The Capacity Region of the L-User Gaussian Inverse Compute-and-Forward Problem

Yanying Chen, Yiwei Song, and Natasha Devroye

***Abstract*— We consider an L-user multiple access channel where transmitter m has access to the linear equation $\mathbf{u}_m = \bigoplus_{l=1}^{L} f_{ml}\mathbf{w}_l$ 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 dependencies between the equations at the different relays 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 shown to be capacity for the additive white Gaussian noise channel.**

***Index Terms*— Channel capacity, multiple access channel, compute-and-forward, correlated sources, joint source-channel coding, multiuser channels.**

I. Introduction

THE recently proposed Compute-and-Forward (CF) framework [3] 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 $\mathbf{w}$ which are vectors of length k of elements over a finite field of size p, $\mathbb{F}_p$, or $\mathbf{w} \in \mathbb{F}_p^k$. 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 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 paper. For more works on multi-source, multi-relay setups, please refer to [4] and [5] 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 paper, 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.

A. Past Work

The problem statement and motivation builds upon the compute-and-forward (CF) framework [3]: 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 [6], an integer-forcing linear receiver framework is developed for a MIMO system and is shown to outperform conventional linear receivers. Papers [7], [8] 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 [8], 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

Manuscript received October 6, 2013; revised August 7, 2016; accepted September 27, 2016. Date of publication October 24, 2016; date of current version November 18, 2016. This work was supported by NSF under Award 1216825 and Award 1053933.

The authors are with The University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: ychen90@uic.edu; sywad87@gmail.com; devroye@uic.edu).

Communicated by M. Langberg, Associate Editor for Coding Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2016.2620466

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

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 [1], where an achievable rate region was presented. Though not formally presented in [1], 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 [9]–[11].

The capacity of an extension of the Slepian-Wolf MAC of [9] to an arbitrary number of users, each of which has access to a subset of independent messages is solved in [12] and simplified in [13].

We note that when going beyond two-users, our L-user ICF problem cannot be mapped into the framework in [12], 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 [14], 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 [9],[2] which also corresponds to the region obtained here for two users. More generally, in [14] 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 [1] 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.

B. Contribution and Outline

Our main contribution 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 II. Before demonstrating the most general results for arbitrary L, in Section III 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 IV, 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 V, an achievable rate region for the general L-user ICF problem is first derived. Our first main contribution, besides the formulation of the problem, is the design of a decoder which exploits the dependencies inherent in the equations available at the transmitters. Achievability is then followed by the derivation of the *capacity* region for the Gaussian MAC channel model, the paper's second main contribution. The converse follows along similar lines to those in [10] and [11], 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.

[2]As shown in the special case d) in [9], 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.

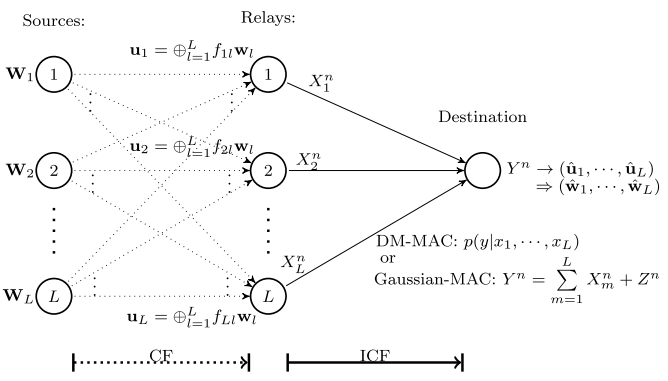


Fig. 1. L-user ICF problem in which L relays each have a linear combination $\mathbf{u}_m = \oplus_{l=1}^{L} f_{ml}\mathbf{w}_l$ of L messages and wish to convey these messages to a single destination.

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.

II. Problem Statement: Definitions and Channel Models

As shown in Fig. 1, 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 [3], 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 1 (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* $R_1 \geq R_2 \geq \cdots \geq R_L$.

Definition 2 (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 [3]):* $\mathbf{u}_m = \bigoplus_{l=1}^{L} f_{ml}\mathbf{w}_l$ *in* $\mathbb{F}_p^k$, *for some given* $f_{ml} \in \mathbb{F}_p$. *In matrix form,*

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

or

$$\mathbf{U} = \mathbf{F} \cdot \mathbf{W},$$

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 3: 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 we need 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 V-E, 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 27 in Appendix A, which enumerates the number of equation sections with different properties and is used in the error analysis.*

Definition 4 (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*

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

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

Definition 6 (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* $\Pr\left[\bigcup_{m=1}^{L}\left\{\hat{\mathbf{u}}_m \neq \mathbf{u}_m\right\}\right] < \epsilon$.

Definition 7 (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* $\varepsilon_1, \cdots, \varepsilon_L$ *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 8 (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 9: Let the computation rate region $\mathcal{R}^{CF}(R_1, \cdots, R_L)$ *defined in [3] 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 we will only be interested in 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 10 (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*

$$\begin{aligned} s_c &:= k_c - k_{c+1}, \\ \rho_c &:= \frac{1}{n}\log_2 p^{s_c} = R_c - R_{c+1}, \end{aligned} \tag{2}$$

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.,* $\mathbf{w}_{l,c}$.

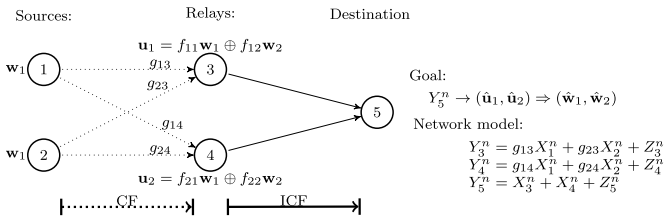


Fig. 2. Two-user ICF problem with Gaussian-MAC channel. Power constraints P_1, P_2, P_3, P_4, respectively.

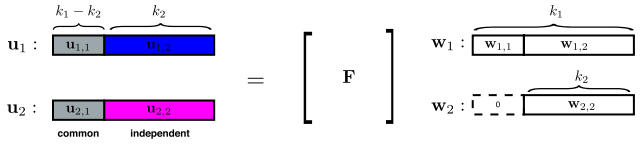


Fig. 3. Two-user ICF message/equation structure. Grey indicates that equation sections $\mathbf{u}_{1,1}$ and $\mathbf{u}_{2,1}$ are fully correlated, while different solid colors indicate that two equation sections $\mathbf{u}_{1,2}$ and $\mathbf{u}_{2,2}$ are independent. All message sections $\mathbf{w}_{i,j}$ are mutually independent; $i, j = 1, 2$.

Define the upper triangular matrix

$$\begin{aligned}&\tilde{\mathbf{W}}_{L\times L}\\ &:= [\tilde{\mathbf{W}}_{*1}, \tilde{\mathbf{W}}_{*2}, \cdots, \tilde{\mathbf{W}}_{*L}]\\ &= \begin{pmatrix} \mathbf{w}_{1,1} & \mathbf{w}_{1,2} & \cdots & \mathbf{w}_{1,c-1} & \mathbf{w}_{1,c} & \cdots & \mathbf{w}_{1,L-1} & \mathbf{w}_{1,L} \\ \mathbf{0} & \mathbf{w}_{2,2} & \cdots & \mathbf{w}_{2,c-1} & \mathbf{w}_{2,c} & \cdots & \mathbf{w}_{2,L-1} & \mathbf{w}_{2,L} \\ \vdots & & & & & & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{w}_{l,c} & \cdots & \mathbf{w}_{l,L-1} & \mathbf{w}_{l,L} \\ \vdots & & & & & & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{w}_{L,L} \end{pmatrix}.\end{aligned} \tag{3}$$

Definition 11 (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

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

Definition 12 (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 $R_{L+1} = 0$.

