Exploiting Structure in Cooperative Networks

BY

YANYING CHEN
B.Sc. Tianjin University, Tianjin, China, 2009

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Chicago, 2015

Chicago, Illinois

Defense Committee:

Natasha Devroye, Chair and Advisor
Dan Schonfeld
Hulya Seferoglu
Daniela Tuninetti
Young-Han Kim, University of California, San Diego

CONTRIBUTION OF AUTHORS

The content of this thesis is the result of a joint effort between my PhD advisor Professor Natasha Devroye and I.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF ABBREVIATIONS

AWGN	Additive White Gaussian Noise
DF	Decode-and-Forward
CF	Compress-and-Forward
i.i.d.	independent and identically distributed
MAC	Multiple access channel
BC	Broadcast channel
RC	Relay channel
ICF	Inverse Compute-and-Forward

SUMMARY

In this thesis we explore two examples of how to exploit structure in networks with cooperative nodes. 1) In the first direction, we explore the impact of *message structure* and how it may (or may not) be exploited to increase capacity depending on how this "matches" the channel's structure. As an example of this concept, the Inverse Compute-and-Forward (ICF) problem is proposed and studied, where we show that K-wise message correlations when $K > 2$, cannot be utilized to improve rate regions in a Gaussian MAC channel. 2) In the second direction we work towards explicitly exploiting *channel structure* in a zero-error primitive relay channel scenario. The problem of communicating over a primitive relay channel without error is for the first time proposed, with the goal of exploring and fulfilling the intuition that the central role of a relay is to *only* deliver *"what the destination needs"*. A novel relaying scheme termed "Colour-and-Foward" is proposed and is shown to be the most efficient way of compressing signals at the relay terminal, for any fixed number of channel uses, when enabling an effectively full cooperation between the relay and the destination terminals, i.e. achieving the single-input multi-output (SIMO) upper bound, is required. This Colour-and-Forward relaying is designed by an explicit exploit of the channel structure and directly embodies the intuition of having relay transmit "only what the destination needs".

CHAPTER 1

INTRODUCTION

Shannon theory studies the fundamental limits of communication: source coding considers how much one may compress data, while channel coding considers how fast one may reliably communicate data. To understand these limits, one needs to understand how to exploit different forms of structure in the problem. In source coding, the structure / form of the source distribution may be exploited to compress the data efficiently. In channel coding, one devises coding schemes to combat the particular channel or noise structure to ensure reliable communication. In this thesis we explore two examples of how to exploit structure in networks with cooperative nodes. 1) In the first direction, we explore the impact of *message structure* and how it may (or may not) be exploited to increase capacity depending on how this "matches" the channel's structure. 2) In the second direction we work towards explicitly exploiting *channel structure* in a zero-error primitive relay channel scenario.

In the first direction, we explore how to exploit *message structure* in a multiple access channel. One could ask whether mimicking message structure to create correlated codewords is always capacity achieving. We show that there is more to it than that: the message structure must somehow be "exploitable" by the channel's structure as well. As an example of this concept, the Inverse Compute-and-Forward (ICF) problem is proposed and studied, where we show that K-wise message correlations when $K > 2$, cannot be utilized to improve rate regions in a Gaussian MAC channel. The ICF problem, considers an L user multiple access channel where transmitter m has access to the linear equation $\mathbf{u}_m = \bigoplus_{l=1}^{L} f_{ml}\mathbf{w}_l$ of independent messages $\mathbf{w}_l \in \mathbb{F}_p^{k_l}$ with $f_{ml} \in \mathbb{F}_p$, and the destination wishes to recover all L messages. This problem may be motivated as the last hop in a network where relay nodes employ the Compute-and-Forward strategy and decode linear equations of messages; we seek to do the reverse and extract messages from sums over a multiple access channel. In particular, we exploit the particular form of correlation between the equations at the different relays – which does not map onto known results for multiple access channels with correlated sources – to improve the reliable communication rates beyond those achievable by simply forwarding all equations to the destination independently. The presented achievable rate region for the discrete memoryless channel model is furthermore shown to be capacity for the additive white Gaussian noise channel.

In the second direction, we focus on *channel structure* in the relay channel. We believe the relay channel highlights the role of channel structure in developing capacity-achieving relaying schemes. We are particularly interested in exploiting channel structure at the relay node to provide "what the destination terminal needs". We are not aware of an explicit attempt to quantify this intuition and hence potentially develop a new relaying strategy beyond the known Amplify, Decode or Compress-and-Forward schemes (and their variations). We start tackling this ambitious problem in a simpler setting than the general relay channel: 1) we focus on the primitive relay channel (PRC), which decouples the multiple access and broadcast components of the relay channel by having the link from the relay to the destination be out of band and of fixed capacity r; and 2) we focus on zero-error communication with finite channel inputs and outputs, which turns our problem into a combinatorial one. We believe it is also somewhat easier to see, and quantify, what the destination terminal "needs" in the zero-error setting. In particular, for zero-error communication over a PRC, we develop an exact quantitative measure of "what the destination needs" and "what the relay should (at least) send". This quantity is related to the minimum number of colors of a coloring on a graph G_R which captures "what the destination needs" and is constructed based on the channel structure: both the source to destination and the source to relay channel structures and how their relations are captured. This graph may also be used to develop associated relaying strategies which are shown to be capacity achieving for several classes of zero-error primitive relay channels.

1.1 Motivational Examples

1.1.1 Message structure, codebook structure and channel structure

In our first direction, we explore how to exploit message structure in a multiple access channel and show that, with some linear constraints (inherited from the lattice implementation of Compute-and-Forward framework) on the messages , K-wise message correlations when $K > 2$, cannot be utilized to improve rate regions in a Gaussian MAC channel. One might initially conjecture that a capacity-achieving scheme should somehow mimic the message structure in the codeword structure, but we show that this is not necessarily true. We illustrate the motivation behind this conjecture: the classical two-user discrete memoryless MAC (1), Slepian-Wolf MAC coding (2) and L-user discrete memoryless MAC (3) in which capacity achieving schemes mimic the message structure. We then point out a counterexample (the ICF problem, Chapter 2) to show the role of channel structure: the message structure must somehow be "exploitable" by the channel's structure as well.

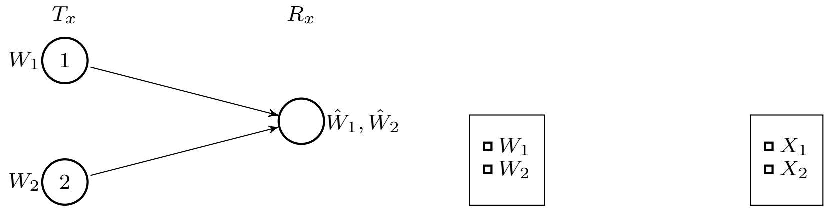
Let $W_1, \cdots, W_M$ denote messages to be communicated and let the dependence / independence among these M random variables be called the *message structure*. A codebook $X_l^n \in \mathcal{X}_l^n$ at terminal l consists of a collection of length-n vectors of alphabet $\mathcal{X}_l$. In the random coding framework, these are i.i.d. generated according to distribution of the random variable X_l with support $\mathcal{X}_l$. The *codebook structure* refers to the relationship among random variables X_l. Finally, the *channel structure* is characterized by the conditional probability mass function in the discrete memoryless channel or any relationship between the channel inputs and channel output. In the following examples, we adopt graphical models [1] to represent the relationship between/among random variables, thus indicating the message/codebook structure.

Example 1 (Two-user MAC with independent messages W_1, W_2)**.** *As shown in Fig. 1(a), two transition terminals with independent message W_1 and W_2, Tx-1 with W_1 and Tx-2 with W_2, want to transmit its own message to the destination terminal Rx simultaneously with a discrete memoryless channel $p(y|x_1, x_2)$ with channel input $\mathcal{X}_1 \times \mathcal{X}_2$ and output $\mathcal{Y}$.*

The capacity region is achieved by constructing random codebooks by P_{X_1} at Tx-1 and P_{X_2} at Tx-2 respectively, where random variable X_1 and X_2 are independent as the message W_2 and W_2 are. Fig. 1(b) and Fig. 1(c) are the graphical models, representing the message and codebook structure. The resulting capacity region (1) is, where we see that the codebook

---

[1] A graphical model is a probabilistic model for which a graph denotes the conditional dependence structure between random variables. Each arrow indicates a dependency.



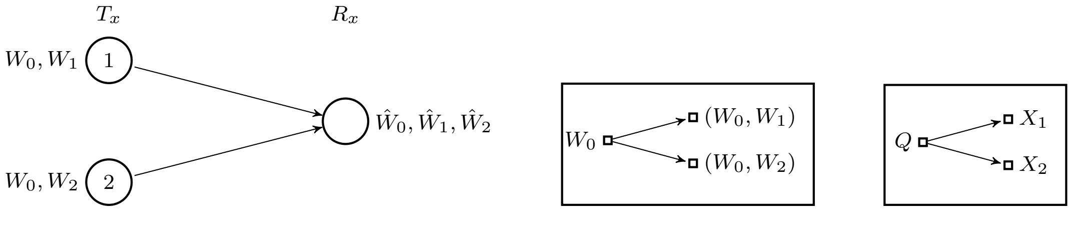
(a) Channel and communication goal (b) Message structure (c) Codebook structure

Figure 1. Two-user MAC with independent message W_1, W_2

structure mimics the message structure (cannot create correlation, or destroy it any further, so this is rather trivial):

$$\left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq I(X_1; Y|X_2) \\ R_2 \leq I(X_2; Y|X_1) \\ R_1 + R_2 \leq I(X_1, X_2; Y) \\ \text{ for } p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2) \end{array} \right\}.$$

Example 2 (Two-user MAC with one common message W_0 and two private messages W_1, W_2)**.** *As shown in Fig. 2(a), three messages W_0, W_1, W_2 are independent and are partially revealed to two terminals: terminals Tx-l has respectively a private message W_l and a common message W_0, $l = 1, 2$. The channel is the same as in the previous example and is characterized by $p(y|x_1, x_2)$.*



(a) Channel and communication goal (b) Message structure (c) Codebook structure

Figure 2. Two-user MAC with one common message W_0 and two private messages W_1, W_2

The capacity region is achieved by constructing random codebooks according to the Markov chain $X_1 \leftrightarrow Q \leftrightarrow X_2$. First, codebook $Q^n(W_0)$ with $\|\mathcal{W}_0\|$ codewords, indexed by common message W_0 is generated by the distribution of random variable Q. Then, codebook $X_l^n(W_l, W_0)$ at Tx-l is generated according to the conditional distribution $P_{X_l|Q}$, $l = 1, 2$. For example, when Terminal Tx-1 wants to communicate message pair $(W_0, W_1) = (w_0, w_1)$, sequence $q^n(w_0)$ is first chosen by codebook $Q^n(W_0)$ and then codeword $x_l^n(w_1, w_0)$ with probability $p(x_1^n(w_1, w_0)) = \prod_{i=1}^{n} p_{X_{1i}|Q=q_i(w_0)}(x_{1i})$ is transmitted. Fig. 2(b) and Fig. 2(c) are the graphical models, representing the message and codebook structure. The resulting capacity region (2) is as follows, where again the codebook structure seems to mimic the message structure:

$$\left\{ (R_1, R_2) : \begin{array}{l} R_1 \leq I(X_1; Y|X_2, Q), \\ R_2 \leq I(X_2; Y|X_1, Q), \\ R_1 + R_2 \leq I(X_1, X_2; Y|Q) \\ R_0 + R_1 + R_2 \leq I(X_1, X_2; Y) \\ \text{for } p(q, x_1, x_2, y) = p(q)p(x_1|q)p(x_2|q)p(y|x_1, x_2) \\ \text{ and } \|\mathcal{Q}\| \leq \min\{\|\mathcal{X}_1\| \cdot \|\mathcal{X}_2\| + 2, \|\mathcal{Y}\| + 3\} \end{array} \right\}.$$

Example 3 (L-user MAC with a Special Message Hierarchy (3))**.** *As shown in Fig. Figure 3 and Figure 4, L messages $W_1, \cdots, W_L$ are independent and are partially revealed to L terminals: terminal Tx-*1 *has access to all L messages $\{W_1, \cdots, W_L\}$; terminal Tx-*2 *knows all messages except for the first one, i.e. $\{W_, \cdots, W_L\}$; terminal Tx-*3 *knows all messages except for the first two, i.e $\{W_3, \cdots, W_L\}$ and so forth. The channel is characterized by $p(y|x_1, \cdots, x_L)$ with channel input $\mathcal{X}_1 \times \cdots \mathcal{X}_L$ and output $\mathcal{Y}$.*

The capacity region can be achieved by random codebooks generated by random variables $X_1, \cdots, X_L$ satisfying $p(x_1, \cdots, x_L) = p(x_L) \cdot p(x_{L-1}|x_L) \cdots p(x_1|x_2)$, as shown in 4(b). First, codebook $X_L^n(W_L)$ is generated according to distribution p_{X_L} and revealed to all terminals. Then, codebook $X_{L-1}^n(W_{L-1}, W_L)$ is generated according to $p_{X_{L-1}|X_L}$, producing codeword $x_{L-1}^n(w_{L-1}, w_L)$ with probability $p(x_{L-1}^n(w_{L-1}, w_L)) = \prod_{i=1}^{n} p_{X_{L-1}|X_L = x_{Li}(w_L)}(x_{(L-1)i})$. The re-

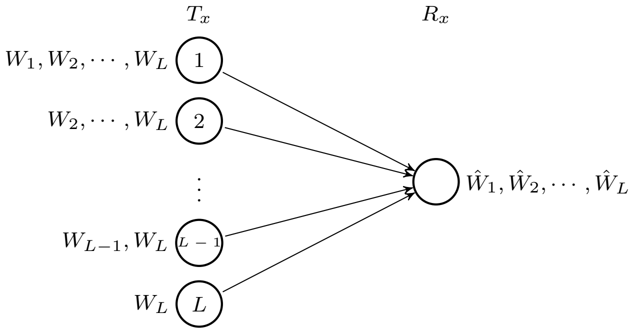
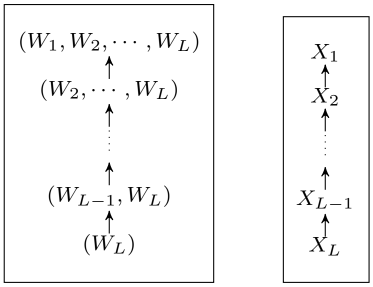


Figure 3. L-user MAC with a Special Message Hierarchy: channel and communication goal

maining codebooks are generated similarly. The resulting capacity region is the closure of the convex hull of all rate tuples satisfying:

$$\left\{ (R_1, \cdots, R_L) : \begin{array}{l} R_1 \leq I(X_1; Y | X_2, \cdots, X_L), \\ R_1 + R_2 \leq I(X_1, X_2; Y | X_3, \cdots, X_L), \\ \quad \vdots \\ R_1 + R_2 + \cdots + R_{L-1} \leq I(X_1, \cdots, X_{L-1}; Y | X_L) \\ R_1 + R_2 + \cdots + R_L \leq I(X_1, \cdots, X_L; Y) \\ \text{for } p(x_1, \cdots, x_L, y) = p(x_1|x_2) \cdots p(x_{L-1}|x_L) p(x_L) p(y|x_1, \cdots, x_L) \end{array} \right\}.$$

Unlike examples 1, 2 and 3, example 4 shows a case where capacity region is achieved while the “codebook structure” (Fig. 5(c)) is different from the “message structure” (Fig. 5(b)).

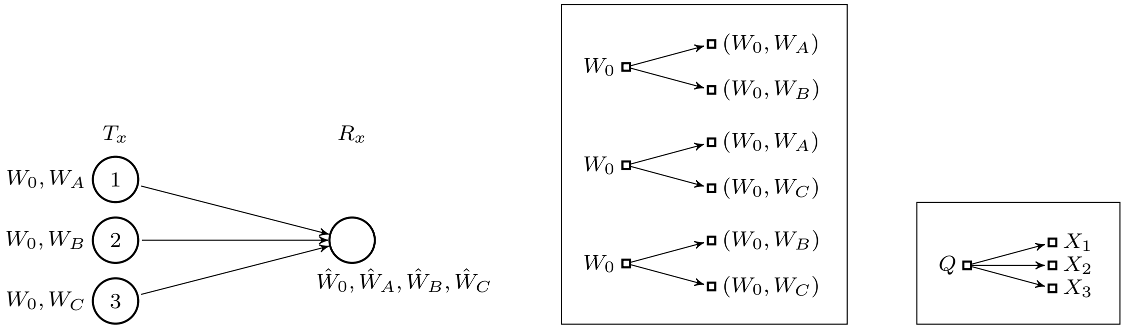


(a) Message structure (b) Codebook structure

Figure 4. L-user MAC with a Special Message Hierarchy

Example 4 (Three-user Gaussian MAC with hybrid message W_0, W_A, W_B, W_C)**.** *As shown in Fig. Figure 5, all terminals share one common message W_0 and terminal Tx-$1, 2, 3$ has access to private messages W_A, W_B and W_C respectively.* ***As shown in Fig. 5(b), messages W_A, W_B, W_C are pairwise independent conditioned on the common message W_0, but in general are not mutually independent given W_0.*** *The channel is specified by the input/output relationship $Y = X_1 + X_2 + X_3 + Z$, where $Z \sim \mathcal{N}(0, 1)$ is a Gaussian random variable.*

By the argument established in Chapter 2, one may show that the random codebooks generated from Q, X_1, X_2, X_3 satisfying $p(q, x_1, x_2, x_3) = p(q)p(x_1|q)p(x_2|q)p(x_3|q)$ as shown in Fig. 5(c) is capacity achieving. Comparing the Fig. 5(b) and Fig. 5(c), it is clear that capacity-achieving codebook structure is not consistent with the message structure. Actually,



(a) Channel and communication goal (b) Message structure (c) Codebook structure

Figure 5. Three-user Gaussian MAC with hybrid message W_0, W_A, W_B, W_C

it is shown that for a Gaussian MAC channel, any type of K-wise message correlations when $K > 2$, are not "exploitable" by the channel.

1.2 Contribution

The contributions center around demonstrating the exploration of message/channel/codebook structures and how to employ them to enhance the communication efficiency, i.e. enlarge the achievable rate region. We identity that codebook structure does not have to fully mimic the message structure to be capacity-achieving and the message structure should be "exploitable" by the channel's structure. In the primitive relay network, we construct a quantitative measure of "what the destination terminal needs" in communication in a zero-error context and develop a novel relaying strategy, termed as Colour-and-Forward, which is shown to be optimum – giving the most efficient compression at the relays while enabling the whole network to achieve its absolute maximal message rate – for any fixed number of channel use n.

1.3 Outline of thesis

In Chapter 2, the Inverse Compute-and-Forward problem is formulated and the capacity region for decoding L independent messages over a Gaussian multiple access channel when the L transmitters each have a linear equation of these L messages, subject to invertibility conditions, is derived. In Chapter 3, 4 and 5, the zero-error and small-error communication over a primitive relay channel are formulated and 0- and ϵ- Colour-and-Forward relaying are developed.

CHAPTER 2

INVERSE COMPUTE-AND-FORWARD

In this chapter, we explore how to exploit *message structure* in a multiple access channel. As an example of this concept, the Inverse Compute-and-Forward (ICF) problem is proposed and studied, where we show that K-wise message correlations when $K > 2$, cannot be utilized to improve rate regions in a Gaussian MAC channel.

2.1 Introduction

The recently proposed Compute-and-Forward (CF) framework (4) enables the decoding of linear combinations of messages at relays over Gaussian channels. The decoding of integer combinations of lattice codewords corresponds to decoding integer combinations of the underlying messages $\mathbf{u}$ which are vectors of length k of elements over a finite field of size p, $\mathbb{F}_p$, or $\mathbf{u} \in \mathbb{F}_p^k$. When decoding sums of messages suffices, this may sometimes be done at higher rates using the CF rates than decoding individual messages.

In the CF model, individual messages are transmitted over a multiple access channel (MAC), and linear combinations of messages are decoded[1]; in the inverse compute-and-forward (ICF) channel model studied here the reverse is done, i.e. a destination node seeks to decode individual messages over a MAC from relays which possess linear combinations of messages. In a larger

---

[1] The CF framework may handle more general cases when combinations of messages are transmitted as well, but this statement was made for the sake of argument/intuitive definition of the ICF model.

network one may envision source nodes having messages, destination nodes wanting to decode these messages, and intermediate relay nodes decoding individual or linear equations of messages according to the CF framework. We determine the rates at which we may extract individual messages from *linear message equations* known at relays over a MAC. This may be combined with CF rates in deriving overall achievable rates in larger networks. We provide some examples for doing so, but this is not the main focus of this study. For more works on multi-source multi-relay setup, please refer to (5), (6) and references therein.

We focus on the general L-user ICF problem where each relay node possesses a linear combination of L messages assumed to have been obtained using the CF framework. These relays transmit over a MAC to a single destination which seeks to decode the L individual messages. In order for the problem to be feasible, the matrix relating the messages to the equations must be invertible. The coefficient matrix is assumed to be non-singular throughout the study, and several additional invertibility constraints, for succinctness, will also be imposed. One might consider sending these L equations to the destination using independent codebooks as in a MAC, and having the destination invert the message equations to obtain the original messages. However, we show that the relays may extract dependencies from the linear equations when message rates are unequal, which allows one to achieve a larger rate region. In particular, we show that when message rates are unequal, 1) a common message may be extracted, 2) knowing some equations limits the number of values other equations may take on, and 3) there is a special pairwise (conditionally) independent structure in the equations.

Past Work. The problem statement and motivation builds upon the compute-and-forward (CF) framework (4): it is assumed that message equations have been previously decoded at the relays, and that messages are length k vectors of elements over a finite field $\mathbb{F}_p$, as in the CF framework. There are many other applications of CF, but they all differ from the ICF problem. For example, in (7), an integer-forcing linear receiver framework is developed for a MIMO system and is shown to outperform conventional linear receivers. Papers (8), (9) study a distributed antenna system (DAS) where antenna terminals, which serve user terminals, are connected to a central processor (CP) via digital error-free links of finite capacity. Both the up- and down-link can be facilitated by CF; we note that the "Reverse Compute and Forward" precoding strategy proposed in (9), should not be confused with the ICF problem proposed here. In these examples, linear equations are known at a single node (for the MIMO scenario) or can be gathered to a central node by some error-free links (in the DAS system). In contrast, the ICF problem studies how to directly extract the original messages over the air from the equations known to distributed nodes.

The ICF problem was first considered for the two-user case in (10), where an achievable rate region was presented. Though not formally presented in (10), one may show, as done here, that the two-user ICF problem may be mapped to sending one common message and two private messages over a MAC. This corresponds to the Slepian-Wolf MAC, whose capacity is known for both the discrete and Gaussian channels (2; 11; 12). The capacity of an extension of the Slepian-Wolf MAC of (2) to an arbitrary number of users, each of which has access to a subset of independent messages is solved in (13) and simplified in (14). We note that when going beyond two-users, this L-user ICF problem cannot be mapped into the framework in (13), as in the latter, the users either have common message(s) or completely independent ones, but do not have for example, the pairwise (but not mutual) independence correlation pattern. We are not aware of any other related problems which explicitly capture the pairwise independent structure. One might attempt to cast this problem into the framework considered by (15), as the transmission of arbitrarily correlated sources over a MAC channel via joint source-channel coding. We first remark that for the two-user case their achievable rate region results in the capacity region of the Slepian-Wolf MAC (2)[1], which also corresponds to the region obtained here for two users. More generally, in (15) only uncomputable multi-letter capacity expressions are presented for L arbitrarily correlated i.i.d. sources. In this work we strengthen the initial results of (10) considerably by obtaining the single-letter and fully-characterized capacity region for the general Gaussian L-user ICF problem rather than an achievable rate region for the two-user problem.

Contribution and Outline. The main contribution of this study is the derivation of the capacity region for decoding L independent messages over a Gaussian multiple access channel when each of L transmitters has a linear combination of these messages, subject to invertibility conditions. we first present the necessary definitions and formally state the general ICF problem in Section 2.2. Before demonstrating the most general results for arbitrary L, in Section 2.3

---

[1] As shown in the special case d) in (2), a channel-coding problem may be seen as a special case of the related joint source-channel coding problem, where messages are extended into information sources with the equivalent entropy rate while the channel model stays the same.

the $L = 2$ user case is used to build intuition. We provide plots of numerical evaluations of the ICF capacity region compared to other possible regions for this model, and an example of how to combine this rate region with a CF rate region to obtain an overall rate region for a relay network. In Section 2.4, the $L = 3$ user case is also outlined to build additional intuition for the new ingredient in moving beyond two users – pairwise independent but not mutually independent components at the transmitters. In Section 2.5, an achievable rate region for the general L-user ICF problem is first derived, followed by the capacity region for the Gaussian MAC channel model, the main contribution of this study. The converse follows along similar lines to those in (11; 12), but differs in an interesting way due to the special pairwise independent component of the message equations. In essence, for Gaussian channels, only pairwise dependency between equations is of concern and any correlations of order higher than 2 cannot be exploited to improve the rate regions.

Notation. Row vectors and matrices are written in bold font in lower and upper case, respectively. Length-n, $n \in \mathbb{N}$, vector codewords are represented by X^n. Define $C(x)$ as $\frac{1}{2}\log_2(1+x)$, $E[\cdot]$ as the expectation operator, and $\Pr[A]$ the probability of event A. Let $A \otimes B$ denote the Cartesian product of the sets A and B, and $\|A\|$ denote the cardinality of set A. $\|X^n\|$ also denotes the Euclidean norm of vector X^n. For p prime, let $\mathbb{F}_p^k \cong \{0, 1, \cdots, p-1\}^k$ ("$\cong$" indicates "is isomorphic to") denote the field of length k vectors of elements in the field $\mathbb{F}_p \cong \{0, 1, \cdots, p-1\}$, under element-wise addition/multiplication modulo p. Let $\mathrm{var}(X)$ denote the variance of X, $R_{\min} = \min\{R_1, \cdots, R_L\}$, and $R_{\max} = \max\{R_1, \cdots, R_L\}$. Let X_A denote the set $\{X_a, a \in A\}$ which contains all X_a with index a from a given set A. Similar notation is used to defined $\mathbf{w}_A$ (the set of messages with indices in the set A) and $\mathbf{u}_A$ (the set of equations with indices in the set A). We use the following indexing convention: l is used for sources ($\mathbf{w}$), m for relays ($\mathbf{u}$), and c for equation/message sections.

2.2 Problem Statement: Definitions and Channel Models

As shown in Figure 6, L source nodes indexed by l ($l = 1, \cdots, L$) would like to communicate with one destination node via L intermediate relay nodes indexed by m ($m = 1, \cdots, L$). The relays have successfully decoded the "message equations" $\mathbf{u}_m = \bigoplus_{l=1}^{L} f_{ml}\mathbf{w}_l$ (to be made precise below). The ICF problem seeks to determine at what rates these message equations may be transmitted over a MAC channel in order to decode the individual messages at a single destination. We make this more precise below, where we note that while definitions such as *messages* and *equations* follow the definitions in (4), new definitions of *message sections* and *equation sections* are needed to rigorously and compactly define the particular dependency structure between the equations, which impacts the description of the capacity region.

Definition 5 (Messages, Message rate)**.** *Source-l has message $\mathbf{w}_l$ ($l = 1, 2, \cdots, L$) which is uniformly drawn from $\mathbb{F}_p^{k_l} \cong \{0, 1, \cdots, p-1\}^{k_l}$, and viewed as a row vector of elements in $\mathbb{F}_p$ of length k_l. The messages of the different sources are independent. Without loss of generality, $k_1 \geq k_2 \geq \cdots \geq k_L$; all messages are zero-padded at the head to a common length $k = \max_l k_l$. For block length n, the message rate R_l of message $\mathbf{w}_l$ at source-l is defined as $R_l := \frac{1}{n}\log_2(p^{k_l})$. Let $\mathbf{W}$ denote the $L \times k$ matrix whose l-th row is the message $\mathbf{w}_l$. Note that $R_1 \geq R_2 \geq \cdots \geq R_L$.*

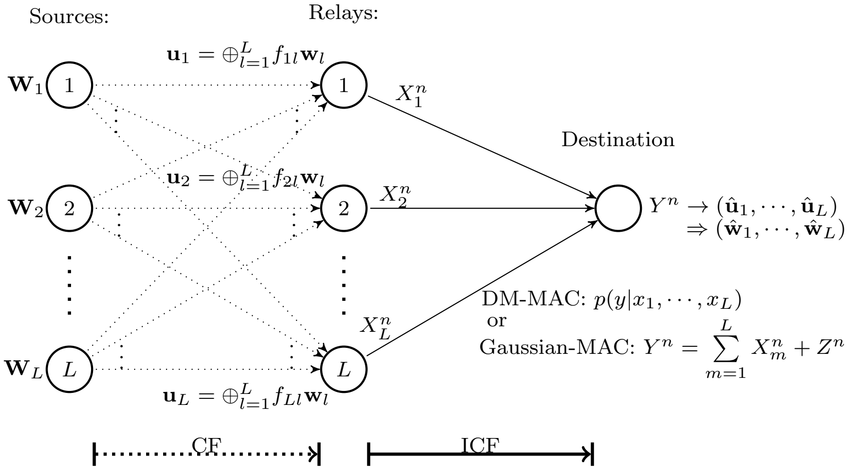

Sources:
Relays:
$\mathbf{u}_1 = \oplus_{l=1}^{L} f_{1l}\mathbf{w}_l$
$\mathbf{W}_1$
1
1
X_1^n
Destination
$\mathbf{u}_2 = \oplus_{l=1}^{L} f_{2l}\mathbf{w}_l$
$\mathbf{W}_2$
2
2
X_2^n
$Y^n \to (\hat{\mathbf{u}}_1, \cdots, \hat{\mathbf{u}}_L)$
$\Rightarrow (\hat{\mathbf{w}}_1, \cdots, \hat{\mathbf{w}}_L)$
DM-MAC: $p(y|x_1, \cdots, x_L)$
or
Gaussian-MAC: $Y^n = \sum_{m=1}^{L} X_m^n + Z^n$
X_L^n
$\mathbf{W}_L$
L
L
$\mathbf{u}_L = \oplus_{l=1}^{L} f_{Ll}\mathbf{w}_l$
CF
ICF

Figure 6. L-user ICF problem in which L relays each have a linear combination $\mathbf{u}_m = \oplus_{l=1}^{L} f_{ml}\mathbf{w}_l$ of L messages and wish to convey these messages to a single destination.

Definition 6 (Equations decoded at relays)**.** *Relay m, $m = 1, \cdots, L$, is assumed to have recovered a linear combination of the messages (as in the Compute-and-Forward framework (4)):* $\mathbf{u}_m = \bigoplus_{l=1}^{L} f_{ml}\mathbf{w}_l$ *in* $\mathbb{F}_p^k$*, for some given* $f_{ml} \in \mathbb{F}_p$*. In matrix form,*

$$\begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_L \end{pmatrix} = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1L} \\ \vdots & & & \vdots \\ f_{L1} & f_{L2} & \cdots & f_{LL} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_L \end{pmatrix}, \quad \text{or} \quad \mathbf{U} = \mathbf{F} \cdot \mathbf{W},$$

where $\mathbf{f}_m = (f_{m1}, \cdots, f_{mL})$*,* $\mathbf{U}^T = \left(\mathbf{u}_1^T, \cdots, \mathbf{u}_L^T\right)$*,* $\mathbf{F}^T = \left(\mathbf{f}_1^T, \cdots, \mathbf{f}_L^T\right)$*, and* $\mathbf{W}^T = \left(\mathbf{w}_1^T, \cdots, \mathbf{w}_L^T\right)$*. We note that each equation can take on* $2^{nR_{\max}} := 2^{n \max\{R_1, \cdots, R_L\}}$ *possible values.*

Remark 1. *Unless otherwise noted, we assume that* $\mathbf{F}$ *and all* c *by* c *sub-matrices from its first* c *columns are of full rank,* $c = 1, \cdots, L$*. This assumption is made to simplify notation and the derivation of the general* L*-user achievable rate region considerably. In particular, to recover all messages at the destination, all one needs is for* $\mathbf{F}$ *to be full rank; requiring specific sub-matrices to be full rank as well is not necessary to derive an achievable rate region. However, as will be outlined in examples in subsection 2.5.5, when some of the sub-matrices are not full rank one must carefully consider which equation sections (formally defined later) are linearly dependent. This in turn will affect the number and form of error events and hence rate region. While the derivation of achievable rate regions for individual cases is relatively straightforward, we have thus far not been able to come up with a compact, non-enumerative rate region for general* $\mathbf{F}$*. The current conditions on* $\mathbf{F}$ *come from the proof of Lemma 62 in Appendix A, which explains the properties of equation sections and leads to Lemma 64, which enumerates the number of equation sections and is used in the error analysis.*

Definition 7 (Memoryless MAC channel)**.** *The last hop of the network is a memoryless multiple access channel (MAC) defined by the conditional probability mass functions* $p(y|x_1, \cdots, x_L)$ *which are identical at each channel use and relate the channel inputs* $X_1^n, X_2^n, \cdots, X_L^n$ *in alphabets* $\mathbb{X}_m^n$ $(m = 1, 2, \cdots, L)$ *and the channel output* Y^n *in alphabet* $\mathbb{Y}^n$ *seen at the destination node. For the memoryless additive white Gaussian noise (AWGN) channel, all input and output alphabets are the real line, and this input/output relationship, over* n *channel uses, may be expressed as*

$$Y^n = \sum_{m=1}^{L} X_m^n + Z^n, \tag{2.1}$$

where Z^n *is i.i.d. Gaussian noise,* $Z^n \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{I}_{n\times n})$*, subject to power constraints* $E\left[\|X_m^n\|^2\right] \leq nP_m$*.*

Definition 8 (Encoding at relays)**.** *Each relay is equipped with an encoder,* $\mathcal{E}_m : \mathbb{F}_p^k \to \mathbb{X}_m^n$*, that maps the decoded equation* $\mathbf{u}_m$*, a length-*k *vector, to a length-*n *codeword, i.e,* $X_m^n = \mathcal{E}_m(\mathbf{u}_m) \in \mathbb{X}_m^n$*. For the Gaussian noise channel the encoders are further subject to power constraints* $E\left[\|X_m^n\|^2\right] \leq nP_m$*.*

Definition 9 (Decoding and probability of error)**.** *The destination wishes to recover the messages in* $\mathbf{W}$*. The decoder* $\mathcal{D}_1$ *at the destination node estimates the set of equations transmitted by the relays from the received signal, i.e.,* $\{\hat{\mathbf{u}}_1, \cdots, \hat{\mathbf{u}}_L\} = \mathcal{D}_1(Y^n)$*. We say that the equation set* $\{\mathbf{u}_1, \cdots, \mathbf{u}_L\}$ *are decoded with average probability of error* ϵ *if* $\Pr\left[\bigcup_{m=1}^{L} \{\hat{\mathbf{u}}_m \neq \mathbf{u}_m\}\right] < \epsilon$*.*

Definition 10 (Achievable, ICF achievable rate region)**.** *A rate tuple* $(R_1, \cdots, R_L)$ *is achievable if for any* $\epsilon > 0$ *and* n *large enough, there exist a sequence of encoders* $\mathcal{E}_1, \cdots, \mathcal{E}_L$ *and a decoder* $\mathcal{D}_1$ *such that the probability of error is bounded by* ϵ*. An ICF achievable rate region* $\mathcal{R}^{ICF}(R_1, \cdots, R_L)$ *is a set of achievable rate tuples for the ICF channel model.*

Definition 11 (ICF capacity region)**.** *The capacity region for the ICF problem* $\mathcal{C}^{ICF}(R_1, \cdots, R_L)$ *is the closure of the set of all achievable rate tuples.*

Remark 2. *Let the computation rate region* $\mathcal{R}^{CF}(R_1, \cdots, R_L)$ *defined in (4) capture the constraints on message rates imposed by the communication from source nodes to the last layer of relays. Then the intersection of* $\mathcal{R}^{CF}(R_1, \cdots, R_L)$ *and the ICF rate region* $\mathcal{R}^{ICF}(R_1, \cdots, R_L)$ *yields an achievable rate region for a larger network in which there is a single destination node* *desiring multiple messages. For succinctness, we omit the superscript ^{ICF} in most of the following as this study will only be focused on the ICF problem (rather than this intersection with CF rates).*

We now break up the messages and equations into sections, which will allow us to succinctly describe the dependency structure between the equations at different nodes.

