertation, University of Hawaii, Honolulu, 1972.

[3] R. Ahlswede, "Multi-way communication channels," in *IEEE Int. Symp. Inf. Theory*, Tsahkadsor, Armenia, Sept. 1971, pp. 23–51.

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

[5] M. Yassaee and M. Aref, "Slepian-Wolf coding over cooperative networks," in *IEEE Int. Symp. Inf. Theory*, Seoul, Korea, June 2009, pp. 879–883.

[6] ——, "Slepian-Wolf coding over cooperative relay networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3462–3482, 2011.

[7] X. Wu and L.-L. Xie, "On the optimal compressions in the compress-and-forward relay schemes," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2613–2628, 2013.

[8] G. Kramer and J. Hou, "Short-message quantize-forward network coding," in *2011 8th Int. Workshop on Multi-Carrier Systems Solutions (MC-SS)*, Herrsching, Germany, May 2011, pp. 1–3.

[9] J. Hou and G. Kramer, "Short message noisy network coding for multiple sources," in *IEEE Int. Symp. Inf. Theory*, Boston, USA, July 2012, pp. 1677–1681.

[10] S. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," in *IEEE Inf. Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010, pp. 1–5.

[11] ——, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.

[12] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[13] A. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 789–804, March 2006.

[14] N. Ratnakar and G. Kramer, "The multicast capacity of deterministic relay networks with no interference," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2425–2432, June 2006.

[15] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.

[16] A. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, March 1975.

[17] T. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

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

[19] J. L. Massey, *Applied Digital Information Theory*, ETH Zurich, Zurich, Switzerland, 1980-1998.

[20] A. Orlitsky and J. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, March 2001.

[21] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[22] B. Schein, "Distributed coordination in network information theory," Ph.D. dissertation, MIT, Cambridge, MA, USA, 2001.

[23] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.

[24] M. Yassaee and M. Aref, "Generalized compress-and-forward strategy for relay networks," in *IEEE Int. Symp. Inf. Theory*, Toronto, Canada, July 2008, pp. 2683–2687.

[25] G. Kramer and J. Hou, "On message lengths for noisy network coding," in *IEEE Inf. Theory Workshop (ITW)*, Paraty, Brazil, Oct. 2011, pp. 430–431.

[26] P. Zhong, A. Haija, and M. Vu, "On compress-forward without Wyner-Ziv binning for relay networks," *submitted to IEEE Trans. Inf. Theory*, 2011. [Online]. Available: http://arxiv.org/abs/1111.2837/

[27] A. Raja and P. Viswanath, "Compress-and-forward scheme for a relay network: Approximate optimality and connection to algebraic flows," *submitted to IEEE Trans. Inf. Theory*. [Online]. Available: http://arxiv.org/abs/1012.0416/

[28] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[29] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

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

[31] L. Sankar, G. Kramer, and N. B. Mandayam, "Offset encoding for multiple-access relay channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3814–3821, 2007.

[32] J. Du, M. Xiao, M. Skoglund, and S. Shamai (Shitz), "Short-message noisy network coding with partial source cooperation," in *IEEE Inf. Theory Workshop (ITW)*, Lausanne, Switzerland, Sept. 2012, pp. 144–147.

[33] L.-L. Xie and P. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1348–1358, April 2005.

[34] L. Ozarow, S. Shamai, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. on Veh. Technol.*, vol. 43, no. 2, pp. 359–378, May 1994.

[35] G. Kramer and A. van Wijngaarden, "On the white gaussian multiple-access relay channel," in *IEEE Int. Symp. Inf. Theory*, Sorrento, Italy, June 2000, p. 40.

[36] L. Sankaranarayanan, G. Kramer, and N. Mandayam, "Capacity theorems for the multiple-access relay channel," in *42nd. Allerton Conf. Communications, Control, and Computing*, Sept. 2004.

[37] G. Kramer, P. Gupta, and M. Gastpar, "Information-theoretic multihopping for relay networks," in *Int. Zurich Seminar on Communications*, Zurich, Switzerland, Feb. 2004, pp. 192–195.

[38] L. Sankaranarayanan, G. Kramer, and N. Mandayam, "Hierarchical sensor networks: capacity bounds and cooperative strategies using the multiple-access relay channel model," in *2004 First Annual IEEE Commun. Soc. Conf. on Sensor and Ad Hoc Commun. and Networks*, Santa Clara, Oct. 2004, pp. 191–199.

[39] N. Gaarder and J. Wolf, "The capacity region of a multiple-access discrete memoryless channel can increase with feedback (corresp.)," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 100–102, Jan. 1975.

[40] T. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 27, no. 3, pp. 292–298, May 1981.

[41] L. Ozarow, "The capacity of the white gaussian multiple access channel with feedback," *IEEE Trans. on Inf. Theory*, vol. 30, no. 4, pp. 623–629, 1984.

[42] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, April 2006.

[43] M. Bloch and J. Kliewer, "On secure communication with constrained randomization," in *IEEE Int. Symp. Inf. Theory*, Boston, MA, USA, July 2012, pp. 1172–1176.

[44] I. Csiszár, "Almost independence and secrecy capacity," *Prob. of Inf. Transmission*, vol. 32, no. 1, pp. 40–47, Jan.–March 1996.

[45] A. Winter, "Secret, public and quantum correlation cost of triples of random variables," in *IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, Sept. 2005, pp. 2270–2274.

[46] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.

[47] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[48] P. Cuff, H. Permuter, and T. Cover, "Coordination capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4181–4206, Sept. 2010.

[49] U. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Advances in Cryptology - Eurocrypt 2000*. Lecture Notes in Computer Science, Springer-Verlag, 2000, pp. 351–368.

[50] M. Bloch and J. Barros, *Physical Layer Security From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[51] A. Wyner, "The wire-tap channel," *Bell Syst. Tech. Journal*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[52] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[53] G. Kramer, "Teaching IT: An identity for the Gelfand-Pinsker converse," *IEEE Inf. Theory Society Newsletter*, vol. 61, no. 4, pp. 4–6, Dec. 2011.

[54] J. Hou and G. Kramer, "Informational divergence approximations to product distributions," in *13th Canadian Workshop on Inf. Theory (CWIT)*, Toronto, Canada, June 2013, pp. 76–81.

[55] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York: Springer, 2005.

[56] B. Chern and A. Ozgur, "Achieving the capacity of the n-relay gaussian diamond network within logn bits," in *IEEE Inf. Theory Workshop (ITW)*, Lausanne, Switzerland, Sept. 2012, pp. 377–380.

[57] R. Kolte and A. Ozgur, "Improved capacity approximations for gaussian relay networks," in *IEEE Inf. Theory Workshop (ITW)*, Seville, Spain, Sept. 2013, pp. 1–5.

[58] ——, "Optimized noisy network coding for gaussian relay networks," in *Int. Zurich Seminar on Communications*, Zurich, Switzerland, Feb. 2014, pp. 140–143.

[59] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.

[60] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[61] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *IEEE Int. Symp. Inf. Theory*, Seattle, USA, July 2006, pp. 1668–1672.