Remark 13: The notation tilde˜ is adopted for depicting the segmented representation of message and equation matrices. Notation $\tilde{\mathbf{W}}_{L\times L}$ and $\mathbf{W}_{L\times k}$ both refer to the same underlying message matrix and only differ in the indexing of its columns. Similar notation is used for $\tilde{\mathbf{U}}_{L\times L}$ and $\mathbf{U}_{L\times k}$.

III. Two-User Case

Before demonstrating the general L-use result, we consider the $L = 2$ user case with message/equation structure shown in Fig. 3. 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 Fig. 2 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 III-A, 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 III-B, 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 III-B.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.

A. Three Achievable Rate Regions for the Two-User Discrete Memoryless ICF Channel

1) Scheme 1 (A Non-Coherent Scheme Without Cardinality Bounding): Ignoring the dependencies between the two equations and communicating the two equation indices (of rates $R_{\max} = \max\{R_1, R_2\}$ each) to the destination as if they were independent messages yields the rate region:

$$\begin{aligned}&\mathcal{R}_{\text{Naive}}(R_1, R_2)\\ &= \Big\{(R_1, R_2) :\\ &\quad R_{\max} \le \min\{I(X_1; Y|X_2), I(X_2; Y|X_1)\}\\ &\quad R_{\max} + R_{\max} \le I(X_1, X_2; Y)\\ &\quad \text{for } p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2)\Big\}.\end{aligned} \tag{4}$$

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

2) 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 $R_1 \neq R_2$.

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 [1], which resulted in the rate region:

$$\begin{aligned}&\mathcal{R}_{\text{CB}}(R_1, R_2)\\ &= \Big\{(R_1, R_2) :\\ &\quad R_{\min} \le \min\{I(X_1; Y|X_2), I(X_2; Y|X_1)\}\\ &\quad R_1 + R_2 \le I(X_1, X_2; Y)\\ &\quad \text{for } p(x_1, x_2, y) = p(x_1)p(x_2)p(y|x_1, x_2)\Big\}.\end{aligned} \tag{5}$$

The region $\mathcal{R}_{\text{CB}}(R_1, R_2)$ improves over $\mathcal{R}_{\text{Naive}}(R_1, R_2)$ 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 $\mathcal{R}_{\text{CB}}(R_1, R_2)$ 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.

3) 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 [9] (which in turn may be seen as Special case d) of joint-source-channel coding over a MAC as studied in [14]) of a two-user MAC with a common message and two private messages. This idea is first expressed in [1], but was not fully explored, and yields the region:

$$\begin{aligned}&\mathcal{R}_{\text{ICF}}(R_1, R_2)\\ &= \Big\{(R_1, R_2):\\ &\quad R_{\min} \le \min\Big\{I(X_1; Y|X_2, Q), I(X_2; Y|X_1, Q),\\ &\quad \frac{1}{2}I(X_1, X_2; Y|Q)\Big\}\\ &\quad R_1 + R_2 \le I(X_1, X_2; Y)\\ &\quad \text{for } p(q, x_1, x_2, y) = p(q)p(x_1|q)p(x_2|q)p(y|x_1, x_2)\Big\}.\end{aligned} \tag{6}$$

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

Remark 14: 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 $\{p(x_1)p(x_2)p(y|x_1, x_2)\}$.

B. Numerical Comparison

Consider the AWGN channel model in Fig. 2 with $g_{13} = g_{23} = g_{24} = 1$, $g_{14} = -1$ in the first hop, and symmetrize the powers as $P_s = P_1 = P_2$. Note that one can easily obtain regions for general g_{ij} and power constraints, but that this is not the focus of this work.

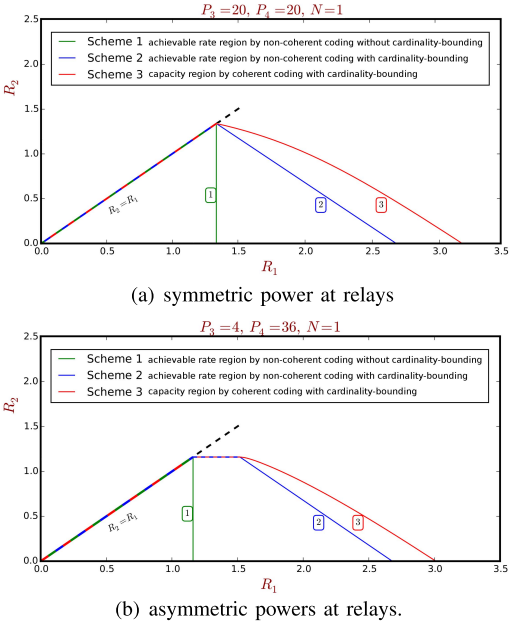
1) Numerical Comparison of Three Two-User ICF Only Rate Regions: We now numerically evaluate the three achievable rate regions of Schemes 1, 2, and 3 for the ICF hop only of an additive Gaussian noise channel as shown in Fig. 2, where we recall that all noises are i.i.d. unit variance Gaussians, i.e.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

$$\begin{aligned}&\mathcal{R}^{\mathcal{G}}_{\text{Naive}}(R_1, R_2)\\ &= \Big\{(R_1, R_2) : R_{\max} \le \min\{C(P_3), C(P_4), \frac{1}{2}\cdot C(P_3+P_4)\}\Big\},\end{aligned} \tag{7}$$

$$\begin{aligned}&\mathcal{R}^{\mathcal{G}}_{\text{CB}}(R_1, R_2)\\ &= \left\{(R_1, R_2) : \begin{array}{l} R_{\min} \le \min\{C(P_3), C(P_4)\}\\ R_1 + R_2 \le C(P_3 + P_4)\end{array}\right\}.\end{aligned} \tag{8}$$

Scheme 3 has been shown to be exhausted by jointly Gaussian inputs [10], 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

$$\begin{aligned}&\mathcal{R}^{\mathcal{G}}_{\text{ICF}}(R_1, R_2 \,|\, b_1, b_2)\\ &= \Big\{(R_1, R_2) : R_{\min} \le \min\Big\{C((1-b_1)P_3), C((1-b_2)P_4),\\ &\quad \frac{1}{2}C((1-b_1)P_3 + (1-b_2)P_4)\Big\}\\ &\quad R_1 + R_2 \le C(P_3 + P_4 + 2\sqrt{b_1 b_2}\sqrt{P_3 P_4})\Big\}.\end{aligned} \tag{9}$$



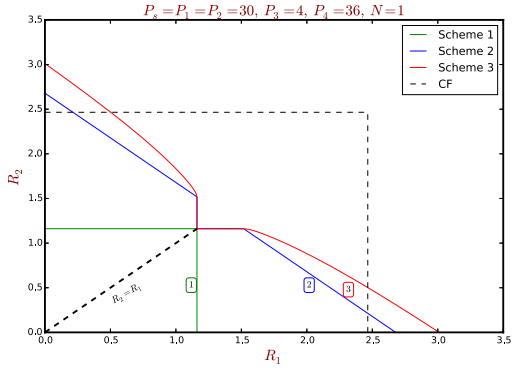
(a) symmetric power at relays

(b) asymmetric powers at relays.