Definition 12 (Message sections, Matrix of message sections)**.** *Message $\mathbf{w}_l \in \mathbb{F}_p^{k_l}$ is, after zero-padding at the head, a length-k row vector and may be partitioned into L segments $\mathbf{w}_{l,c}$ (the cth message section of message $\mathbf{w}_l$), $c = 1, \cdots, L$ (from head to tail) of lengths s_c and rates ρ_c where*

$$\begin{aligned} s_c &:= k_c - k_{c+1}, \\ \rho_c &:= \frac{1}{n} \log_2 p^{s_c} = R_c - R_{c+1}, \end{aligned} \tag{2.2}$$

with $k_{L+1} = 0$ and $R_{L+1} = 0$. Notice that $\sum_{c=1}^{L} s_c = k$ and $\sum_{c=1}^{L} \rho_c = R_{\max}$.

The matrix of the c-th message section is a matrix of dimension $L \times s_c$*, denoted by* $\tilde{\mathbf{W}}_{*c}$*. The l-th row of matrix* $\tilde{\mathbf{W}}_{*c}$ *is the c-th message section of message* $\mathbf{w}_l$*, i.e.,* $\mathbf{w}_{l,c}$*. Define the upper triangular matrix*

$$\begin{aligned}\tilde{\mathbf{W}}_{L\times L} &:= [\tilde{\mathbf{W}}_{*1}, \tilde{\mathbf{W}}_{*2}, \cdots, \tilde{\mathbf{W}}_{*L}] \\ &= \begin{pmatrix} \mathbf{w}_{1,1} & \mathbf{w}_{1,2} & \cdots & \mathbf{w}_{1,c-1} & \mathbf{w}_{1,c} & \cdots & \mathbf{w}_{1,L-1} & \mathbf{w}_{1,L} \\ \mathbf{0} & \mathbf{w}_{2,2} & \cdots & \mathbf{w}_{2,c-1} & \mathbf{w}_{2,c} & \cdots & \mathbf{w}_{2,L-1} & \mathbf{w}_{2,L} \\ \vdots & & & & & & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{w}_{l,c} & \cdots & \mathbf{w}_{l,L-1} & \mathbf{w}_{l,L} \\ \vdots & & & & & & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{w}_{L,L} \end{pmatrix}.\end{aligned} \tag{2.3}$$

Definition 13 (Equation sections, Matrix of equation sections)**.** *Similarly,* $\mathbf{u}_{m,c}$ *denotes the c-th section of equation* $\mathbf{u}_m$*, i.e.,* $\mathbf{u}_{m,c} := \mathbf{f}_m \cdot \tilde{\mathbf{W}}_{*c}$*. The matrix of c-th equation section* $\tilde{\mathbf{U}}_{*c}$ *has* $\mathbf{u}_{m,c}$ *as its m-th row, i.e.* $\tilde{\mathbf{U}}_{*c}^T := (\mathbf{u}_{1,c}^T, \mathbf{u}_{2,c}^T, \cdots, \mathbf{u}_{L,c}^T)$*. We have*

$$\tilde{\mathbf{U}}_{L\times L} := [\tilde{\mathbf{U}}_{*1}, \tilde{\mathbf{U}}_{*2}, \cdots, \tilde{\mathbf{U}}_{*L}] = \mathbf{F} \cdot [\tilde{\mathbf{W}}_{*1}, \tilde{\mathbf{W}}_{*2}, \cdots, \tilde{\mathbf{W}}_{*L}] \quad .$$

Definition 14 (Section rates)**.** *We denote as* ρ_c *the rate of section* $\mathbf{w}_{l,c}$ *or* $\mathbf{u}_{m,c}$*. Recall that* $\rho_c := \frac{1}{n}\log_2 p^{s_c} = R_c - R_{c+1}$ *with* $s_c := k_c - k_{c+1}$*,* $k_{L+1} = 0$ *and* $R_{L+1} = 0$*.*

2.3 Two-user Case

Before demonstrating the general L-user result, we consider the $L = 2$ user case in Figure 7 to build intuition. Recall that the matrix $\mathbf{F}$ is assumed to be non-singular and the first column should not have zeros, i.e., $f_{11}, f_{21} \neq 0$. In Figure 7 the first hop corresponds to the CF hop, and in the Gaussian case, at each channel use, $Y_3 = g_{13}X_1+g_{23}X_2+Z_3$ and $Y_4 = g_{14}X_1+g_{24}X_2+Z_4$.

In subsection 2.3.1, we briefly walk through three achievability schemes to show how dependency patterns may be created by the presence of interference at the relays, and how these may be exploited by different schemes in the ICF hop. In subsection 2.3.2, we numerically evaluate these three achievable rate regions for the Gaussian-MAC channel. An illustrative example of how CF and ICF rate regions may be combined – an interesting problem in itself but not the focus here – is provided in subsection 2.3.2.2. The takeaways are that 1) linear equations of messages create dependencies at the relays that may be exploited, and 2) in combining CF and ICF in a larger network, interference is not necessarily harmful and allows for the creation of such dependencies.

2.3.1 Three achievable rate regions for the two-user discrete memoryless ICF channel

Scheme 1: a non-coherent scheme without cardinality bounding. Ignoring the dependencies between the two equations and communicating the two equation indices (of rates

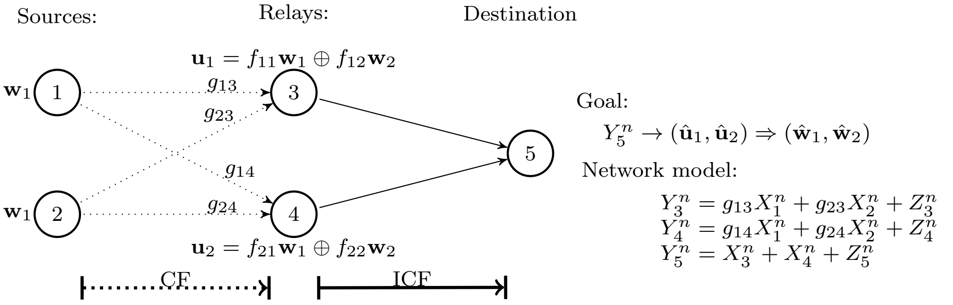

Sources:
Relays:
Destination
$\mathbf{u}_1 = f_{11}\mathbf{w}_1 \oplus f_{12}\mathbf{w}_2$
$\mathbf{w}_1$
1
g_{13}
3
g_{23}
Goal:
$Y_5^n \to (\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2) \Rightarrow (\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2)$
5
g_{14}
Network model:
$\mathbf{w}_1$
2
g_{24}
4
$Y_3^n = g_{13}X_1^n + g_{23}X_2^n + Z_3^n$
$Y_4^n = g_{14}X_1^n + g_{24}X_2^n + Z_4^n$
$Y_5^n = X_3^n + X_4^n + Z_5^n$
$\mathbf{u}_2 = f_{21}\mathbf{w}_1 \oplus f_{22}\mathbf{w}_2$
CF
ICF

Figure 7. Two-user ICF problem with Gaussian-MAC channel. Power constraints P_1, P_2, P_3, P_4, respectively.

$R_{\max} = \max\{R_1, R_2\}$ each) to the destination as if they were independent messages yields the rate region:

$$\mathcal{R}_{\text{Naive}}(R_1, R_2) = \left\{ (R_1, R_2) : \begin{array}{l} R_{\max} \le \min\{I(X_1; Y|X_2), I(X_2; Y|X_1)\} \\ R_{\max} + R_{\max} \le I(X_1, X_2; Y) \\ \text{for } p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2) \end{array} \right\}. \tag{2.4}$$

This region may be improved upon by properly accounting for the correlations between the two equations.

Scheme 2: a non-coherent scheme with cardinality bounding. Assuming $R_1 \ge R_2$, each equation may take on R_1 values. However, as $\mathbf{U} = \mathbf{F} \cdot \mathbf{W}$ and $\mathbf{F}$ is full rank, $(\mathbf{u}_1, \mathbf{u}_2)$ and $(\mathbf{w}_1, \mathbf{w}_2)$ are in one-to-one correspondence, and there are only $R_1 + R_2 \le 2R_1$ possibilities. Hence, sending the two equation indices independently is redundant whenever $R_1 \ne R_2$.

To exploit this, note that when one equation is fixed, the other may not take on all possible values in $\mathbb{F}_p^{k_1}$; this observation led to the "cardinality based approach" of (10), which resulted in the rate region:

$$\mathcal{R}_{\rm CB}(R_1,R_2) = \left\{ (R_1,R_2) : \begin{array}{c} R_{\min} \leq \min\{I(X_1;Y|X_2), I(X_2;Y|X_1)\} \\ R_1+R_2 \leq I(X_1,X_2;Y) \\ \text{for } p(x_1,x_2,y) = p(x_1)p(x_2)p(y|x_1,x_2) \end{array} \right\}. \tag{2.5}$$

The region $\mathcal{R}_{\rm CB}(R_1,R_2)$ improves over $\mathcal{R}_{\rm Naive}(R_1,R_2)$ as the error events are more carefully bounded (i.e. if one equation is correct, this limits the number of choices of the other equation). Inspection of $\mathcal{R}_{\rm CB}(R_1,R_2)$ reveals that the codewords are still independently generated which does not exploit the common messages present in the problem, and is generally not capacity achieving.

Scheme 3: a capacity-achieving coherent coding scheme with cardinality bounding. The relays, which have $\mathbf{u}_1$ and $\mathbf{u}_2$, actually share a common message – the message section $\mathbf{w}_{1,1}$ of the rate ρ_1 message $\mathbf{w}_1$, in addition to each having a private, independent message of rate ρ_2 ($\mathbf{u}_{1,2} = f_{11}\mathbf{w}_{1,2} + f_{12}\mathbf{w}_{2,2}$ or $\mathbf{u}_{2,2} = f_{21}\mathbf{w}_{1,2} + f_{22}\mathbf{w}_{2,2}$). We may map the two-user ICF problem into the Slepian-Wolf MAC problem (2) (which in turn may be seen as Special case d) of joint-source-channel coding over a MAC as studied in (15)) of a two-user MAC with a common message and two private messages. This idea is first expressed in (10), but was not fully explored, and yields the region:

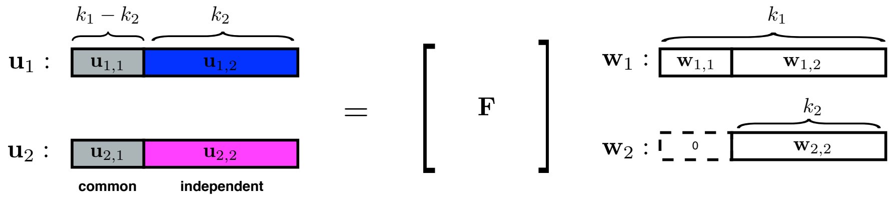


Figure 8. Two-user ICF message/equation structure. Grey indicates that equation sections $\mathbf{u}_{1,1}$ and $\mathbf{u}_{2,1}$ are fully correlated, while different solid colors indicate that two equation sections $\mathbf{u}_{1,2}$ and $\mathbf{u}_{2,2}$ are independent. All message sections $\mathbf{w}_{i,j}$ are mutually independent; $i, j = 1, 2$.

$$\mathcal{R}_{\mathrm{ICF}}(R_1, R_2) = \left\{ (R_1, R_2) : \begin{array}{l} R_{\min} \leq \min\left\{ I(X_1; Y|X_2, Q), I(X_2; Y|X_1, Q), \frac{1}{2} I(X_1, X_2; Y|Q) \right\} \\ R_1 + R_2 \leq I(X_1, X_2; Y) \\ \text{for } p(q, x_1, x_2, y) = p(q)p(x_1|q)p(x_2|q)p(y|x_1, x_2) \end{array} \right\}. \tag{2.6}$$

The cardinality of the alphabet of Q may be bounded as $||\mathcal{Q}|| \leq \min\left\{ ||\mathcal{X}_1|| \cdot ||\mathcal{X}_2|| + 2, ||\mathcal{Y}|| + 3 \right\}$.

Remark 3. *Any rate pair achieved by Scheme 2 can be achieved by the capacity-achieving Scheme 3 by setting $Q = \emptyset$. Comparing these two regions, the left hand sides of the inequalities are identical, but the right hand sides have increased due to the possible correlation of the codewords created through Q, i.e. $I(X_1, X_2; Y)$ maximized over $\{p(q)p(x_1|q)p(x_2|q)p(y|x_1, x_2)\}$ is generally larger than the maximum evaluated over $\{p(x_1)p(x_2)p(y|x_1, x_2)\}$.*

2.3.2 Numerical comparison

Consider the AWGN channel model in Fig. Figure 7 with $g_{13} = g_{23} = g_{24} = 1$, $g_{14} = -1$ in the first hop, and symmetrize the powers as $P_{\rm s} = P_1 = P_2$. Note that one can easily obtain regions for general g_{ij} and power constraints, but that this is not the focus of this work.

2.3.2.1 Numerical comparison of three two-user ICF only rate regions

We now numerically evaluate the three achievable rate regions of Schemes 1, 2, and 3 for the ICF hop only of an additive Gaussian noise channel as shown in Fig. Figure 7, where we recall that all noises are i.i.d. unit variance Gaussians, i.e. $Z_i^n \sim \mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{I}_{n\times n})$, $i = 3, 4, 5$. Scheme 1 and 2 lead to the regions $\mathcal{R}^{\mathcal{G}}_{\rm Naive}(R_1, R_2)$ and $\mathcal{R}^{\mathcal{G}}_{\rm CB}(R_1, R_2)$, which correspond to those in (Equation 2.4) and (Equation 2.5) for Gaussian inputs:

$$\mathcal{R}^{\mathcal{G}}_{\rm Naive}(R_1, R_2) = \left\{ (R_1, R_2) : R_{\max} \leq \min\{C(P_3), C(P_4), \frac{1}{2} \cdot C(P_3 + P_4)\} \right\}, \tag{2.7}$$

$$\mathcal{R}^{\mathcal{G}}_{\rm CB}(R_1, R_2) = \left\{ (R_1, R_2) : \begin{array}{c} R_{\min} \leq \min\{C(P_3), C(P_4)\} \\ \\ R_1 + R_2 \leq C(P_3 + P_4) \end{array} \right\}. \tag{2.8}$$

Scheme 3 has been shown to be exhausted by jointly Gaussian inputs (11), yielding the region

$\bigcup_{b_1,b_2\in[0,1]} \mathcal{R}^{\mathcal{G}}_{\text{ICF}}(R_1, R_2 \,|\, b_1, b_2)$, where for each pair of constants $b_1, b_2 \in [0,1]$ we define

$$\mathcal{R}^{\mathcal{G}}_{\text{ICF}}(R_1, R_2 \,|\, b_1, b_2) = \left\{ (R_1, R_2) : \begin{array}{c} R_{\min} \leq \min\bigl\{C((1-b_1)P_3), C((1-b_2)P_4), \\ \frac{1}{2}C((1-b_1)P_3 + (1-b_2)P_4)\bigr\} \\ R_1 + R_2 \leq C(P_3 + P_4 + 2\sqrt{b_1 b_2}\sqrt{P_3 P_4}) \end{array} \right\}. \tag{2.9}$$

Fig. 9(a) demonstrates the relative rate regions of the three schemes for equal relay power $P_3 = P_4 = 20$, while Fig. 9(b) demonstrates the regions for asymmetric powers $P_3 = 4, P_4 = 36$. From the figure, one can see how Scheme 3 improves upon Scheme 2 (coherent gains), that in turn improves upon Scheme 1 (proper accounting of dependencies in error events). Coherent gains are most useful for unequal R_1 and R_2; when $R_1 = R_2$, all regions degrade to the same line segment depicted using thick black dots. This is intuitive: at equal rates there are no common messages and the two linear equations known to the relays are independent and no dependencies may be extracted or exploited. One may also observe that when the powers at the relays (nodes 3,4) are asymmetric but sum to the same value, the gains of Scheme 2 over Scheme 1 increase while the gains of Scheme 3 over Scheme 2 decrease. The region of Scheme 1 decreases as the powers become more asymmetric as the regular MAC channel region is constrained by the minimum of the powers at the relays. The region of Scheme 3 also decreases with increasing asymmetry in powers: the coherent gain manifests itself in the sum-rate as an additional term $\sqrt{P_3P_4}$. For fixed sum $P_3 + P_4$ this is maximized when they are equal.

2.3.2.2 An example: combining CF and ICF in a network

We now illustrate how ICF may be combined with the CF rate region to provide an overall achievable rate region in an AWGN relay network.

In the first hop, or the CF stage, since the channel gain to receiver 3 is $Y_3 = X_1 + X_2 + Z_3$ and that to receiver 4 is $Y_4 = X_1 - X_2 + Z_4$, the relay nodes 3 and 4 may decode equations $\mathbf{u}_1 = \mathbf{w}_1 \oplus \mathbf{w}_2$ and $\mathbf{u}_2 = \mathbf{w}_1 \ominus \mathbf{w}_2$ (which intuitively match the channel gains) using the CF framework at rates (4). Next, in the ICF stage, destination node 5 recovers $(\mathbf{w}_1, \mathbf{w}_2)$ from $(\mathbf{u}_1, \mathbf{u}_2)$ at rates:

$$\text{First hop: } \begin{cases} R_1 \leq \frac{1}{2}\log\left(\frac{1}{2} + P_{\mathrm{s}}\right) \\ \\ R_2 \leq \frac{1}{2}\log\left(\frac{1}{2} + P_{\mathrm{s}}\right) \end{cases} \qquad \text{Second hop: region (Equation 2.9).} \tag{2.10}$$

To obtain an achievable rate region for the entire network, first intersect the CF and ICF rate regions in (Equation 2.10) and then take the convex hull of the resulting regions. As one can see in Fig. 10(b), the achievable rate region for the whole network when using CF + ICF Scheme 3, improves upon Scheme 2, that in turns improves upon Scheme 1. Note that when looking at only the ICF rate region, at equal rates Scheme 3 does *not* outperform the other schemes. However, when combined with the CF region in a larger network, using CF + ICF (scheme 3) outperforms the other schemes. This is because source nodes 1,2 may transmit at unequal rates (which maximizes the benefits of Scheme 3's coherent gains in the ICF phase), and then use time sharing between this and the reverse unequal rates to achieve the larger rate region.

2.3.2.3 Comparison with the scheme of decode and forward and full cooperation (DF+FCo)

One alternative approach for the two-hop network is to have both relays in the first hop decode and forward (DF) the two messages $\mathbf{w}_1$ and $\mathbf{w}_2$. This allows them to fully cooperate (FCo) in the second hop. This leads to the following achievable rate regions, which again must be intersected and then convex-hulled:

$$\text{First hop: } \begin{cases} R_1 \leq \frac{1}{2}\log\left(1+P_\mathrm{s}\right) \\ R_2 \leq \frac{1}{2}\log\left(1+P_\mathrm{s}\right) \\ R_1+R_2 \leq \frac{1}{2}\log\left(1+2P_\mathrm{s}\right) \end{cases} \qquad \text{Second hop: } \begin{cases} R_1+R_2 \leq \frac{1}{2}\log\left(1+P_3+P_4+2\sqrt{P_3P_4}\right). \end{cases} \tag{2.11}$$

As one can see from the expressions in equation (Equation 2.10) and (Equation 2.11), the extra sum rate constraint, which is due to treating the first hop as two MAC channels in the DF stage, could potentially[1] render DF+FCo inferior to CF+ICF. This is confirmed by the simulations shown in Fig. Figure 11. One misleading thought is that the superiority of CF+ICF comes *solely* from the CF stage and that ICF is immaterial here. To clarify the role of ICF scheme, we also plot the overall network rate region by adopting CF and the naive ICF (ICF Scheme 1) in green in Fig. Figure 11, where one can see that ignoring the correlations between the equations (ICF Scheme 1) could reduce the gains significantly. Thus, a proper ICF

---

[1] This is true when the powers at the relay nodes are not too much smaller than those at the source nodes; otherwise, the second hop rate constraints will dominate.

scheme is needed for the overall superior performance of the CF+ICF scheme. We also note that in some extreme scenarios, as shown in Fig. 11(b), the gain of CF+ICF over DF+FCo can be substantial.

Remark 4. *We do not claim that CF+ICF generally leads to larger rates than DF+FCo. For example, when the powers at the source nodes are abundant while those at the relay nodes are scarce, the overall rate region will be dominated by the rate constraints of the second hop. In this scenario, CF+ICF and DF+FCo will have exactly the same performance. Also, our simulations assume that the channel coefficients are integers (with absolute value* 1*), which is well suited to the Compute-and-Forward scheme. When the channel coefficients are not as assumed here, one needs to carefully choose the equation to decode, which is outside of the scope of this study.*

2.4 Three-user Case

We now move to the three-user ICF problem to build additional intuition. Recall the following assumptions placed on coefficient matrix $\mathbf{F}$: (1) full rank; (2) any 2 by 2 submatrix from its first two columns is non-singular; and (3) all entries in its first column are non-zero. As shown in Fig. Figure 12, recall that $\mathbf{w}_{l,c}$ denotes a *message section* of length $s_c := k_c - k_{c+1}$ (for $k_4 := 0$) which corresponds to the c-th segment of message $\mathbf{w}_l$ for $c \in \{1, 2, 3\}$. Let $\tilde{\mathbf{W}}_{*c}$ be the matrix of dimension $3 \times s_c$ whose l-th row is $\mathbf{w}_{l,c}$. Following the notation of Section 2.2: $\begin{bmatrix} \tilde{\mathbf{U}}_{*1} & \tilde{\mathbf{U}}_{*2} & \tilde{\mathbf{U}}_{*3} \end{bmatrix} = \begin{pmatrix} \mathbf{F} \end{pmatrix} \cdot \begin{bmatrix} \tilde{\mathbf{W}}_{*1} & \tilde{\mathbf{W}}_{*2} & \tilde{\mathbf{W}}_{*3} \end{bmatrix}$, or, breaking this into *message sections* and *equation sections*, as shown in Fig. Figure 12.

It can be checked that:

(I) $\tilde{\mathbf{U}}_{*1}$, or $\mathbf{u}_{1,1}, \mathbf{u}_{2,1}, \mathbf{u}_{3,1}$ are completely correlated, and may be used to reconstruct $\mathbf{w}_{1,1}$, a common message known to all relays.

(II) $\mathbf{u}_{1,2}, \mathbf{u}_{2,2}, \mathbf{u}_{3,2}$ are pairwise independent and have the property that the third is a deterministic function of the other two. These three are not mutually independent.

(III) $\mathbf{u}_{1,3}, \mathbf{u}_{2,3}, \mathbf{u}_{3,3}$ are mutually independent.

In moving to three users one interesting new aspect arises: in addition to extracting a common message and two independent messages from the equations as in the two-user case, in the three-user case we also extract three pairwise independent messages. One may wonder if/how this kind of dependency may be exploited. We show that for the Gaussian MAC channel model, no coherent power gains may be obtained from such pairwise independent correlation. This is at least partially due to the linearity and second moment constraints of the AWGN channel where Gaussians maximize entropy, and the second moment of a linear sum of random variables depends only on the pairwise correlation between its elements. We conjecture that, for fixed source/message dependencies, coherent encoding is *possible* or *valuable* only when these dependencies are not destroyed by the channel.

Remark 5. *One might ask whether this problem maps onto an extension of the two-user Slepian-Wolf problem. An extension of the Slepian-Wolf MAC is considered in (13), where each transmitter in an L-MAC has access to an arbitrary subset of messages from a set of independent messages. Our problem cannot be mapped into the framework in (13), as in the latter, the users either have common message(s) or completely independent ones, but do not have the pairwise* *(but not mutual) independence property seen here. We are not aware of any other related problems which capture the pairwise independent structure. The ICF problem might be a special case of the problem considered in (15), which obtains an uncomputable multi-letter expression for the capacity region for sending arbitrarily correlated sources* $(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3) \sim \prod_{i=1}^{n} p(s_{1i}, s_{2i}, s_{3i})$ *over a MAC channel. It is easy to pull out the fully common and the conditionally independent components, but how to cast the pairwise independent but not mutually independent components as a source of this form* $(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3) \sim \prod_{i=1}^{n} p(s_{1i}, s_{2i}, s_{3i})$ *is an open problem. Unfortunately, even if one were able to cast our constraints into a source of that form, the capacity region is not computable. We will next show a simple achievability scheme, which turns out to be the explicitly computable capacity region in the Gaussian case.*

Theorem 15 (Memoryless three-user ICF achievability)**.** *Assume that* $\mathbf{F}$ *and all* c *by* c *submatrices from its first* c *columns are of full rank,* $c = 1, \cdots, L$*. The messages* $(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3)$ *at rates* $(R_1 \geq R_2 \geq R_3)$ *may be recovered from* $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ *sent over a MAC if the rates lie in*

$$\mathcal{R}_{IN} := \bigcup_{p(q)p(x_1|q)p(x_2|q)p(x_3|q)} \mathcal{R} \tag{2.12}$$

for $\|Q\| \leq \min\{\|\mathcal{X}_1\| \cdot \|\mathcal{X}_2\| \cdot \|\mathcal{X}_3\| + 3, \|\mathcal{Y}\| + 4\}$*, where* $\mathcal{R}$ *is the set of* (R_1, R_2, R_3) *with* $(R_1 \geq R_2 \geq R_3)$:

$$R_1 + R_2 + R_3 \leq I(X_1, X_2, X_3; Y) \tag{2.13a}$$

$$2R_2 + R_3 \leq I(X_1, X_2, X_3; Y|Q) \tag{2.13b}$$

$$R_2 + R_3 \leq \min\{I(X_1, X_2; Y|X_3, Q), I(X_1, X_3; Y|X_2, Q), I(X_2, X_3; Y|X_1, Q)\} \tag{2.13c}$$

$$R_3 \leq \min\{I(X_1; Y|X_2, X_3, Q), I(X_2; Y|X_1, X_3, Q), I(X_3; Y|X_1, X_2, Q)\}. \tag{2.13d}$$

Remark 6. *To understand the form, consider for example* (Equation 2.13b)*. This results from the error event that all message sections except the common message (*$\mathbf{w}_{1,1}$ *or* $\tilde{\mathbf{U}}_{*1}$*) are incorrect. The rate of these incorrect message sections is* $2(R_2 - R_3) + 3(R_3) = 2R_2 + R_3$*. Similarly,* (Equation 2.13c) *corresponds to when the common message portion and one of the codewords is correct and thus the rates of the incorrect message portions is* $1(R_2 - R_3) + 2(R_3) = R_2 + R_3$*. Finally,* (Equation 2.13d) *corresponds to when the common message and two entire codewords are correct: only the independent message section of rate* R_3 *is wrong.*

An alternative interpretation is the following: (Equation 2.13a) corresponds to the overall sum rate constraint and (Equation 2.13b) corresponds to the sum constraint apart from the cooperative or common message of rate $R_1 - R_2$ (see Fig. Figure 12). Any single link cannot help the destination distinguish between more than 2^{nR_1} possibilities for the equations (or messages), because knowing one $\mathbf{u}$, say $\mathbf{u}_1$, can at most resolve 2^{nR_1} uncertainties. Hence, the other two links must help the destination to distinguish between at least $2^{n(R_2+R_3)}$ values so that overall, it may distinguish between the $2^{n(R_1+R_2+R_3)}$ possible equation or message values. This explains (Equation 2.13c). Analogously, any *two* links cannot help the destination distinguish between more than $2^{n(R_1+R_2)}$ values; the third link distinguishes between the remaining 2^{nR_3} choices. For the Gaussian channel, the above achievable rate region is the capacity region, given in Theorem 17 for general L.

Remark 7. *The above theorem holds for $R_1 \geq R_2 \geq R_3$; other relative orderings may be obtained similarly. We do not claim the convex hull of the rate regions for different orderings to be achievable as the relative values of R_1, R_2, R_3 are fixed as part of the ICF problem setting. When deriving an achievable rate region for a larger network, one takes the convex hull after intersecting the CF and ICF rate regions.*

2.5 Main result: L-user ICF achievable rate region

We now present the main technical contributions: 1) an achievable rate region for the general L-user ICF problem of extracting L independent messages from linear equations of these messages over a multiple access channel, and 2) the capacity region for the L-user Gaussian ICF channel. Both regions are enlarged with respect to a MAC with independent messages as the relays extract and exploit a special form of dependency from the linear equations they possess. The extraction of a common message allows for coherent gains, while knowing some equations limits the values other equations may take on and hence reduces the number of error events.

The main theorem is stated in terms of message rates R_l, while its proof in the Appendix B is argued via section rates ρ_c (Definition 14, Section 2.2). The use of section rates not only facilitates the error analysis but also helps to reveal the effect of dependency patterns among the equations at the relays. There is a one-to-one mapping between $\rho_1, \cdots, \rho_L$ and $R_1, \cdots, R_L$ given by $\rho_c = R_c - R_{c+1}$, $R_{L+1} = 0$.

2.5.1 An ICF achievable rate region for the Memoryless ICF channel

Our main achievability result for the L-user ICF channel model follows.

Theorem 16 (Achievable rate region for Memoryless ICF Channels)**.** *Assume that* $\mathbf{F}$ *and all* c *by* c *sub-matrices from its first* c *columns are of full rank,* $c = 1, \cdots, L$*. The messages* $(\mathbf{w}_1, \cdots, \mathbf{w}_L)$ *may be recovered from the equations* $\mathbf{u}_1, \cdots, \mathbf{u}_L$ *over the memoryless MAC channel* $p(y|x_1, \cdots, x_L)$ *if:*

$$\sum_{l=1}^{L} R_l \le I(X_1, \cdots, X_L; Y) \tag{2.14a}$$

$$2R_2 + \sum_{l=3}^{L} R_l \le I(X_1, \cdots, X_L; Y|Q) \tag{2.14b}$$

$$\sum_{l=\nu+1}^{L} R_l \le I(X_{A^C}; Y|X_A, Q) \quad \textit{for } \nu = 1, 2, \cdots, L-1 \tag{2.14c}$$

for all $A \subset \{1, 2, \cdots, L\}$, $\|A\| = \nu$*, taken over* $p(q) \cdot p(x_1|q) \cdot \cdots \cdot p(x_L|q) \cdot p(y|x_1, \cdots, x_L)$*.*

First, it may be verified that the two-user region in (Equation 2.6) and the three-user achievability scheme in Theorem 15 may be obtained as special cases of this theorem by selecting $L = 2$ and $L = 3$ respectively. Note that there are 2^L inequalities in total in (Equation 2.14), compared to the $2^L - 1$ in a classical MAC.

We may interpret (Equation 2.14c) as follows. Take for example $L = 5$, $\nu = 2$, $A = \{2, 3\}$ and $A^C = \{1, 4, 5\}$. Then (Equation 2.14c) works out to

$$(0)R_1 + (0)R_2 + (1)R_3 + (1)R_4 + (1)R_5 \leq I(X_1, X_4, X_5; Y|X_2, X_3, Q).$$

In this case, the correctly decoded codewords X_2^n and X_3^n can at most help the destination distinguish between $2^{n(R_1+R_2)}$ possible values of the messages $\mathbf{w}_1, \cdots, \mathbf{w}_5$. Hence, the remaining codewords must help distinguish at least $2^{n(R_3+R_4+R_5)}$ of the remaining message tuples, and these may be communicated at a rate up to $I(X_1, X_4, X_5; Y|X_2, X_3, Q)$ if X_2^n and X_3^n are correct (and hence also the common message encoded into Q is correct). Alternatively, from a linear algebra perspective, given the correct estimation of codewords X_2^n and X_3^n, i.e., $\mathbf{u}_2$ and $\mathbf{u}_3$, we may completely remove variables $\mathbf{w}_1$ and $\mathbf{w}_2$ from the set of remaining equations, i.e., $\mathbf{u}_1, \mathbf{u}_4, \mathbf{u}_5$. Thus, we have a new equation set $\mathbf{U}' = \mathbf{F} \cdot \mathbf{W}'$, which relates $(\mathbf{u}_1, \mathbf{u}_4, \mathbf{u}_5)$ to $(\mathbf{w}_3, \mathbf{w}_4, \mathbf{w}_5)$, with at most $2^{n(R_3+R_4+R_5)}$ different solutions.

The proof is provided in Appendix B. The achievability scheme generates a common codebook for the common message $\mathbf{w}_{1,1}$ (or equivalently equation section matrix $\tilde{\mathbf{U}}_{*1}$) and conditionally independent (conditioned on this common part) codebooks at each transmitter for the remaining equation sections. We index everything by the equation sections and use a joint typicality decoder to estimate these directly.

Remark 8. *As noted in Remark 5, whether the above presented achievability scheme may be cast as a special case of the L-user problem of sending arbitrarily correlated sources over a*

MAC as considered in (15) is an interesting open question. One might suspect so, but it is not clear how to express the dependencies induced by the ICF problem as an i.i.d. (but correlated) source of the form $(\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_L) \sim \prod_{i=1}^{n} p(s_{1i}, s_{2i}, \cdots, s_{Li})$. *We furthermore go beyond an achievability scheme and in the following section show capacity explicitly by obtaining a general converse.*

2.5.2 The ICF Capacity Region for the Linear Gaussian-MAC model

We now turn our attention to AWGN channels. In moving towards capacity, the difficulty lies not in deriving rate bounds which match the general achievable rate region but rather in showing that restriction to input distributions of the form $p(q)p(x_1|q)\cdots p(x_L|q)$ and Gaussian is without loss of generality. In general, given the message equations, it may appear that all relay node inputs could be arbitrarily correlated and hence outer bounds would need to be evaluated over all joint $p(x_1, x_2, \cdots, x_L)$. However, for the AWGN channel we show that the form of the equations dictates a particular dependency structure. This structure, for Gaussian channels, results in an achievable outer bound exhausted by Gaussian inputs.

Theorem 17 (The ICF Capacity Region for Linear Gaussian MAC)**.** *Assume that* $\mathbf{F}$ *and all* c *by* c *submatrices from its first* c *columns are of full rank,* $c = 1, \cdots, L$. *One can fully* *recover messages* $\mathbf{w}_1, \cdots, \mathbf{w}_L$ *from the equations* $\mathbf{u}_1, \cdots, \mathbf{u}_L$ *transmitted by the relays via a linear Gaussian MAC channel in* (Equation 2.1) *if and only if the message rates* R_l *satisfy:*

$$\begin{cases} \sum_{l=1}^{L} R_l \leq \frac{1}{2}\log_2\left(1+\sum_{j=0}^{L} d_j^2\right) \\ 2R_2 + \sum_{l=3}^{L} R_l \leq \frac{1}{2}\log_2\left(1+\sum_{j=1}^{L} d_j^2\right) \\ \sum_{l=\nu+1}^{L} R_l \leq \frac{1}{2}\log_2\left(1+\sum_{j\in A^C} d_j^2\right) \end{cases} \tag{2.15}$$

for $\nu = 1, \cdots, L-1$, $R_{L+1} := 0$, *and all* A *such that* $\|A\| = \nu$, $A \subset \{1, 2, \cdots, L\}$, *with some* $\{d_0, \cdots, d_L\}$ *such that* $d_0 = \sqrt{b_1} + \sqrt{b_2} + \cdots + \sqrt{b_L}$, $d_j = \sqrt{P_j - b_j}$, *and* $0 \leq b_j \leq P_j$, *for* $j = 1, \cdots, L$.

Proof. **Achievability:** Let $Q, Q_1, Q_2, \cdots, Q_L \sim \mathcal{N}(0,1)$, and all independent, be used to generate i.i.d. length n sequences Q^n, $Q_1^n, \cdots, Q_L^n$. Relay m sends:

$$X_m^n(\mathbf{u}_m) = \sqrt{b_m}Q^n(\mathbf{u}_{m,1}) + \sqrt{P_m - b_m}Q_m^n(\mathbf{u}_{m,2}, \cdots, \mathbf{u}_{m,L}), \;\; 0 \leq b_m \leq P_m.$$

Thus, $p(q)$ is Gaussian, and every $p(x_m|q)$ is again Gaussian. Then, at each channel use,

$$\begin{aligned} Y &= X_1 + \cdots + X_L + Z \\ &= \sqrt{b_1}Q + \sqrt{P_1 - b_1}Q_1 + \cdots + \sqrt{b_L}Q + \sqrt{P_L - b_L}Q_L + Z \\ &:= d_0 Q + d_1 Q_1 + \cdots + d_L Q_L + Z \end{aligned}$$

where $d_0 = \sqrt{b_1} + \cdots + \sqrt{b_L}$ and $d_m = \sqrt{P_m - b_m}$, $m = 1, 2, \cdots, L$ as in the Theorem statement. Evaluating the bounds of Theorem 16, we obtain the achievable rate region specified by inequalities (Equation 2.15).

Converse:

The converse uses Lemmas 18, 19 and 20 to upper bound the capacity region as follows

$$\mathcal{C} \overset{\text{Lemma 18}}{\subseteq} \mathcal{R}_{\text{out}} \overset{\text{Lemma 19}}{\subseteq} \bigcup \mathcal{R}' \overset{\text{Lemma 20}}{\subseteq} \bigcup \mathcal{R}'' .$$

We first state the lemmas, explain the intuition and show how they are used to establish the converse. We defer the proofs of Lemmas 18 and 19 to the following subsections, while the proof of Lemma 20 is inline.