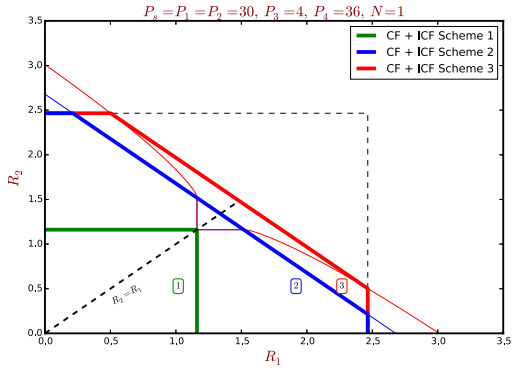
Fig. 4. Numerical evaluation for two-user Gaussian-MAC ICF problem. In (a) $P_3 = P_4 = 20$, and in (b) $P_3 = 4$, $P_4 = 36$. Only the $R_1 \ge R_2$ scenario is plotted.

Fig. 4(a) demonstrates the relative rate regions of the three schemes for equal relay power $P_3 = P_4 = 20$, while Fig. 4(b) demonstrates the regions for asymmetric powers $P_3 = 4$, $P_4 = 36$. From the figure, we 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 $\sqrt{P_3 P_4}$. For fixed sum $P_3 + P_4$ this is maximized when they are equal.

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.



(a) asymmetric powers at relays + combining CF and ICF



(b) the convex hull of the intersection region.

Fig. 5. 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 19. (a) also contains the first CF hop explained in equation (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 $R_2 = R_1$.

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 [3]. Next, in the ICF stage, destination node 5 recovers $(\mathbf{w}_1, \mathbf{w}_2)$ from $(\mathbf{u}_1, \mathbf{u}_2)$ at rates:

$$\text{First hop: } \begin{cases} R_1 \leq \frac{1}{2}\log\left(\frac{1}{2} + P_s\right) \\ R_2 \leq \frac{1}{2}\log\left(\frac{1}{2} + P_s\right), \end{cases}$$
$$\text{Second hop: region (9).} \tag{10}$$

To obtain an achievable rate region for the entire network, first intersect the CF and ICF rate regions in (10) and then take the convex hull of the resulting regions. As we can see in Fig. 5(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.

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:

$$\text{First hop: } \begin{cases} R_1 \leq \frac{1}{2}\log\left(1 + P_s\right) \\ R_2 \leq \frac{1}{2}\log\left(1 + P_s\right) \\ R_1 + R_2 \leq \frac{1}{2}\log\left(1 + 2P_s\right) \end{cases}$$
$$\text{Second hop: } R_1 + R_2 \leq \frac{1}{2}\log\left(1 + P_3 + P_4 + 2\sqrt{P_3 P_4}\right). \tag{11}$$

As we can see from the expressions in equation (10) and (11), the extra sum rate constraint, which is due to treating the first hop as two MAC channels in the DF stage, could potentially[3] render DF+FCo inferior to CF+ICF. This is confirmed by the simulations shown in Fig. 6. 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. 6, where we see that ignoring the correlations between the equations (ICF Scheme 1) could reduce the gains significantly. Thus, a proper ICF 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. 6(b), the gain of CF+ICF over DF+FCo can be substantial.

Remark 15: 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 paper.*

IV. 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. 7, 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

[3]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.

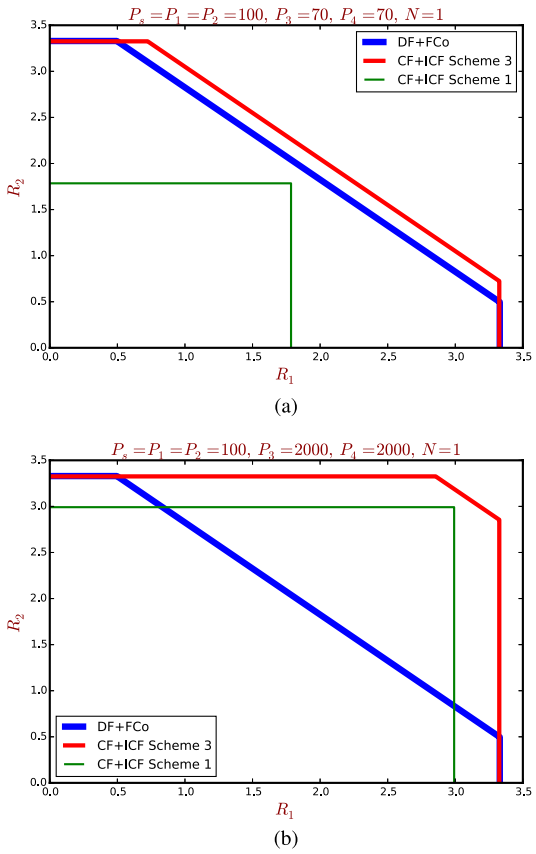


Fig. 6. Examples of CF+ICF outperforming DF+FCo.

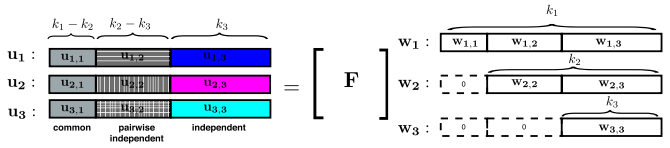


Fig. 7. Three user ICF message/equation structure. The grey color indicates that these equation sections ($\mathbf{u}_{*,1}$) are fully correlated; shading indicates that these three equation sections ($\mathbf{u}_{*,2}$) are pairwise independent, while different solid colors indicate that these three equation sections ($\mathbf{u}_{*,3}$) are mutually independent. All message sections $\mathbf{w}_{i,j}$ are mutually independent.

is $\mathbf{w}_{l,c}$. Following the notation of Section II: $\left[\tilde{\mathbf{U}}_{*1}\ \tilde{\mathbf{U}}_{*2}\ \tilde{\mathbf{U}}_{*3}\right] = \left(\mathbf{F}\right)\cdot\left[\tilde{\mathbf{W}}_{*1}\ \tilde{\mathbf{W}}_{*2}\ \tilde{\mathbf{W}}_{*3}\right]$, or, breaking this into *message sections* and *equation sections*, as shown in Fig. 7.

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 16: Our problem cannot be mapped into the framework in [12], which considered an extension of the two-user Slepian-Wolf MAC to an L-user MAC in which each transmitter has access to an arbitrary subset of messages from a set of independent messages. In [12], 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 explicitly capture the pairwise independent structure, but do note that the generality of Cover's problem formulation for the MAC with arbitrarily correlated sources [14] can capture the special dependence structure seen here. However, while [14] finds a 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, a computable expression is currently unknown. We will next show a simple achievability scheme for our specific problem, which turns out to be the explicitly computable capacity region in the Gaussian case.*

Theorem 17 (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*

$$\mathcal{R}_{IN} := \bigcup_{p(q)p(x_1|q)p(x_2|q)p(x_3|q)} \mathcal{R} \tag{12}$$

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* $(R_1 \geq R_2 \geq R_3)$:

$$R_1 + R_2 + R_3 \leq I(X_1, X_2, X_3; Y) \tag{13a}$$
$$2R_2 + R_3 \leq I(X_1, X_2, X_3; Y|Q) \tag{13b}$$
$$R_2 + R_3 \leq \min\{I(X_1, X_2; Y|X_3, Q), \tag{13c}$$
$$I(X_1, X_3; Y|X_2, Q), \tag{13d}$$
$$I(X_2, X_3; Y|X_1, Q)\} \tag{13e}$$
$$R_3 \leq \min\{I(X_1; Y|X_2, X_3, Q), \tag{13f}$$
$$I(X_2; Y|X_1, X_3, Q), \tag{13g}$$
$$I(X_3; Y|X_1, X_2, Q)\}. \tag{13h}$$

Remark 18: To understand the form, consider for example (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,* (13e) *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,* (13h) *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: (13a) corresponds to the overall sum rate constraint and (13b) corresponds to the sum constraint apart from the cooperative or common message of rate $R_1 - R_2$ (see Fig. 7). 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 (13e). 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 21 for general L.

Remark 19: 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.

V. 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 VI-A is argued via section rates ρ_c (Definition 12, Section II). 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 $\rho_c = R_c - R_{c+1}$, $R_{L+1} = 0$.

A. An ICF Achievable Rate Region for the Memoryless ICF Channel

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

Theorem 20 (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:

$$\sum_{l=1}^{L} R_l \leq I(X_1, \cdots, X_L; Y) \tag{14a}$$

$$2R_2 + \sum_{l=3}^{L} R_l \leq I(X_1, \cdots, X_L; Y|Q) \tag{14b}$$

$$\sum_{l=\nu+1}^{L} R_l \leq I(X_{A^C}; Y|X_A, Q) \text{ for } \nu = 1, 2, \cdots, L-1 \tag{14c}$$

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 (6) and the three-user achievability scheme in Theorem 17 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 (14), compared to the $2^L - 1$ in a classical MAC.

We may interpret (14c) as follows. Take for example $L = 5$, $\nu = 2$, $A = \{2, 3\}$ and $A^C = \{1, 4, 5\}$. Then (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 VI-A. 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.

B. 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 our 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 21 (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* (1) *if and only if the message rates* R_l *satisfy:*

$$\begin{cases} \sum_{l=1}^{L} R_l \le \frac{1}{2}\log_2\left(1+\sum_{j=0}^{L} d_j^2\right) \\ 2R_2 + \sum_{l=3}^{L} R_l \le \frac{1}{2}\log_2\left(1+\sum_{j=1}^{L} d_j^2\right) \\ \sum_{l=\nu+1}^{L} R_l \le \frac{1}{2}\log_2\left(1+\sum_{j\in A^C} d_j^2\right) \end{cases} \tag{15}$$

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* $0 \le b_j \le P_j$, *for* $j = 1, \cdots, L$.

Proof: Achievability: Achievability follows directly from Theorem 20 by selecting input distributions $p(q)$, and every $p(x_m|q)$ to be Gaussian as follows:

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 Gaussian codewords:

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

Then, at each channel use,

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

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 20, we obtain the achievable rate region specified by inequalities (15).

Converse: The converse uses Lemmas 22, 23 and 24 to upper bound the capacity region as follows

$$\mathcal{C} \overset{\text{Lemma 22}}{\subseteq} \mathcal{R}_{\text{out}} \overset{\text{Lemma 23}}{\subseteq} \bigcup \mathcal{R}' \overset{\text{Lemma 24}}{\subseteq} \bigcup \mathcal{R}''.$$

We first state the lemmas, explain the intuition and show how they are used to establish the converse. We defer the proofs of Lemmas 22 and 23 to the following subsections, while the proof of Lemma 24 is inline.

First, Lemma 22 provides an outer bound $\mathcal{R}_{\text{out}}$ valid for any memoryless channel. Define

$$\mathcal{P} := \{p(q, x_1, \cdots, x_L) : X_m \to Q \to X_{m'}, \quad \forall m \ne m',\ m, m' \in \{1, 2, \cdots, L\}\} \tag{16}$$

Lemma 22: $\mathcal{C} \subseteq \mathcal{R}_{out}$, *where* $\mathcal{R}_{out}$ *is defined as*

$$\mathcal{R}_{out} := \bigcup_{p(q,x_1,\cdots,x_L)\in\mathcal{P}} \mathcal{R}(Q, X_1, \cdots, X_L), \tag{17}$$

where $\mathcal{R}(Q, X_1, \cdots, X_L)$ *denotes the set of rate tuples* $(R_1, \cdots, R_L)$ *that satisfy inequalities* (14).

Lemma 23 further loosens the outer bound $\mathcal{R}_{\text{out}}$ in Lemma 22 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 V-D 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 23: 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:*

$$\begin{cases} \sum_{l=1}^{L} R_l \le C(\sum_{m=1}^{L} E[X_m^2] + \sum_{m\ne m'} E[X_m X_{m'}]) \\ 2R_2 + \sum_{l=3}^{L} R_l \le C(\sum_{m=1}^{L} var[X_m|Q]) \\ \sum_{l=\nu+1}^{L}(l-\nu)(R_l - R_{l+1}) \le C(\sum_{m\in A^C} var[X_m|Q]) \end{cases} \tag{18}$$

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 24. This lemma is based on the power constraints and the Markov chains $X_m \to Q \to X_{m'}$, $\forall m \ne m'$, $m, m' \in \{1, 2, \cdots, L\}$. To show Lemma 24, note that it follows from [10, Lemma B.3] that $E[X_m X_{m'}] \le \sqrt{E[X_m^2] - \text{var}(X_m|Q)}\sqrt{E[X_{m'}^2] - \text{var}(X_{m'}|Q)}$. 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 24: The region $\mathcal{R}' \subseteq \mathcal{R}''$, *where* $\mathcal{R}''$ *consists of the rate tuples that satisfy*

$$\begin{cases} \sum_{l=1}^{L} R_l \le C(\sum_{m=1}^{L} E[X_m^2] \\ \qquad + \sum_{m\ne m'} \sqrt{t_m t_{m'}}\sqrt{E[X_m^2]E[X_{m'}^2]}) \\ 2R_2 + \sum_{l=3}^{L} R_l \le C(\sum_{m=1}^{L}(1-t_m)E[X_m^2]) \\ \sum_{l=\nu+1}^{L} R_l \le C(\sum_{m\in A^C}(1-t_m)E[X_m^2]) \end{cases} \tag{19}$$

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

Combining Lemma 22, Lemma 23 and Lemma 24, 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 21 with b_j replaced by $t_j P_j$ – i.e. may be achieved by jointly Gaussian inputs which are conditionally independent given Gaussian $p(q)$. ■