First, Lemma 18 provides an outer bound $\mathcal{R}_{\text{out}}$ valid for any memoryless channel. Define

$$\mathcal{P} := \{p(q, x_1, \cdots, x_L) : X_m \to Q \to X_{m'},\ \forall m \neq m',\ m, m' \in \{1, 2, \cdots, L\}\} \tag{2.16}$$

Lemma 18. *$\mathcal{C} \subseteq \mathcal{R}_{out}$, where $\mathcal{R}_{out}$ is defined as*

$$\mathcal{R}_{out} := \bigcup_{p(q,x_1,\cdots,x_L)\in\mathcal{P}} \mathcal{R}(Q, X_1, \cdots, X_L), \tag{2.17}$$

where $\mathcal{R}(Q, X_1, \cdots, X_L)$ denotes the set of rate tuples $(R_1, \cdots, R_L)$ that satisfy inequalities (Equation 2.14).

Lemma 19 further loosens the outer bound $\mathcal{R}_{\text{out}}$ in Lemma 18 for the Gaussian-MAC model $Y = X_1 + \cdots + X_L + Z$ and shows $\mathcal{R}_{\text{out}} \subseteq \bigcup \mathcal{R}'$. The essence of its proof in Section 2.5.4 is to note that for Gaussian channels subject to power constraints, only second moment constraints are of interest and the variance of a linear sum of random variables does not depend on correlations of order higher than 2.

Lemma 19. *For the Gaussian-MAC model, $Y = X_1 + \cdots + X_L + Z$, for any given $p(q, x_1, \cdots, x_L) \in \mathcal{P}$, region $\mathcal{R}(Q, X_1, \cdots, X_L)$ can be outer bounded by region $\mathcal{R}'$, where $\mathcal{R}'$ consists of the rate tuples:*

$$\begin{cases} \sum_{l=1}^{L} R_l \leq C(\sum_{m=1}^{L} E[X_m^2] + \sum_{m \neq m'} E[X_m X_{m'}]) \\ 2R_2 + \sum_{l=3}^{L} R_l \leq C(\sum_{m=1}^{L} var[X_m|Q]) \\ \sum_{l=\nu+1}^{L} (l-\nu)(R_l - R_{l+1}) \leq C(\sum_{m \in A^C} var[X_m|Q]) \end{cases} \tag{2.18}$$

for $\nu = 1, 2, \cdots, L-1$, $R_{L+1} := 0$, and all possible A such that $A \subset \{1, 2, \cdots, L\}$ and $\|A\| = \nu$.

We outer bound the outer bound $\mathcal{R}'$ one more time in Lemma 20. This lemma is based on the power constraints and the Markov chains $X_m \to Q \to X_{m'}$, $\forall m \neq m'$, $m, m' \in \{1, 2, \cdots, L\}$. To show Lemma 20, note that it follows from (11, Lemma B.3) that $E[X_m X_{m'}] \leq \sqrt{E[X_m^2] - \text{var}(X_m|Q)}\sqrt{E[X_{m'}^2] - \text{var}(X_{m'}|Q)}$. This, together with $t_m = \frac{E[X_m^2] - \text{var}(X_m|Q)}{E[X_m^2]} \in [0, 1]$, $m = 1, \cdots, L$ immediately lead to the following Lemma.

Lemma 20. *The region* $\mathcal{R}' \subseteq \mathcal{R}''$*, where* $\mathcal{R}''$ *consists of the rate tuples that satisfy*

$$\begin{cases} \sum_{l=1}^{L} R_l \le C(\sum_{m=1}^{L} E[X_m^2] + \sum_{m\neq m'} \sqrt{t_m t_{m'}} \sqrt{E[X_m^2]E[X_{m'}^2]}) \\ 2R_2 + \sum_{l=3}^{L} R_l \le C(\sum_{m=1}^{L} (1-t_m)E[X_m^2]) \\ \sum_{l=\nu+1}^{L} R_l \le C(\sum_{m\in A^C} (1-t_m)E[X_m^2]) \end{cases} \tag{2.19}$$

for $\nu = 1, 2, \cdots, L-1,$ *and all possible* A *such that* $A \subset \{1, 2, \cdots, L\}$ *and* $\|A\| = \nu$.

Combining Lemma 18, Lemma 19 and Lemma 20, we have

$$\begin{aligned} \mathcal{C} \;\subseteq\; \mathcal{R}_{\text{out}} \;&\subseteq \bigcup_{p(q,x_1,\cdots,x_L)\in\mathcal{P}} \mathcal{R}'|_{Y=X_1+\cdots+X_L+Z,\ p(q,x_1,\cdots,x_L)\in\mathcal{P}} \\ &\subseteq \bigcup_{t_1,t_2,\cdots,t_L\in[0,1]} \mathcal{R}''|_{t_1,\cdots,t_L} \end{aligned} .$$

where the last region may be verified to be that stated in Theorem 17 with b_j replaced by $t_j P_j$ – i.e. may be achieved by jointly Gaussian inputs which are conditionally independent given Gaussian $p(q)$. □

2.5.3 Proof of Lemma 18

Proof. We have the Markov chain $\mathbf{W} \to \mathbf{U} \to (X_1, \cdots, X_L) \to Y \to \hat{\mathbf{U}}$. Recall that $\tilde{\mathbf{U}}_{*c}$ stands for the cth column of the equation matrix $\tilde{\mathbf{U}}_{L\times L}$, which is equivalent to $\mathbf{U}_{L\times k}$, and that $\rho_c := R_c - R_{c+1}$:

$$
\begin{aligned}
n(\sum_{l=1}^{L} R_l) = n\sum_{c=1}^{L} c\rho_c &\overset{(a)}{=} H(\mathbf{U}) \\
&\overset{(b)}{\leq} I(\mathbf{U};Y^n) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(\mathbf{U};Y_i) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} I(\mathbf{U}, X_{1i}, \cdots, X_{Li}; Y_i) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} I(X_{1i}, \cdots, X_{Li}; Y_i) + n\epsilon_n \\
&\overset{(e)}{\leq} nI(X_1, \cdots, X_L; Y) + n\epsilon_n
\end{aligned}
\tag{2.20}
$$

$$
\begin{aligned}
n(2R_2 + \sum_{l=3}^{L} R_l) = n\sum_{c=2}^{L} c\rho_c &\overset{(a)}{=} H([\tilde{\mathbf{U}}_{*2}, \tilde{\mathbf{U}}_{*3}, \cdots, \tilde{\mathbf{U}}_{*L}]) \overset{(a)}{=} H(\mathbf{U}|\tilde{\mathbf{U}}_{*1}) \\
&= I(\mathbf{U};Y^n|\tilde{\mathbf{U}}_{*1}) + H(\mathbf{U}|Y^n, \tilde{\mathbf{U}}_{*1}) \\
&\overset{(b)}{\leq} I(\mathbf{U};Y^n|\tilde{\mathbf{U}}_{*1}) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(\mathbf{U};Y_i|\tilde{\mathbf{U}}_{*1}) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} I(\mathbf{U}, X_{1i}, \cdots, X_{Li}; Y_i|\tilde{\mathbf{U}}_{*1}) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} I(X_{1i}, \cdots, X_{Li}; Y_i|\tilde{\mathbf{U}}_{*1}) + n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n} I(X_{1i}, \cdots, X_{Li}; Y_i|Q_i) + n\epsilon_n \qquad (Q_i := \tilde{\mathbf{U}}_{*1}) \\
&\overset{(e)}{\leq} nI(X_1, \cdots, X_L; Y|Q) + n\epsilon_n
\end{aligned}
\tag{2.21}
$$

$$
\begin{aligned}
n(\sum_{l=\nu+1}^{L}(l-\nu)(R_l-R_{l+1}) = n\sum_{c=\nu+1}^{L}(c-\nu)\rho_c &\overset{(a)}{=} H(\mathbf{U}|\mathbf{u}_A) \overset{(a)}{=} H(\mathbf{u}_{A^C}|\tilde{\mathbf{U}}_{*1},\ \mathbf{u}_{A,}) \\
&= I(\mathbf{u}_{A^C};Y^n|\tilde{\mathbf{U}}_{*1},\mathbf{u}_A) + H(\mathbf{u}_{A^C}|Y^n,\tilde{\mathbf{U}}_{*1},\ \mathbf{u}_A) \\
&\overset{(b)}{\leq} I(\mathbf{u}_{A^C};Y^n|\tilde{\mathbf{U}}_{*1},\ \mathbf{u}_A) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(\mathbf{u}_{A^C};Y_i|\tilde{\mathbf{U}}_{*1},\ \mathbf{u}_A) + n\epsilon_n \qquad (2.22) \\
&\overset{(c)}{=} \sum_{i=1}^{n} I(\mathbf{u}_{A^C},X_{A^C i};Y_i|\mathbf{U}_{,1},\mathbf{u}_A,X_{Ai}) + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} I(X_{A^C i};Y_i|\tilde{\mathbf{U}}_{*1},X_{Ai}) + n\epsilon_n
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(d)}{=} \sum_{i=1}^{n} I(X_{A^C i};Y_i|Q_i,X_{Ai}) + n\epsilon_n \qquad (Q_i := \tilde{\mathbf{U}}_{*1}) \\
&\overset{(e)}{\leq} nI(X_{A^C};Y|Q,X_A) + n\epsilon_n
\end{aligned}
$$

The equalities in (a) all follow by definitions, and Lemma 62 and 64 in the Appendix. This is where we use that $\mathbf{F}$ and all c by c sub-matrices from its first c columns are of full rank – if not Lemmas 62 and 64, and hence the relationships between rates and entropies would change. Inequalities (b) follow from Fano's Inequality, where $\epsilon_n \to 0$ as $n \to \infty$. Steps (c) follow from the encoding functions, the Markov chain at the start of this proof, and the memoryless channel properties. In steps (d), we set $Q_i := \tilde{\mathbf{U}}_{*1}$. In steps (e), by further time-sharing arguments and Jensen's inequality we obtain the form in (Equation 2.14) as $n \to \infty$.

Notice that since the $\mathbf{u}_m$ are conditionally pairwise independent given $\tilde{\mathbf{U}}_{*1}$ and since X_m^n is a function of $\mathbf{u}_m$, then X_m^n (and hence also X_m) are conditionally pairwise independent given Q. $\square$

2.5.4 Proof of Lemma 19

Proof. The key is to first apply the Max-Entropy therorem conditioned on $Q = q$. The proof of $I(X_{A^C}; Y|X_A, Q) \leq C\left(\sum_{m\in A^C} \mathrm{var}[X_m|Q]\right)$ is shown as an example.

$$\begin{aligned} I(X_{A^C}; Y|X_A, Q) = E_Q[I(X_{A^C}; Y|X_A, Q = q)] &\overset{(a)}{=} E_Q[h(\sum_{m\in A^C} X_m + Z|Q = q) - h(Z)] \\ &\overset{(b)}{\leq} E_Q\left[\frac{1}{2}\log\left(\frac{\mathrm{var}(\sum_{m\in A^C} X_m + Z|Q = q)}{\mathrm{var}(Z)}\right)\right] \\ &\overset{(c)}{=} E_Q\left[\frac{1}{2}\log\left(1 + \sum_{m\in A^C} \mathrm{var}(X_m|Q = q)\right)\right] \\ &\overset{(d)}{\leq} \frac{1}{2}\log\left(1 + \sum_{m\in A^C} \mathrm{var}(X_m|Q)\right), \end{aligned} \tag{2.23}$$

where (a) follows by definition of Y and the linearity of the AWGN channel model, (b) follows by the fact that Gaussians maximize entropy subject to second moment constraints (c) is the critical step and follows from 1) the linearity of the AWGN channel model, 2) the variance of a linear sum of random variables is defined by the pairwise relationships between these random variables, and does not depend on any higher order correlations such as for example

$E[X_1X_2X_3|Q=q]$, and 3) the fact that X_i's are conditionally independent conditioned on Q. Since this is the crucial step, note that

$$\begin{aligned}\mathrm{var}(\sum_{m\in A^C} X_m + Z|Q=q) &= \sum_{m\in A^C}\mathrm{var}(X_m|Q=q) + 2\sum_{i,j\in A^C, i\neq j}\mathrm{cov}(X_i,X_j|Q=q)+\mathrm{var}(Z)\\ &= \sum_{m\in A^C}\mathrm{var}(X_m|Q=q)+\mathrm{var}(Z), \qquad (2.24)\end{aligned}$$

where 'cov' denotes the covariance between two random variables. Note that since X_i, X_j are conditionally independent given $Q=q$, $\mathrm{cov}(X_i,X_j|Q=q)=0$. Step (d) follows from Jensen's inequality. □

2.5.5 On the assumptions placed on F

As commented in Remark 1, the assumption that **F** and all c by c sub-matrices from its first c columns, $c=1,2,\cdots,L$, are of full rank is made for the succinctness of presentation. Without the requirements on sub-matrices, one could further exploit the specific dependencies between the equations $\mathbf{u}_m$ for each specific coefficient matrix **F**. We provide examples of how to proceed in this direction for $L=2$ and 3 next. We note that **F** must always be full rank in order for the ICF problem to be feasible. However, no further requirements need to be imposed on sub-matrices to do so.

Two-user example: Recall that we require $\mathbf{F}$ to be full rank and its first column entries f_{11} and f_{21} to be non-zero. However, there are four types of 2 by 2 matrices (upto scalings on rows) that yield invertible $\mathbf{F}$ (feasible) but violate the assumptions on sub-matrices:

$$\mathbf{F} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \ \mathbf{F} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \ \mathbf{F} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \ \mathbf{F} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Consider

$$\mathbf{F} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \text{ and hence } \begin{cases} \mathbf{u}_1 = 0 \cdot \mathbf{w}_1 \oplus 1 \cdot \mathbf{w}_2 \\ \mathbf{u}_2 = 1 \cdot \mathbf{w}_1 \oplus 1 \cdot \mathbf{w}_2 \end{cases}.$$

In this case, the two equations $\mathbf{u}_1$ and $\mathbf{u}_2$ are actually independent. Although $\mathbf{F}$ is still full rank and may be inverted to recover the original messages $\mathbf{W}$, knowing $\mathbf{u}_1$, for example, can only resolve $\mathbf{w}_2$ and the number of possible choices of $\mathbf{u}_2$ is 2^{nR_1}. Thus, the cardinality bounding arguments in Scheme 2 in Section 2.3 fails. The achievable rate region shrinks to

$$\left\{ (R_1, R_2) : \begin{array}{l} R_{\max} \leq I(X_2; Y|X_1) \\ R_{\min} \leq I(X_1; Y|X_2) \\ R_1 + R_2 \leq I(X_1, X_2; Y) \\ \text{for } p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2) \end{array} \right\}.$$

When $\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, following similar arguments, one can check that the region (Equation 2.5) should be modified to

$$\left\{ (R_1, R_2) : \begin{array}{l} R_{\min} \leq I(X_2; Y|X_1) \\ R_{\max} \leq I(X_1; Y|X_2) \\ R_1 + R_2 \leq I(X_1, X_2; Y) \\ \text{for } p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2) \end{array} \right\}.$$

We omit the other cases for brevity. This is an example of how, in contrast to (16), we do not require all square sub-matrices of $\mathbf{F}$ to be full rank. Nevertheless, the format of the rate region varies.

Three-user example: Recall that we require $\mathbf{F}$ to be full rank and further assume that (1) its first column entries f_{11}, f_{21} and f_{31} are all non-zero; (2) any 2 by 2 submatrix from its first two columns is nonsingular. There are many (but finite) realizations of $\mathbf{F}$ such that it satisfies the feasibility constraint (full rank) but violates the assumptions on sub-matrices. We consider one example to show that the derivation of achievable rate region for each individual case is a relatively straightforward extension of the work presented in Appendix B, but the format of corresponding rate region differs from case to case. Let

$$\mathbf{F} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 3 \\ 1 & 2 & 3 \end{bmatrix} \text{ and hence } \begin{cases} \mathbf{u}_1 = 1 \cdot \mathbf{w}_1 \oplus 1 \cdot \mathbf{w}_2 \oplus 1 \cdot \mathbf{w}_3 \\ \mathbf{u}_2 = 1 \cdot \mathbf{w}_1 \oplus 1 \cdot \mathbf{w}_2 \oplus 3 \cdot \mathbf{w}_3 \\ \mathbf{u}_3 = 1 \cdot \mathbf{w}_1 \oplus 2 \cdot \mathbf{w}_2 \oplus 3 \cdot \mathbf{w}_3 \end{cases} .$$

It may be checked that:

1. coefficient matrix **F** is invertible but sub-matrix $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ is singular [1];
2. equation sections $\mathbf{u}_{1,1}$, $\mathbf{u}_{2,1}$ and $\mathbf{u}_{3,1}$ share the same information;
3. equation sections $\mathbf{u}_{1,2}$ and $\mathbf{u}_{2,2}$ are exactly the same instead of being (pairwise) independent;
4. equation sections $\mathbf{u}_{1,3}$, $\mathbf{u}_{2,3}$ and $\mathbf{u}_{3,3}$ are mutually independent.

We now ask whether the derived achievable rate region in the Appendix B for the discrete memoryless MAC still holds in this case. The analyses of *Error event type I, II III* remain valid while the analysis of *Error event type IV* is unable to proceed, and must be modified as follows:

---

[1] Note that is submatrix $\begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix}$ is also singular but it does not violate our sub-matrix assumption.

Let β_2 represent the number of correctly estimated equation sections among $\tilde{\mathbf{U}}_{*2}$, i.e., $\mathbf{u}_{1,2}, \mathbf{u}_{2,2}, \cdots, \mathbf{u}_{L,2}$, and β_3 for $\tilde{\mathbf{U}}_{*3}$. Let $\nu \in \{1,2\}$, $A \subset \{1,2,3\}$, and $\|A\| = \nu$. For detailed definitions of these parameters/indicators, please refer to the proof in Appendix B.

- The analysis of *Error event type I* holds due to the definition of the jointly typical set.
- The analysis of *Error event type II*, *Error event type III*, and case $\nu = 1$ of *Error event type IV* remains valid even though $\beta_2 = 2$ is possible, which violates Lemma 64. This results from the fact that the *most demanding* constraints among these error event cases do not change. For example, when $\nu = 1$, $A = \{1\}$, $A^C = \{2,3\}$, we have $\mathbf{u}_{1,2} = \mathbf{u}_{2,2} =$ the correct value. Thus, β_2 cannot be 1 as expected but is actually 2. Surprisingly, this does not disagree with equation (Equation B.3) when $\nu = 1$ since $\gamma_2 = 1, \gamma_3 = 2$ stays true.
- The singularity of sub-matrix $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ does affect the case when $\nu = 2$ in *Error event type IV*. For example, when $A = \{1,2\}$, $A^C = \{3\}$, we have $\beta_2 = 2, \beta_3 = 2$ and thus $\gamma_2 = 1, \gamma_3 = 1$, i.e., $1 \cdot \rho_2 + 1 \cdot \rho_3 = R_2 \leq I(X_3; Y|X_1, X_2, Q)$ instead of $0 \cdot \rho_2 + 1 \cdot \rho_3 = R_3 \leq I(X_3; Y|X_1, X_2, Q)$.

In summary, the achievable rate region for this particular choice of $\mathbf{F}$ would lead to the same region as in (Equation 2.13) except for the third term in inequality (Equation 2.13d) which becomes the new $R_2 \leq I(X_3; Y|X_1, X_2, Q)$.

Remark 9. *Note that if two rows are exchanged in matrix* $\mathbf{F}$*, say the 2nd and 3rd rows, then inequality* $R_3 \leq I(X_2; Y|X_1, X_3, Q)$ *in region* (Equation 2.13) *will be replaced by* $R_2 \leq$

$I(X_2; Y|X_1, X_3, Q)$. *Thus, we note that the assumption that all* $c \times c$ *sub-matrices of the first* c *columns of* $\mathbf{F}$ *be non-singular is not necessary for our coding scheme, but makes a succinct and consistent presentation of rate regions possible.*

While achievable rate regions could be naturally extended using the above techniques, we note that for the Gaussian model, the converse as currently written would not naturally follow. The Markov inequalities (pairwise independent conditioned on the common message) no longer naturally follow and the current argument that mutually independent (conditioned on Q) Gaussians maximize the outer bound would fail.

2.5.6 On the generalization of the ICF result

The ICF problem and particular message structure is motivated by relay networks in which CF is used at relay nodes. An abstract generalization of our capacity result holds for the following channel model.

Abstract Gaussian ICF model. Consider again an L-user Gaussian channel model as in (Equation 2.1). Consider a set of $1+2+3+\cdots+L$ independent messages and a set of $L \times L$ functions satisfying:

1. One message $W_{1,1}$ is of rate ρ_1, two messages $W_{1,2}, W_{2,2}$ of rate ρ_2, three messages $W_{1,3}, W_{2,3}, W_{3,3}$ of rate ρ_3, $\cdots$, L messages $W_{1,L}, \cdots, W_{L,L}$ of rate ρ_L.
2. All users know message $W_{1,1}$ (or a one-to-one function $T_{i,1}$ thereof).

3. Each user $i = 1, 2, \cdots, L$, for each $l = 2, 3, \cdots, L$ knows a function say $T_{i,l}$ of the messages $W_{1,l}, \cdots, W_{l,l}$ such that given any l of L functions $T_{i,l}$, $i = 1, 2, \cdots, L$, it is possible to reconstruct the original l messages.
4. For $l = 2, 3, \cdots, L$, any two $T_{i,l}$ for different i are independent.

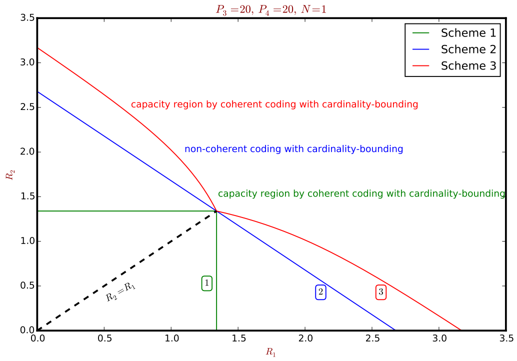
Constraints 2) and 3) allow us to relate message rates to the entropy (or conditional entropy) of some sets of equations, needed in Lemma 18 in Subsection 2.5.3. Furthermore, since all messages are independent, together with constraint 4) in particular, the set of Markov chains $X_m \to Q \to X_{m'}$, $\forall m \neq m',\ m, m' \in \{1, 2, \cdots, L\}$, presented in Lemma 18 are ensured. Thus, Lemma 19 and Lemma 20 may be derived, and the converse for the Gaussian channel follows.

The remainder of the necessary definitions follow by extension of those in Section 2.2. Then the next Corollary is easy to obtain from the proof of Theorem 17.

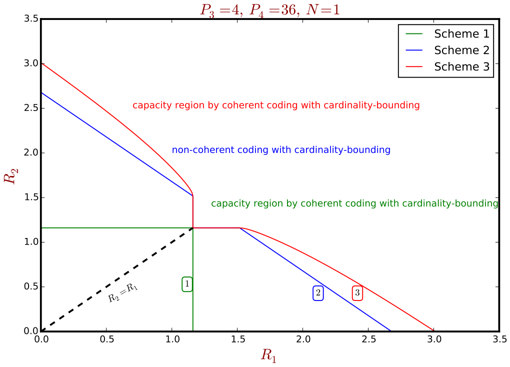
Corollary 21. *The capacity region of Theorem 17 is the capacity region for the Abstract Gaussian ICF model described above, with the convention that* $\rho_c = R_c - R_{c+1}$ *and* $R_{L+1} = 0$.

2.6 Conclusion

We consider an L-user multiple access channel where transmitter m has access to the linear equation $\mathbf{u}_m = \bigoplus_{l=1}^{L} f_{ml}\mathbf{u}_l$ of independent messages $\mathbf{u}_l \in \mathbb{F}_p^{k_l}$ with $f_{ml} \in \mathbb{F}_p$, and the destination wishes to recover all L messages. The dependency patterns among these given equations are explored and exploited to enlarge the achievable rate region relative to sending these equations independently as in a classical MAC channel. In the discrete memoryless MAC channel model, a tighter achievable rate region than (10) is obtained by adopting a coherent encoding scheme which exploits the fact that given equations at unequal message rates, common messages are in fact shared by the transmitters. In the Gaussian MAC channel, the general L-user capacity region is derived. All derived results assume invertibility constraints on the coefficient matrix of the decoded message equations, which is discussed. The outer bound relies heavily on the the linearity and second moment constraints of the AWGN channel, in addition to careful accounting of the dependency structure between the equations. In essence, only pairwise dependency between equations is of concern in Gaussian channels. This ICF capacity region may be used as a building block for the "last hop" in relay networks where CF is employed at relay nodes, besides being of independent interest. As such, capacity is also obtained for a generalized abstraction of our model. Whether the achievable rate region presented for a general, non-Gaussian memoryless channel is capacity remains an interesting open question; We are currently not able to find an example of a channel where this type of message dependency would enlarge the achievable rate region.

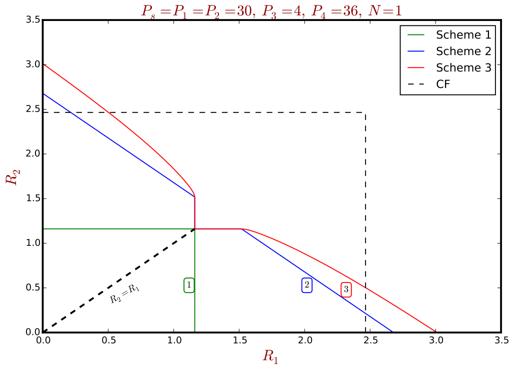


(a) symmetric power at relays

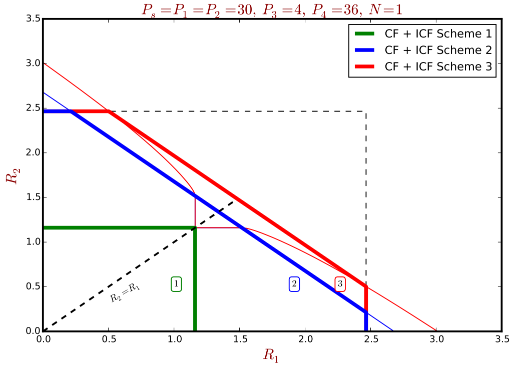


(b) asymmetric powers at relays.

Figure 9. Numerical evaluation for two-user Gaussian-MAC ICF problem. In (a) $P_3 = P_4 = 20$, and in (b) $P_3 = 4$, $P_4 = 36$. The union of the two orderings $R_1 \geq R_2$ and $R_2 \geq R_1$ (each convex) is plotted rather than their convex hull, as elaborated on in Remark 7.

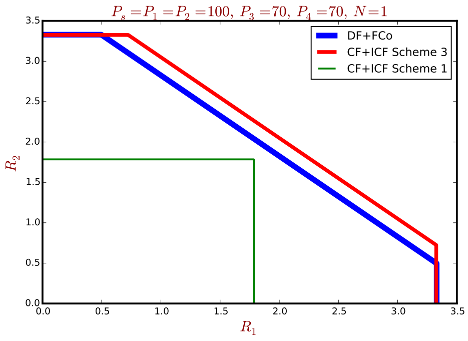


(a) asymmetric powers at relays. + combinning CF and ICF

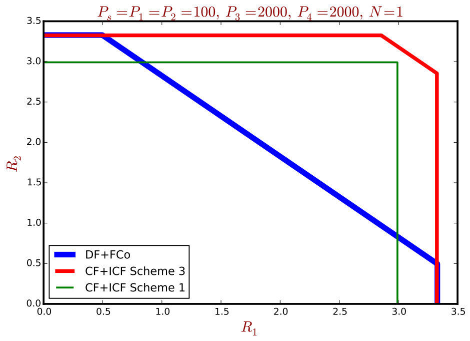


(b) the convex hull of the intersection region.

Figure 10. An example: combining CF and ICF in a network. Powers at the source nodes are $P_s = P_1 = P_2 = 30$; Powers at the relay nodes are $P_3 = 4$, $P_4 = 36$; I.id noises are with variance $N = 1$. In (a), the union of the two orderings $R_1 \geq R_2$ and $R_2 \geq R_1$ (each convex) is plotted rather than their convex hull, as elaborated on in Remark 7. (a) also contains the first CF hop explained in equation (Equation 2.10). In (b), we show the convex hull of the intersection of each scheme with the CF rate region. We use the convention: thin dotted lines for the first hop, thin solid lines for the second hop, thick solid lines for the rate regions for the whole network and thick dotted lines to depict the line $R_2 = R_1$.



(a)



(b)

Figure 11. Examples of CF+ICF outperforming DF+FCo.

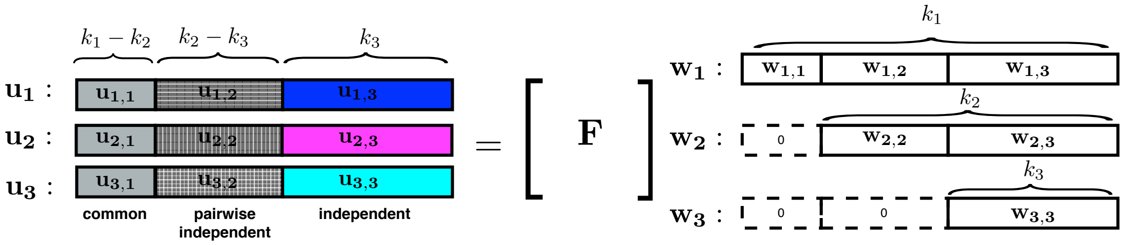


Figure 12. Three user ICF message/equation structure. The grey color indicates that these equation sections ($\mathbf{u}_{*,1}$) are fully correlated; shading indicates that these three equation sections ($\mathbf{u}_{*,2}$) are pairwise independent, while different solid colors indicate that these three equation sections ($\mathbf{u}_{*,3}$) are mutually independent. All message sections $\mathbf{w}_{i,j}$ are mutually independent.

CHAPTER 3

BACKGROUND ON COLOUR-AND-FORWARD RELAYING AND PRIMITIVE RELAY CHANNELS

From this chapter onwards, we will present our work on how to exploit channel structure to improve communication efficiency. In particular, the primitive relay channel is adopted to study how to optimally operate the relay terminal using the smallest conference rate that enables the whole network to achieve its absolute maximum message rate, i.e. the single-input multiple-output upper bound. The (zero-error) Colour-and-Forward and ϵ-Colour-and- Forward relaying algorithms will be represented in Chapter 4 and Chapter 5 respectively. In this chapter, we will provide background on the primitive relay channel, our motivation for developing Colour-and-Forward relaying algorithms, and how we was inspired to study zero- error communication over the primitive relay channel.

3.1 Primitive relay channels

As shown in Figure 13, a primitive relay channel (PRC) $((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R),\ r)$ consists of a source terminal S that wants to communicate a message W to a destination terminal D aided by a relay terminal R. The broadcasting links $(\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R)$ from the source to the relay and destination terminals are orthogonal to the error-free conference link with maximum rate r bits / channel use from the relay to the destination terminal. This channel model is motivated when a relay terminal cannot simultaneously transmit and receive signals or when the relay has an out-of-band link to the destination.

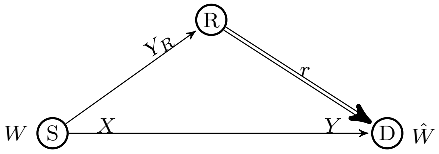


Figure 13. A primitive relay channel $((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R), r)$.

Clearly, a primitive relay channel is much simpler than a classical relay channel $(\mathcal{X} \times \mathcal{X}_R, p(y, y_R|x, x_R), \mathcal{Y} \times \mathcal{Y}_R)$: It decouples the multiple-access component and the broadcasting component in a classical relay channel. Studying PRCs can help to better understand the classical relay channels. As pointed out in (17), there are two perspectives on the study of primitive relay channels. From the transmitter's point of view, it can be seen as the simplest channel coding problem with a source coding constraint. At the same time, from the relay's point of view, it is the simplest source coding problem for a channel code; the relay wishes to compress Y_R^n to help the receiver decode X^n.

My endeavour on primitive relay channels follow the second perspective: Seeing communicating over a PRC as a source coding problem *for a channel code.* [1] Note that for each input symbol $X = x$, what the relay terminal shall observe is a random variable that is defined by the conditional pmf $p(y_R|x)$; Once a codebook is chosen, what the relay terminal shall observe would be a random process, each random variable of which depends on the transmitted symbol. The terminology "for a channel code" is adopted to emphasize the phenomenon that what needs to be compressed at the relay terminal depends on the channel codebook. This is also reflected in the optimization over all possible codebooks in Theorem 31 and equation (Equation 4.10) for resolving the minimum required conference rate, which is the main focus of this study.

In particular, the question we are interested in is how to operate the relay terminal to achieve the maximal possible network message rate while using the least number of bits on the conference link. we are driven by the straightforward intuition that the core function of the relay is to help the destination in disambiguating the channel inputs, i.e. to provide "what the destination needs". The relay need not decipher the channel inputs (messages) nor transmit what the destination can infer about the channel inputs from its own received signals. A relay's goal is not to decode the message - this is why Decode-and-Forward fails in general; it is not to provide "what the destination does not want", i.e. the noise, - this is why Amplify-and-Forward fails in general; nor is it desirable to waste its communication to send "what destination already possesses". One might argue that Partial Decode-and-Forward and

---

[1] Surely, the source coding problem and the design of the channel codebooks will be coupled and/or entangled.

Compress-and-Forward embody the idea of providing "what the destination needs" to some extent. However, we are not aware of any *explicit* attempt to characterize and quantify this intuition, which could potentially lead to a new relaying strategy with improved rates.

What the relay should forward depends on both the broadcasting links and the allowable conference rate r. When r is infinite or *large enough*, the relay can simply forward everything it has observed to the destination terminal. Thus, the primitive relay channel effectively turns into a point-to-point channel with single input and two outputs, say $(\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R)$, whose capacity is known. The natural question to ask is how large the conference link capacity r should be to ensure that the PRC network can achieve the capacity of the point-to-point channel $(\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R)$. We denote this capacity as the *single-input multi-output (SIMO) upper bound* for the given PRC channel. When conference rate r is big enough such that the SIMO upper bound can be achieved, We say that an "effectively full cooperation" between the relay and destination terminals can be established.

The *small-error* [1] version of this question was first proposed in (17) and remains open. We propose the zero-error version of this problem and obtain the exact solution for any fixed number of channel use n. Next, let us see a toy problem, which explains the motivation of this study: the advocation of having the relay terminal deliver "only what the destination needs" as well as the introduction of studying the "zero-error" communication over a PRC.

---

[1] Communication allowing a vanishing probability of error is called *small-error* or $\epsilon-$*error* communication, while communication without error is called *zero-error* or 0-*error* communication.

3.2 A motivating example

Take for example a PRC with $p(y, y_R|x) = p(y|x)p(y_R|x)$ as in Figure 14. The destination, upon receiving Y can tell whether $\{1, 2\}$ or $\{3, 4\}$ were sent, but not which message within those sets. The relay can "provide the destination what it needs" by forwarding E or O, i.e. whether the X was even or odd. This amounts to considerable savings for the conference link capacity with respect to sending Y_R directly, and allows the destination to fully resolve which X was sent as long as the conference link capacity is at least 1 bit. Please check the detailed explanation in Figure 17.

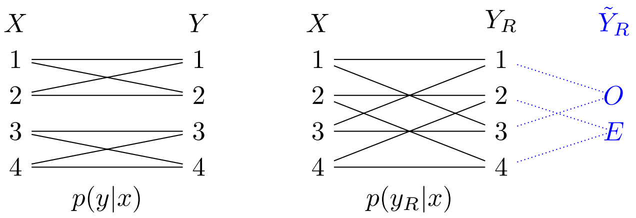


Figure 14. Toy Problem: $p(y, y_R|x) = p(y|x)p(y_R|x)$. A solid link indicates the probability value $p(*|x)$ is positive, where $*$ indicates y or y_R.

It may be checked that this simple channel does not fall into a class of PRCs for which capacity is known, i.e. it is not a degraded, semideterministic, orthogonal-component, or semideterministic PRC.

The next question is how one interpret this toy example and decide if it is possible and how to generalize the insights that its solution embodies, if any.

First try: take this toy problem as a small-error communication problem. Assume that the probability value on each solid link in Figure 14 is equal to $\frac{1}{2}$. Assume $r = 1$ bit.

It can be checked that regardless of the distribution on X, we have $H(Y|X) = H(Y_R|X) = 1$ bit. It can also be checked that the $H(Y) \leq 2$ bits and the equality is achieved when $\Pr[X \in \{1,3\}] = \Pr[X \in \{2,4\}] = \frac{1}{2}$. Similarly, $H(Y_R) \leq 2$ bits and the equality is achieved when $\Pr[X \in \{1,2\}] = \Pr[X \in \{3,4\}] = \frac{1}{2}$. It is also clear that the capacity of this channel, denoted as C_ϵ, should be upper bounded by the maximal entropy of X, i.e., $C_\epsilon \leq \max_{p(x)} H(X)$. That is, $C_\epsilon \leq \log 4 = 2$ bits.