C. Proof of Lemma 22

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{(a1)}{=} 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 \qquad (20) \\
n(2R_2 + \sum_{l=3}^{L} R_l) & \\
&= n\sum_{c=2}^{L} c\rho_c \\
&\overset{(a2)}{=} H([\tilde{\mathbf{U}}_{*2}, \tilde{\mathbf{U}}_{*3}, \cdots, \tilde{\mathbf{U}}_{*L}]) \overset{(a3)}{=} 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) \\
&\quad + n\epsilon_n \qquad (Q_i := \tilde{\mathbf{U}}_{*1}) \\
&\overset{(e)}{\leq} nI(X_1, \cdots, X_L; Y|Q) + n\epsilon_n \qquad (21) \\
n(\sum_{l=v+1}^{L} R_l) &= n(\sum_{l=v+1}^{L} (l-v)(R_l - R_{l+1})) \\
&= n\sum_{c=v+1}^{L} (c-v)\rho_c \\
&\overset{(a4)}{=} H(\mathbf{U}|\mathbf{u}_A) \overset{(a5)}{=} 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 \\
&\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 \\
&\overset{(d)}{=} \sum_{i=1}^{n} I(X_{A^C i}; Y_i|Q_i, X_{Ai}) \\
&\quad + n\epsilon_n \qquad (Q_i := \tilde{\mathbf{U}}_{*1}) \\
&\overset{(e)}{\leq} nI(X_{A^C}; Y|Q, X_A) + n\epsilon_n \qquad (22)
\end{aligned}$$

The equalities in (a) all follow by definitions and the linear algebraic arguments in Lemma 27 in the Appendix. Equalities $(a4)$, $(a5)$ follow from $H(\mathbf{U}|\mathbf{u}_A) = H(\mathbf{u}_A, \mathbf{u}_{A^C}|\mathbf{u}_A) = H(\mathbf{u}_{A^C}|\mathbf{u}_A) = H(\mathbf{u}_{A^C}|\tilde{\mathbf{U}}_{*1},\ \mathbf{u}_A)$. 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 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 (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. ■

D. Proof of Lemma 23

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

$$\begin{aligned}
&I(X_{A^C}; Y|X_A, Q) \\
&\quad = E_Q[I(X_{A^C}; Y|X_A, Q = q)] \\
&\quad \overset{(a)}{=} E_Q[h(\sum_{m\in A^C} X_m + Z|Q = q) - h(Z)] \\
&\quad \overset{(b)}{\leq} E_Q\left[\frac{1}{2}\log\left(\frac{\text{var}(\sum_{m\in A^C} X_m + Z|Q = q)}{\text{var}(Z)}\right)\right] \\
&\quad \overset{(c)}{=} E_Q\left[\frac{1}{2}\log\left(1 + \sum_{m\in A^C} \text{var}(X_m|Q = q)\right)\right] \\
&\quad \overset{(d)}{\leq} \frac{1}{2}\log\left(1 + \sum_{m\in A^C} \text{var}(X_m|Q)\right), \qquad (23)
\end{aligned}$$

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

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

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

E. On the Assumptions Placed on $\mathbf{F}$

As commented in Remark 3, the assumption that $\mathbf{F}$ 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 $\mathbf{F}$. We provide examples of how to proceed in this direction for $L=2$ and 3 next. We note that $\mathbf{F}$ 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.

1) 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:

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

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 III fails. The achievable rate region shrinks to

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

When $\mathbf{F}=\begin{bmatrix}1&0\\0&1\end{bmatrix}$, following similar arguments, one can check that the region (5) should be modified to

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

We omit the other cases for brevity. This is an example of how, in contrast to [2], we do not require all square sub-matrices of $\mathbf{F}$ to be full rank. Nevertheless, the format of the rate region varies.

2) 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 VI-A, 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 $\mathbf{F}$ is invertible but sub-matrix $\begin{bmatrix}1&1\\1&1\end{bmatrix}$ is singular[4];
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 VI-A for the discrete memoryless MAC still holds in this case. The analyses of error events related to $\mathcal{E}_0,\mathcal{E}_1,\mathcal{E}_{2,0}$ remain valid while the analysis of for example $\mathcal{E}_{2,A}$ for the set $A=\{1,2\}$ must be altered. In particular, when $A=\{1,2\}$, Lemma 27 would yield $||\mathcal{U}_{2,A}||\le 2^{nR_3}$ rather than what it should be, which is $||\mathcal{U}_{2,A}||\le 2^{nR_2}$. When all 2 by 2 submatrices from the first two columns are nonsingular, given two equation values, there will be no uncertainty about the second message sections, i.e. $\mathbf{w}_{1,2}$ and $\mathbf{w}_{2,2}$ (note that $\mathbf{w}_{3,2}$ is null.) But because of the singularity of sub-matrix $\begin{bmatrix}1&1\\1&1\end{bmatrix}$, knowing $\mathbf{u}_1$ and $\mathbf{u}_2$ does not fully resolve the second message sections but leaves one degree of freedom. Note that there is always one degree of freedom among the third message sections, so we have $||\mathcal{U}_{2,A}||=2^{n\rho_2}\cdot 2^{n\rho_3}=2^{nR_2}$ instead of $2^{n\rho_3}=2^{nR_3}$. In summary, the achievable rate region for this particular choice of $\mathbf{F}$ would lead to the same region as in (13) except for the third term in inequality (13h) which becomes the new $R_2\le I(X_3;Y|X_1,X_2,Q)$.

Remark 25: Note that if two rows are exchanged in matrix $\mathbf{F}$*, say the 2nd and 3rd rows, then inequality* $R_3\le I(X_2;Y|X_1,X_3,Q)$ *in region* (13) *will be replaced by* $R_2\le I(X_2;Y|X_1,X_3,Q)$*. Thus, we note that the assumption that all* $c\times c$ *sub-matrices of the first* c *columns of* $\mathbf{F}$ *be*

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

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.

F. On the Generalization of Our 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.

1) Abstract Gaussian ICF Model: Consider again an L-user Gaussian channel model as in (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 $\rho_3, \cdots$, 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 22 in Subsection V-C. 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 22 are ensured. Thus, Lemma 23 and Lemma 24 may be derived, and the converse for the Gaussian channel follows.

The remainder of the necessary definitions follow by extension of those in Section II. Then the next Corollary is easy to obtain from the proof of Theorem 21.

Corollary 26: The capacity region of Theorem 21 is the capacity region for the Abstract Gaussian ICF model described above, with the convention that $\rho_c = R_c - R_{c+1}$ and $R_{L+1} = 0$.

VI. Conclusion

We consider an L-user multiple access channel where transmitter m has access to the linear equation $\mathbf{u}_m = \bigoplus_{l=1}^{L} f_{ml}\mathbf{w}_l$ of independent messages $\mathbf{w}_l \in \mathbb{F}_p^{kl}$ 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 [1] 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.

Appendix

Recall that notation $\tilde{\mathbf{W}}_{L\times L}$ and $\mathbf{W}_{L\times k}$ both refer to the same underlying message matrix and only differ in the indexing of its columns; similarly for the notation $\tilde{\mathbf{U}}_{L\times L}$ and $\mathbf{U}_{L\times k}$.

A. Proof of Theorem 20

Proof: 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 wants to decode the underlying set of messages, i.e. $\mathbf{W}_{L\times k}$, and can do so by decoding and inverting the corresponding set of message equations, i.e. $\mathbf{U}_{L\times k}$ or $\tilde{\mathbf{U}}_{L\times L}$, because the coefficient matrix $\mathbf{F}$ and all c by c submatrices from its first c columns are of full rank.

For a given coefficient matrix $\mathbf{F}$,[5] each *set* of messages $\{\mathbf{w}_1, \mathbf{w}_2, \cdots \mathbf{w}_L\}$ uniquely define a *set* of message equations

[5]Recall that in the ICF problem the coefficient matrix $\mathbf{F}$ is revealed to the destination node before the communication starts.