We next try to apply various communication schemes on this toy channel and compute the maximal achievable rate under each scheme. Note that these achievable rates serve as lower bounds on the capacity C_ϵ.

- By direct transmission, i.e. using only the direct link from the source terminal to the destination terminal, we have the achievable rate $R_{\epsilon,1} = \max_{p(x)} I(X;Y)$. It can be checked that $R_{\epsilon,1} = 1$ bit and the maximum is achieved when $\Pr[X \in \{1,3\}] = \Pr[X \in \{2,4\}] = \frac{1}{2}$.
- When the direct link is not used, we have $R_{\epsilon,2} = \max_{p(x)} \min\{r, I(X;Y_R)\}$. $R_{\epsilon,2} = \min\{r, 1\} = 1$ bit, because $\max_{p(x)} I(X;Y_R) = 1$ bit.

- When the relay adopts the Decode-and-Forward strategy, we have $R_{\epsilon,3} = \max_{p(x)} \min\{I(X;Y) + r, I(X;Y_R)\}$. $R_{\epsilon,3} = \max_{p(x)} I(X;Y_R)$, because $I(X;Y_R) \leq I(X;Y) + r$. Thus, $R_{\epsilon,3} = 1$ bit. The equality is achieved when $\Pr[X \in \{1,2\}] = \Pr[X \in \{3,4\}] = \frac{1}{2}$.
- When Partial Decode-and-Forward is adopted, we have $R_{\epsilon,4} = \max_{p(u,x),\|\mathcal{U}\| \leq \|\mathcal{X}\|} \min\{I(X;Y) + r, I(U;Y_R) + I(X;Y|U)\}$. It can be checked that $R_{\epsilon,4} = 2$ bits and the equality is achieved when $\Pr[X = 1] = \Pr[X = 4]$ and $\Pr[X = 2] = \Pr[X = 3]$, and $U = 1_{Y_R \in \{1,3\}}$, meaning $U = 1$ when $Y_R \in \{1,3\}$ and $U = 0$ otherwise.

Note that the achievable rate by Partial Decode-and-Forward strategy coincides with the upper bound of the capacity. Thus, we can claim that the capacity of this channel is 2 bits/channel use.

We also remark that this capacity $C_\epsilon = 2$ bits per channel use is achieved in one-shot, i.e. by using only one channel use and requires no block coding. Furthermore, the probability of error is exactly 0.

We can see that the intuitive "Even/Odd" mapping in Figure 14 from Y_R's to the two labels E, O coincides the Partial Decode-and-Forward relaying strategy. But since the Partial Decode-and-Forward scheme involves the axillary random variable U, it is in general not clear how to construct a proper communication scheme.

A second try: take this toy problem as a zero-error communication problem. We are excited about the effectiveness and the efficiency of the "Even/Odd" mapping, but we are frustrated to see that this straightforward and intuitive mapping seems to just be an application of the traditional Partial Decode-and-Forward relaying scheme, which is not constructive.

We note that the essence of this mapping is that Y_R needs not to decode X, it just needs to say whether X is even or odd (E or O), which is exactly the information that the destination terminal lacks about the transmitted symbol. This motivates us to realize that

the essential role of a relay is to only provide "what the destination needs"

To define "what the destination needs", we need to define the destination terminal's goal. A natural setting is to request the destination terminal to obtain as much information about X as if the genie pair (Y, Y_R) was given, based on its own observation and the message sent by the relay terminal through the conference link. That is to let the whole network to achieve the SIMO bound.

Given the goal of achieving the SIMO bound, we still need to represent and quantify "what the destination does and does not know about X". This line of thinking naturally leads to the zero-error communication setting and the proposal of the study on communicating over a PRC without error. In Chapter 4, Colour-and-Forward relaying is developed for the zero-error PRC communication problem and is shown to be optimum for any fixed number of channel use. Based on the insights gained in the zero-error scenario, the ϵ-Colour-and-Forward relaying that mimics the construction of that in the zero-error scenario is proposed in Chapter 5.

3.3 Notation convention

Throughout Chapter 4 and Chapter 5, we will use subscripts $_z$ adn $_\epsilon$ to denote the zero-error and small-error context respectively. we use upper and lower cases to differentiate the overall network message rate R_z and the conference rate r_z. The superscripts of a graph or set indicate the vertex nodes or the elements of the set, while their subscripts denote the needed parameters. Bold font, as well as the superscript ^{n} are both used to denote a sequence of length n. Let random pair $(X, Y) \sim p(x, y)$ and $(X, Y) \in \mathcal{X} \times \mathcal{Y}$. Denote the marginals for X and Y by $p(x)$ and $p(y)$ respectively. When a conditional joint pmf $p(y, y_R|x)$ with support $\mathcal{X}$ and output $\mathcal{Y} \times \mathcal{Y}_R$ is restricted to input $\mathcal{K}$, we denote its *induced conditional pmf, support and output* by $p_{\mathcal{K}}(y, y_R|x)$, $\mathcal{K}$ and $\mathcal{Y}|_{\mathcal{K}} \times \mathcal{Y}_R|_{\mathcal{K}}$ respectively. All logarithms are base 2.

3.3.1 Graph theoretic notation

A graph $G(V, E)$ consists of a set V of vertices or nodes together with a set E of edges, which are two-element subsets of V. Two nodes connected by an edge are called *adjacent.* we will usually drop the V, E indices in $G(V, E)$.

An *independent set* of a graph G is a set of vertices, no two of which are adjacent. Let *independence number* $\alpha(G)$ be the maximum cardinality of all independent sets. A *maximum independent set* is an independent set that has $\alpha(G)$ vertices. Note that one graph can have multiple maximum independent sets. A *colouring* of graph G is any function c over the vertex set such that c^{-1} induces a partition of the vertex set into independent sets of G. The *chromatic number* $\chi(G)$ of the graph G is the least number of colours in any colouring. A *minimum colouring* of graph G uses $\chi(G)$ colours.

The *strong product* $G \boxtimes H$ of two graphs G and H is defined as the graph with vertex set $V(G \boxtimes H) = V(G) \times V(H)$, in which two distinct vertices (g, h) and (g', h') are adjacent iff g is adjacent or equal to g' in G and h is adjacent or equal to h' in H. $G^{\boxtimes n}$ denotes the strong product of n copies of G.

A *confusability graph* $G^X_{p(y|x)}$ of X given Y, specified by conditional probability function $p(y|x)$ with support $\mathcal{X}$ and output $\mathcal{Y}$, is a graph whose vertex set is $\mathcal{X}$ and an edge is placed when two different nodes $x, x' \in \mathcal{X}$ may be "confused", that is, if $\exists y \in \mathcal{Y} : p(y|x) \cdot p(y|x') > 0$. For a given conditional probability function $p(y|x)$, we denote $S^{X|Y}_{p(y|x)}(y) := \{x : p(y|x) > 0\}$ as the *conditional support of* $Y = y$. Thus, the confusability graph $G^X_{p(y|x)}$ can be equivalently constructed by fully connecting the nodes inside each conditional support $S^{X|Y}_{p(y|x)}(y)$, for all $y \in \mathcal{Y}$.

Graph $G(A)$ is the induced subgraph of graph G, with vertex set $A \subseteq V(G)$ and edge set $(A \times A) \cap E(G)$.

3.4 Contribution

The main contribution of this study:

- Zero-error communication over a primitive relay channel is for the first time proposed and studied.
- Colour-and-Forward relaying is designed and serves as an example of how one may explicitly exploit channel structure with the intuitive goal of having relay transmit "only what the destination needs"
- Show that Color-and-Forward algorithm is optimal– for any fixed number of channel uses, this relaying scheme requires the smallest conference link capacity if one desires to achieve the SIMO upper bound.
- Develop various bounds on the minimum required conference rate r^*_z

- Develop an alternative capacity achieving scheme for the small-error point-to-point channel, which embodies an explicit codebook construction using graph theoretic notation.
- Derive the ϵ-Colour-and-Forward relaying for small-error PRC communication problems, which serves a good example of how to transfer insights gained from studying the zero-error problem into the small-error domain.

CHAPTER 4

0-COLOUR-AND-FORWARD

In this chapter, the zero-error communication over a primitive relay channel is first defined before the Colour-and-Forward algorithm is introduced. We will show that the Colour-and-Forward relaying defined in Definition 28 is optimum for any fixed number of channel uses n. We do not have a solution for the asymptotic scenario, but some bounds for it are obtained based on the n-shot Colour-and-Forward relaying scheme.

4.1 Zero-error preliminaries

The zero-error capacity of a point-to-point discrete memoryless channel was initially studied by Shannon in (18) in 1956; see (19; 20) for further zero-error capacity details.

Consider zero-error communication over a point-to-point channel $(\mathcal{X}, p(y|x), \mathcal{Y})$. First, note that only whether $p(y|x)$ is zero or not matters for communication without error. Next, consider first communicating over a single channel use: the maximal number of channel inputs the destination can distinguish without error is $\alpha(G^X_{p(y|x)})$, the maximum number of vertices that are non-adjacent, or pairwise distinguishable. When multiple channel uses are allowed, we know that $\alpha([G^X_{p(y|x)}]^{\boxtimes n})$ is the number of distinguishable channel inputs X^n, where $[G^X_{p(y|x)}]^{\boxtimes n}$ is the strong product of n copies of graph $G^X_{p(y|x)}$.[1] Thus, the zero-error capacity of a point-

---

[1] Note that the n-fold strong product graph $[G^X_{p(y|x)}]^{\boxtimes n}$ is equivalent to graph $G^{X^n}_{p(y^n|x^n)}$, which is the confusability graph directly constructed from *the compound channel* $(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y}^n)$ with $p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i)$.

to-point channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ is defined as the supremum of all achievable message rates, i.e., $\sup_n \frac{1}{n} \log \alpha([G^X_{p(y|x)}]^{\boxtimes n})$.

The zero-error capacity is then characterized as (19)

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha([G^X_{p(y|x)}]^{\boxtimes n}) = \lim_{n\to\infty} \log \sqrt[n]{\alpha([G^X_{p(y|x)}]^{\boxtimes n})},$$

which may be upper and lower bounded as (18; 19):

$$\log \alpha(G^X_{p(y|x)}) \le \lim_{n\to\infty} \log \sqrt[n]{\alpha([G^X_{p(y|x)}]^{\boxtimes n})} \le \log \|\mathcal{X}\|$$

where $\|\mathcal{X}\|$ is the cardinality of the input alphabet, which is the maximal number of possible inputs per channel use. Note the limit exists by Lemma 22.

Lemma 22. *Let G denote the confusability graph specified by $p(y|x)$. Then the sequence $\{\log \sqrt[n]{\alpha(G^{\boxtimes n})}\}_{n=1}^{\infty}$ converges to $\sup\{\log \sqrt[n]{\alpha(G^{\boxtimes n})}, n = 1, 2, \cdots\}$.*

Proof. It can be checked that the sequence $\{\alpha(G^{\boxtimes n})\}_{n=1}^{\infty}$ is super-multiplicative, i.e. $\alpha(G^{\boxtimes (n_1+n_2)}) \ge \alpha(G^{\boxtimes n_1}) \cdot \alpha(G^{\boxtimes n_2})$ for any indices n_1, n_2. Thus, the sequence $\{\log \alpha(G^{\boxtimes n})\}_{n=1}^{\infty}$ is super-additive and each item is non-negative. By Fekete's Lemma, the limit $\lim_{n\to\infty} \log \sqrt[n]{\alpha(G^{\boxtimes n})}$ exists and is equal to $\sup\{\log \sqrt[n]{\alpha(G^{\boxtimes n})}, n = 1, 2, \cdots\}$. □

Remark 10. *The behavior of the sequence of independence numbers for strong product graphs, say $\{\alpha(G^{\boxtimes n})\}_{n=1}^{\infty}$, is a long standing open question and attracts attention in research fields like graph theory and combinatorics (21), (22), and is notoriously difficult. As discussed in the*

Introduction section, it took 23 *years for the researchers to prove that Shannon's conjecture that the Shannon capacity for the pentagon graph/channel is* $\frac{1}{2}\log 5$*, while the value of the Shannon capacity for the* 7*-cycle graph remains unresolved (23). Given this, it is understandable that researchers are reluctant to approach zero-error communication problems over a multiple-terminal network. We thus would like to emphasize that the formation and proposal of zero-error communication over a PRC is a real contribution by itself.*

4.2 Zero-error communication over a primitive relay channel and the minimum conference rates r_z^* and $r_z^{*(n)}$

We first define zero-error communication over a PRC, then we introduce cut-set bounds, SIMO bounds and the minimum conference rates r_z^* and $r_z^{*(n)}$, which are the quantities and optimization problems of interest in this study.

4.2.1 Zero-error communication over a primitive relay channel

As shown in Figure 15, an *n-shot protocol* $(n, \underline{\mathcal{X}}, h, g)$ for zero-error communication over a PRC $((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R),\ r_z)$ is composed of a codebook $\underline{\mathcal{X}} \subseteq \mathcal{X}^n$, a r_z-admissible relaying function $h : \mathcal{Y}_R^n \to \mathcal{W}_R$ which satisfies $\|\mathcal{W}_R\| \leq 2^{n \cdot r_z}$ and a decoding function $g : \mathcal{Y}^n \times \mathcal{W}_R \to \underline{\mathcal{X}}$. Let $\hat{\underline{X}}$ and $\hat{W}$ demote the estimate for codeword $\underline{X}$ and the message W respectively. Note that $\hat{W} = \phi^{-1}(\hat{\underline{X}}) = \phi^{-1}(g(Y^n, W_R))$. Because the mapping $\phi(\,)$ is bijective, decoding message $W \in \mathcal{W}$ is equivalent to decoding codeword $\underline{X} \in \underline{\mathcal{X}}$. We will not distinguish these two concepts and abuse notation $\hat{w} \in \underline{\mathcal{X}}$ and $\hat{W} = g(Y^n, W_R)$ for the decoding result at the destination.

A message rate $R_z := \frac{1}{n}\log\|\mathcal{W}\| = \frac{1}{n}\log\|\underline{\mathcal{X}}\|$ is *achievable* if there exists an n-shot protocol $(n, \underline{\mathcal{X}}, h, g)$ over a PRC $((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R),\ r_z)$ achieving zero error, i.e. $\Pr[g(y, w_R) \neq$

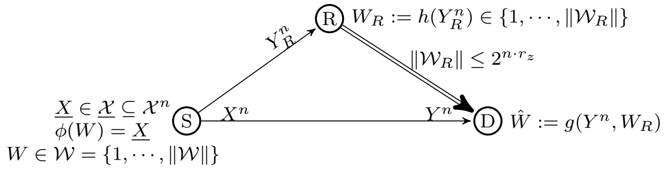

R
$W_R := h(Y_R^n) \in \{1, \cdots, \|\mathcal{W}_R\|\}$
Y_R^n
$\|\mathcal{W}_R\| \le 2^{n \cdot r_z}$
$\underline{X} \in \underline{\mathcal{X}} \subseteq \mathcal{X}^n$
$\phi(W) = \underline{X}$
$W \in \mathcal{W} = \{1, \cdots, \|\mathcal{W}\|\}$
S
X^n
Y^n
D
$\hat{W} := g(Y^n, W_R)$

Figure 15. An n-shot protocol $(n, \underline{\mathcal{X}}, h, g)$ for zero-error communication over a PRC $((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R), r_z)$, with an encoder ϕ, a codebook $\underline{\mathcal{X}}$, a relaying function h and a decoding function g.

$w] = 0$ for all values $w \in \underline{\mathcal{X}}$. The capacity C_z of zero-error communication over a PRC $((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R), r_z)$ is the supremum of all possible achievable rates R_z for any n. Clearly, C_z is at most $\log \|\mathcal{X}\|$.

As indicated in Chapter 3 , the goal of this study is to study what is the best compression algorithm that leads to the most efficient summary of relay's observations while enabling the relay to help the destination terminal to its maximum capability. Basically, we keep the broadcasting links $(\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R)$ unchanged in a PRC and ask how C_z changes as r_z varies. So in the rest of theis chapter, we will use $C_z(r_z)$ to denote the zero-error capacity of a PRC $((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R), r_z)$.

4.2.2 “Cut-set” bound for C_z

Before formally proposing the SIMO bound for a PRC channel, we state the cut-set bounds first. Note that the inequality in (Equation 4.1) involves the value of r, which does not show up in the SIMO bound in (Equation 4.2).

Proposition 23 (Zero-error capacity cut-set bound)**.** *The capacity $C_z(r_z)$ of the 0-error PRC is upper bounded by*

$$\begin{aligned} C_z(r_z) \leq \min\{ & \log \lim_{n\to\infty} \sqrt[n]{\alpha([G^X_{p(y|x)}]^{\boxtimes n})} + r_z, \\ & \log \lim_{n\to\infty} \sqrt[n]{\alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})} \}. \end{aligned} \tag{4.1}$$

Proof. Note that $\log \lim_{n\to\infty} \sqrt[n]{\alpha([G^X_{p(y|x)}]^{\boxtimes n})}$ is the zero-error capacity of the direct link from the source to the destination terminal; if this is orthogonal to what is received from the relay, we obtain the first bound. The second bound is obtained by recognizing $\log \lim_{n\to\infty} \sqrt[n]{\alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})}$ as the zero-error capacity of a point-to-point channel $p(\tilde{y^n}|x)$ with $\tilde{y^n} = (y^n, y_R^n)$, obtained by giving (genie) y_R^n to the destination. $\square$

4.2.3 SIMO bounds and the minimum conference rates r_z^* and $r_z^{*(n)}$

As discussed in Chapter 3, the question we are after is how to operate the relay terminal so that with the minimum conference rate one can achieve the maximal possible network message rate, i.e., the capacity of the virtual point-to-point channel $(\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R)$. Formally, we propose

Proposition 24 (Zero-error capacity SIMO upper bound)**.** *The zero-error capacity of a PRC channel*

$((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R),\ r_z)$ is upper bounded by

$$C_z(r_z) \leq SIMO := \log \lim_{n\to\infty} \sqrt[n]{\alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})} \}. \tag{4.2}$$

This is established by allowing full cooperation between the relay and destination terminals. Proposition 24 can as well be derived from the cut-set bounds in Proposition 23 by setting r_z equal to infinity, making the first quantity in (Equation 4.1) irrelevant for the minimization.

Clearly, $C_z(r_z)|_{r_z=\infty} = SIMO$. Also, $C_z(r_z) = SIMO$, when $r_z \geq \log \|\mathcal{Y}_R\|$, implying that the conference link can afford letting the relay terminal transmit its complete observation without compression. We are interested in the minimum value of r_z such that $C_z(r_z) = SIMO$, i.e., what is the best compression rate at the relay terminal. Formally, we define the minimum conference rate r_z^* that can enable an effectively full cooperation between the relay and destination terminals as:

Definition 25 (The minimum conference rate r_z^*)**.**

$$r_z^* := \inf\{r_z : C_z(r_z) = SIMO\}\,. \tag{4.3}$$

When restricted to the n channel uses only, $n = 1, 2, \cdots$, let $C_z^{(n)}(r_z)$ denote the supremum of messages rates that are achievable by using n channel uses. Similarly, we can derive and define the the corresponding SIMO bound $SIMO(n)$ and the minimum conference rate $r_z^{*(n)}$:

Proposition 26 (The n-shot zero-error capacity SIMO upper bound)**.** *The n-shot zero-error capacity of a PRC channel $((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R),\, r_z)$ is upper bounded by*

$$C_z^{(n)}(r_z) \leq SIMO(n) := \log \sqrt[n]{\alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})}\,\}, \tag{4.4}$$

where $SIMO(n)$ *denotes the maximum achievable rate of the virtual channel* $(\mathcal{X}^n, p(y^n, y_R^n|x^n), \mathcal{Y}^n \times \mathcal{Y}_R^n)$.

Definition 27 (The n-shot minimum conference rate $r_z^{*(n)}$)**.**

$$r_z^{*(n)} := \inf\{r_z : C_z^{(n)}(r_z) = SIMO(n)\} . \tag{4.5}$$

Remark 11. *We emphasize again that*

$$SIMO = \sup_{n=1,2,\cdots} SIMO(n) = \lim_{n\to\infty} SIMO(n) . \tag{4.6}$$

The first equality comes from the definition of the capacity of a point-to-point channel to be the superemum of its all achievable rates. The second equality is established by by Lemma 22.

The plan is to first derive an upper bound on $r_z^{*(n)}$*, for any fixed number of channel use* n*. We then derive various upper bounds on* r_z^* *based on this sequence of upper bounds of* $r_z^{*(n)}$*. For details, please check Chapter 4.7.*

4.3 Colour-and-Forward relaying: from an intuition to an algorithm

4.3.1 Colour-and-Forward relaying: from an intuition to an algorithm

The Colour-and-Forward relaying strategy that will be presented is based on the intuition of providing “what the destination needs”, i.e. remaining information lossless, while trying to minimize the number of bits needed to do so. In this subsection, we demonstrate how to transform this intuition into a practical and executable construction of a relaying algorithm using 1-shot case.

Sitting at the destination terminal, for a given a conditional joint pmf $p(y, y_R|x)$ with support $\mathcal{X}$ and output $\mathcal{Y} \times \mathcal{Y}_R$, we consider an arbitrary observation $Y = y$. Given this observation $Y = y$, the destination knows that the channel input symbol lies in the corresponding conditional support $S^X_{p(y|x)}(y)$. What the destination needs is to resolve the ambiguity among which x out of $S^X_{p(y|x)}(y)$ was sent. Furthermore, according to the joint pmf $p(y, y_R|x)$, the destination knows what the relay could have observed when the channel input symbol is $\mathbf{X} = \mathbf{x}$ given observation $Y = y$, i.e.

$$B^{Y_R}_{p(y,y_R|x)}(x, y) := \{y_R : p(y, y_R|x) > 0 \text{ for given } x \text{ and } y\}.$$

In order to help D distinguish which channel input symbol x was actually transmitted, the relay terminal needs to differentiate different collections of y_R, i.e., $B^{Y_R}_{p(y,y_R|x)}(x, y)$ in terms of the first index x for a given second index y. We propose to do so through the **Construction of the graph** $G_R(V, E)$ as shown in Table I.

For example, in the toy problem in 3.2 discussed in the Introduction (??) section, when the destination terminal observes $Y = 1$, it knows that the transmitted symbol can be either $X = 1$ or $X = 2$ (assuming that they can both show up in the codebook), as shown in Table II. When the true transmitted symbol is indeed $X = 1$, the relay terminal would have observed a $Y_R = y_R$ where $y_R \in \{1, 3\}$ based on the conditional joint pmf $p(y, y_R|x) = p(y|x)p(y_R|x)$. On the other hand, when the true transmitted symbol is indeed $X = 2$, the relay terminal would have observed a $Y_R = y_R$ where $y_R \in \{2, 4\}$. Thus, to help the destination terminal to further

1) Vertices: $V = \mathcal{Y}_R := \{y_{R1}, y_{R2}, \cdots y_{R\|\mathcal{Y}_R\|}\}$;
2) Edges: for every $y \in \mathcal{Y}$, construct a sequence of subsets of $\mathcal{Y}_R$, $B^{Y_R}_{p(y,y_R|x)}(x, y)$, indexed by x, where $x \in S^{X|Y}_{p(y|x)}(y)$. Edges are placed by fully connecting any two subsets $B^{Y_R}_{p(y,y_R|x)}(x, y)$ and $B_{Y_R}(x', y)$, where $x \neq x'$ (i.e. put an edge between every pair (y_R, y'_R) where $y_R \in B^{Y_R}_{p(y,y_R|x)}(x, y)$ and $y'_R \in B_{Y_R}(x', y)$.) Note that for a given $Y = y$, the y_R vertices that are inside one $B^{Y_R}_{p(y,y_R|x)}(x, y)$ need not be connected.

TABLE I

CONSTRUCTION OF THE GRAPH $G_R(V, E)$

decide which symbol X out of the two-symbol set $\{1, 2\}$ is the true transmitted symbol, it suffices for the relay terminal to tell group $\{1, 3\}$ from group $\{2, 4\}$. Note that it is immaterial, from the perspective of helping the destination terminal, if the relay will distinguish its observations one from another within a given group; for example, whether $Y_R = 1$ and $Y_R = 3$ (from group $\{1, 3\}$) will be distinguished or not by the relay terminal has no interest to the destination terminal (for classifying the ambiguity when $Y = 1$ was observed).

We listed in Figure 16 the needs for other observations, say, $Y = 2, 3, 4$ and *superimpose all the edge constraints* to get the compression graph G_R. We can label (or color) the 4 vertices by E and O, standing for *Even* and *Odd*, and satisfy the edge constraints, meaning no two connected

	The ambiguity at destination $S^X_{p(y\|x)}(y)$	What relay has observed $B^{Y_R}_{p(y,y_R\|x)}(x,y)$	What the destination wants from relay expressed as edge constraints
$Y=1$	$X=1$ $X=2$	$\{1,3\}$ $\{2,4\}$	$1-2, 1-4, 3-2, 3-4$

TABLE II

EXPRESS “WHAT THE DESTINATION TERMINAL NEEDS” WHEN IT OBSERVES $Y=1$ AS EDGE CONSTRAINTS IN THE COMPRESSION GRAPH.

vertices have the same label (or color). Thus, all the “needs” or request of the destination terminal is met. As discussed earlier in the Introduction section (??), forwarding E or O to the destination terminal will help it to fully identify what X symbol has been transmitted.

As one can see from the three confusability graphs listed in Figure 17: (1) the destination terminal cannot fully distinguish all four transmitted symbols based on its own observations, i.e. graph $G^X_{p(y|x)}$ is not edge-free; (2) Graph $G^X_{p(y,\tilde{y}_R|x)}$ is free of edges, meaning the destination terminal can fully distinguish four transmitted symbols based on its own observations and two labels (or colours), say $\tilde{Y}_R = E, O$; (3) Two graphs $G^X_{p(y,y_R|x)}$ and $G^X_{p(y,\tilde{y}_R|x)}$ are the same, termed as the property of being informationloss-less in Theorem 30, which We will show to be in general true.

	$S^X_{p(y\|x)}(y)$	$B^{Y_R}_{p(y,y_R\|x)}(x,y)$	edges
$Y=1$	$X=1$ $X=2$	$\{1,3\}$ $\{2,4\}$	$1-2, 1-4, 3-2, 3-4$
$Y=2$	$X=1$ $X=2$	$\{1,3\}$ $\{2,4\}$	$1-2, 1-4, 3-2, 3-4$
$Y=3$	$X=3$ $X=4$	$\{1,3\}$ $\{2,4\}$	$1-2, 1-4, 3-2, 3-4$
$Y=4$	$X=3$ $X=4$	$\{1,3\}$ $\{2,4\}$	$1-2, 1-4, 3-2, 3-4$

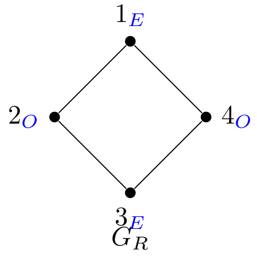


Figure 16. Express "what destination needs" by imposing compression constraints, i.e. the edge constraints on compression graph G_R. The relay can simply "provide the destination what it needs" by forwarding E or O, i.e. whether the X was even or odd.

4.3.2 Formal definitions of Colour-and-Forward algorithm

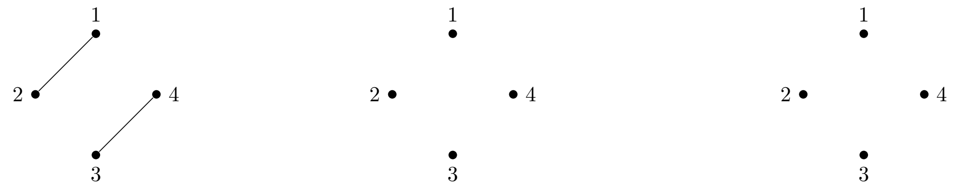
We now formally define the Colour-and-Forward graph and Colour-and-Forward relaying for any fixed number of channel use n. Consider the compound or symbol-extended broadcasting channel $(\mathcal{X}^n, p(y^n, y_R^n|x^n), \mathcal{Y}^n \times \mathcal{Y}_R^n)$ which is represented by a conditional joint pmf $p(y^n, y_R^n|x^n)$ with support $\mathcal{X}^n$ and output $\mathcal{Y}^n \times \mathcal{Y}_R^n$. The compression graph can be analogously defined and the vertex nodes are y_R^n's. Recall that bold font, as well as the superscript ^{n} are both used to denote a sequence of length n. The n-shot Colour-and-Forward graph is defined as:

Definition 28 (Colour-and-Forward graph $G_R^{(n)}$)**.** *Given a conditional joint pmf* $p(\mathbf{y}, \mathbf{y}_R|\mathbf{x})$ *with support* $\mathcal{X}^n$ *and output* $\mathcal{Y}^n \times \mathcal{Y}_R^n$*, graph* $G_R^{(n)}$ *is an undirected graph with vertex set* $\mathcal{Y}_R^n$ *and an edge* $\mathbf{y}_{R1} - \mathbf{y}_{R2}$ *is imposed when for some* $\mathbf{y}$*,* $\mathbf{x}_1 \neq \mathbf{x}_2$*,* $\Pr(\mathbf{Y} = \mathbf{y}, \mathbf{Y}_R = \mathbf{y}_{R1}|\mathbf{X} = \mathbf{x}_1) \cdot \Pr(\mathbf{Y} = \mathbf{y}, \mathbf{Y}_R = \mathbf{y}_{R2}|\mathbf{X} = \mathbf{x}_2) > 0$.

Y	$S^X_{p(y\|x)}(y)$	edges
1	$\{1,2\}$	$1-2$
2	$\{1,2\}$	$1-2$
3	$\{3,4\}$	$3-4$
4	$\{3,4\}$	$3-4$

(Y, Y_R)	$S^X_{p(y,y_R\|x)}(y, y_R)$	edges
$(1,1)$	$\{1\}$	$\emptyset$
$(1,3)$		
$(2,1)$		
$(2,3)$		
$(1,2)$	$\{2\}$	$\emptyset$
$(1,2)$		
$(2,4)$		
$(2,4)$		
$(3,1)$	$\{3\}$	$\emptyset$
$(3,3)$		
$(4,1)$		
$(4,3)$		
$(3,2)$	$\{4\}$	$\emptyset$
$(3,2)$		
$(4,4)$		
$(4,4)$		

$(Y, \tilde{Y}_R)$	$S^X_{p(y,\tilde{y}_R\|x)}(y, \tilde{y}_R)$	edges
$(1,O)$	$\{1\}$	$\emptyset$
$(1,O)$		
$(2,O)$		
$(2,O)$		
$(1,E)$	$\{2\}$	$\emptyset$
$(1,E)$		
$(2,E)$		
$(2,E)$		
$(3,O)$	$\{3\}$	$\emptyset$
$(3,O)$		
$(4,O)$		
$(4,O)$		
$(3,E)$	$\{4\}$	$\emptyset$
$(3,E)$		
$(4,E)$		
$(4,E)$		



(a) $G^X_{p(y|x)}$. (b) $G^X_{p(y,y_R|x)}$. (c) $G^X_{p(y,\tilde{y}_R|x)}$.

Figure 17. Three confusability graphs in the toy problem in Figure 14.

Note that $G_R^{(n)}$ is short for the standard notation $G^{Y_R^n}_{p(y^n,y_R^n|x^n)}$, which explicitly indicates that the vertex nodes of the graph are relay's observation y_R^n's and the edges are determined by the broadcasting links $(\mathcal{X}^n, p(y^n, y_R^n|x^n), \mathcal{Y}^n \times \mathcal{Y}_R^n)$.

It can be checked that the construction of the graph $G_R(V, E)$ described in Table Table I is consistent with the Definition 28 when n is 1.

Remark 12. *Note that graph $G_R^{(2)}$, constructed from $(\mathcal{X}^2, p(y^2, y_R^2|x^2), \mathcal{Y}^2 \times \mathcal{Y}_R^2)$, cannot be derived from graph $G_R^{(1)}$, which is constructed from $(\mathcal{X}, p(y, y_R|x), \mathcal{Y}\times\mathcal{Y}_R)$, by any standard graph* *product operations. In particular, the relationship between graphs* $G_R^{(2)}$ *and* $G_R^{(1)}$ *is different from any of the four standard graph products surveyed in (24). To emphasize this fact, we adopt open parenthesis for* n *in the superscript, i.e.* $^{(n)}$*, to denote the* n*-shot Colour-and-Forward graph* $G_R^{(n)}$*. The complexity of graph* $G_R^{(n)}$ *is a reflection of the channel structure, of which one can make use to provide a more efficient compression algorithm by adopting a proper block coding length* n*.*

We now propose a novel relaying function $W_R^{*(n)}$, based on the Colour-and-Forward graph $G_R^{(n)}$.

Definition 29 (Colour-and-Forward relaying $W_R^{*(n)}$)**.** *Given a conditional joint pmf* $p(\mathbf{y}, \mathbf{y}_R|\mathbf{x})$ *with support* $\mathcal{X}^n$ *and output* $\mathcal{Y}^n \times \mathcal{Y}_R^n$*, we define the Colour-and-Forward relaying* $W_R^{*(n)}$ *as a function of* $\mathbf{Y}_R$ *by a minimum colouring* c *with* $\chi(G_R^{(n)})$ *colours on graph* $G_R^{(n)}$*:*

$$W_R^{*(n)} := c(\mathbf{Y}_R)$$

where graph $G_R^{(n)}$ *is defined in Definition 28. (Note that* c *is not unique.)*

4.4 Colour-and-Forward relaying is information-lossless

Recall that the Colour-and-Forward relaying $W_R^{*(n)}$ is defined as a minimum coloring function on the Colour-and-Forward graph, as shown in Definition 29 and Definition 28, for any n channel uses. In this section, we will show that this relaying $W_R^{*(n)}$ is *infomrationloss-less* and leads to a complete characterization of the minimum required conference rate $r_z^{*(n)}$, for any fixed number of channel uses, proposed as an optimization problem in (27). In subsection We will also discuss the connection between the Colour-and-Forward relaying and the Witsenhausen's source coding problem (25).

4.4.1 Information-lossless Theorem

The *information-lossless* property demonstrated in the toy problem in Figure Figure 17 can generalized as:

Theorem 30 (Information-lossless)**.** *Colour-and-Forward relaying* $W_R^{*(n)}$ *is information-lossless in the sense that, together with* Y^n*, the destination terminal can infer as much information about the transmitted symbol* X^n *as if the genie* Y_R^n *was received. Mathematically,*

$$\{S^{X^n}_{p(y^n,w_R^{*(n)}|x^n)}(y^n,w_R^{*(n)}):\ (y^n,w_R^{*(n)})\in\mathcal{Y}^n\times\mathcal{W}_R^{*(n)}\}=\{S^{X^n}_{p(y^n,y_R^n|x^n)}(y^n,y_R^n):\ (y^n,y_R^n)\in\mathcal{Y}^n\times\mathcal{Y}_R^n\} \tag{4.7}$$

Furthermore, it holds that $G^{X^n}_{p(y^n,y_R^n|x^n)}=G^{X^n}_{p(y^n,w_R^{*(n)}|x^n)}$*, i.e. the confusability graph on* $\mathcal{X}^n$ *from* $p(y^n,y_R^n|x^n)$ *equals that from* $p(y^n,w_R^{*(n)}|x^n)$*.* $W_R^{*(n)}$ *is generated by Definition 29 from* $p(y^n,y_R^n|x^n)$ *with support* $\mathcal{X}^n$ *and output* $\mathcal{Y}^n\times\mathcal{Y}_R^n$*.*

Recall that a confusability graph, for example $G^{X^n}_{p(y^n,y_R^n|x^n)}$, by definition is characterized by the collection of conditional joint supports, for example $\{S^{X^n}_{p(y^n,y_R^n|x^n)}(y^n,y_R^n):(y^n,y_R^n)\in\mathcal{Y}^n\times\mathcal{Y}_R^n\}$, and does not depend on the actual probability values. Thus, once the equality in (Equation 4.7) is established, it immediately follows that $G^{X^n}_{p(y^n,y_R^n|x^n)}=G^{X^n}_{p(y^n,w_R^{*(n)}|x^n)}$ holds. We will show the proof in the next subsection (subsection 4.4.2).

Recall that in Subsection 4.3.1 we showed that the "Even/Odd" mapping for the toy problem indicated in Figure 14 in Chapter 3.2, can be derived by creating a Colour-and-Forward graph and is indeed a Colour-and-Forward relaying, by recognizing $\tilde{Y}_R$ as $W_R^{*(1)}$. It can be checked in Figure 17 that two sets of conditional supports, indicated in equation (Equation 4.7) are the same and two corresponding confusability graphs are equal.

4.4.2 Proof of Theorem 30

We next show the proof of Theorem 30.

Proof. We first rewrite equation (Equation 4.7) in a more compact way:

$$\{S^{\mathbf{X}}_{p(\mathbf{y},w_R^{*(n)}|\mathbf{x})}(\mathbf{y},w_R^{*(n)}):\ (\mathbf{y},w_R^{*(n)})\in\mathcal{Y}^n\times\mathcal{W}_R^{*(n)}\}=\{S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y},\mathbf{y}_R):\ (\mathbf{y},\mathbf{y}_R)\in\mathcal{Y}^n\times\mathcal{Y}_R^n\} \tag{4.8}$$

where the bold font is used to denote a sequence of length n for succinctness. Note that the superscript $^{(n)}$ as denoted in Remark 12 is kept to emphasize that Colour-and-Forward graph $G_R^{(n)}$ cannot be constructed from $G_R^{(1)}$ via any standard graph product operations. Thus, $W_R^{*(n)}$ needs not to be related to $W_R^{*(1)}$.

Note $W_R^{*(n)}=c(\mathbf{Y}_R)$ is a deterministic function of $\mathbf{Y}_R$ by Definition 29. Thus given the conditional joint pmf $p(\mathbf{y},\mathbf{y}_R|\mathbf{x})$, the introduced conditional joint pmf $p(\mathbf{y},w_R^{*(n)}|\mathbf{x})$ is computable. We justify equality (Equation 4.8) by pointing out

$$S^{\mathbf{X}}_{p(\mathbf{y},w_R^{*(n)}|\mathbf{x})}(\mathbf{y},w_R^{*(n)})=\bigcup_{\mathbf{y}_R\in c^{-1}(w_R^{*(n)})}S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y},\mathbf{y}_R) \tag{4.9}$$

and showing that every non-empty $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0,\mathbf{y}_{R0})$ is equal to $S^{\mathbf{X}}_{p(\mathbf{y},w_R^{*(n)}|\mathbf{x})}(\mathbf{y}_0,w_{R0}^{*(n)})$, where $w_{R0}^{*(n)}=c(\mathbf{y}_{R0})$.

For every $(\mathbf{y}_0, \mathbf{y}_{R0})$ such that $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0}) \neq \emptyset$, we denote $c(\mathbf{y}_{R0}) = w^{*(n)}_{R0}$ and let $c^{-1}(w^{*(n)}_{R0}) = \{\mathbf{y}_{R0}, \mathbf{y}_{R1}, \cdots, \mathbf{y}_{R(K-1)}\}$, where $K \geq 1$ is the number of $\mathbf{y}_R$'s that are mapped to the same colour index $w^{*(n)}_{R0}$. When $K = 1$, $S^{\mathbf{X}}_{p(\mathbf{y},w^{*(n)}_R|\mathbf{x})}(\mathbf{y}_0, w^{*(n)}_{R0}) = S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0})$. When $K \geq 2$, $S^{\mathbf{X}}_{p(\mathbf{y},w^{*(n)}_R|\mathbf{x})}(\mathbf{y}_0, w^{*(n)}_{R0}) = S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0}) \cup S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R1}) \cup \cdots \cup S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R(K-1)})$. Note that $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0})$ is non-empty:

- when $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0})$ has only one element, say $\mathbf{x}_0$, it is true that $\Pr(\mathbf{Y} = \mathbf{y}_0, \mathbf{Y}_R = \mathbf{y}_{R0}|\mathbf{X} = \mathbf{x}_0) > 0$. By the construction of $W^{*(n)}_R$ in Definition 29, it holds that $\Pr(\mathbf{Y} = \mathbf{y}_0, \mathbf{Y}_R = \mathbf{y}_{Rt}|\mathbf{X} = \mathbf{x}_q) = 0$ for all $t = 1, \cdots, K-1$ and $\mathbf{x}_q \neq \mathbf{x}_0$. Otherwise, the presumption that $\mathbf{y}_{R0}$ and $\mathbf{y}_{Rt}$ share the same colour index $w^{*(n)}_{R0}$ leads to a contradiction. As a result, for all $t = 1, \cdots, K-1$, $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{Rt}) = \{\mathbf{x}_0\}$ when $\Pr(\mathbf{Y} = \mathbf{y}_0, \mathbf{Y}_R = \mathbf{y}_{Rt}|\mathbf{X} = \mathbf{x}_0) > 0$ and $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{Rt}) = \emptyset$ otherwise. Thus, we have $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0}) \cup S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R1}) \cup \cdots \cup S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R(K-1)}) = S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0})$, and $S^{\mathbf{X}}_{p(\mathbf{y},w^{*(n)}_R|\mathbf{x})}(\mathbf{y}_0, w^{*(n)}_{R0}) = S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0})$.
- when $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0})$ has more than one element, i.e., $\mathbf{x}_0, \mathbf{x}'_0 \in S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0})$ and $\mathbf{x}_0 \neq \mathbf{x}'_0$. Applying the argument above twice, we have $\Pr(\mathbf{Y} = \mathbf{y}_0, \mathbf{Y}_R = \mathbf{y}_{Rt}|\mathbf{X} = \mathbf{x}_q) = 0$ for all $t = 1, \cdots, K-1$ when $\mathbf{x}_q \neq \mathbf{x}_0$ and $\mathbf{x}_q \neq \mathbf{x}'_0$. Thus, $\Pr(\mathbf{Y} = \mathbf{y}_0, \mathbf{Y}_R = \mathbf{y}_{Rt}|\mathbf{X} = \mathbf{x}) = 0$ for all $t = 1, \cdots, K-1$ and all $\mathbf{x} \in \mathcal{X}$, i.e., $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{Rt}) = \emptyset$ for all $t = 1, \cdots, K-1$. So, $S^{\mathbf{X}}_{p(\mathbf{y},w^{*(n)}_R|\mathbf{x})}(\mathbf{y}_0, w^{*(n)}_{R0}) = S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0})$.

Thus, every non-empty $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y}_0, \mathbf{y}_{R0})$ is equal to $S^{\mathbf{X}}_{p(\mathbf{y},w^{*(n)}_R|\mathbf{x})}(\mathbf{y}_0, w^{*(n)}_{R0})$, where $w^{*(n)}_{R0} = c(\mathbf{y}_{R0})$ and hence equation (Equation 4.8) and equation (Equation 4.7) hold. $\square$

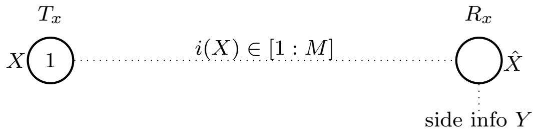
4.4.3 Connection with the Witsenhausen's source coding problem

Readers might be reminded of Witsenhausen's work[1] in (25) shown in Figure 18, where a point-to-point zero-error source coding problem with correlated side information available only at the receiver's end is studied. Witsenhausen's graph G_X is on vertex set $\mathcal{X}$ and x_1 is joined to x_2 by an edge when, for some y, $p(x_1, y)p(x_2, y) > 0$. The minimum signal alphabet size, for encoding a sequence of n independent pairs is shown to be the chromatic number $\chi(G_X^n)$ of the product of n copies of the graph G_X. Witsenhausen's graph G_X solves a standard compression problem for exact recovery of the source random variable X given side information Y. In contrast, the relaying function in Definition 29 is not a standard compression problem as the destination is not seeking to reconstruct y_R, but rather is a channel-aware form of compression: two relay observations y_{R1} and y_{R2} cannot be "combined", i.e. are connected in graph G_R only when the destination, based on its own observation and the color index, is incapable of resolving the ambiguity between x and x', i.e. for some y and $x \neq x'$, $p(y, y_{R1}|x)p(y, y_{R2}|x') > 0$.

In the problem of zero-error communication over a PRC, consider the extreme channel realization when $Y_R = X$ with probability 1. In this scenario, graph $G^X_{p(y,y_R|x)}$ becomes edge-free and with a large enough conference rate, one can achieve overall network message rate $\log \|\mathcal{X}\|$. This also implies that the channel input codebook has to be the full channel input alphabet $\mathcal{X}$, say$\underline{\mathcal{X}} = \mathcal{K} = \mathcal{X}$. In this specific case, finding r_z^* may be mapped to the source coding problem with receiver side information, which was solved by Witsenhausen (25) and

---

[1] In this subsection, we adopt the notation for confusability graphs from the Witsenhausen's work (25).



Given P_{XY}, what is $M_{\min} := \min M$, subject to $\Pr[X \neq \hat{X}] = 0$?

Figure 18. Transmit X to a receiver which has knowledge of Y (the side information) by means of a discrete signal taking as few values as possible.

in this case, coincides with the result presented here. It is noted that the construction of the G_R graph (on $\mathcal{Y}_R$) in Table I or Definition 29 gives the same graph as that constructed by Witsenhausen (on $\mathcal{X}$). This can be seen by realizing that Witsenhausen's graph G_X can be constructed by fully connecting $S_{X|Y}(y)$ for each y, which is exactly what the iterative algorithm does in Step 2.(a) in this scenario.

But we note that in general, finding the minimum conference rate r_z^* is different from Witsenhausen's source coding problem with receiver side information. This is because: 1) not all PRCs have $Y_R = X$; 2) when the SIMO bound is not the absolute maximum $\log \|\mathcal{X}\|$, only some subset of the channel input alphabet can be transmitted and there may be more than one choice of channel input codebook, as seen in the example in Tables I and II; and 3) in general, $G_R^{(n)}$ is not a n-fold strong product of graph $G_R^{(1)}$, i.e. $G_R^{(n)} \neq [G_R^{(1)}]^{\boxtimes n}$, and cannot be constructed via any standard graph product operations surveyed in (24).

4.5 Characterizing $r_z^{*(n)}$ by $ColourRate(n)$

In this section, we present an exact characterization of the minimum required conference rate $r_z^{*(n)}$ as defined in Definition (27), based on the Colour-and-Forward relaying $W_R^{*(n)}$ defined in Definition 29. Recall that when a conditional joint pmf $p(y, y_R|x)$ with support $\mathcal{X}$ and output $\mathcal{Y} \times \mathcal{Y}_R$ is restricted to input $\mathcal{K}$, we denote its *induced conditional pmf, support and output* by $p_{\mathcal{K}}(y, y_R|x)$, $\mathcal{K}$ and $\mathcal{Y}|_{\mathcal{K}} \times \mathcal{Y}_R|_{\mathcal{K}}$ respectively.

Theorem 31 (Colour-and-Forward relaying is optimal)**.** *Colour-and-Forward relaying in Definition 29 leads to the minimum required conference rate $r_z^{*(n)}$ for any fixed number of channel use n. That is, $r_z^{*(n)} = ColourRate(n)$, where $r_z^{*(n)}$ is specified in* (Equation 4.5) *and $ColourRate(n)$ is defined as*

$$ColourRate(n) := \min_{\mathcal{K} \text{ is a maximum independent set of graph } G^{X^n}_{p(y^n, y_R^n|x^n)}} \log \sqrt[n]{\chi(G_R^{(n)}|_{\mathcal{K}})}, \tag{4.10}$$

where $\chi(G_R^{(n)}|_{\mathcal{K}})$ is the chromatic number of graph $G_R^{(n)}|_{\mathcal{K}}$, constructed via the algorithm described in Definition 28 with restricted input / codebook $\mathcal{K}$.

Letting $n = 1$ in Theorem 31, we have Corollary 32. We will use this 1-shot scenario to illustrate the minimization involved in equations (Equation 4.10) and (Equation 4.11), in Remark 13. We also compare $ColourRate(1)$, in 1-shot scenario, to other trivial bounds, in Remark 14.

Corollary 32 (Colour-and-Forward relaying is optimal, $n = 1$)**.** *Colour-and-Forward relaying* $W_R^{*(1)}$ *in Definition 29 characterizes the exact value of the minimum required conference rate* $r_z^{*(1)}$*, specified in* (Equation 4.5)*. That is,* $r_z^{*(1)} = ColourRate(1)$ *and*

$$ColourRate(1) := \min_{\mathcal{K} \text{ is a maximum independent set of graph } G^X_{p(y,y_R|x)}} \log \chi(G_R|_{\mathcal{K}}) , \tag{4.11}$$

where $\chi(G_R|_{\mathcal{K}})$ *is the chromatic number of graph* $G_R|_{\mathcal{K}}$*, constructed via the algorithm described in Table I from the induced conditional joint pmf* $p_{\mathcal{K}}(y, y_R|x)$*.*

Remark 13. *We will provide an example of how to compute* $ColourRate(1)$ *in detail in Subsection (XX). But to give one a sense of the optimization involved, we provide a summary in this remark. We note that the minimization is over the different maximum independent sets of the graph* $G^X_{p(y,y_R|x)}$ *and that different maximum independent sets may yield different conference link rates. To illustrate this, consider the PRC described by the joint distribution* $p(y, y_R|x)$ *provided in Table I. Its confusability graph, and compression graphs* G_R *constructed by the Colour-and-Forward algorithm (for inputs in* $\mathcal{X}$ *or some maximum independent subsets* $\mathcal{K}_1$ *and* $\mathcal{K}_2$ *of the confusabilty graph* $G^X_{p(y,y_R|x)}$*) when* $n = 1$*, are shown in Table II. We note that in order to have the smallest number of colours for the conference link one must use* $\mathcal{K}_2$ *and not* $\mathcal{K}_1$*.*

Remark 14. *Note that the vertex set of graph* $G_R|_{\mathcal{K}}$ *is* $\mathcal{Y}_R|_{\mathcal{K}}$*, which is a subset of* $\mathcal{Y}_R$*, i.e.* $\mathcal{Y}_R|_{\mathcal{K}} \subseteq \mathcal{Y}_R$*. Thus,*

$$\chi(G_R|_{\mathcal{K}}) \overset{(a)}{\leq} \|\mathcal{Y}_R|_{\mathcal{K}}\| \overset{(b)}{\leq} \|\mathcal{Y}_R\| .$$

By Brooks' Theorem (26), the chromatic number of a graph is at most the maximum degree Δ *(the largest vertex degree), unless the graph is complete or an odd cycle. So* $\chi(G_R|_{\mathcal{K}})$ *is at most* Δ *and can be as low as* 1. *Inequality* (a) *can be strict. The equality in* (b) *is obtained only when the restriction of support from* $\mathcal{X}$ *to* $\mathcal{K}$ *does not prohibit any* $Y_R = y_R$ *from showing up. One extreme case is when graph* $G^X_{p(y,y_R|x)}$ *is edge free, then* $\mathcal{K}$ *equals to the whole vertex set and* $\mathcal{Y}_R|_{\mathcal{K}} = \mathcal{Y}_R$. *Please refer to the examples in the case study in Chapter 4.6.*

4.5.1 Achievability proof for Theorem 31

To show the achievability of Theorem 31 is to show $r_z^{*(n)} \leq ColourRate(n)$.

Proof for the achievability of Theorem 31. We establish the achievability by explicitly constructing an n-shot protocol $(n, \underline{\mathcal{X}}, h, g)$ that can achieve $SIMO(n)$ when $r_z \geq ColourRate(n)$, implying $r_z^{*(n)} \leq ColourRate(n)$.

Choose codebook $\underline{\mathcal{X}}$ to be the (or any) maximum independent set $\mathcal{K}$ of graph $G^{X^n}_{p(y^n,y^n_R|x^n)}$ that achieves the minimum in equation (Equation 4.10), i.e., $\log \sqrt[n]{\chi(G_R^{(n)}|_{\underline{\mathcal{X}}})} = ColourRate(n)$. Note that $\|\underline{\mathcal{X}}\|$ is equal to $\alpha(G^{X^n}_{p(y^n,y^n_R|x^n)})$, i.e., $\log \|\underline{\mathcal{X}}\| = SIMO(n)$.

Consider the restricted conditional joint pmf $p|_{\underline{\mathcal{X}}}(y^n, w_R^{*(n)}|x^n)$, from which we construct the Colour-and-Forward graph $G_R^{(n)}|_{\underline{\mathcal{X}}}$, according to Definition 28, and the Colour-and-Forward relaying $W_R^{*(n)}|_{\underline{\mathcal{X}}}$, according to Definition 29. Choose the relaying function $h(\mathbf{y}_R) := W_R^{*(n)}|_{\underline{\mathcal{X}}} = c(\mathbf{y}_R)$ as in Definition 29. Note that $\log \|\mathcal{W}_R^{*(n)}|_{\underline{\mathcal{X}}}\| = \log \sqrt[n]{\chi(G_R^{(n)}|_{\underline{\mathcal{X}}})} = ColourRate(n)$. So when $r_z \geq ColourRate(n)$, $W_R^{*(n)}|_{\underline{\mathcal{X}}}$ can be successfully transmitted to the destination terminal.

By Theorem 30, we have $G^{X^n}_{p|\underline{\mathcal{X}}(y^n,w_R^{*(n)}|x^n)} = G^{X^n}_{p|\underline{\mathcal{X}}(y^n,y_R^n|x^n)}$. Thus, the decoding function g specified in equation (Equation 4.12) is valid and completes the construction of the desired n-shot protocol $(n, \underline{\mathcal{X}}, h, g)$ that can achieve $SIMO(n)$ when $r_z \geq ColourRate(n)$.

$$g(y^n, w_R^{*(n)}) := S^{X^n}_{p|\underline{\mathcal{X}}(y^n,w_R^{*(n)}|x^n)}(y^n, w_R^{*(n)}) = \bigcup_{y_R \in c^{-1}(w_R^{*(n)})} S^{X^n}_{p|\underline{\mathcal{X}}(y^n,y_R^n|x^n)}(y^n, y_R^n)\,. \tag{4.12}$$

$\square$

4.5.2 Converse proof for Theorem 31

To prove optimality of Colour-and-Forward relaying $W_R^{*(n)}$ is to show $r_z^{*(n)} \geq ColourRate(n)$, which requires the following zero-error data-processing inequality.

Lemma 33 (Data-Processing Inequality)**.** *Given a conditional pmf* $p(y|x)$*, let* $Z = f(Y)$ *be any deterministic mapping* $f : \mathcal{Y} \to \mathcal{Z}$ *and denote* $p(z|x)$ *the induced conditional pmf from* $p(y|x)$*. Then the confusability graph* $G^X_{p(y|x)}$ *specified by* $p(y|x)$ *has no more edges than the confusability graph* $G_{X|Z}$ *specified by* $p(z|x)$*; i.e.,* $E(G_{X|Y}) \subseteq E(G_{X|Z})$*.*

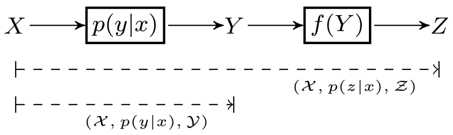


Figure 19. Data-processing cannot increase the zero-error capacity: the zero-error capacity of the induced channel $(\mathcal{X}, p(z|x), \mathcal{Z})$ is no larger than the original channel $(\mathcal{X}, p(y|x), \mathcal{Y})$.

Recall that the zero-error capacity of a point-to-point channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ is fully characterized by the confusability graph $G^X_{p(y|x)}$ and is directly related to the independence number of its n-fold strong product. Lemma 33 implies $\alpha(G^X_{p(y|x)}) \geq \alpha(G^X_{p(z|x)})$, because the more densely a graph is connected, the smaller its independence number becomes. Equivalently, as shown in Figure 19, the zero-error capacity of the induced channel $(\mathcal{X}, p(z|x), \mathcal{Z})$ is no larger than the original channel $(\mathcal{X}, p(y|x), \mathcal{Y})$. Lemma 33 and its validity follows directly from the definition of confusability graph and the nature of zero-error communication.

Now we are ready to present the proof for the converse part of Theorem 31.

Proof for the converse of Theorem 31. It suffices to prove $r_z^{*(n)} \geq ColourRate(n)$. Let $(n, \underline{\mathcal{X}}, h, g)$ denote any n-shot protocol that can achieve the SIMO upper-bound message rate $SIMO(n)$ without error, say $R_z^{(n)} = \frac{1}{n} \log \|\underline{\mathcal{X}}\| = \log \sqrt[n]{\alpha(G^X_{p(y^n, y^n_R|x^n)})}$. We will show that $\|\mathcal{W}_R^{(n)}\| \geq \|\mathcal{W}_R^{*(n)}\| = 2^{n \cdot ColourRate(n)}$ must hold for any such relaying function $h : \mathcal{Y}_R^n \to \mathcal{W}_R^{(n)}$ (for any n).

Because rate $\frac{1}{n} \log \|\underline{\mathcal{X}}\|$ can be achieved by the given n-shot protocol $(n, \underline{\mathcal{X}}, h, g)$, we know that the induced subgraph $G^{X^n}_{p(y^n, w_R^{(n)}|x^n)}(\underline{\mathcal{X}})$ [1] [2] must be edge-free. Recall that $W_R^{(n)}$ is a deterministic function of Y_R^n, so $(Y^n, W_R^{(n)})$ is a deterministic function of (Y^n, Y_R^n). According to the data-processing inequality in Lemma 33, we have $E(G^{X^n}_{p(y^n, y^n_R|x^n)}(\underline{\mathcal{X}})) \subseteq E(G^{X^n}_{p(y^n, w_R^{(n)}|x^n)}(\underline{\mathcal{X}})) =$

---

[1] Graph $G(A)$ is the induced subgraph of graph G, with vertex set $A \subseteq V(G)$ and edge set $(A \times A) \cap E(G)$.

[2] $G^{X^n}_{p(y^n, w_R^{(n)}|x^n)}(\underline{\mathcal{X}})$ is the same as $G^{X^n}_{p|\underline{\mathcal{X}}(y^n, w_R^{(n)}|x^n)}$.

$\emptyset$. Thus, we know that two induced subgraphs $G^{X^n}_{p(y^n,y^n_R|x^n)}(\underline{\mathcal{X}})$ and $G^{X^n}_{p(y^n,w^{(n)}_R|x^n)}(\underline{\mathcal{X}})$ must both be free of edges. Consider any triple $(\mathbf{X}=\mathbf{x},\mathbf{Y}=\mathbf{y},W^{(n)}_R=w^{(n)}_R)\in\underline{\mathcal{X}}\times\mathcal{Y}^n|_{\underline{\mathcal{X}}}\times\mathcal{W}^{(n)}_R$, we have

$$\begin{aligned}&\Pr[\mathbf{Y}=\mathbf{y},W^{(n)}_R=w^{(n)}_R|\mathbf{X}=\mathbf{x}]\\=&\sum_{\mathbf{y}_R:\,h(\mathbf{y}_R)=w^{(n)}_R}\Pr[\mathbf{Y}=\mathbf{y},\mathbf{Y}_R=\mathbf{y}_R|\mathbf{X}=\mathbf{x}]\end{aligned}$$

Thus,

$$\begin{aligned}&S^{\mathbf{X}}_{p(\mathbf{y},w^{(n)}_R|\mathbf{x})}(\mathbf{y},w^{(n)}_R)=\{\mathbf{x}\in\underline{\mathcal{X}}:p|_{\underline{\mathcal{X}}}(\mathbf{y},w^{(n)}_R|\mathbf{x})>0\}\\=&\{\mathbf{x}\in\underline{\mathcal{X}}:\sum_{\mathbf{y}_R:\,h(\mathbf{y}_R)=w^{(n)}_R}\Pr[\mathbf{Y}=\mathbf{y},\mathbf{Y}_R=\mathbf{y}_R|\mathbf{X}=\mathbf{x}]>0\}\\=&\bigcup_{\mathbf{y}_R:\,h(\mathbf{y}_R)=w^{(n)}_R}\{\mathbf{x}\in\underline{\mathcal{X}}:\Pr[\mathbf{Y}=\mathbf{y},\mathbf{Y}_R=\mathbf{y}_R|\mathbf{X}=\mathbf{x}]>0\}\\=&\bigcup_{\mathbf{y}_R:\,h(\mathbf{y}_R)=w^{(n)}_R}S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y},\mathbf{y}_R)\end{aligned}\tag{4.13}$$

$S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y},\mathbf{y}_R)$ has zero or one element because graph $G^{X^n}_{p(y^n,y^n_R|x^n)}(\underline{\mathcal{X}})$ has no edges. Similarly, since graph $G^{X^n}_{p(y^n,w^{(n)}_R|x^n)}(\underline{\mathcal{X}})$ is edge-free, $S^{\mathbf{X}}_{p(\mathbf{y},w^{(n)}_R|\mathbf{x})}(\mathbf{y},w^{(n)}_R)$ shall also at most have one element. So in equation (Equation 4.13), the sets to be unioned can have 0 or 1 element and all non-empty sets shall be the same, i.e., containing one same element. This means that for any fixed $\mathbf{Y}=\mathbf{y}$, any two different $\mathbf{y}_R$'s such that $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y},\mathbf{y}_{R1})$ and $S^{\mathbf{X}}_{p(\mathbf{y},\mathbf{y}_R|\mathbf{x})}(\mathbf{y},\mathbf{y}_{R2})$ (which are both either an empty set or a single-element set) have different elements, say $\mathbf{x}_1$ and $\mathbf{x}_2$, are prohibited to be mapped into the same color $W^{(n)}_R$. That is, requiring two $\mathbf{y}_R$'s to be differentiated (via the relaying function h) if for some $\mathbf{y}$, $\mathbf{x}_1\neq\mathbf{x}_2$, $\Pr(\mathbf{Y}=\mathbf{y},\mathbf{Y}_R=\mathbf{y}_{R1}|\mathbf{X}=\mathbf{x}_1)\cdot\Pr(\mathbf{Y}=\mathbf{y},\mathbf{Y}_R=\mathbf{y}_{R2}|\mathbf{X}=\mathbf{x}_2)>0$, is necessary. Equivalently, all edges in the compression graph $G_R^{(n)}$ constructed in the Colour-and-Forward algorithm are necessary; any other valid relay mapping $W_R^{(n)} = h(\mathbf{Y}_R)$ would result in equally or more strict edge constraints than Colour-and-Forward or graph $G_R^{(n)}$. Note that as more edges are added to a graph, its chromatic number cannot decrease. Therefore, for any valid relay mapping $W_R^{(n)}$, we have $\|\mathcal{W}_R\| \geq \|\mathcal{W}_R^{*(n)}\|$, implying $r_z^{*(n)} \geq ColourRate(n)$. $\square$

4.6 An example of computing $ColourRate(1)$

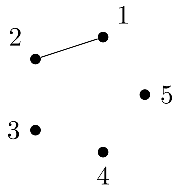
In this section, taking the example of the 1-shot case, we will demonstrate: (1) how to compute $ColourRate(1)$ in Corollary 32 and illustrate in details the process of minimization over all possible maximum independent sets; (2) how to interpret the information-lossless property of the Colour-and-Forward relaying $W_R^{*(1)}$ stated in Theorem 30. For succinctness, we drop the superscript $^{(1)}$ and use W_R^* to denote $W_R^{*(1)}$.

Throughout the section, we will be studying the PRC, whose broadcasting component is defined in Table IV, and try to compute the minimum required conference rate $r_z^{*(1)}$, a threshold below which the PRC will fail to achieve the SIMO upper bound message rate $SIMO(1)$. Recall that $r_z^{*(1)}$ is equal to $ColourRate(1)$ by Corollary 32. Table IV enumerates a conditional joint probability mass function: $p(y, y_R|x)$, where $\|\mathcal{X}\| = \|\mathcal{Y}\| = \|\mathcal{Y}_R\| = 5$. An entry at position (x, y, y_R) is denoted by "$*$" (the actual value does not matter), when its probability $p(y, y_R|x)$ is positive and by "0" when $p(y, y_R|x) = 0$.

4.6.1 Compute all possible codebooks

Recall that valid codebooks are the maximum independent sets of the confusability graph $G^X_{p(y,y_R|x)}$. So we first compute confusability graph $G^X_{p(y,y_R|x)}$ from the conditional joint pmf illustrated on Table IV, by enumerating the collection of conditional supports, i.e., $\{S^X_{p(y,y_R|x)}(y,y_R) : (y,y_R) \in \mathcal{Y} \times \mathcal{Y}_R\}$, and obtaining edge constraints by fully connecting the X symbols within each conditional support $S^X_{p(y,y_R|x)}(y,y_R)$. This process and the resulting confusability graph $G^X_{p(y,y_R|x)}$ are provided in detail in Figure 20.

(Y, Y_R)	$S^X_{p(y,y_R\|x)}(y,y_R)$	edges
$(1,3)$	$\{1\}$	$\emptyset$
$(1,4)$		
$(2,1)$		
$(3,3)$	$\{2\}$	$\emptyset$
$(3,1)$	$\{3\}$	$\emptyset$
$(4,3)$		
$(2,4)$	$\{4\}$	$\emptyset$
$(4,4)$		
$(5,3)$		
$(1,1)$	$\{5\}$	$\emptyset$
$(5,5)$		
$(2,2)$	$\{1,2\}$	$1-2$



confusability graphs: $G^X_{p(y,y_R|x)}$

Figure 20. Construct confusability graph $G^X_{p(y,y_R|x)}$.

As shown in Figure 20, graph $G^X_{p(y,y_R|x)}$ has two maximum independent sets: $\mathcal{K}_1 = \{1,3,4,5\}$ and $\mathcal{K}_2 = \{2,3,4,5\}$.

4.6.2 Compute Colour-and-Forward graph G_R

After obtaining all possible codebooks, one can consider the restricted conditional joint pmf under each chosen codebook, say $(\mathcal{K}_1, p_{\mathcal{K}_1}(y, y_R|x), \mathcal{Y}|_{\mathcal{K}_1} \times \mathcal{Y}_R|_{\mathcal{K}_1})$ and $(\mathcal{K}_2, p_{\mathcal{K}_2}(y, y_R|x), \mathcal{Y}|_{\mathcal{K}_2} \times \mathcal{Y}_R|_{\mathcal{K}_2})$, and compute the corresponding Colour-and-Forward graphs, say $G_R|_{\mathcal{K}_1}$ and $G_R|_{\mathcal{K}_2}$, according to Definition 29 or the construction algorithm shown in Table I. But before computing $G_R|_{\mathcal{K}_1}$ and $G_R|_{\mathcal{K}_2}$, we would like to compute $G_R|_{\mathcal{X}}$ based on $(\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R)$ first as a benchmark for comparison. Note that the construction of Colour-and-Forward graph G_R, in Definition 28, and Colour-and-Forward relaying W_R^*, in Definition 29, applies to any given broadcasting component, regardless of whether the conditional joint pmf is a restricted one or not, say, $(\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R)$, $(\mathcal{K}_1, p_{\mathcal{K}_1}(y, y_R|x), \mathcal{Y}|_{\mathcal{K}_1} \times \mathcal{Y}_R|_{\mathcal{K}_1})$ and $(\mathcal{K}_2, p_{\mathcal{K}_2}(y, y_R|x), \mathcal{Y}|_{\mathcal{K}_2} \times \mathcal{Y}_R|_{\mathcal{K}_2})$.

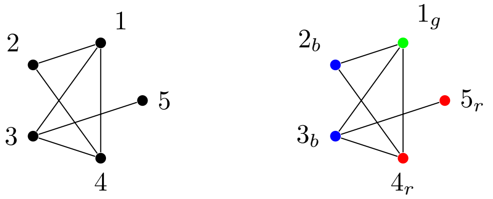
Figure 21(a) illustrates the iterative algorithm: for each $Y \in [1:5]$, construct a sequence of $B^{Y_R}_{p(y,y_R|x)}(x, y) \subseteq \mathcal{Y}_R = [1:5]$, where $x \in S^X_{p(y|x)}(y) = \{x : p(y|x) > 0\}$ and put an edge between every pair (y_R, y_R') where $y_R \in B^{Y_R}_{p(y,y_R|x)}(x, y)$ and $y_R' \in B^{Y_R}_{p(y,y_R|x)}(x', y)$. Superimposing these edges, we have the compression graph $G_R|_{\mathcal{X}}$ as shown in Figure 21(b).

In graph $G_R|_{\mathcal{X}}$ in Figure 21(b), different colours are used to denote one choice of minimum colouring function c. These colours specify the relay's mapping $W_R^* = c(Y_R)$.

As shown in Figure 22, two collections of conditional supports are equal, say, $\{S^X_{p(y,y_R|x)} : (y, y_R) \in \mathcal{Y} \times \mathcal{Y}_R\} = \{S^X_{p(y,w_R^*|x)} : (y, w_R^*) \in \mathcal{Y} \times \mathcal{W}_R^*\}$, and two corresponding confusability graphs are equal, say, $G^X_{p(y,y_R|x)} = G^X_{p(y,w_R^*|x)}$. That is, compressing Y_R into W_R^* is information

	$S^X_{p(y\|x)}(y)$	$B^{Y_R}_{p(y,y_R\|x)}(x,y)$	edges
$Y=1$	$X=1$ $X=5$	$\{3,4\}$ $\{1\}$	$1-3, 1-4$
$Y=2$	$X=1$ $X=2$ $X=4$	$\{1,2\}$ $\{2\}$ $\{4\}$	$1-2, 1-4, 2-4$
$Y=3$	$X=2$ $X=3$	$\{3\}$ $\{1\}$	$1-3$
$Y=4$	$X=3$ $X=4$	$\{3\}$ $\{4\}$	$3-4$
$Y=5$	$X=4$ $X=5$	$\{3\}$ $\{5\}$	$3-5$

(a) The iterative algorithm for constructing compression graph G_R.



Compression graph G_R One minimum colouring c

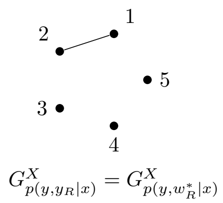
(b) The compression graph G_R and one choice of minimum colouring function c.

Figure 21. Constructing compression graph G_R from $p(y, y_R|x)$ in Table IV. Note that the least number of colours required is: $\chi(G_R) = 3$.

lossless in the sense that together with Y, W_R^* provides the same ability to distinguish different $X = x$'s as Y_R, as stated in Theorem 30.

(Y,Y_R)	$S^X_{p(y,y_R\mid x)}(y,y_R)$	edges
$(1,3)$ $(1,4)$ $(2,1)$	$\{1\}$	$\emptyset$
$(3,3)$	$\{2\}$	$\emptyset$
$(3,1)$ $(4,3)$	$\{3\}$	$\emptyset$
$(2,4)$ $(4,4)$ $(5,3)$	$\{4\}$	$\emptyset$
$(1,1)$ $(5,5)$	$\{5\}$	$\emptyset$
$(2,2)$	$\{1,2\}$	$1-2$

(Y,W_R^*)	$S^X_{p(y,w_R^*\mid x)}(y,w_R^*)$	edges
$(1,b)$ $(1,r)$ $(2,g)$	$\{1\}$	$\emptyset$
$(3,b)$	$\{2\}$	$\emptyset$
$(3,g)$ $(4,b)$	$\{3\}$	$\emptyset$
$(2,r)$ $(4,r)$ $(5,b)$	$\{4\}$	$\emptyset$
$(1,g)$ $(5,r)$	$\{5\}$	$\emptyset$
$(2,b)$	$\{1,2\}$	$1-2$



(a) Conditional supports $S^X_{p(y,y_R|x)}(y,y_R)$.

(b) Conditional supports $S^X_{p(y,w_R^*|x)}(y,w_R^*)$.

(c) Confusability graphs.

Figure 22. Compressing Y_R into W_R^* according to the minimum colouring function c on graph G_R in 21(b), is information lossless in the sense that together with Y, W_R^* provides as much information of about X as Y_R. Note that the independence number is: $\alpha(G^X_{p(y,y_R|x)}) = 4$.

4.6.3 Compute Colour-and-Forward graph $G_R|_{\mathcal{K}_1}$

When codebook $\mathcal{K}_1 = \{1,3,4,5\}$ is chosen, the *induced* broadcasting component is $(\mathcal{K}_1, p_{\mathcal{K}_1}(y,y_R|x), \mathcal{Y}|_{\mathcal{K}_1} \times \mathcal{Y}_R|_{\mathcal{K}_1})$.

- The iterative algorithm: in 23(a).[1]
- The compression graph $G_R|_{\mathcal{K}_1}$: in 23(b).[2]

---

[1] We retain the cross-out items in Tables 23(a), 25(a) to serve a comparison with the construction algorithm in Figure 21(a).

[2] We retain the dotted edges in 23(b), 25(b) to serve as a comparison with the compression graph in 21(b).

- One choice of minimum colouring on compression graph $G_R|_{\mathcal{K}_1}$: in 23(b) with chromatic number $\chi(G_R|_{\mathcal{K}_1}) = 3$.
- The information-lossless property stated in Theorem 30 is verified in Figure 24.

4.6.4 Compute Colour-and-Forward graph $G_R|_{\mathcal{K}_2}$

When codebook $\mathcal{K}_2 = \{2, 3, 4, 5\}$ is chosen and the *induced* broadcasting component is $(\mathcal{K}_2, p_{\mathcal{K}_2}(y, y_R|x), \mathcal{Y}|_{\mathcal{K}_2} \times \mathcal{Y}_R|_{\mathcal{K}_2})$.