$\{\mathbf{u}_1, \mathbf{u}_2, \cdots \mathbf{u}_L\}$. When the message rates $R_1, \cdots, R_L$ or equivalently the equation rates $\rho_1, \cdots, \rho_L$ are fixed, there are only $2^{n\cdot(R_1+\cdots+R_L)}$ possible *sets* of message equations. Let $\mathcal{U}$ be the collection of sets of message equations that satisfy $\mathbf{U} = \mathbf{F}\cdot\mathbf{W}$.

The decoder enumerates all possible sets of message equations in $\mathcal{U}$[6] and looks for a unique equation set $(\mathbf{u}_1, \cdots, \mathbf{u}_L)$ 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 $A_\epsilon^{(n)}(Q, X_1, \cdots, X_L, Y)$. If none, or more than one set of equation sections are jointly typical with the given Y^n, the decoder sets the estimated $\hat{\tilde{\mathbf{U}}}_{L\times L}$ to null and declares an error.

Error analysis: For a given codebook $\mathcal{C}$, we are interested in the averaged probability of error $\bar{P}_e^n(\mathcal{C})$ when this particular codebook is adopted:

$$\begin{aligned}&\bar{P}_e^{(n)}(\mathcal{C})\\ &:= \frac{1}{2^{n\cdot(R_1+\cdots+R_L)}} \sum_{\{\mathbf{u}_1,\mathbf{u}_2,\cdots\mathbf{u}_L\}\in\mathcal{U}}\\ &\quad\times \Pr\left(\hat{\tilde{\mathbf{U}}}_{L\times L} \neq \tilde{\mathbf{U}}_{L\times L} \mid \tilde{\mathbf{U}}_{L\times L} \text{ is sent, codebook} = \mathcal{C}\right)\end{aligned} \tag{24}$$

where $\tilde{\mathbf{U}}_{L\times L}$ is the segmented message equation matrix representation of the transmitted set of equations.

Similar to Shannon's random coding argument in the point-to-point channel, we are not directly computing the averaged probability of error for a particular codebook. We compute the expected probability of error $P_e^{(n)}$ with respect to all possible codebooks generated through the mechanism described earlier:

$$\begin{aligned}P_e^{(n)} :=& \sum_{\text{codebook } \mathcal{C} \in \left\{\substack{\text{codebooks generated through the codebook}\\ \text{generation mechanism}}\right\}}\\ &\times \Pr[\text{codebook} = \mathcal{C}]\cdot \bar{P}_e^{(n)}(\mathcal{C}).\end{aligned} \tag{25}$$

If $P_e^{(n)}$ can be made arbitrarily small then there must exist at least one codebook whose averaged probability of error $\bar{P}_e^{(n)}(\mathcal{C})$ goes to zero when the number of channel uses goes to infinity.

Combining equations (24) and (25) and noting the symmetric role of all possible equation sets in the communication scheme, it follows that

$$\begin{aligned}P_e^{(n)} &= \frac{1}{2^{n\cdot(R_1+\cdots+R_L)}} \sum_{\{\mathbf{u}_1,\mathbf{u}_2,\cdots\mathbf{u}_L\}\in\mathcal{U}}\\ &\quad\times \Pr\left(\hat{\tilde{\mathbf{U}}}_{L\times L} \neq \tilde{\mathbf{U}}_{L\times L} \mid \tilde{\mathbf{U}}_{L\times L} \text{ is sent}\right)\\ &= \Pr\left(\hat{\tilde{\mathbf{U}}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L} \mid \tilde{\mathbf{U}}^0_{L\times L} \text{ is sent}\right),\end{aligned} \tag{26}$$

where without loss of generality, given the symmetry, a particular set of message equations

$$\begin{aligned}\tilde{\mathbf{U}}^0_{L\times L} &= [\mathbf{u}_1^0, \mathbf{u}_2^0, \cdots \mathbf{u}_L^0]^T := [\tilde{\mathbf{U}}^0_{*1}, \tilde{\mathbf{U}}^0_{*2}, \cdots, \tilde{\mathbf{U}}^0_{*L}]\\ &= \mathbf{F}\cdot[\tilde{\mathbf{W}}^0_{*1}, \tilde{\mathbf{W}}^0_{*2}, \cdots, \tilde{\mathbf{W}}^0_{*L}]\end{aligned} \tag{27}$$

may be assumed to have been sent. Let $X_m^n(\mathbf{u}_m^0) := (x_{m,1}(\mathbf{u}_m^0), x_{m,2}(\mathbf{u}_m^0), \cdots, x_{m,n}(\mathbf{u}_m^0))$ be the codeword sent by relay m $(m = 1, 2, \cdots L)$ according to our encoder for this particular set of equations $\tilde{\mathbf{U}}^0_{L\times L}$. The receiver receives output Y^n randomly generated according to the channel $p(y^n|x_1^n, \cdots x_L^n) = \prod_{i=1}^n p(y_i|x_{1,i}(\mathbf{u}_1^0), x_{2,i}(\mathbf{u}_2^0), \cdots x_{L,i}(\mathbf{u}_L^0))$.

The decoder outputs either a null value or a valid estimate $\tilde{\mathbf{U}}_{L\times L}$, which may or may not be the true one sent. Define the random event $E\left(\tilde{\mathbf{U}}_{L\times L}\right)$ indexed by a message equation set as:

$$\begin{aligned}E\left(\tilde{\mathbf{U}}_{L\times L}\right) := \Big\{&\left(Q^n(\tilde{\mathbf{U}}_{*1}), X_1^n(\mathbf{u}_1), \cdots, X_L^n(\mathbf{u}_L), Y^n\right)\\ &\in A_\epsilon^{(n)} \mid \tilde{\mathbf{U}}^0_{L\times L} \text{ sent}\Big\}, \quad \text{where } \tilde{\mathbf{U}}_{L\times L} \in \mathcal{U}.\end{aligned}$$

Further define the event

$$\begin{aligned}E^c\left(\tilde{\mathbf{U}}_{L\times L}\right) := \Big\{&\left(Q^n(\tilde{\mathbf{U}}_{*1}), X_1^n(\mathbf{u}_1), \cdots, X_L^n(\mathbf{u}_L), Y^n\right)\\ &\notin A_\epsilon^{(n)} \mid \tilde{\mathbf{U}}^0_{L\times L} \text{ sent}\Big\}, \quad \text{where } \tilde{\mathbf{U}}_{L\times L} \in \mathcal{U}.\end{aligned}$$

Notice that both definitions include conditioning on the message equations $\tilde{\mathbf{U}}^0_{L\times L}$ being sent. A precise account of the possible erroneous scenarios that are disjoint would be as follows, where we remind the reader that here the E and E^c notation is meant to denote an event rather than an expectation:

1) None of the events $E\left(\tilde{\mathbf{U}}_{L\times L}\right)$ happen, i.e. $\left[\cap_{\tilde{\mathbf{U}}_{L\times L}\in\mathcal{U}} E^c(\tilde{\mathbf{U}}_{L\times L})\right]$.
2) More than one event of the type $E\left(\tilde{\mathbf{U}}_{L\times L}\right)$ happens.
3) Only one event happens, but it is not $E\left(\tilde{\mathbf{U}}^0_{L\times L}\right)$. That is $\left[E(\tilde{\mathbf{U}}_{L\times L}) \cap \left(\bigcap_{\tilde{\mathbf{U}}'_{L\times L}\neq\tilde{\mathbf{U}}_{L\times L}} E^c(\tilde{\mathbf{U}}'_{L\times L})\right)\right]$, for any $\tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L}$.