- The iterative algorithm: in 25(a).[1]
- The compression graph $G_R|_{\mathcal{K}_2}$: in 25(b).[2]
- One choice of minimum colouring on compression graph $G_R|_{\mathcal{K}_2}$: in 25(b) with chromatic number $\chi(G_R|_{\mathcal{K}_2}) = 2$.
- The information-lossless property stated in Theorem 30 is verified in Figure 26.

4.6.5 The value of $ColourRate(1)$

So $ColourRate(1) = \log \min\{\chi(G_R|_{\mathcal{K}_1}), \chi(G_R|_{\mathcal{K}_2})\} = \log \min\{3, 2\} = 1$ bit/channel use. This means that when the conference rate is smaller than 1 bit/channel use, there exists no communication scheme that can achieve the message rate $SIMO(1) = \log 4 = 2$ bits/channel use. Note that in order to have the smallest number of colours for the conference link one must use $\mathcal{K}_2$ and not $\mathcal{K}_1$ as the codebook.

---

[1] See footnote 1.

[2] See footnote 2.

4.7 Properties of $r_z^{*(n)}$ and its connection with r_z^*

We note that in general, the Colour-and-Forward graph or the compression graph $G_R^{(n)}$, which determines the value of $r_z^{*(n)}$, cannot be constructed via any standard graph product operations surveyed in (24). The Colour-and-Forward compression graphs $G_R^{(n)}$s' behavior and their chromatic numbers as a function of n are not obvious, and are interesting open questions that requires more work. We show in Subsection 4.7.2 a class of PRC channels where $r_z^{*(n)}$ is known as an exemplary exploration of charactering $G_R^{(n)}$ and $r_z^{*(n)}$.

In Subsection 4.7.1, we discuss how to go beyond the limitation of n-shot channel use and try to infer r_z^* based on $r_z^{*(n)}$'s. As demonstrated in Figure 27, the PRC capacity $C_z^{(n)}(r_z)$ depends on two parameters: the number of channel uses n and the conference link capacity r_z. For any fixed number of channel use n, PRC capacity $C_z^{(n)}(r_z)$ is non-decreasing w.r.t. r_z and when $r_z \geq r_z^{*(n)}$, $C_z^{(n)}(r_z)$ stays at $SIMO(n)$. For the sequence $SIMO(n)$, we have the asymptomatic conclusion that $\sup_n SIMO(n) = \lim_{n\to\infty} SIMO(n)$. But for finite n's, how $SIMO(n)$, i.e. the sequence of independence numbers of the strong product graphs $\log \sqrt[n]{\alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})}$, behaves is a long-standing open question (21; 22).

Figure 27 is also very helpful for understanding Lemma 34 and the corresponding proofs.

4.7.1 Bounds on r_z^*

In this section, we discuss the relationship of $r_z^{*(n)}$ and r_z^* in general and provide the following lower and upper bounds of r_z^*.

Firstly, similar to what was done (17), based on the cut-set bounds on PRC capacity, as presented in Proposition 23, we can have a lower bound for r_z^*:

$$\log \frac{\lim_{n\to\infty} \sqrt[n]{\alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})}}{\lim_{n\to\infty} \sqrt[n]{\alpha([G^X_{p(y|x)}]^{\boxtimes n})}} \leq r_z^* \,. \tag{4.14}$$

Next we show several upper bounds on r_z^* based on the value of $r_z^{*(n)}$'s, which can be fully characterized by the Colour-and-Forward relaying algorithm.

Lemma 34 (Upper bounds on r_z^*). *r_z^* can be lower and upper bounded by:*

(1) r_z^* is no greater than the supremum of $r_z^{*(n)}$. That is, $r_z^* \leq U_1$, where $U_1 := \sup_n r_z^{*(n)}$.

(2) Let $U_2 := \sup_{n\in J} r_z^{*(n)}$, where $J \subseteq \{1, 2, \cdots\}$. As long as J has infinitely many elements, U_2 is an upper bound on r_z^*, i.e., $r_z^* \leq U_2$.

(3) When $\{SIMO(n)\}_{n=1}^{\infty}$ is a closed set, $r_z^* = U_3$, where $U_3 := \min_{t\colon SIMO(t)=SIMO} r_z^{*(t)}$.

It is clear that $U_2 \leq U_1$ and U_2 is a more strict upper bound on r_z^* than U_1. One of the reasons why we can restrict the set on which the superum is defined, from natural numbers $\{1, 2, \cdots\}$ to its any subset J that has infinite many elements, is due to the fact that the sequence $\{SIMO(n)\}_{n=1}^{\infty}$ has a limit, which equals to its supremum.

Note that to infer the behavior of r_z^* from the upper bounds on $r_z^{*(n)}$ requires knowing how $SIMO$ depends on $SIMO(n)$. By the super-multiplicity of the independence number sequence of the strong product graphs and Fekete's lemma, we know that sequence $SIMO(n)$ converges to its supremum. In general, the maximum need not exist.

When $\{SIMO(n)\}_{n=1}^{\infty}$ is an open set, meaning $SIMO = \lim_n SIMO(n) = \sup_n SIMO(n)$ cannot be obtained by any $SIMO(n)$, we conservatively conjecture that

Conjecture 35. $r_z^* = \lim_{n\to\infty} r_z^{*(n)}$.

Remark 15. *We remark that* r_z^* *can be strictly smaller than* U_1. *If* $r_z < U_1$, *it means that* $r_z < r_z^{*(n)}$ *holds for at least one* n. *Otherwise,* r_z *will be an upper bound on* $r_z^{*(n)}$ *and the assumption* $r_z < U_1$ *will contradict with* U_1 *being the least upper bound (the supremum) of* $r_z^{*(n)}$. *We further note that, for any* n *that satisfies* $r_z < r_z^{*(n)}$, *we have* $C_z^{(n)}(r_z) < C_z^{(n)}(r_z^{*(n)}) = SIMO(n)$ *by the definition of* $r_z^{*(n)}$. *But this does not necessarily harm the validity of* $\sup_n C_z^{(n)}(r_z) = \sup_n SIMO(n)$. *In the case where* $\sup_n C_z^{(n)}(r_z) = \sup_n SIMO(n)$ *indeed holds,* r_z^* *will be smaller or equal to* r_z *and thus strictly smaller than* U_1.

Remark 16. r_z^* *can be smaller than the infimum of* $r_z^{*(n)}$. *That is, there may exist PRC channels where* $r_z^* < L_1$, *where* $L_1 := \inf_n r_z^{*(n)}$. *Consider* $C_z^{(n)}(r_z)$ *when* $r_z < L_1$. *Clearly,* $r_z < r_z^{*(n)}$ *holds for any* n, *because* L_1 *equals to the infimum of* $r_z^{*(n)}$. *So for all* n, $C_z^{(n)}(r_z) < SIMO(n)$ *is true. Thus,* $\sup_n C_z^{(n)}(r_z) \leq \sup_n SIMO(n)$ *holds, equivalently,* $\sup_n C_z^{(n)}(r_z) \leq SIMO$ *holds.*

- *When* $\sup_n C_z^{(n)}(r_z) < SIMO$, *we can conclude that* $r_z^* > r_z$.

  Note the assumption is $L_1 > r_z$, *from which one cannot cannot tell if* r_z^* *is bigger or smaller than* L_1.

- *But when* $\sup_n C_z^{(n)}(r_z) = SIMO$ *happens, we can conclude that* $r_z^* \leq r_z$.

  Note the assumption is $r_z < L_1$, *which implies* $r_z^* < L_1$.

So we have shown that $r_z^* < L_1$ *is possible.*

Next we present the proofs for the three statements in Lemma 34.

Proof of Lemma 34 - (1). Consider $C_z^{(n)}(r_z)$ when $r_z \geq U_1$. First $r_z \geq U_1 \geq r_z^{*(n)}$ holds for any n, because U_1 is the supremum of all $r_z^{*(n)}$'s. It is also true that $C_z^{(n)}(r_z) \geq C_z^{(n)}(r_z^{*(n)})$, because $C_z^{(n)}(r_z)$ is non-decreasing with respect to r_z. Note that $C_z^{(n)}(r_z^{*(n)})$ by definition achieves the SIMO upper bound, i.e. $C_z^{(n)}(r_z^{*(n)}) = SIMO(n)$. So $C_z^{(n)}(r_z) \geq SIMO(n)$ holds for any n. Thus we have $\sup\limits_n C_z^{(n)}(r_z) \geq \sup\limits_n SIMO(n)$, equivalently, $C_z(r_z) \geq SIMO$. This implies that $r_z^* \leq U_1$. $\square$

Proof of Lemma 34 - (2). It suffices to show when $r_z \geq U_2$, $C_z(r_z) = SIMO$ holds. Recall that $C_z(r_z)$ is defined as $\sup\limits_n C_z^{(n)}(r_z)$, and $SIMO$ is defined to be $\sup\limits_n SIMO(n)$, which is equal to $\lim\limits_{n\to\infty} SIMO(n)$.

First, $r_z \geq U_2$ implies $r_z \geq r_z^{*(n)}$ for any $n \in J$. Because $C_z^{(n)}(r_z)$ is non-decreasing, we have $C_z^{(n)}(r_z) \geq C_z^{(n)}(r_z^{*(n)})$, equivalently $C_z^{(n)}(r_z) \geq SIMO(n)$, for any $n \in J$. So, $\sup\limits_{n\in J} C_z^{(n)}(r_z) \geq \sup\limits_{n\in J} SIMO(n)$ holds.

Secondly, because $\sup\limits_n C_z^{(n)}(r_z) \geq \sup\limits_{n\in J} C_z^{(n)}(r_z)$, we have $\sup\limits_n C_z^{(n)}(r_z) \geq \sup\limits_{n\in J} SIMO(n)$, equivalently, $C_z(r_z) \geq \sup\limits_{n\in J} SIMO(n)$. If $\sup\limits_{n\in J} SIMO(n) = SIMO$ holds, then $C_z(r_z) \geq SIMO$ holds, which establishes $C_z(r_z) = SIMO$ because $C_z(r_z) \leq SIMO$ is always true.

We next show that $\sup\limits_{n\in J} SIMO(n) = SIMO$ holds by showing $\sup\limits_{n\in J} SIMO(n) = \lim\limits_{n\to\infty} SIMO(n)$. Assume $\sup\limits_{n\in J} SIMO(n) < \lim\limits_{n\to\infty} SIMO(n)$. Let $\sup\limits_{n\in J} SIMO(n) = \lim\limits_{n\to\infty} SIMO(n) - \delta$, where δ is some given positive real number that can be arbitrarily small. By the definition of the existence of a limit, we know that there exists some big number N such that when $n > N$, $SIMO(n) \geq \lim\limits_{n\to\infty} SIMO(n) - \frac{\delta}{2} > \lim\limits_{n\to\infty} SIMO(n) - \delta$. This means that there are at most N

items of sequence $SIMO(n)$, which are smaller than or equal to $\lim\limits_{n\in J} SIMO(n) - \delta$. Because J has infinitely many elements and thus has more than N elements, J has to contain n's that are greater than N. This implies that, set $\{SIMO(n), n \in J\}$ contains $SIMO(n)$'s that are greater than $\lim\limits_{n\to\infty} SIMO(n) - \delta$. So, $\sup\limits_{n\in J} SIMO(n) > \lim\limits_{n\to\infty} SIMO(n) - \delta$ holds and contradicts with the assumption $\sup\limits_{n\in J} SIMO(n) < \lim\limits_{n\to\infty} SIMO(n) - \delta$. Thus, $\sup\limits_{n\in J} SIMO(n) \geq \lim\limits_{n\to\infty} SIMO(n)$. Recall that $\lim\limits_{n\to\infty} SIMO(n) = \sup\limits_{n\to\infty} SIMO(n)$, which is greater or equal to $\sup\limits_{n\in J} SIMO(n)$. So, we have $\sup\limits_{n\in J} SIMO(n) = \lim\limits_{n\to\infty} SIMO(n)$.

☐

Proof of Lemma 34 - (3). When $\{SIMO(n)\}_{n=1}^{\infty}$ is a closed set, its supremum matches its maximum. That is, $\max\limits_{n} C_z^{(n)}(r_z) = \sup\limits_{n} C_z^{(n)}(r_z) = SIMO$. Recall that $r_z^* = \inf\{r_z : C_z(r_z) = SIMO\}$. Clearly, for t that satisfies $SIMO(t) = SIMO$, $r_z^{*(t)}$ is an upper bound on r_z^*. So U_3 is an upper bound on r_z^*. U_3 is tight, because of the optimality of the Colour-and-Forward relaying, i.e., $r_z^{*(n)}$ is the minimum required conference link that can enable the whole network to achieve message rate $SIMO(n)$. ☐

4.7.2 A class of PRC channels where r_z^* is known

Lemma 36. *For PRC channels where $p(y, y_R|x)$ satisfies (a) $p(y, y_R|x) = p(y|x) \cdot p(y_R|x)$, and (b) $p(y|x) > 0$ for all $y \in \mathcal{Y}$ and $x \in \mathcal{X}$,*

1.

$$r_z^* = \lim_{n\to\infty} SIMO(n) \tag{4.15}$$

2. $r_z^{*(n)} = SIMO(n)$*, which implies*

$$\min_{\mathcal{K}\in\tilde{\mathcal{K}}} \mathcal{X}(G_R^{(n)}|_{\mathcal{K}}) = \alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n}) \tag{4.16}$$

Proof of Lemma 36. We first prove statement 1) and statement 2) will follow immediately after 1) has been established.

Recall that we have the following lower bound on r_z^* by inequality (Equation 4.14):

$$\log \frac{\lim\limits_{n\to\infty} \sqrt[n]{\alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})}}{\lim\limits_{n\to\infty} \sqrt[n]{\alpha([G^X_{p(y|x)}]^{\boxtimes n})}} \leq r_z^* \,.$$

Condition (b) implies that $G^X_{p(y|x)}$ is fully-connected and $\alpha([G^X_{p(y|x)}]^{\boxtimes n}) = 1$ for any n. Thus,

$$\lim_{n\to\infty} \log \sqrt[n]{\alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})} \leq r_z^* \,. \tag{4.17}$$

Statement 1) follows by combining (Equation 4.17) with (Equation 4.18) and (Equation 4.19). So we next just need to show (Equation 4.18) and (Equation 4.19).

$$r_z^{*(n)} \leq \lim_{n\to\infty} \log \sqrt[n]{\alpha([G^X_{p(y_R|x)}]^{\boxtimes n})} \tag{4.18}$$

and

$$\lim_{n\to\infty} \log \sqrt[n]{\alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})} = \lim_{n\to\infty} \log \sqrt[n]{\alpha([G^X_{p(y_R|x)}]^{\boxtimes n})} \tag{4.19}$$

Equality (Equation 4.19) holds because $S^X_{p(y,y_R|x)}(y, y_R) = \{x : p(y, y_R|x) > 0\} = \{x : p(y|x) > 0, p(y_R|x) > 0\} = \{x : p(y_R|x) > 0\} = S^X_{p(y_R|x)}(y_R)$, for any $y \in \mathcal{Y}$ and $y_R \in \mathcal{Y}_R$, implying $G^X_{p(y,y_R|x)} = G^X_{p(y_R|x)}$.

Inequality (Equation 4.18) can be shown by considering the following communication scheme where

- The destination terminal ignore its own observation and only utilize what it receives from relay terminal
- The relay terminal adopts the Decode-and-Forward (DF) relaying strategy, which is constrained by the source-to-relay capacity $\lim_{n\to\infty} \log \sqrt[n]{\alpha([G^X_{p(y_R|x)}]^{\boxtimes n})}$ and the relay-to-destination conference link capacity r_z.

So the maximal achievable rate R_z for this communication scheme is $\min\{\lim_{n\to\infty} \log \sqrt[n]{\alpha([G^X_{p(y_R|x)}]^{\boxtimes n})}, r_z\}$. Noting equality (Equation 4.19), it can be checked that when $r_z \geq \lim_{n\to\infty} \log \sqrt[n]{\alpha([G^X_{p(y_R|x)}]^{\boxtimes n})}$, the SIMO bound message rate $\lim_{n\to\infty} \log \sqrt[n]{\alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})}$ can be achieved. By definition of r_z^*, this implies that inequality (Equation 4.18) holds and completes the proof for statement 1).

The above argument for establishing statement 1) is valid, when only n channel uses are allowed. Thus, we have $r_z^{*(n)} = SIMO(n)$, i.e. $\min_{\mathcal{K}\in\tilde{\mathcal{K}}} \mathcal{X}(G_R^{(n)}|_{\mathcal{K}}) = \alpha([G^X_{p(y,y_R|x)}]^{\boxtimes n})$.

□

Remark 17. *Equality* (Equation 4.16) *relates independence and chromatic numbers of two different graphs, which are both derived from the same underlying channel* $p(y, y_R|x)$*. This equality might have meaningful implication or interpretation in graph theory as well as understanding* *the intrinsic channel structure. It might be another example of bridging the closely intertwined research fields like combinatorics, graph theory and information theory. We thus propose the following open question:*

Open question: How to interpret the equality in (Equation 4.16)***?***

4.8 The perfect PRCs and examples

4.8.1 Zero-error capacity of a special class of primitive relay channels

Applying Theorem 31 using the Colour-and-Forward relay strategy W_R^* defined in Definition 29, We may obtain the zero-error capacity of a special class of primitive relay channels. We term these *perfect primitive relay channels* as (1) like in the point-to-point channel, we can characterize the zero-error capacity exactly, and not because any of the associated graphs are *perfect graphs* necessarily; (2) the zero-error capacity of such PRCs is the maximal possible rate – the logarithm of the channel input alphabet size $\|\mathcal{X}\|$.

Definition 37. *A PRC channel* $((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R), r_z)$ *is perfect if*

1. $G^X_{p(y,y_R|x)}$ *is edge free;*
2. $r_z \geq U_3$*, where* U_3 *is defined in Lemma 34.*

Theorem 38. *The zero-error capacity of the perfect primitive relay channel satisfying conditions in Definition 37, is*

$$C_{z,perfect} = \log ||\mathcal{X}||.$$

Proof. The converse is trivial: the zero-error capacity is always upper bounded by $\log \|\mathcal{X}\|$. The achievability follows by default:

- graph $G^X_{p(y,y_R|x)}$ being edge-free implies that block coding brings no gain than a single-shot coding scheme. That is, $SIMO(n) = \log \alpha(G^X_{p(y,y_R|x)}) = \log \|\mathcal{X}\|$ for any n. So U_3 in Lemma 34 is well-defined.
- $SIMO = \log \|\mathcal{X}\|$ is achievable because $r_z \geq r_z^*$ is guaranteed by $r_z \geq U_3$.

To be explicit, the n-shot protocol $(n,\, \underline{\mathcal{X}} = \mathcal{X},\, h, g)$ achieves zero error when $r_z \geq U_3$, with the codebook $\underline{\mathcal{X}}$ being the whole channel input alphabet as desired, the relaying $h(y_R) := c(y_R^n)$ as in Definition 29 and the decoding function $g(y^n, w_R^{*(n)})$ as in the proof of Theorem 31:

$$g(y^n, w_R^{*(n)}) := S^{X^n}_{p|\underline{\mathcal{X}}(y^n, w_R^{*(n)}|x^n)}(y^n, w_R^{*(n)}) = \bigcup_{y_R \,\in\, c^{-1}(w_R^{*(n)})} S^{X^n}_{p|\underline{\mathcal{X}}(y^n, y_R^n|x^n)}(y^n, y_R^n)$$

□

Note that while the overall message rate remains as a constant, we still adopt n-shot protocols or block-coding schemes; we do so to minimize the required conference rate at the relay-to-destination conference link. An interesting open question is to find an example where $SIMO(1) = SIMO(2)$ and $r_z^{*(1)} > r_z^{*(2)}$.

4.8.2 More examples

We provide in the following another three conventional examples to further illustrate the intuition and potential benefit of relaying to provide “what the destination needs”. For the sake of simplifying the description for a conditional joint pmf, we let $p(y, y_R|x) = p(y_R|x)p(y|x)$ in the these three examples. An edge in a bipartite graph between X and Y (or X and Y_R) indicates $p(y|x) > 0$ (or $p(y_R|x) > 0$).

4.8.2.1 The pentagon problem

We now consider a channel where the direct link between the source and destination consists of Shannon's "pentagon problem", which was notoriously difficult to solve. If the relay link is such that the corresponding channel forms a perfect PRC (this relay link described by $p(y_R|x)$ is not unique) an example of which is shown in Figure 28, we have $\alpha(G^X_{p(y,y_R|x)}) = 5$ and rate $\log 5$ can be achieved, when $r_z \geq \log 3$ (in a 1-shot scheme). Note that smaller values of r_z might still be able to guarantee the maximal rate $\log \|\mathcal{X}\| = \log 5$ when multiple channel uses are allowed, but this is left open.

We compare the rate achieved by our strategy with that achieved by a "Decode-and-Forward" (DF) relaying strategy. In a DF strategy, the relay would like to decode every codeword $w \in \underline{\mathcal{X}}$, in which case the message rate is constrained by $R_z \leq \log \alpha(G^X_{p(y_R|x)})$. In this example, $\alpha(G^X_{p(y_R|x)}) = 3$. Thus, $R_z \leq \log 3$ is a hard constraint on the message rates that can be achieved by Decode-and-Forward, which is clearly inferior to that achieved by our scheme. This scheme might be seen as a "channel-aware" (depends on the conditional $p(y, y_R|x)$) compression of Y_R, and thus might be seen as a smart way of implementing Compress-and-Forward.

4.8.2.2 An example where no compression is possible

We now provide an example in Figure 29 to show that there exist channels for which no information lossless compression is possible at the relay and the relay has to forward everything that it has observed, i.e, $r^*_z = \log \|\mathcal{Y}_R\|$. The relaying scheme W^*_R captures this phenomenon by requiring 8 different colours for 8 y_R's, as shown in Figure 29.

4.8.2.3 An example where much compression is possible

Finally, in Figure 30 We show an example of a channel where Y_R may be highly compressed without losing the *needed* information about X – i.e. we do *not* need to reconstruct Y_R at the destination, but only need to use the conferencing link to resolve any remaining ambiguity from the direct link. Here, one may verify that by sending only one of the two colours over the conferencing link, that a capacity of $\log 8$ may be achieved when $r_z \geq \log 2$.

4.9 Conclusion

In this chapter, the problem of communicating over a primitive relay channel without error is for the first time proposed, with the goal of exploring and fulfilling the intuition that the central role of a relay is to *only* deliver *"what the destination needs"*. Next, a novel relaying scheme termed "Colour-and-Foward" is proposed and is shown to be the most efficient way of compressing signals at the relay terminal, for any fixed number of channel uses, when enabling an effectively full cooperation between the relay and the destination terminals, i.e. achieving the single-input multi-output (SIMO) upper bound, is required. This Colour-and-Forward relaying is designed by an explicit exploit of the channel structure and directly embodies the intuition of having relay transmit "only what the destination needs". We also provide various non-trivial bounds on the asymptotic case, say r_z^* – the minimum required conference rate such that the given PRC channel can achieve the SIMO upper bound message rate. But the general relationship of $r_z^{*(n)}$ and r_z^* depends on the behavior of SIMO bounds, say $SIMO(n)$, which is the sequence of independence numbers of the graph products of a graph specified by the broadcasting component of the PRC channel in discussion, and is not clear. We believe understanding the behavior of the Colour-and-Forward graphs $G_R^{(n)}$ plays a central role in deciding the limit of how much and how efficiently the relay terminal can contribute to the overall communication. Probably, a characterization of the Colour-and-Forward graphs $G_R^{(n)}$ will help in identifying certain new categories of PRC channels whose capacity characterize may be more computable/tractable. Also, graph theory and combinatorics could potentially benefit from the connection among the graphs in Lemma 36, introduced by the study of zero-error communication over a PRC and the Colour-and-Forward relaying algorithm. In next chapter, we will show how to apply the Colour-and-Forward relaying scheme to the small-error communication problem over a PRC channel.

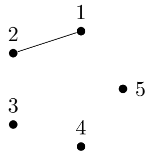
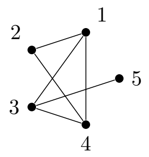
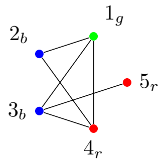
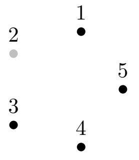
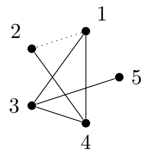
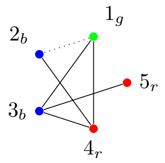
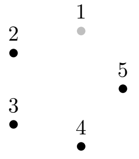
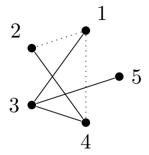
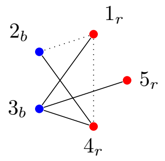
confusability graph	compression graph	colored compression graph
 $\mathcal{X} = [1:5]$	 Compression graph $G_R\|_{\mathcal{X}}$	 $\chi(G_R\|_{\mathcal{X}}) = 3$
 $\mathcal{K}_1 = [1, 3:5]$	 Compression graph $G_R\|_{\mathcal{K}_1}$	 $\chi(G_R\|_{\mathcal{K}_1}) = 3$
 $\mathcal{K}_2 = [2, 3:5]$	 Compression graph $G_R\|_{\mathcal{K}_2}$	 $\chi(G_R\|_{\mathcal{K}_2}) = 2$

TABLE III

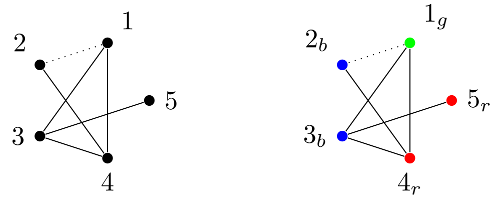
AN EXAMPLE TO SHOW THE IMPACT OF THE CHOICE OF INDEPENDENT SETS IN THEOREM 31. THE CONDITIONAL JOINT PMF $P(Y, Y_R|X)$ IN DISCUSSION IS SHOWN IN IV.

$s(p(y,y_R\|x))$		Y_R					Y_R					Y_R					Y_R					Y_R				
		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Y	1	0	0	$*$	$*$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	$*$	0	0	0	0
	2	$*$	$*$	0	0	0	0	$*$	0	0	0	0	0	0	0	0	0	0	0	$*$	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	$*$	0	0	$*$	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	$*$	0	0	0	0	0	$*$	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	$*$	0	0	0	0	0	0	$*$
		$X=1$					$X=2$					$X=3$					$X=4$					$X=5$				

TABLE IV. Conditional joint probability mass function: $p(y, y_R|x)$, where $\|\mathcal{X}\| = \|\mathcal{Y}\| = \|\mathcal{Y}_R\| = 5$. Note that $s(p(y, y_R|x))$ equals to $*$ when $p(y, y_R|x) > 0$ (actual value is unimportant) and 0, otherwise.

	$S^X_{p\|_{\mathcal{K}_1}(y\|x)}(y)$	$B^{Y_R}_{p\|_{\mathcal{K}_1}(y,y_R\|x)}(x,y)$	edges
$Y=1$	$X=1$ $X=5$	$\{3,4\}$ $\{1\}$	$1-3, 1-4$
$Y=2$	$X=1$ ~~$X=2$~~ $X=4$	$\{1,2\}$ ~~$\{2\}$~~ $\{4\}$	~~$1-2, 1-4, 2-4$~~ $1-4, 2-4$
$Y=3$	~~$X=2$~~ $X=3$	~~$\{3\}$~~ $\{1\}$	~~$1-3$~~ $\emptyset$
$Y=4$	$X=3$ $X=4$	$\{3\}$ $\{4\}$	$3-4$
$Y=5$	$X=4$ $X=5$	$\{3\}$ $\{5\}$	$3-5$

(a) The iterative algorithm for constructing compression graph $G_R|_{\mathcal{K}_1}$.



Compression graph $G_R|_{\mathcal{K}_1}$ One minimum colouring c

(b) The compression graph $G_R|_{\mathcal{K}_1}$ and one choice of minimum colouring function c.

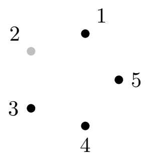
Figure 23. Constructing compression graph $G_R|_{\mathcal{K}_1}$ from induced conditional joint pmf $(\mathcal{K}_1, p_{\mathcal{K}_1}(y, y_R|x), \mathcal{Y}|_{\mathcal{K}_1} \times \mathcal{Y}_R|_{\mathcal{K}_1})$. Note that the least number of colours required is: $\chi(G_R|_{\mathcal{K}_1}) = 3$.

(Y, Y_R)	$S^X_{p(y,y_R\|x)}(y, y_R)$	edges
$(1,3)$ $(1,4)$ $(2,1)$	$\{1\}$	$\emptyset$
~~$(3,3)$~~	~~$\{2\}$~~	~~$\emptyset$~~
$(3,1)$ $(4,3)$	$\{3\}$	$\emptyset$
$(2,4)$ $(4,4)$ $(5,3)$	$\{4\}$	$\emptyset$
$(1,1)$ $(5,5)$	$\{5\}$	$\emptyset$
~~$(2,2)$~~	~~$\{1,2\}$~~ $\{1\}$	~~$1 — 2$~~ $\emptyset$

(a) Conditional supports $S^X_{p_{\mathcal{K}_1}(y,y_R|x)}(y, y_R)$.

(Y, W^*_R)	$S^X_{p(y,w^*_R\|x)}(y, w^*_R)$	edges
$(1,b)$ $(1,r)$ $(2,g)$	$\{1\}$	$\emptyset$
~~$(3,3)$~~	~~$\{2\}$~~	~~$\emptyset$~~
$(3,g)$ $(4,r)$	$\{3\}$	$\emptyset$
$(2,r)$ $(4,r)$ $(5,b)$	$\{4\}$	$\emptyset$
$(1,g)$ $(5,r)$	$\{5\}$	$\emptyset$
~~$(2,2)$~~	~~$\{1,2\}$~~ $\{1\}$	~~$1 — 2$~~ $\emptyset$

(b) Conditional supports $S^X_{p_{\mathcal{K}_1}(y,w^*_R|x)}(y, w^*_R)$.



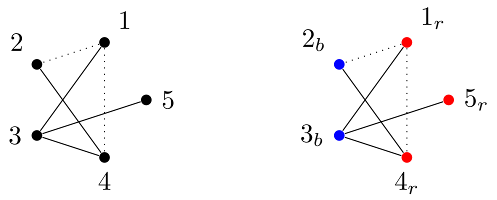
$G^X_{p_{\mathcal{K}_1}(y,y_R|x)} = G^X_{p_{\mathcal{K}_1}(y,w^*_R|x)}$

(c) Confusability graphs.

Figure 24. Compressing Y_R into W^*_R according to the minimum colouring function c on graph G_R in 23(b), is information lossless in the sense that together with Y, W^*_R provides as much information of about X as Y_R.

	$S^X_{p\|_{\mathcal{K}_2}(y\|x)}(y)$	$B^{Y_R}_{p\|_{\mathcal{K}_2}(y,y_R\|x)}(x,y)$	edges
$Y=1$	~~$X=1$~~ $X=5$	~~$\{3,4\}$~~ $\{1\}$	~~$1-3, 1-4$~~ $\emptyset$
$Y=2$	~~$X=1$~~ $X=2$ $X=4$	~~$\{1,2\}$~~ $\{2\}$ $\{4\}$	~~$1-2, 1-4, 2-4$~~ $2-4$
$Y=3$	$X=2$ $X=3$	$\{3\}$ $\{1\}$	$1-3$
$Y=4$	$X=3$ $X=4$	$\{3\}$ $\{4\}$	$3-4$
$Y=5$	$X=4$ $X=5$	$\{3\}$ $\{5\}$	$3-5$

(a) The iterative algorithm for constructing compression graph $G_R|_{\mathcal{K}_2}$.



Compression graph $G_R|_{\mathcal{K}_2}$ One minimum colouring c

(b) The compression graph $G_R|_{\mathcal{K}_2}$ and one choice of minimum colouring function c.

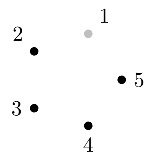
Figure 25. Constructing compression graph $G_R|_{\mathcal{K}_2}$ from induced conditional joint pmf $(\mathcal{K}_2, p_{\mathcal{K}_2}(y, y_R|x), \mathcal{Y}|_{\mathcal{K}_2} \times \mathcal{Y}_R|_{\mathcal{K}_2})$. Note that the least number of colours required is: $\chi(G_R|_{\mathcal{K}_2}) = 2$.

(Y, Y_R)	$S^X_{p(y,y_R\|x)}(y, y_R)$	edges
~~$(1,3)$~~ ~~$(1,4)$~~ ~~$(2,1)$~~	~~$\{1\}$~~	~~$\emptyset$~~
$(3,3)$	$\{2\}$	$\emptyset$
$(3,1)$ $(4,3)$	$\{3\}$	$\emptyset$
$(2,4)$ $(4,4)$ $(5,3)$	$\{4\}$	$\emptyset$
$(1,1)$ $(5,5)$	$\{5\}$	$\emptyset$
~~$(2,2)$~~	~~$\{1,2\}$~~ $\{2\}$	~~$1 — 2$~~ $\emptyset$

(a) Conditional supports $S^X_{p_{\mathcal{K}_2}(y,y_R|x)}(y, y_R)$.

(Y, W^*_R)	$S^X_{p(y,w^*_R\|x)}(y, w^*_R)$	edges
~~$(1,3)$~~ ~~$(1,4)$~~ ~~$(2,1)$~~	~~$\{1\}$~~	~~$\emptyset$~~
$(3,b)$	$\{2\}$	$\emptyset$
$(3,r)$ $(4,b)$	$\{3\}$	$\emptyset$
$(2,r)$ $(4,r)$ $(5,b)$	$\{4\}$	$\emptyset$
$(1,r)$ $(5,r)$	$\{5\}$	$\emptyset$
~~$(2,2)$~~	~~$\{1,2\}$~~ $\{2\}$	~~$1 — 2$~~ $\emptyset$

(b) Conditional supports $S^X_{p_{\mathcal{K}_2}(y,w^*_R|x)}(y, w^*_R)$.



$G^X_{p_{\mathcal{K}_2}(y,y_R|x)} = G^X_{p_{\mathcal{K}_2}(y,w^*_R|x)}$

(c) Confusability graphs.

Figure 26. Compressing Y_R into W^*_R according to the minimum colouring function c on graph G_R in 23(b), is information lossless in the sense that together with Y, W^*_R provides as much information of about X as Y_R.

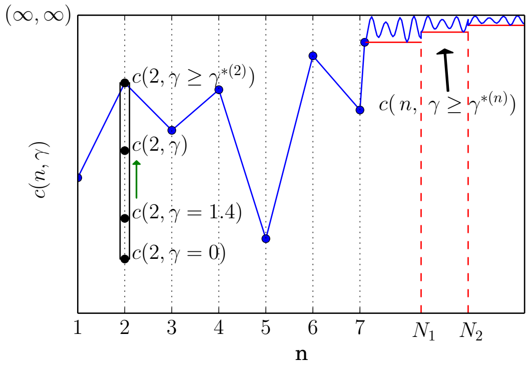


Figure 27. A symbolic graph for showing PRC capacity $C_z^{(n)}(r_z)$ as a function of n and r_z. Note that for succinctness, subscripts $_z$ are omitted and $C(n,r)$ is adopted to indicate $C_z^{(n)}(r_z)$. $C(n, r \geq r^{*(n)})$ equals to the SIMO bound $SIMO(n)$. Red solid and dashed lines are depicted to indicate that the SIMO bound sequence $SIMO(n)$ converges. That is, there exists big enough n (N_1 or N_2) after which all SIMO bounds will be within certain divergence from the limit, which equals to the supremum of the sequence.

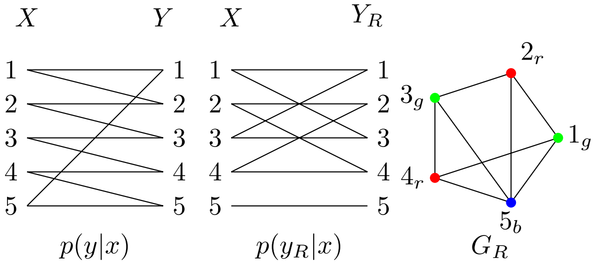


Figure 28. Pentagon problem: marginals and G_R graph used for relaying. The capacity is $\log 5$ and may be achieved if $r_z \geq \log 3$, in 1-shot.

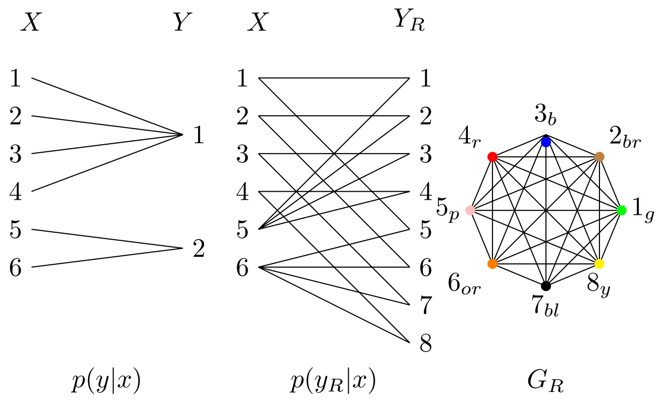


Figure 29. An example where information lossless compression at the relay is impossible.

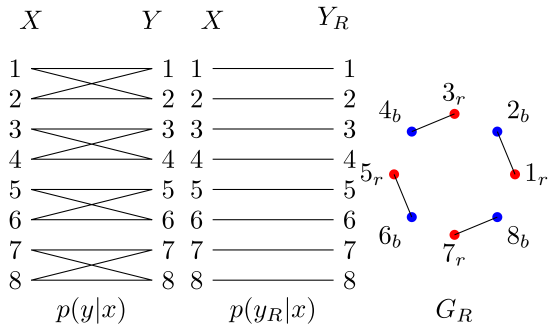


Figure 30. An example where much compression is possible.

CHAPTER 5

ϵ-COLOUR-AND-FORWARD RELAYING IN SMALL-ERROR PRIMITIVE RELAY CHANNELS

In this chapter, we will apply Colour-and-Forward relaying, originally proposed for the communication over a primitive relay channel (PRC) channel without error, to the conventional small-error communication case, as studied in (17). A new ϵ-Colour-and-Forward relaying algorithm is analogously proposed for the problem of communicating over a PRC channel allowing arbitrarily small probability of error. To facilitate the transfer of insights from zero-error Colour-and-Forward relaying to small-error PRCs, an alternative ϵ-error capacity-achieving scheme for a discrete memoryless point-to-point channel is proposed. This scheme resembles that for obtaining the 0-error capacity, utilizing a joint typicality graph which parallels the role of the confusability graph in 0-error communications. In this new framework, an explicit codebook construction approach is provided for the ϵ-error scenario using graph theoretic notation: the codebook is constructed as some maximum independent set of a *joint-typicality confusability graph* under the optimal input distributions. Although the codebook – which involves finding the maximum independent sets of a graph is not easily computable and is in general NP-complete, it provides a unified way to analyze both the 0 and ϵ-error capacities of a point-to-point channel. Consequently, when the relay and destination can fully cooperate, the small-error PRC can be analogized as a virtual zero-error PRC; any zero-error Colour-and-Forward relaying developed for this virtual zero-error PRC serves as one ϵ-error Colour-and-Forward relaying for the original small-error PRC.

5.1 Shannon's channel coding theorems: small-error versus zero-error

Communication allowing a vanishing probability of error is called *small-error* or $\epsilon-$*error* communication, while communication without error is called *zero-error* or 0-*error* communication. Let $(\mathcal{X}, p(y|x), \mathcal{Y})$ denote a discrete memoryless point-to-point channel, where finite sets $\mathcal{X}$ and $\mathcal{Y}$ are respectively the channel input and output alphabets, and the conditional probability mass function (pmf) $p(y|x)$ describes the channel. The small-error capacity and the zero-error capacity of a point-to-point discrete memoryless channel were initially studied by Claude E. Shannon, in (27) in 1948 and in (18) in 1956.

5.1.1 The small-error / ϵ-error scenario

Capacity for the small-error scenario is given below.

Theorem 39 (Shannon's channel coding theorem: small error scenario (27))**.** *The supremum of all message rates that one can communicate via channel* $(\mathcal{X}, p(y|x), \mathcal{Y})$*, allowing a vanishing probability of block-error, is* $\max\limits_{p(x)\in\mathcal{P}_X} I(X;Y)$*.* $\mathcal{P}_X$ *denotes the set of all possible pmfs defined on set* $\mathcal{X}$*.*

This is established by a random coding argument, which generates codebooks randomly, utilizes joint typicality decoding and computes the average (expected) probability of block-error over all possible codebooks. When the block-error probability averaged over all possible codebooks goes to zero as the block length goes to infinity, one may conclude that there exists at least one codebook under which the block-error probability can be made arbitrarily small. This shows the existence, rather than construction, of a good codebook.

5.1.2 The zero-error scenario

The expression for the capacity of a point to point channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ when one wishes to recover the sent message *exactly* looks quite different:

Theorem 40 (Shannon's channel coding theorem: zero error scenario (18))**.** *The supremum of all message rates that one can communicate via channel* $(\mathcal{X}, p(y|x), \mathcal{Y})$ *without error is* $\sup_n \frac{1}{n} \log \alpha([G^X_{p(y|x)}]^{\boxtimes n})$.

The proof for this theorem is very intuitive and straightforward. Note that for zero-error communication, only whether $p(y|x)$ is zero or not matters and the point-to-point channel can be fully described by its *confusability graph* $G^X_{p(y|x)}$ where an edge indicates that the receiver cannot distinguish the two vertices it connects. To build intuition, first consider communicating over a single channel use: the maximal number of channel inputs the destination can distinguish without error is $\alpha(G^X_{p(y|x)})$, the maximum number of vertices that are non-adjacent, or pairwise distinguishable. When multiple channel uses are allowed, by extension, $\alpha([G^X_{p(y|x)}]^{\boxtimes n})$ is the number of distinguishable channel inputs X^n, where $[G^X_{p(y|x)}]^{\boxtimes n}$ is the strong product of n copies of graph $G^X_{p(y|x)}$.[1] [2]

---

[1] Note that the n-fold strong product graph $[G^X_{p(y|x)}]^{\boxtimes n}$ is equivalent to graph $G^{X^n}_{p(y^n|x^n)}$, which is the confusability graph directly constructed from *the compound channel* $(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y}^n)$ with $p(y^n|x^n) = \prod_{i=1}^n p(y_i|x_i)$.

[2] One can show that the supremum of sequence $\{\frac{1}{n} \log \alpha([G^X_{p(y|x)}]^{\boxtimes n})\}_{n=1}^{\infty}$ is equal to its limit.

5.2 Definitions

We now make several definitions for both 0- and ϵ-error communication so as to make the parallel evident. In particular, we define the 0-error and ϵ-error version of the confusability graphs.

5.2.1 Definitions for zero-error communication

Definition 41 (Conditional support)**.** *For a given triple* $(\mathcal{X}, p(y|x), \mathcal{Y})$*, the conditional support of* $Y = y$ *is* $S^X_{p(y|x)}(y) := \{x : p(y|x) > 0\}$.

Definition 42 (Confusability graph)**.** *The confusability graph* $G^X_{p(y|x)}$ *with respect to* Y *and* $p(y|x)$ *for point to point channel* $(\mathcal{X}, p(y|x), \mathcal{Y})$*, is defined on* $\mathcal{X}$ *and two different symbols* $x' \neq x'' \in \mathcal{X}$ *are connected by an edge if there exists some* $y \in \mathcal{Y}$ *such that* $p(y|x') \cdot p(y|x'') > 0$.

5.2.2 Robust typicality

We adopt the form of *robust typicality* proposed in (28); several lemmas based on this definition provided or easily derived from (28) are provided in the Appendix for completeness.

Definition 43. *For a given random variable* $X \sim p(x)$*, its typical set* $T^{\mathbf{X}}_{p(x),\epsilon}$ *is defined as*

$$\{\mathbf{x} \in \mathcal{X}^n : \forall a \in \mathcal{X}, |v_{\mathbf{x}}(a) - p(a)| < \epsilon \cdot p(a)\}$$

where $v_{\mathbf{x}}(a)$ *denotes the empirical distribution on* $\mathcal{X}$ *based on the sample* $\mathbf{x}$.

Definition 44. *For a given random variable pair* $(X, Y) \sim p(x, y)$*, their joint typical set* $T^{\mathbf{XY}}_{p(x,y),\epsilon}$ *is defined as*

$$\{(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n : \forall (a, b) \in \mathcal{X} \times \mathcal{Y}, |v_{\mathbf{x},\mathbf{y}}(a, b) - p(a, b)| < \epsilon \cdot p(a, b)|\}$$

where $v_{\mathbf{x},\mathbf{y}}(a, b)$ *denotes the empirical distribution on* $\mathcal{X} \times \mathcal{Y}$ *based on the sample pair* $(\mathbf{x}, \mathbf{y})$*.*

5.2.3 Definitions for ϵ-error communication

Definition 45 (ϵ-conditional support)**.** *For a given triple* $(\mathcal{X}, p(y|x), \mathcal{Y})$*, the n-shot joint-typicality conditional support of* $\mathbf{Y} = \mathbf{y}$ *is* $S^{\mathbf{X}}_{p(x,y),\epsilon}(\mathbf{y}) := \{\mathbf{x} \in \mathcal{X}^n : (\mathbf{x}, \mathbf{y}) \in T^{\mathbf{XY}}_{p(x,y),\epsilon}\}$*.*

Definition 46 (ϵ-confusability graph)**.** *For a point to point channel* $(\mathcal{X}, p(y|x), \mathcal{Y})$ *and a chosen input distribution* $p(x)$*, its n-shot joint-typicality confusability graph* $G^{\mathbf{X}}_{p(x,y),\epsilon}$ *is defined on* $\mathcal{X}^n$*, where two vertices* $\mathbf{x}'$ *and* $\mathbf{x}''$ *are connected if there exists one* $\mathbf{y} \in \mathcal{Y}^n$ *such that* $(\mathbf{x}', \mathbf{y}) \in T^{\mathbf{XY}}_{p(x,y),\epsilon}$ *and* $(\mathbf{x}'', \mathbf{y}) \in T^{\mathbf{XY}}_{p(x,y),\epsilon}$*. The parameter* $0 < \epsilon < 1$*.*

Definition 47 (Restricted ϵ-confusability graph)**.** *For a point to point channel* $(\mathcal{X}, p(y|x), \mathcal{Y})$ *and a chosen input distribution* $p(x)$*, choose parameters* $0 < \epsilon_1 < \epsilon_2 < 1$ *such that typical set* $T^{\mathbf{X}}_{p(x),\epsilon_1}$ *is a subset of the vertex set of the* ϵ*-confusability graph* $G^{\mathbf{X}}_{p(x,y),\epsilon_2}$*, defined in Definition 46. Then, the restricted* ϵ*-confusability graph is defined as* $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2} = G^{\mathbf{X}}_{p(x,y),\epsilon_2}(T^{\mathbf{X}}_{p(x),\epsilon_1})$*, i.e. the induced subgraph of graph* $G^{\mathbf{X}}_{p(x,y),\epsilon_2}$ *on vertex subset* $T^{\mathbf{X}}_{p(x),\epsilon_1}$*.*

5.3 A new capacity-achieving scheme for Shannon's small-error coding theorem

The link between the 0-error and ϵ-error channel coding theorems lies in using the ϵ-confusability graph in Def. 46 to mimic a zero-error communication protocol whose confusability graph is provided in Def. 42. We formally define the new communication protocol in Subsection 5.3.1, evaluate its probability of error in Subsection 5.3.2, demonstrate the maximal achievable rate of this protocol in Subsection 5.3.3. This will show that that this new communication protocol is indeed capacity-achieving. That is,

Theorem 48. *The communication protocol in Definition 49 can achieve message rates that are arbitrarily close to* $\max\limits_{p(x)} I(X;Y)$*, i.e., it is capacity-achieving.*

Remark 18. *We acknowledge that the communication protocol in Definition 49 is essentially equivalent to the standard random coding protocol, rephrased in the graph theoretic notation. One might think that this is a trivial "rephrasing". But it is noted that this representation via graph theoretic notation is meaningful, because it provides a novel perspective of viewing the small-error communication and thus allows and facilitates the usage of mathematical tools in graph theory and combinatorics.*

5.3.1 The ϵ-error point-to-point communication protocol

We now propose a protocol whose codebook and decoding are based on the restricted ϵ-confusability graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$ of Definition 47.

Definition 49 (Communication protocol)**.** *For any given input distribution $p(x)$, the protocol consists of a codebook L, which is an independent set of the restricted ϵ-confusability graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$, and a decoder:*

$$L \cap S^{\mathbf{X}}_{p(x,y),\epsilon_2}(\mathbf{y}) \,. \tag{5.1}$$

If the intersection is an empty set, the receiver declares an error. If the intersection is non-empty, the receiver claims the only element it contains to be the transmitted one.

Remark 19. *It is noted that there is a one-to-one mapping from the message set to the codebook L and the transmitter sends the codeword corresponding to a given message.*

The communication protocol in Definition 49 essentially view the graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$ as its confusability graph, just like in the zero-error scenario. Intuitively, when only some independent set of graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$ is sent, the destination terminal can distinguish all codewords pairwise *(with arbitrarily small probability of error)* given its side information (its own observation $\mathbf{Y} = \mathbf{y}$). Thus, the maximum number of codewords (channel input sequences) that the destination terminal can pairwise distinguish is equal to the independence number of graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$.

Recall that graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$ is a restricted graph of $G^{\mathbf{X}}_{p(x,y),\epsilon_2}$ over the vertex subset $T^{\mathbf{X}}_{p(x),\epsilon_1}$. Thus, being an independent set of graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$, codebook L simultaneously qualifies for an independent set of graph $G^{\mathbf{X}}_{p(x,y),\epsilon_2}$ and a subset of $T^{\mathbf{X}}_{p(x),\epsilon_1}$.

Considering the first aspect, i.e. codebook L being an independent set of the ϵ-confusability graph $G^{\mathbf{X}}_{p(x,y),\epsilon_2}$, it follows that the intersection of a clique $S^{\mathbf{X}}_{p(x,y),\epsilon_2}(\mathbf{y})$ and an independent set L can at most have one element. In the zero-error communication scenario, the intersection

$L \cap S^{\mathbf{X}}_{p(y|x),\epsilon_2}(\mathbf{y})$ has exactly one element, which is exactly the sent codeword. In the small-error communication scenario, the intersection $L \cap S^{\mathbf{X}}_{p(x,y),\epsilon_2}(\mathbf{y})$ has at most one element, which may or may not be the sent codeword. As we will show, when taking into the second aspect, i.e. codebook L being a subset of $T^{\mathbf{X}}_{p(x),\epsilon_1}$ and $\epsilon_1 < \epsilon_2$, the probability that the intersection has the sent codeword as the unique element can be made arbitrarily close to 1.

Remark 20. *Having two parameters ϵ_1, ϵ_2 and the requirement $\epsilon_1 < \epsilon_2$ are necessary to bound the probability of error in* (Equation 5.2)*. This subtlety comes from the necessity of establishing Lemma 65. We note that weak typicality is insufficient to establish the Lemma 65 (or consequently Theorem 48). Overall, we believe that adopting robust typicality and using the restricted ϵ-confusability graph in Definition 47 are both critical, and novel, components of this protocol.*

5.3.2 Analysis of probability of error

The probability of error under this protocol is:

$$\Pr[\text{error}] := \frac{1}{\|L\|} \sum_{\mathbf{x} \in L} \Pr[\text{ error } \mid \mathbf{x} \in L \text{ was sent }] \leq \delta_{\epsilon_1,\epsilon_2}(n), \tag{5.2}$$

where $\delta_{\epsilon_1,\epsilon_2}(n) := 2\|S^{XY}_{p(x,y)}\| \cdot e^{\frac{-(\epsilon_2-\epsilon_1)^2\cdot\frac{1}{1+\epsilon_1}\cdot(n\cdot p_{\min}(a,b))}{3}}$ and $p_{\min}(a,b) := \min_{(a,b)\in S^{XY}_{p(x,y)}} p(a,b)$ goes to 0 as $n \to \infty$, as

$$
\begin{aligned}
&\Pr[\text{ error } \mid \mathbf{x} \in L \text{ was sent }] \\
=&\Pr[L \cap S^{X}_{p(x,y),\epsilon_2,n}(\mathbf{Y}) = \emptyset \text{ or} \\
&L \cap S^{X}_{p(x,y),\epsilon_2,n}(\mathbf{Y}) \neq \{\mathbf{x}\} \mid \mathbf{x} \in L \text{ was sent }] \\
=&1 - \Pr[L \cap S^{X}_{p(x,y),\epsilon_2,n}(\mathbf{Y}) = \{\mathbf{x}\} \mid \mathbf{x} \in L \text{ was sent }] \\
=&1 - \Pr[(\mathbf{x},\mathbf{Y}) \in T^{\mathbf{XY}}_{p(x,y),\epsilon_2} \mid \mathbf{x} \in L \text{ was sent }] \\
\leq&1 - (1 - \delta_{\epsilon_1,\epsilon_2}(n)) = \delta_{\epsilon_1,\epsilon_2}(n)
\end{aligned}
\tag{5.3}
$$

where the inequality and the value of $\delta_{\epsilon_1,\epsilon_2}(n)$ follow from Lemma 65 in the Appendix.

5.3.3 The maximal achievable rate

It has been shown that the probability of error can be made small, but now discuss how large the codebook L may be while guaranteeing this.

We claim that for any given $p(x)$, the communication protocol in Definition 49 can achieve rates arbitrarily close to $I(X;Y)$ evaluated at $p(x)$ (we will maximize over $p(x)$ in the next subsection). This will be shown by investigating the maximum cardinality of codebook L. Recall that it is desired that

1. L is a subset of the vertex set of graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$, i.e, $L \subseteq T^{\mathbf{X}}_{p(x),\epsilon_1}$.
2. L is an independent set of graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$, denoted by $E\left(G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}(L)\right) = \emptyset$.

We show that such an L exists using a random coding argument. First generate a random set S consisting of s i.i.d. sequences of X of length n, i.e. $S = \{\mathbf{x}(w) : w \in [s]\}$. Next, compute the probability that this random set S satisfies requirements 1) and 2) above. The probability being positive indicates the existence of some "good" realizations of set S satisfying 1) and 2). Thus, rate $\frac{1}{n}\log s$ would be achievable.

Mathematically, we are interested in:

$$\begin{aligned}
&\Pr[S \subseteq T^{\mathbf{X}}_{p(x),\epsilon_1} \text{ and } E\left(G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}(S)\right) = \emptyset] \\
=&\Pr[S \subseteq T^{\mathbf{X}}_{p(x),\epsilon_1} \text{ and } E\left(G^{\mathbf{X}}_{p(x,y),\epsilon_2}(S)\right) = \emptyset] \\
\geq&\Pr[S \subseteq T^{\mathbf{X}}_{p(x),\epsilon_1}] - \Pr[E\left(G^{\mathbf{X}}_{p(x,y),\epsilon_2}(S)\right) \neq \emptyset] \qquad (5.4)\\
\geq&\left(1 - 2\cdot\|S^{X}_{p(x)}\|\cdot e^{-\frac{\epsilon_1^2\cdot n\cdot p_{\min}(a)}{3}}\right)^{s} - 2^{n(2\cdot(\frac{1}{n}\log s - I(X;Y)) + \epsilon_2')} \\
=:&\Delta_{p(x,y),\epsilon_1,\epsilon_2}(n)
\end{aligned}$$

where $\epsilon_2' := \epsilon_2 \cdot (H(XY) + H(X|Y) + 2H(X) + H(Y))$. The first equality follows by Definition 47 that graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$ is an induced subgraph of graph $G^{\mathbf{X}}_{p(x,y),\epsilon_2}$. The first inequality in the above equation follows from $\Pr[A \cap B] \geq \Pr[A] - \Pr[\bar{B}]$.

We defer the proof of the second inequality for now and first interpret this lower bound.

- When $\frac{1}{n}\log s \geq I(X;Y)$, the lower bound $\Delta_{p(x,y),\epsilon_1,\epsilon_2}(n) < 0$, which is not useful.
- For any $\frac{1}{n}\log s < I(X;Y)$ (but arbitrarily close), for any $p(x,y), \epsilon_1, \epsilon_2$, there exists n large enough such that $\Delta_{p(x,y),\epsilon_1,\epsilon_2}(n) > 0$. Thus, $\Pr[S \subseteq T^{\mathbf{X}}_{p(x),\epsilon_1} \text{ and } E\left(G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}(S)\right) = \emptyset] > 0$.

Now we show the computation of probabilities $\Pr[S \subseteq T^{\mathbf{X}}_{p(x),\epsilon_1}]$ and $\Pr[E\left(G^{\mathbf{X}}_{p(x,y),\epsilon_2}(S)\right) \neq \emptyset]$.

5.3.3.1 compute $\Pr[S \subseteq T^{\mathbf{X}}_{p(x),\epsilon_1}]$

By Lemma 66 in the Appendix and the independence of $\mathbf{x}(w), w \in [s]$, we have

$$\begin{aligned} \Pr[S \subseteq T^{\mathbf{X}}_{p(x),\epsilon_1}] &= \Pr[\forall w \in [s],\, \mathbf{x}(w) \in T^{\mathbf{X}}_{p(x),\epsilon_1}] \\ &\geq \left(1 - 2 \cdot \|S^{X}_{p(x)}\| \cdot e^{-\frac{\epsilon_1^2 \cdot n \cdot p_{\min}(a)}{2+\epsilon_1}}\right)^{s} \end{aligned} \tag{5.5}$$

5.3.3.2 compute $\Pr[E\left(G^{\mathbf{X}}_{p(x,y),\epsilon_2}(S)\right) \neq \emptyset]$

$$\begin{aligned} &\Pr[E\left(G^{\mathbf{X}}_{p(x,y),\epsilon_2}(S)\right) \neq \emptyset] \\ =&\Pr[G^{\mathbf{X}}_{p(x,y),\epsilon_2}(S) \text{ has some edge(s) }] \\ \overset{(a)}{\leq}& s \cdot (s-1) \cdot \Pr[\mathbf{x}(w) \text{ and } \mathbf{x}(w') \text{ is connected}] \\ =& s \cdot (s-1) \cdot \Pr[\{(\mathbf{x}, \mathbf{x}', \mathbf{y}) :\ \mathbf{x}(w) = \mathbf{x},\, \mathbf{x}(w') = \mathbf{x}',\, (\mathbf{x}(w), \mathbf{y}) \in T^{\mathbf{XY}}_{p(x,y),\,\epsilon_2}, (\mathbf{x}(w'), \mathbf{y}) \in T^{\mathbf{XY}}_{p(x,y),\epsilon_2}\}] \\ =& s \cdot (s-1) \cdot \sum_{(\mathbf{x},\mathbf{y}) \in T^{\mathbf{XY}}_{p(x,y),\epsilon_2}} \sum_{\mathbf{x}' \in S^{X}_{p(x,y),\epsilon_2,n}(\mathbf{y})} p(\mathbf{x}, \mathbf{x}', \mathbf{y}) \\ \overset{(b)}{\leq}& s \cdot (s-1) \cdot 2^{n(1+\epsilon_2)H(X,Y)} \cdot 2^{n(1+\epsilon_2)H(X|Y)} \\ &\qquad \cdot \frac{1}{2^{n(1-\epsilon_2)H(X)}} \cdot \frac{1}{2^{n(1-\epsilon_2)H(X)}} \cdot \frac{1}{2^{n(1-\epsilon_2)H(Y)}} \\ :=& s \cdot (s-1) \cdot 2^{n(-2 \cdot I(X;Y) + \epsilon_2')} < 2^{n(2 \cdot (\frac{1}{n}\log s - I(X;Y)) + \epsilon_2')} \end{aligned} \tag{5.6}$$

The inequality (a) follows from the union bound and (b) follows from Lemmas 67, 68, 69.

5.4 A unified framework for small-error and zero-error coding

In this section, we will introduce a "matrix" $P^{(n)}$ representation of a point-to-point discrete memoryless channel $(\mathcal{X}, p(y|x), \mathcal{Y})$. Depending on the type of communication, i.e. zero-error or small-error, this matrix $P^{(n)}$ is defined as $P_\epsilon^{(n)}$ and $P_z^{(n)}$ respectively. The zero-error / small-error capacity of the given channel can be correspondingly defined by supremum of the independence numbers of the confusability graphs, determined through the same mechanism as in Definition 56 by matrix $P_\epsilon^{(n)}$ or $P_z^{(n)}$.

5.4.1 Matrix representation for small-error coding scheme

Definition 50 (Matrix $P_\epsilon^{(n)}$)**.** *For a given point-to-point discrete memoryless channel* $(\mathcal{X}, p(y|x), \mathcal{Y})$ *and a fixed channel input distribution* $p(x)$*, define a two-dimensional matrix* $P_\epsilon^{(n)}$ *by*

- *The dimensions are* $\|\mathcal{X}^n\| \times \|\mathcal{Y}^n\|$*.*
- *Each entry is either* 0 *or* 1 *and is specified by* $P_\epsilon^{(n)}(\mathbf{x}, \mathbf{y}) = I_{p(\mathbf{x},\mathbf{y})>0} \cdot I_{(\mathbf{x},\mathbf{y}) \in T^{\mathbf{XY}}_{p(x,y),\epsilon_{2,n}}} \cdot I_{\mathbf{x} \in T^{\mathbf{X}}_{p(x),\epsilon_{1,n}}}$*, where* $p(x,y) = p(x)p(y|x)$*.*

For each matrix $P_\epsilon^{(n)}$, define a graph:

Definition 51. *For any* n*-shot use of point-to-point discrete memoryless channel allowing arbitrarily small probability of error, represented by* $P_\epsilon^{(n)}$*, define graph* $G^{\mathbf{X}}_{P_\epsilon^{(n)}}$*:*

- *The vertex set consists of all row indices* $\mathbf{x}$*, where there is at least one non-zero entry.*
- *Consider every column of the matrix and fully connect row indices of the non-zero entries in that column. The edge set is a union of edges resulting from all columns.*

It can be checked that the confusability graph $G^{\mathbf{X}}_{p(x,y),\epsilon_1,\epsilon_2}$ is equivalent to graph $P^{(n)}_\epsilon$. The decoding function, say $L \cap S^{\mathbf{X}}_{p(x,y),\epsilon_2}$, in the communication protocol in Definition 49 is equivalent to looking at the column indexed by $\mathbf{Y} = \mathbf{y}$ in matrix $P^{(n)}_\epsilon$ and trying to see if there is only one non-zero entry in that column. As n increases, the probability that there is exactly one non-zero entry in the column can be arbitrarily close to 1.

Because the communication protocol in Definition 49 is capacity-achieving (48), it follows that

Lemma 52. *The small-error capacity of a point-to-point discrete memoryless channel* $(\mathcal{X}, p(y|x), \mathcal{Y})$ $P^{(n)}$ *is characterized by the supremum of independence numbers of graph* $G^{\mathbf{X}}_{P^{(n)}_\epsilon}$*, i.e.*

$$C(P^{(n)}_\epsilon) = \sup_n \max_{p(x)} \frac{1}{n} \log \alpha(G^{\mathbf{X}}_{P^{(n)}_\epsilon}) \tag{5.7}$$

where $G^{\mathbf{X}}_{P^{(n)}_\epsilon}$ *is defined by* $P^{(n)}_\epsilon$ *via Definition 51.*

5.4.2 Matrix representation for zero-error coding scheme

Definition 53 (Matrix $P^{(n)}_z$)**.** *For a given point-to-point discrete memoryless channel* $(\mathcal{X}, p(y|x), \mathcal{Y})$ *and a fixed number of channel use* n*, define a two-dimensional matrix* $P^{(n)}_\epsilon$ *by*

- *The dimensions are* $\|\mathcal{X}^n\| \times \|\mathcal{Y}^n\|$.
- *Each entry is either* 0 *or* 1 *and is specified by* $P^{(n)}_z(\mathbf{x}, \mathbf{y}) = I_{p(\mathbf{y}|\mathbf{x})>0}$*, where* $p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i|x_i)$.

For each matrix $P^{(n)}_z$, define a graph:

Definition 54. *For any n-shot use of point-to-point discrete memoryless channel allowing arbitrarily small probability of error, represented by $P_z^{(n)}$, define graph $G^{\mathbf{X}}_{P_z^{(n)}}$:*

- *The vertex set consists of all row indices $\mathbf{x}$, where there is at least one non-zero entry.*
- *Consider every column of the matrix and fully connect row indices of the non-zero entries in that column. The edge set is a union of edges resulting from all columns.*

Consider $n = 1$. It is straightforward that the conditional support of $Y = y$ in Definition 41 is just the x-indices of non-zero entries in the column y in matrix P_z, i.e. $S^X_{p(y|x)}(y) = \{x : P_z(x, y) > 0\}$. Futhermore, the confusability graph $G^X_{p(y|x)}$ is a union of cliques, where each clique is define on the support of a column. It can be checked that graph $G^{\mathbf{X}}_{P_z^{(n)}}$ is the same as graph $[G^X_{p(y|x)}]^{\boxtimes n}$.

Thus, it follows by definition that

Lemma 55. *The zero-error capacity of a point-to-point discrete memoryless channel $P_z^{(n)}$ is characterized by the supremum of independence numbers of graph $G^{\mathbf{X}}_{P_z^{(n)}}$, i.e.*

$$C(P_z^{(n)}) = \sup_n \frac{1}{n} \log \alpha(G^{\mathbf{X}}_{P_z^{(n)}}) \tag{5.8}$$

where $G^{\mathbf{X}}_{P_z^{(n)}}$ is defined by $P_z^{(n)}$ via Definition 54.

5.4.3 A unified framework for small-error and zero-error coding

Thus, we can represent a n-shot usage of a point-to-point discrete memoryless channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ by matrix $P_\epsilon^{(n)}$ or $P_z^{(n)}$. We define the confusability graph $G^{\mathbf{X}}_{P(n)}$ as

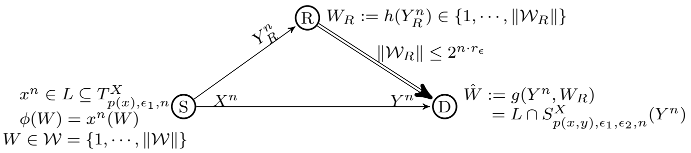
Definition 56. *For a given point-to-point discrete memoryless channel* $P^{(n)}$*, define graph* $G^{\mathbf{X}}_{P^{(n)}}$*:*

- *The vertex set consists of all row indices* $\mathbf{x}$*, where there is at least one non-zero entry.*
- *Consider every column of the matrix and fully connect row indices of the non-zero entries in that column. The edge set is a union of edges resulting from all columns.*

5.5 Application to the primitive relay channel

In this section, we will utilize the matrix representation of the capacity-achieving communication protocol in Definition 49 to mimic the construction of the Colour-and-Forward relaying algorithm, originally developed for zero-error communication, to construct the ϵ-Colour-and-Forward relaying .

5.5.1 Construct ϵ-Colour-and-Forward



R
S
D
$W_R := h(Y_R^n) \in \{1, \cdots, \|\mathcal{W}_R\|\}$
Y_R^n
$\|\mathcal{W}_R\| \leq 2^{n \cdot r_\epsilon}$
$x^n \in L \subseteq T^X_{p(x),\epsilon_1,n}$
$\phi(W) = x^n(W)$
$W \in \mathcal{W} = \{1, \cdots, \|\mathcal{W}\|\}$
X^n
Y^n
$\hat{W} := g(Y^n, W_R)$
$= L \cap S^X_{p(x,y),\epsilon_1,\epsilon_2,n}(Y^n)$

Figure 31. A capacity-achieving communication protocol by joint-typicality confusability graph, when $r_\epsilon = \infty$. An n-shot protocol (n, L, h, g) for small-error communication over a PRC $((\mathcal{X}, p(y, y_R|x), \mathcal{Y} \times \mathcal{Y}_R),\ r_\epsilon)$, with an encoder ϕ, a codebook L, a relaying function h and a decoding function g.

Let $p^*(x)$ be the optimal input distribution that maximizes $I(X;Y,Y_R)$. By Theorem 48, we can use the communication protocol with $p^*(x)$ in Definition 49 to achieve the maximum (or cut-set) information rate $\max\limits_{p(x)} I(X;Y,Y_R)$.

It is noted that for any fixed number of channel use n, the communication protocol in Definition 49 is fully specified by the restricted ϵ-confusability graph $G^{\mathbf{X}|\mathbf{Y},\mathbf{Y}_R}_{p(x,y,y_R),\epsilon_1,\epsilon_2}$ or equivalently, the collection of *restricted* ϵ-conditional supports

$$T := \left\{T^{\mathbf{X}}_{p(x),\epsilon_1} \cap S^{\mathbf{X}|\mathbf{Y},\mathbf{Y}_R}_{p(x,y,y_R),\epsilon_2}(\mathbf{y},\mathbf{y}_R) : (\mathbf{y},\mathbf{y}_R) \in \mathcal{Y}^n \times \mathcal{Y}^n_R\right\} \tag{5.9}$$

Based on collection T, we can implement the Colour-and-Forward relaying algorithm in (29; 30).

We first state the ϵ-Colour-and-Forward graph and relaying and will explain the motivation in the next subsection.

Definition 57 (ϵ-Colour-and-Forward graph $G^{(n)}_{R,\epsilon_2}$)**.** *For a joint distribution $p(x,y,y_R)$, define $G^{(n)}_{R,\epsilon_2}$ as a graph with*

- *Vertex set* $\{\mathbf{y}_R : \exists(\mathbf{x},\mathbf{y})$ *s.t.* $(\mathbf{x},\mathbf{y},\mathbf{y}_R) \in T^{\mathbf{X}\mathbf{Y}\mathbf{Y}_R}_{p(x,y,y_R),\epsilon_2}\}$
- *Edges constructed by connecting two vertices* $\mathbf{y}_{R1}$ *and* $\mathbf{y}_{R2}$ *if there exist* $(\mathbf{x}_1,\mathbf{y})$*,* $(\mathbf{x}_2,\mathbf{y})$*, and* $\mathbf{x}_1 \neq \mathbf{x}_2$ *such that* $(\mathbf{x}_1,\mathbf{y},\mathbf{y}_{R1}) \in T^{\mathbf{X}\mathbf{Y}\mathbf{Y}_R}_{p(x,y,y_R),\epsilon_2}$ *and* $(\mathbf{x}_2,\mathbf{y},\mathbf{y}_{R2}) \in T^{\mathbf{X}\mathbf{Y}\mathbf{Y}_R}_{p(x,y,y_R),\epsilon_2}$.

For the Colour-and-Forward relaying function, we define

Definition 58 (ϵ-Colour-and-Forward relaying $W_{R,\epsilon}^{(n)}$)**.** *Let c be a minimum colouring on the ϵ-Colour-and-Forward graph $G_{R,\epsilon_2}^{(n)}$, defined in Definition 57. Let*

$$W_{R,\epsilon}^{(n)} = \begin{cases} c(\mathbf{y}_R), & \text{when } \mathbf{y}_R \in V(G_{R,\epsilon_2}^{(n)}) \\ 0, & \text{otherwise} \end{cases}. \tag{5.10}$$

5.5.2 ϵ-Colour-and-Forward is information-lossless

ϵ-Colour-and-Forward is information-lossless, in the sense that

Lemma 59 (Information-lossless)**.** *The following two graphs are the same:*

$$G_{p(\mathbf{x},\mathbf{y},\mathbf{y}_R),\epsilon_2}^{\mathbf{X}|\mathbf{Y},\mathbf{Y}_R} = G_{p(\mathbf{x},\mathbf{y},w_R),\epsilon_2}^{\mathbf{X}|\mathbf{Y},W_R} \tag{5.11}$$

where graphs $G_{p(\mathbf{x},\mathbf{y},\mathbf{y}_R),\epsilon_2}^{\mathbf{X}|\mathbf{Y},\mathbf{Y}_R}$ and $G_{p(\mathbf{x},\mathbf{y},w_R),\epsilon_2}^{\mathbf{X}|\mathbf{Y},W_R}$ are defined according to Definition 46, and $W_{R,\epsilon}^{(n)}$ is defined by the ϵ-Colour-and-Forward algorithm.

Corollary 60.

$$G_{p(x,y,y_R),\epsilon_1,\epsilon_2}^{\mathbf{X}} = G_{p(\mathbf{x},\mathbf{y},w_R),\epsilon_1,\epsilon_2}^{\mathbf{X}} \tag{5.12}$$

because

$$G_{p(x,y,y_R),\epsilon_2}^{\mathbf{X}|\mathbf{Y},\mathbf{Y}_R}(T_{p(x),\epsilon_1}^{\mathbf{X}}) = G_{p(\mathbf{x},\mathbf{y},w_R),\epsilon_2}^{\mathbf{X}|\mathbf{Y},W_R}(T_{p(x),\epsilon_1}^{\mathbf{X}}) \tag{5.13}$$

and graph $G^{\mathbf{X}}_{p(x,y,y_R),\epsilon_1,\epsilon_2}$ can be viewed as an induced subgraph of $G^{X}_{p(x,y,y_R),\epsilon_2,n}$ by restricting the vertex set to $T^{\mathbf{X}}_{p(x),\epsilon_1}$. That is,

$$G^{\mathbf{X}}_{p(x,y,y_R),\epsilon_1,\epsilon_2} = G^{\mathbf{X}|\mathbf{Y},\mathbf{Y}_R}_{p(\mathbf{x},\mathbf{y},\mathbf{y}_R),\epsilon_2}(T^{\mathbf{X}}_{p(x),\epsilon_1}) \tag{5.14}$$

Proof of Lemma 59. $G^{\mathbf{X}|\mathbf{Y},\mathbf{Y}_R}_{p(\mathbf{x},\mathbf{y},\mathbf{y}_R),\epsilon_2} = G^{\mathbf{X}|\mathbf{Y},W_R}_{p(\mathbf{x},\mathbf{y},w_R),\epsilon_2}$ or ϵ-Colour-and-Forward relaying $W^{(n)}_{R,\epsilon}$ is information lossless follows the same proof in the zero-error Colour-and-Forward relaying, nicked named as 0-Colour-and-Forward.