However, to simplify, it can be checked that the union of these three scenarios is a subset of the event $\left[E^c(\tilde{\mathbf{U}}^0_{L\times L}) \cup \left(\bigcup_{\tilde{\mathbf{U}}_{L\times L}\neq\tilde{\mathbf{U}}^0_{L\times L}} E(\tilde{\mathbf{U}}_{L\times L})\right)\right]$, i.e. either the true event does not happen or one or more of the false events happens. Thus, equation (26) can be upper bounded by:

$$\begin{aligned}P_e^{(n)} &= \Pr\left(\hat{\tilde{\mathbf{U}}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L} \mid \tilde{\mathbf{U}}^0_{L\times L} \text{ is sent}\right)\\ &\leq \Pr\left(E^c(\tilde{\mathbf{U}}^0_{L\times L}) \cup \left[\bigcup_{\tilde{\mathbf{U}}_{L\times L}\neq\tilde{\mathbf{U}}^0_{L\times L}} E(\tilde{\mathbf{U}}_{L\times L})\right]\right)\\ &\leq \Pr\left(E^c(\tilde{\mathbf{U}}^0_{L\times L})\right) + \Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L}\neq\tilde{\mathbf{U}}^0_{L\times L}} E(\tilde{\mathbf{U}}_{L\times L})\right)\end{aligned} \tag{28}$$

where the second inequality comes from the union bound.

The first term in (28) corresponds to the probability that the true transmitted equation set does not pass the decoding test, which vanishes by properties of the jointly typical set $A_\epsilon^{(n)}$ using standard arguments.

The core of the achievability proof lies in properly bounding the second term: it has to be carried out in the ICF context, i.e. taking into account the given pairwise independent but not mutually independent structure of the equations known at the different transmitters, which is more involved than in a MAC channel with independent messages.

[6]We note that in conference versions of this work [1], [2] the error analysis assumed the decoder produced elements in $\mathcal{U}$ without enforcing this condition in the decoder; this is now fixed in the current decoder.

We now further subdivide the set $\{\tilde{\mathbf{U}}_{L\times L} : \tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L}\}$ into orthogonal parts: $\mathcal{U}_1$, $\mathcal{U}_{2,0}$ and $\mathcal{U}_{2,A}$ as indicated below. Such subdivisions allow for a simple bounding of their cardinalities and error event probabilities, as shown in Lemma 27 and Lemma 28.

Define

$$\begin{aligned}\mathcal{U}_1 &:= \{\tilde{\mathbf{U}}_{L\times L} \in \mathcal{U} : \tilde{\mathbf{U}}_{*1} \neq \tilde{\mathbf{U}}^0_{*1}\}\\ &= \{\tilde{\mathbf{U}}_{L\times L} \in \mathcal{U} : \text{wrong 1st equation section}\\ &\qquad \text{(common equation section)}\} \qquad (29)\\ \mathcal{U}_{2,0} &:= \{\tilde{\mathbf{U}}_{L\times L} \in \mathcal{U} : \tilde{\mathbf{U}}_{*1} = \tilde{\mathbf{U}}^0_{*1}, \mathbf{u}_i \neq \mathbf{u}^0_i,\\ &\qquad \forall i \in \{1, 2, \cdots L\}\}\\ &= \{\tilde{\mathbf{U}}_{L\times L} \in \mathcal{U} : \text{correct 1st equation section,}\\ &\qquad \text{but wrong overall equations for all relays}\} \qquad (30)\end{aligned}$$

For some $A \subset \{1, 2, \cdots L\}, 0 < ||A|| < L$,

$$\begin{aligned}\mathcal{U}_{2,A} &:= \{\tilde{\mathbf{U}}_{L\times L} \in \mathcal{U} : \tilde{\mathbf{U}}_{*1} = \tilde{\mathbf{U}}^0_{*1}, \mathbf{u}_a = \mathbf{u}^0_a, \forall a \in A\}\\ &= \{\tilde{\mathbf{U}}_{L\times L} \in \mathcal{U} : \text{correct 1st equation section}\\ &\qquad \text{and } ||A|| \text{ right equations, indexed by set } A\} \qquad (31)\end{aligned}$$

Then

$$\{\tilde{\mathbf{U}}_{L\times L} : \tilde{\mathbf{U}}_{L\times L} \neq \tilde{\mathbf{U}}^0_{L\times L}\} = \mathcal{U}_1 \cup \mathcal{U}_{2,0} \cup \left(\bigcup_{A\subset\{1,2,\cdots L\}, 0<||A||<L} \mathcal{U}_{2,A}\right),$$

and hence

$$\begin{aligned}&\Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L}\neq\tilde{\mathbf{U}}^0_{L\times L}} E(\tilde{\mathbf{U}}_{L\times L})\right)\\ &\leq \Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L}\in\mathcal{U}_1} E(\tilde{\mathbf{U}}_{L\times L})\right) + \Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L}\in\mathcal{U}_{2,0}} E(\tilde{\mathbf{U}}_{L\times L})\right)\\ &\quad + \sum_{A\subset\{1,2,\cdots L\}, 0<||A||<L} \Pr\left(\bigcup_{\tilde{\mathbf{U}}_{L\times L}\in\mathcal{U}_{2,A}} E(\tilde{\mathbf{U}}_{L\times L}).\right) \qquad (32)\end{aligned}$$

Further define the following events, for $A \subset \{1, 2, \cdots L\}, 0 < ||A|| < L$,

$$\begin{aligned}\mathcal{E}_0 &:= E(\tilde{\mathbf{U}}^0_{L\times L}), \quad \mathcal{E}^c_0 = E^c(\tilde{\mathbf{U}}^0_{L\times L}) \qquad (33)\\ \mathcal{E}_1 &:= \left\{\tilde{\mathbf{U}}_{L\times L} \in \mathcal{U}_1 \cap E(\tilde{\mathbf{U}}_{L\times L})\right\} \qquad (34)\\ \mathcal{E}_{2,0} &:= \left\{\tilde{\mathbf{U}}_{L\times L} \in \mathcal{U}_{2,0} \cap E(\tilde{\mathbf{U}}_{L\times L})\right\} \qquad (35)\\ \mathcal{E}_{2,A} &:= \left\{\tilde{\mathbf{U}}_{L\times L} \in \mathcal{U}_{2,A} \cap E(\tilde{\mathbf{U}}_{L\times L})\right\}. \qquad (36)\end{aligned}$$

Then,

$$\begin{aligned}P_e^{(n)} &\leq \Pr\left(\mathcal{E}^c_0\right) + ||\mathcal{U}_1|| \cdot \Pr\left(\mathcal{E}_1\right) + ||\mathcal{U}_{2,0}|| \cdot \Pr\left(\mathcal{E}_{2,0}\right)\\ &\quad + \sum_{A\subset\{1,2,\cdots L\}, 0<||A||<L} ||\mathcal{U}_{2,A}|| \cdot \Pr\left(\mathcal{E}_{2,A}\right) \qquad (37)\end{aligned}$$

In Lemma 27, we upper bound the cardinalities of sets $\mathcal{U}_1$, $\mathcal{U}_{2,0}$ and $\mathcal{U}_{2,A}$. In Lemma 28, we upper bound the probability items $\Pr(\mathcal{E}_1)$, $\Pr(\mathcal{E}_{2,0})$ and $\Pr(\mathcal{E}_{2,A})$. We prove these in the next two sections.