Recall that graph $G^{\mathbf{X}|\mathbf{Y},\mathbf{Y}_R}_{p(\mathbf{x},\mathbf{y},\mathbf{y}_R),\epsilon_2}$ is fully specified by the collection $\{S^{\mathbf{X}}_{p(x,y,y_R),\epsilon_2}(\mathbf{y},\mathbf{y}_R) : (\mathbf{y},\mathbf{y}_R) \in \mathcal{Y}^n \times \mathcal{Y}^n_R\}$ and graph $G^{\mathbf{X}|\mathbf{Y},W_R}_{p(\mathbf{x},\mathbf{y},w_R),\epsilon_2}$ by $\{S^{\mathbf{X}}_{p(\mathbf{x},\mathbf{y},w_R),\epsilon_2}(\mathbf{y},w_R) : (\mathbf{y},w^n_R) \in \mathcal{Y}^n \times \mathcal{W}_R\}$.

The lemma follows if these two collections are equivalent, say,

$$\begin{aligned} &\{S^{\mathbf{X}}_{p(x,y,y_R),\epsilon_2}(\mathbf{y},\mathbf{y}_R) : (\mathbf{y},\mathbf{y}_R) \in \mathcal{Y}^n \times \mathcal{Y}^n_R\} \\ &= \{S^{\mathbf{X}}_{p(\mathbf{x},\mathbf{y},w_R),\epsilon_2}(\mathbf{y},w_R) : (\mathbf{y},w_R) \in \mathcal{Y}^n \times \mathcal{W}_R\} \end{aligned} \tag{5.15}$$

We do not state the proof explicitly here, but showing that the above equivalence statement follows naturally from the proof for the 0-Colour-and-Forward being information lossless, the *Theorem 2* in (29).

Firstly, it has been shown in Section 5.4 that the small-error and zero-error communication over the point-to-point single-input multiple-output channel, $(\mathcal{X}, p(y,y_R|X), \mathcal{Y} \times \mathcal{Y}_R)$, can be unitedly represented in the matrix form $P^{(n)}$. Recall that $p^*(x)$ denotes the optimal input distribution that maximizes $I(X;Y,Y_R)$. So we have $\sup\limits_n \frac{1}{n} \log \alpha(G_{P^n_{\epsilon,p^*(x)}})$.

Furthermore, the 0-Colour-and-Forward graph (the Colour-and-Forward graph G_R in *Definition 12* in (29)) and ϵ-Colour-and-Forward graph can also be defined in a unified way:

- Given a three dimensional matrix $\mathbf{P}(\mathbf{x}, \mathbf{y}, \mathbf{y}_R)$
- The vertex set is $\{\mathbf{y}_R : \exists \text{ some } (\mathbf{x}, \mathbf{y}) \text{ s.t. } P(\mathbf{x}, \mathbf{y}, \mathbf{y}_R) > 0\}$
- The edges are constructed by connecting $\mathbf{y}_{R1}$ and $\mathbf{y}_{R2}$ if there exist $(\mathbf{x}_1, \mathbf{y})$ and $(\mathbf{x}_2, \mathbf{y})$, where $\mathbf{x}_1 \neq \mathbf{x}_2$ such that $P(\mathbf{x}_1, \mathbf{y}, \mathbf{y}_{R1}) \cdot P(\mathbf{x}_2, \mathbf{y}, \mathbf{y}_{R2}) > 0$.

Thus, the argument for 0-Colour-and-Forward graph being information lossless applies for ϵ-Colour-and-Forward graph. $\square$

5.6 Conclusion and Discussion

In this chapter, we have presented a new framework for viewing the ϵ-error channel coding theorem in terms of zero-error notions such as confusability graphs, unifying the 0 and ϵ-error communication strategies. The motivation is to provide a tool for translating results in one domain to the other. We show that this new coding scheme is capacity achieving and furthermore *explicitly constructs* the codebooks as independent sets of confusability graphs (rather than random generation). It is acknowledged that this new communication scheme is essentially equivalent to the standard random coding protocol and might at first appear to be a mere rephrasing of the Shannon's coding scheme using graph theoretic notations. It is noted that this representation via graph theoretic notation is meaningful, because it provides a novel perspective on how to view small-error communication and thus allows and facilitates the usage of mathematical tools in graph theory and combinatorics. Further connections with coding for computing is also an open question that is worth investigating.

Based on the alternative capacity-achieving scheme for the point-to-point discrete memoryless channel, the ϵ-Colour-and-Forward relaying scheme for the small-error communication over PRC channels is established, adopting the insights from the Colour-and-Forward relaying scheme originally developed for the zero-error PRC channels. This transfer of insights sets a good example of relating problems in different domains, in particular connecting the zero-error and small-error communication problems. The zero-error Colour-and-Forward has been shown to be optimal for any fixed number of channel uses. It remains a open problem to determine whether ϵ-Colour-and-Forward is optimum for small-error PRCs to achieve its SIMO bounds.

APPENDICES

Appendix A

COPYRIGHT PERMISSION

This Appendix includes the copyright permission granted from the IEEE to use published work in thesis. The following statement has been copied from the CopyRight Clearance Center (Rightslink)

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

1) The following IEEE copyright/credit notice should be placed prominently in the references: [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication].

2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.

3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. However permission to reprint/republish IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution to

Appendix A (Continued)

Appendix B

PROOF OF THEOREM 16

Proof. We first present the definition for *Error matrix* $\mathbf{T}$ and lemmas regarding the structure of the equations at the relays which will aid in the error analysis.

Definition 61 (Error matrix $\mathbf{T}$)**.** *Define the* $L \times L$ *binary matrix* $\mathbf{T}_{L\times L}$ *as:*

$$\mathbf{T}_{L\times L} \quad := \begin{pmatrix} I_{\mathbf{u}^0_{1,1}}(\mathbf{u}_{1,1}) & I_{\mathbf{u}^0_{1,2}}(\mathbf{u}_{1,2}) & \cdots & I_{\mathbf{u}^0_{1,L}}(\mathbf{u}_{1,L}) \\ I_{\mathbf{u}^0_{2,1}}(\mathbf{u}_{2,1}) & I_{\mathbf{u}^0_{2,2}}(\mathbf{u}_{2,2}) & \cdots & I_{\mathbf{u}^0_{2,L}}(\mathbf{u}_{2,L}) \\ \vdots & & & \vdots \\ I_{\mathbf{u}^0_{L,1}}(\mathbf{u}_{L,1}) & I_{\mathbf{u}^0_{L,2}}(\mathbf{u}_{L,2}) & \cdots & I_{\mathbf{u}^0_{L,L}}(\mathbf{u}_{L,L}) \end{pmatrix}_{L\times L} \qquad \in \{0,1\}^{L\times L},$$

where $I_a(b)$ *is the indicator function which has value* 1 *if* $a = b$ *and 0 otherwise. Entries of* $\mathbf{T}_{L\times L}$ *from different columns are independent, while entries of the same column are correlated, which follows from the message matrix* $\mathbf{W}_{L\times k}$ *being block upper triangular. We present properties of* $\mathbf{T}$ *in Lemmas 62 and 64, which will be used to enumerate the different error events for the ICF problem. Lemma 62 states that given arbitrary* c *or more elements of* $\tilde{\mathbf{U}}_{*c}$*, the remaining entries of* $\tilde{\mathbf{U}}_{*c}$ *can be reconstructed deterministically; this will limit the number of error events in decoding the equation sections as outlined in Lemma 64.*

Lemma 62 (Properties of equation sections)**.** *Assume the matrix* $\mathbf{F}$ *and all* c *by* c *sub-matrices from its first* c *columns are of full rank. Then, for any column* $c = 1, \cdots, L$,

$$\textit{row space of } \tilde{\mathbf{U}}_{*c} := \textit{ span } [\{\mathbf{u}_{1,c}, \cdots, \mathbf{u}_{L,c}\}] = \textit{ span } [\mathbf{U}_{A,c}]$$

where $\mathbf{U}_{A,c} = \{\mathbf{u}_{a,c} : a \in A\}$, *and* $A \subset \{1, 2, \cdots, L\}, \|A\| \geq c$.

Proof. Recall that we zero-pad at the head of each message to make them of equal length k. Thus,

$$\tilde{\mathbf{U}}_{*c} = \mathbf{F} \cdot \tilde{\mathbf{W}}_{*c} = \mathbf{F} \cdot \begin{pmatrix} \mathbf{w}_{1,c} \\ \vdots \\ \mathbf{w}_{c,c} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix} = \mathbf{F}_{[1,\cdots,L]\times[1,\cdots,c]} \cdot \begin{pmatrix} \mathbf{w}_{1,c} \\ \vdots \\ \mathbf{w}_{c,c} \end{pmatrix}$$

where sub-matrix $\mathbf{F}_{[1,\cdots,L]\times[1,\cdots,c]}$ contains the first c columns of matrix $\mathbf{F}$. Given any c rows of $\tilde{\mathbf{U}}_{*c}$, we can solve for $\{\mathbf{w}_{1,c}, \cdots, \mathbf{w}_{c,c}\}$, i.e., $\tilde{\mathbf{W}}_{*c}$, since any square sub-matrix of $\mathbf{F}_{[1,\cdots,L]\times[1,\cdots,c]}$ is guaranteed to be of full rank. This is why we need only all $c \times c$ sub-matrices of the first c columns to be full rank. □

Definition 63. *Let* $\beta_c := \sum_{m=1}^{L} I_{\mathbf{u}^0_{m,c}}(\mathbf{u}_{m,c})$, $c = 1, 2, \cdots, L$. β_c *indicates the number of "1" entries in cth column of error matrix* $\mathbf{T}$, *i.e., the number of correctly estimated cth equation sections.*

Lemma 64 (Cardinality lemma for L-user ICF problem)**.** *We have*

1. $\beta_c \in \{0, 1, \cdots, c-1, L\}$*,* $\forall c = 1, \cdots, L$*.*

2. *The number of possible values the equation matrix* $\mathbf{U}_{L\times k}$*, i.e.* $\tilde{\mathbf{U}}_{L\times L}$*, may take on is* $2^{n\sum_{c=1}^{L}\gamma_c\rho_c}$*, where* $\rho_c = \frac{1}{n}\log_2 p^{s_c}$ *is the rate of equation section* $\mathbf{u}_{m,c}$*, and* γ_c *is:*

$$\gamma_c = \begin{cases} c-\beta_c & \text{if } \beta_c = 0, 1, \cdots, c-1 \\ 0 & \text{if } \beta_c = L \end{cases}.$$

Proof. First, note that $\beta_c \in \{0, 1, \cdots, c-1, L\}, \forall c = 1, \cdots, L$. To see this, from Lemma 62, we know that, if there are c or more occurrences of $I_{\mathbf{u}^0_{m,c}}(\mathbf{u}_{m,c}) = 1$ in the cth column of error matrix $\mathbf{T}_{L\times L}$, i.e., $\beta_c \geq c$, then the row space of $\tilde{\mathbf{U}}_{*c}$ is determined and the remaining $L - \beta_c$ rows of $\tilde{\mathbf{U}}_{*c}$ may be deterministically computed from the rows for which $\mathbf{u}_{m,c} = \mathbf{u}^0_{m,c}$. Then the remaining $L - \beta_c$ rows of $\tilde{\mathbf{U}}_{*c}$ are guaranteed to equal those of $\tilde{\mathbf{U}}^0_{*c}$. So, $\beta_c \geq c$ implies $\beta_c = L$.

Next, γ_c is equal to the number of rows of $\tilde{\mathbf{U}}_{*c}$ that remain free/unresolved given a particular value of β_c. To see this, from Lemma 62, there are c degrees of freedom for the row space of $\tilde{\mathbf{U}}_{*c}$. When $\beta_c \neq L$, i.e., when $\beta_c < c$, there are at most $c - \beta_c$ remaining choices for the $L - \beta_c$ rows of $\tilde{\mathbf{U}}_{*c}$. So $\gamma_c = c - \beta_c$. When $\beta_c = L$, $\tilde{\mathbf{U}}_{*c}$ is fixed and there is no freedom in specifying its rows, i.e, $\gamma_c = 0$.

Finally, to show the number of choices of $\mathbf{U}_{L\times k}$, i.e. $\tilde{\mathbf{U}}_{L\times L}$, there are $2^{n\gamma_c\rho_c}$ choices of equation section matrix $\tilde{\mathbf{U}}_{*c}$ with dimension $L \times s_c$. As the L equation sections are independent, when $\{\beta_1, \beta_2, \cdots, \beta_L\}$ are given, the number of possible choices of $\mathbf{U}_{L\times k}$ is $\prod_{c=1}^{L} 2^{n\gamma_c\rho_c} = 2^{n\sum_{c=1}^{L}\gamma_c\rho_c}$. □

We now proceed with the proof of achievability. Fix $p(q)\cdot p(x_1|q)\cdot\cdots\cdot p(x_L|q)\cdot p(y|x_1,\cdots,x_L)$.

Codebook generation:

1. Generate $2^{n\rho_1}$ sequences q^n i.i.d. $\sim p(q)$, indexed by $\mathbf{u}_{1,1}$ or equivalently by $\mathbf{u}_{m,1}$, $m = 2,\cdots,L$.
2. At each relay m, $m = 1,\cdots,L$, for each sequence q^n, generate $2^{n(\rho_2+\cdots+\rho_L)}$ sequences $X_m^n(\mathbf{u}_m) := X_m^n(\mathbf{u}_{m,2},\cdots,\mathbf{u}_{m,L}|\mathbf{u}_{m,1})$ i.i.d. according to $\Pr(X_m^n(\mathbf{u}_m)) = \prod_{t=1}^{n} p(x_{mt}|q_t(\mathbf{u}_{m,1}))$, where x_{mt} denotes the t-th position in the row vector/sequence x_m^n, and q_t denotes the t-th position in the sequence q^n.

Notice that we index codebooks by the message equations; this differs somewhat from more standard codebooks indexed by a message $\in \{1,2,\cdots,2^{nR}\}$ for coding rate R. Codebooks $Q^n(\mathbf{u}_{1,1})$ and $X_m^n(\mathbf{u}_m), m = 1,\cdots,L$ are revealed to the relays and destination. Codebook Q^n can be equivalently indexed by $\mathbf{u}_{1,1}, \mathbf{u}_{2,1},\cdots,\mathbf{u}_{L,1}$ as needed or even $\tilde{\mathbf{U}}_{*1}$, i.e. this common portion is available to all relays.

Encoder: Relay m sends signal $X_m^n(\mathbf{u}_m)$.

Decoder: The destination node looks for unique $\mathbf{u}_1,\cdots,\mathbf{u}_L$ such that $(Y^n, Q^n(\tilde{\mathbf{U}}_{*1}), X_1^n(\mathbf{u}_1),\cdots,X_L^n(\mathbf{u}_L))$ are ϵ-jointly typical according to $p(q,x_1,\cdots,x_L,y)$, or lie in the set $A_\epsilon^{(n)}(Q,X_1,\cdots,X_L,Y)$. If none, or more than one tuple of equation sections are jointly typical with the given Y^n, an error is declared.

Appendix B (Continued)

Error analysis: Define the event $E\left(\tilde{\mathbf{U}}_{L\times L}\right) = \left\{\left(Q^n(\tilde{\mathbf{U}}_{*1}), X_1^n(\mathbf{u}_1), \cdots, X_L^n(\mathbf{u}_L), Y^n\right) \in A_\epsilon^{(n)}\right\}$. Assume WLOG that the equation set $\{\mathbf{u}_1^0, \mathbf{u}_2^0, \cdots, \mathbf{u}_L^0\}$, also represented as $\tilde{\mathbf{U}}^0_{L\times L}$, is the true, or transmitted set of message equations. Recall the following definitions, and further define α_m, ν as follows, which will help in succinctly enumerating the error events:

$$\begin{aligned}
\mathbf{T}_{L\times L} &= I_{\tilde{\mathbf{U}}^0_{L\times L}}\left(\tilde{\mathbf{U}}_{L\times L}\right) \quad &&\in \{0,1\}^{L\times L}\\
\beta_c &= \sum_{m=1}^{L} I_{\mathbf{u}^0_{m,c}}\left(\mathbf{u}_{m,c}\right) \quad &&\in \{0,1,\cdots,c-1,L\}\\
\alpha_m &= \prod_{c=1}^{L} I_{\mathbf{u}^0_{m,c}}\left(\mathbf{u}_{m,c}\right) \quad &&\in \{0,1\}\\
\nu &= \text{the number of ones in } \{\alpha_1,\cdots,\alpha_L\} \quad &&\in \{0,1,\cdots,L\}.
\end{aligned}$$

We may then denote four different types of error events; event $E^c(\tilde{\mathbf{U}}^0_{L\times L})$ indicates the case when the correct codeword is not jointly typical with the received signal Y^n and is denoted as error event type I. Error event types II - IV are provided in Table Table V. Note that we use the error matrix $\mathbf{T}$ to facilitate the classification of error events and the analysis of their corresponding probabilities. Recall that, for each $c = 1,\cdots,L$, that β_c denotes the number of correct items in the c-th column of $\mathbf{T}$, i.e, denotes the number of correctly estimated equation sections $\mathbf{u}_{m,c}$. Observe that values of $\beta_1,\cdots,\beta_L$ are mutually independent by default due to the uniform and i.i.d. generation of the message vector ($\mathbf{w}_l \in \mathbb{F}_p^{k_l}$) over the finite field $\mathbb{F}_p$. This allows us to first categorize the error events according to the two possible values of β_1 (Table Table V). Conditioned on the value of β_1, we may then enumerate all possible values for

Appendix B (Continued)

$\beta_2, \cdots, \beta_L$. Such an enumeration is correct but may be simplified as follows: note that greater than or equal to one zeros in the mth row of $\mathbf{T}$ will lead to the same probabilistic dependence of the received signal y^n on codewords x_m^n. That is, for each $m = 1, \cdots, L$, α_m denotes whether the m-th equation $\mathbf{u}_m$ is correct. We can then more succinctly classify the error event types using α_m, reflected as variable ν = the number of ones in $\{\alpha_1, \cdots, \alpha_L\}$ in Table Table V. Further explanations for each error event type follow.

Possible error events		Error event type
$\beta_1 \in \{0, L\}$	$\nu \in \{0, 1, \cdots, L\}$	
0	$\forall\, \nu \in \{0, 1, \cdots, L\}$	II
L	$\nu = 0$	III
L	some fixed ν, $\nu \in \{1, \cdots, L-1\}$ some fixed A, $\|A\| = \nu$	IV
L	$\nu = L$	correct

TABLE V

TABLE DENOTING THE TYPES OF ERROR EVENTS.

By symmetry, the probability of error may be expressed as

$$\begin{aligned}
P_e^{(n)} &= \Pr\left(\hat{\tilde{\mathbf{U}}}^0_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L} \mid \tilde{\mathbf{U}}^0_{L\times L} \text{ is sent }\right)\\
&\le \Pr\left(E^c(\tilde{\mathbf{U}}^0_{L\times L})\right) + \Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L}\neq\tilde{\mathbf{U}}^0_{L\times L},\, \beta_1=0,} E(\tilde{\mathbf{U}}_{L\times L})\right)\\
&\quad + \Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L}\neq\tilde{\mathbf{U}}^0_{L\times L},\, \beta_1=L,\, \nu=0} E(\tilde{\mathbf{U}}_{L\times L})\right) + \Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L}\neq\tilde{\mathbf{U}}^0_{L\times L},\, \beta_1=L,\, \nu\in\{1,\cdots,L-1\}} E(\tilde{\mathbf{U}}_{L\times L})\right)
\end{aligned}$$

We consider the conditions which will drive $P_e^{(n)} \to 0$ as $n \to \infty$ for each of the events separately.

- *Error event type I:* $\Pr\left(E^c(\tilde{\mathbf{U}}^0_{L\times L})\right)$ vanishes by properties of the jointly typical set $A_\epsilon^{(n)}$.
- *Error event type II:* Consider $\Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L}\neq\tilde{\mathbf{U}}^0_{L\times L},\, \beta_1=0} E(\tilde{\mathbf{U}}_{L\times L})\right)$. The constraint $\beta_1 = 0$ indicates that $\mathbf{U}_{*1} \neq \mathbf{U}^0_{*1}$ (common message is incorrect) but says nothing about the remaining $\mathbf{U}_{*c}$, $c = 2, 3, \cdots, L$. This incorrect first column serves as the index for sequence $q^n(\mathbf{U}_{*1})$, on which codewords $X_m^n(\mathbf{u}_m)$ are conditioned. Thus the incorrectness of first column implies that the observed y^n is independent of the true $(q^n, x_1^n, \cdots, x_L^n)$. Thus, for each error event in this category,

$$\Pr\left(E(\tilde{\mathbf{U}}_{L\times L})\,|\beta_1 = 0\right) \le 2^{-n\cdot(I(Q,X_1,\cdots,X_L;Y)-\epsilon)} \overset{(a)}{\le} 2^{-n\cdot(I(X_1,\cdots,X_L;Y)-\epsilon)}.$$

where (a) follows by the Markov chain $Q \to (X_1, \cdots, X_L) \to Y$, By the union bound,

$$\Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\ \beta_1 = 0} E(\tilde{\mathbf{U}}_{L\times L}) \right) \leq \parallel \left\{ \tilde{\mathbf{U}}_{L\times L} \,|\, \tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\ \beta_1 = 0 \right\} \parallel \ \times 2^{-n\cdot(I(X_1,\cdots,X_L;Y)-\epsilon)}.$$

By Lemma 64, the maximal value of the cardinality term (which in turn yields the dominant constraint) occurs when $\beta_c = 0$ for all remaining $c = 2, \cdots, L$, yielding the maximum value of $2^{n\sum_{c=1}^{L}\gamma_c\rho_c} = 2^{n\sum_{c=1}^{L}c\rho_c}$, and hence the following is needed to ensure $P_e \to 0$:

$$\sum_{c=1}^{L} c\rho_c \leq I(X_1, \cdots, X_L; Y). \tag{B.1}$$

- *Error event type III:* Consider $\Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\ \beta_1 = L,\ \nu = 0} E(\tilde{\mathbf{U}}_{L\times L})\right)$. This implies that the first column of $\mathbf{T}$ are all ones (index $\tilde{\mathbf{U}}_{*1}$ of Q^n is correctly decoded) and there exists at least one zero in every row of $\mathbf{T}$ (at least one equation section of each equation is wrong). Thus,

$$\Pr\left(E(\tilde{\mathbf{U}}_{L\times L}) \,|\, \text{ some } \{\beta_1, \cdots, \beta_L\} \text{ s.t. } \tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\ \beta_1 = L,\ \nu = 0 \right) \leq 2^{-n\cdot(I(X_1,\cdots,X_L;Y|Q)-\epsilon)}.$$

By the union bound

$$\begin{aligned} &\Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\ \beta_1 = L,\ \nu = 0} E(\tilde{\mathbf{U}}_{L\times L}) \right) \\ &\leq \parallel \left\{ \tilde{\mathbf{U}}_{L\times L} \,|\, \tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\ \beta_1 = L,\ \nu = 0 \right\} \parallel \ \times 2^{-n\cdot(I(X_1,\cdots,X_L;Y|Q)-\epsilon)}. \end{aligned}$$

By Lemma 64, the maximal cardinality occurs when $\beta_1 = L$ and $\beta_c = 0$ for $c = 2 \cdots, L$, with a maximum value of $2^{n \sum_{c=1}^{L} \gamma_c \rho_c} = 2^{n \sum_{c=2}^{L} c\rho_c}$. This yields the dominant constraint

$$\sum_{c=2}^{L} c\rho_c \leq I(X_1, \cdots, X_L; Y|Q). \tag{B.2}$$

- *Error event type IV:* Consider $\Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\, \beta_1 = L,\, \nu \in \{1,\cdots,L-1\}} E(\tilde{\mathbf{U}}_{L\times L})\right)$. This implies that all entries of the first column and ν rows of matrix $\mathbf{T}$ are correct. We further sub-categorize this type of error event. Let $A \subset \{1, \cdots, L\}$, $\|A\| = \nu$, denote the indices of rows that are all ones, i.e, $\alpha_m = 1, \forall m \in A$. Let $A^C = \{1, \cdots, L\} \setminus A$ denote the rows that contain at least one zero, i.e., $\alpha_m = 0, \forall m \in A^C$. For each A, we note that X_m^n for $m \in A$ are correct, and the remaining $X_{m'}^n$ for $m' \in A^C$ are incorrect. Thus, for each type of error event within this sub-category, we have

$$\Pr\left(E(\tilde{\mathbf{U}}_{L\times L}) \,|\, \text{ some } \{\beta_1, \cdots, \beta_L\} \text{ s.t. } \tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\ \beta_1 = L, A, \|A\| = \nu\right) \leq 2^{-n\cdot\left(I(X_{A^C};Y|X_A,Q)-\epsilon\right)}$$

By the union bound

$$\Pr\left(\bigcup_{\{\beta_1,\cdots,\beta_L\} \text{ s.t. } \tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\, \beta_1 = L,\, A,\, \|A\| = \nu} E(\tilde{\mathbf{U}}_{L\times L})\right)$$
$$\leq \left\| \left\{ \tilde{\mathbf{U}}_{L\times L} \,|\, \{\beta_1, \cdots, \beta_L\} \text{ s.t. } \tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\ \beta_1 = L,\ A, \|A\| = \nu \right\} \right\| \times 2^{-n\cdot\left(I(X_{A^C};Y|X_A,Q)-\epsilon\right)}.$$

By Lemma 64, the maximum value of the cardinality term occurs when $\beta_1 = L$ and

$$\beta_c = \begin{cases} L & c < \nu \\ \\ \nu & c \geq \nu \end{cases} \qquad \text{for } c = 2, \cdots, L,$$

yielding the upper bound, by Lemma 64,

$$\| \left\{ \tilde{\mathbf{U}}_{L\times L} \,|\, \{\beta_1, \cdots, \beta_L\} \text{ s.t. } \tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L},\ \beta_1 = L,\ A, \|A\| = \nu \right\} \| \leq 2^{n \sum_{c=1}^{L} \gamma_c \rho_c} = 2^{n \sum_{c=\nu+1}^{L} (c-\nu)\rho_c}.$$

Thus, the most constraining condition needed to ensure $P_e \to 0$ is

$$\sum_{c=\nu+1}^{L} (c-\nu)\rho_c \leq I(X_{A^C}; Y | X_A, Q), \text{ where } A \subset \{1, \cdots, L\}, \|A\| = \nu\,. \tag{B.3}$$

This constraint holds for all $\nu \in \{1, \cdots, L-1\}$.

Substituting $\rho_c = R_c - R_{c+1}$ $(R_{L+1} = 0)$ yields Theorem 16. $\square$

Appendix C

LIST OF LEMMAS ON THE ROBUST TYPICALITY

The following lemmas may be shown using standard techniques following the presentation in (28) and are included for completeness.

Lemma 65. *Let* $0 < \epsilon_1 < \epsilon_2 < 1$. *For every* $\mathbf{x} \in T^{\mathbf{X}}_{p(x),\epsilon_1}$,

$$\Pr[(\mathbf{x}, \mathbf{y}) \in T^{\mathbf{XY}}_{p(x,y),\epsilon_2} | \mathbf{x} = \mathbf{x}] \geq 1 - \delta_{\epsilon_1,\epsilon_2}(n)\,, \tag{C.1}$$

where $\delta_{\epsilon_1,\epsilon_2}(n) := 2\|S^{XY}_{p(x,y)}\| \cdot e^{\frac{-(\epsilon_2-\epsilon_1)^2 \cdot \frac{1}{1+\epsilon_1} \cdot (n \cdot p_{\min}(a,b))}{3}}$ *and* $p_{\min}(a, b) := \min_{(a,b) \in S^{XY}_{p(x,y)}} p(a, b)$.

Lemma 66.

$$\Pr[\mathbf{x} \in T^{X}_{p(x),\epsilon,n}] \geq 1 - \delta_{\epsilon}(n)\,, \tag{C.2}$$

where $\delta_{\epsilon}(n) := 2 \cdot \|S^{X}_{p(x)}\| \cdot e^{-\frac{\epsilon^2 \cdot n \cdot p_{\min}(a)}{3}}$ *goes to zero as* $n \to \infty$ *and* $p_{\min}(a) := \min_{a \in S^{X}_{p(x)}} p(a)$.

Lemma 67. *For every* $\mathbf{x} \in T^{\mathbf{X}}_{p(x),\epsilon}$,

$$2^{-(1+\epsilon)H(X)n} \leq p(\mathbf{x}) \leq 2^{-(1-\epsilon)H(X)n}\,. \tag{C.3}$$

Lemma 68.

$$(1 - \delta^{XY}_{\epsilon}(n)) \cdot 2^{(1-\epsilon)H(X,Y)n} \leq \|T^{\mathbf{XY}}_{p(x,y),\epsilon}\| \leq 2^{(1+\epsilon)H(X,Y)n} \tag{C.4}$$

$\delta^{XY}_{\epsilon}(n) := 2 \cdot \|S^{XY}_{p(x,y)}\| \cdot e^{-\frac{\epsilon^2 \cdot n \cdot p_{\min}(a,b)}{2+\epsilon}}$ *and* $p_{\min}(a, b) := \min_{(a,b) \in S^{XY}_{p(x,y)}} p(a, b)$.

Appendix C (Continued)

Lemma 69. *Random variable pair* $(X, Y) \sim p(x, y)$. $S^{X}_{p(x,y),\epsilon,n}(\mathbf{y}) := \{\mathbf{x} \in \mathcal{X}^n : (\mathbf{x}, \mathbf{y}) \in T^{\mathbf{XY}}_{p(x,y),\epsilon}\}$. *Then,*

$$\|S^{X}_{p(x,y),\epsilon,n}(\mathbf{y})\| \leq 2^{(1+\epsilon)H(X|Y)n}$$

CITED LITERATURE

1. Cover, T. and Thomas, J.: Elements of Information Theory: Second Edition. Wiley, 2006.

2. Slepian, D. and Wolf, J.: A coding theorem for multiple access channels with correlated sources. Bell Syst. Tech. Journal, 52(7):1037–1076, 1973.

3. Prelov, V. V.: Transmission over a multiple-access channel with a special source hierarchy. Problems Inform. Transmission, 20(4):233–239, 1984.

4. Nazer, B. and Gastpar, M.: Compute-and-forward: Harnessing interference through structured codes. IEEE Trans. Inf. Theory, 57(10):6463–6486, 2011.

5. Hong, S.-N. and Caire, G.: Two-unicast two-hop interference network: Finite-field model. In Information Theory Workshop (ITW), 2013 IEEE, pages 1–5, Sept 2013.

6. Hong, S.-N. and Caire, G.: Structured lattice codes for 2 x 2 x 2 mimo interference channel. In Proc. IEEE Int. Symp. Inf. Theory, pages 2229–2233, July 2013.

7. Zhan, J., Nazer, B., Erez, U., and Gastpar, M.: Integer-Forcing Linear Receivers. `http://arxiv.org/abs/1003.5966/`.

8. Hong, S.-N. and Caire, G.: Reverse compute and forward: A low-complexity architecture for downlink distributed antenna systems. In Proc. IEEE Int. Symp. Inf. Theory,

CITED LITERATURE (Continued)

pages 1147–1151, July 2012.

9. Hong, S.-N. and Caire, G.: Compute-and-forward strategies for cooperative distributed antenna systems. IEEE Trans. Inf. Theory, 59(9):5227–5243, Sept 2013.

10. Song, Y., Devroye, N., and Nazer, B.: Inverse compute-and-forward: Extracting messages from simultaneously transmitted equations. In Proc. IEEE Int. Symp. Inf. Theory, St.Petersburg, Russia, August 2011.

11. Wigger, M. A.: Cooperation on the multiple-access channel. Doctoral dissertation, ETH, 2008.

12. Willems, F. M.: Information theoretical Results for Multiple Access Channels. Doctoral dissertation, K.U. Leuven, 1982.

13. Han, T.: The capacity region of general multiple-access channel with certain correlated sources. Information and Control, 40(1):37–60, 1979.

14. Gunduz, D. and Simeone, O.: On the capacity region of a multiple access channel with common messages. In Proc. IEEE Int. Symp. Inf. Theory, pages 470–474, Austin, June 2010.

15. Cover, T., El Gamal, A., and Salehi, M.: Multiple access channels with arbitrarily correlated sources. IEEE Trans. Inf. Theory, IT-26(6):648–657, November 1980.

CITED LITERATURE (Continued)

16. Chen, Y., Song, Y., and Devroye, N.: The capacity region of three user gaussian inverse-compute-and-forward channels. In Proc. IEEE Int. Symp. Inf. Theory, pages 1476–1480, July 2013.

17. Kim, Y.: Coding techniques for primitive relay channels. In Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput, 2007.

18. Shannon, C.: The zero error capacity of a noisy channel. Information Theory, IRE Transactions on, 2(3):8–19, September 1956.

19. Lovasz, L.: On the shannon capacity of a graph. IEEE Trans. Inf. Theory, 25(1):1–7, 1979.

20. Korner, J. and Orlitsky, A.: Zero-error information theory. IEEE Trans. Inf. Theory, 44(6):2207–2229, 1998.

21. Sonnemann, E. and Krafft, O.: Independence numbers of product graphs. Journal of Combinatorial Theory, Series B, 17(2):133 – 142, 1974.

22. Jha, P. and Slutzki, G.: Independence numbers of product graphs. Applied Mathematics Letters, 7(4):91 – 94, 1994.

23. Vesel, A. and Zerovnik, J.: Improved lower bound on the shannon capacity of C_7. Information Processing Letters, 81(5):277 – 282, 2002.

CITED LITERATURE (Continued)

24. Klavzar, S.: Coloring graph products: A survey. Discrete Mathematics, 155(1 - 3):135 – 145, 1996.

25. Witsenhausen, H.: The zero-error side information problem and chromatic numbers (corresp.). IEEE Trans. Inf. Theory, 22(5):592–593, 1976.

26. Brooks, R. L.: On colouring the nodes of a network. Mathematical Proceedings of the Cambridge Philosophical Society, 37:194–197, 4 1941.

27. Shannon, C.: A mathematical theory of communication. Bell Syst. Tech. J., 27(379-423, 623-656), Jul., Oct. 1948.

28. Orlitsky, A. and Roche, J.: Coding for computing. IEEE Trans. Inf. Theory, 47(3):903–917, 2001.

29. Chen, Y., Shahi, S., and Devroye, N.: Colour-and-forward: relaying "what the destination needs" in the zero-error primitive relay channel. In Allerton Conference on Communication, Control, and Computing, 2014.

30. Chen, Y. and Devroye, N.: The optimality of colour-and-forward relaying for a class of zero-error primitive relay channels. In submitted to the Int. Con. on Inf. Theory (ISIT), 2015.

VITA

Name	Yanying Chen
Education	**University of Illinois at Chicago** Ph.D. in Electrical and Computer Engineering (2009 - 2015) M.Sc. in Electrical and Computer Engineering (2009 - 2015) **Tianjin University** B.Sc. in Electrical Engineering (2005 -2009)
Experience	**Ricoh Innovations Corporation** Summer Research Intern (05/2011 - 08/2011)
Publications	Y. Chen, Y. Song and N. Devroye, “The capacity region of three user Gaussian inverse-compute-and-forward channels,” International Symposium on Information Theory (ISIT), Istanbul, pp. 1476-1480, July 2013.

VITA (Continued)

Publications Y. Chen, Y. Song and N. Devroye, “The capacity region of the L-user Gaussian inverse-compute-and-forward problem,” submitted to Transactions on Information Theory, 10/06/2013.

Publications Y. Chen, S. Shahi and N. Devroye, “Colour-and-Forward: relaying “what the destination needs” in the zero-error primitive relay channel,” 52nd Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, October 2014.

Publications Y. Chen and N. Devroye, “On the optimality of Colour-and-Forward relaying for a class of zero-error primitive relay channels,” to appear in International Symposium on Information Theory (ISIT), 2015.

Publications Y. Chen and N. Devroye, “Colour-and-Forward: relaying only “what the destination needs” in primitive relay channels,” submitted to Transactions on Information Theory, 2015.