Lemma 27 (Cardinality Lemma for Error Events): Assume the matrix $\mathbf{F}$ *and all* c *by* c *sub-matrices from its first* c *columns are of full rank, for every* $c = 1, 2, \cdots L$. *The cardinalities of the different sets* $\mathcal{U}_1$, $\mathcal{U}_{2,0}$, $\mathcal{U}_{2,A}$ *for* $A \subset \{1, 2, \cdots L\}, 0 < ||A|| < L$ *may be upper bounded by:*

$$||\mathcal{U}_1|| \leq 2^{n\sum_{c=1}^{L} c\rho_c} = 2^{n\sum_{l=1}^{L} R_l} \qquad (38)$$

$$||\mathcal{U}_{2,0}|| \leq 2^{n\sum_{c=2}^{L} c\rho_c} = 2^{n(2R_2+\sum_{l=3}^{L} R_l)} \qquad (39)$$

$$||\mathcal{U}_{2,A}|| \leq 2^{n(\sum_{c=||A||+1}^{L}(c-||A||)\rho_c)} = 2^{n(\sum_{l=||A||+1}^{L} R_l)}, \qquad (40)$$

where we recall that $\rho_c = \frac{1}{n}\log_2 p^{s_c}$ *is the rate of equation section* $\mathbf{u}_{m,c}, m \in \{1, 2, \cdots, L\}$.

Lemma 28 (Probability Bounding Lemma for Error Events):

$$\Pr\left(\mathcal{E}_1\right) \leq 2^{-n\cdot(I(X_1,\cdots,X_L;Y)-\epsilon)} \qquad (41)$$

$$\Pr\left(\mathcal{E}_{2,0}\right) \leq 2^{-n\cdot(I(X_1,\cdots,X_L;Y|Q)-\epsilon)} \qquad (42)$$

$$\Pr\left(\mathcal{E}_{2,A}\right) \leq 2^{-n\cdot\left(I(X_{A^C};Y|X_A,Q)-\epsilon\right)}. \qquad (43)$$

To complete the proof of Theorem 20, note that $\Pr(\mathcal{E}^c_0)$ vanishes by properties of the jointly typical set $A^{(n)}_\epsilon$. Combining Lemma 27 and Lemma 28, substituting $\rho_c = R_c - R_{c+1}$ ($R_{L+1} = 0$) and requiring all exponential terms in expression (37) to have a negative exponent yields Theorem 20. ■

B. Proof of Lemma 27

Proof: First, recall the relationship $\rho_c = R_c - R_{c+1}$ ($R_{L+1} = 0$), which yields the equalities in the Lemma. We will proceed using the ρ_c notation.

Next, recall that notation $\tilde{\mathbf{U}}_{L\times L}$ and $\mathbf{U}_{L\times k}$ both refer to the same equation matrix and only differ in the indexing of its columns; similarly for the notation $\tilde{\mathbf{W}}_{L\times L}$ and $\mathbf{W}_{L\times k}$.

Finally, recall that messages $\mathbf{w}_l$'s ($l = 1, 2, \cdots, L$) are independently and uniformly drawn from $\mathbb{F}_p^{k_l} \cong \{0, 1, \cdots, p-1\}^{k_l}$, and should be viewed as a row vectors of elements in $\mathbb{F}_p$ of length k_l. They are zero-padded at the head to be of equal length k. Thus, all columns are independent and so are the message sections. When bounding the cardinalities of sets $\mathcal{U}_1, \mathcal{U}_{2,0}, \mathcal{U}_{2,A}$, we will handle one equation section at a time, say for $\mathbf{U}_{*c}$ with $c = 1, \cdots, L$.

The proof of this lemma hinges on linear algebra and carefully keeping track of the dependencies between the different equations. I.e., when some of the L equations are known to be correct, it affects the number of possible values for the remaining equations.

Take the c-th equation sections for example, where $c \in \{1, \cdots L\}$. We have

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

Note that sub-matrix $\mathbf{F}_{[1,\cdots,L]\times[1,\cdots,c]}$ contains the first c columns of matrix $\mathbf{F}$. Also note that by the problem assumption (in Theorem 20), any square sub-matrix of $\mathbf{F}_{[1,\cdots,L]\times[1,\cdots,c]}$ is guaranteed to be of full rank. This implies that:

- the number of values the equation sections $\tilde{\mathbf{U}}_{*c}$ may take on is equal to the cardinality of the corresponding message sections:

$$||\{\tilde{\mathbf{U}}_{*c}\}|| = ||\{[\mathbf{w}_{1,c}^T, \cdots, \mathbf{w}_{c,c}^T]^T\}|| = 2^{n\rho_c} \cdot 2^{n\rho_c} \cdots 2^{n\rho_c} = 2^{n(c\rho_c)}; \quad (44)$$

- given any c rows of $\tilde{\mathbf{U}}_{*c}$, we can solve for $\{\mathbf{w}_{1,c}, \cdots, \mathbf{w}_{c,c}\}$, i.e., $\tilde{\mathbf{W}}_{*c}$, and all the remaining $L-c$ rows of $\tilde{\mathbf{U}}_{*c}$ as well. Thus, when c or more rows of $\tilde{\mathbf{U}}_{*c}$ are known, $||\{\tilde{\mathbf{U}}_{*c}\}|| = 1$, i.e. all rows are fixed;
- when ν rows where $\nu < c$, are known, $(c-\nu)$ rows of $\tilde{\mathbf{U}}_{*c}$ remain free to take any values. Thus, $||\{\tilde{\mathbf{U}}_{*c}\}|| \le 2^{n(c-\nu)\rho_c}$.

Proof of $||\mathcal{U}_1|| \le 2^{n\sum_{c=1}^{L} c\rho_c} = 2^{n\sum_{l=1}^{L} R_l}$*:* All elements in this set have $\tilde{\mathbf{U}}_{*1} \neq \tilde{\mathbf{U}}_{*1}^0$ (common message is incorrect, hence one possibility must be removed from this column's possible values) but the remaining sections $\tilde{\mathbf{U}}_{*c}$ for $c = 2, \cdots L$ may take on any value. Hence, by (44) and the independence of the different sections we see that

$$||\mathcal{U}_1|| = ||\{\tilde{\mathbf{U}}_{*1}\}|| \cdot ||\{\tilde{\mathbf{U}}_{*2}\}|| \cdots ||\{\tilde{\mathbf{U}}_{*L}\}|| \quad (45)$$
$$\le (2^{n\rho_1} - 1) \cdot 2^{n2\rho_2} \cdots 2^{nL\rho_L} \quad (46)$$
$$\le 2^{n\sum_{c=1}^{L} c\rho_c} = 2^{n\sum_{l=1}^{L} R_l}. \quad (47)$$

Proof of $||\mathcal{U}_{2,0}|| \le 2^{n\sum_{c=2}^{L} c\rho_c} = 2^{n(2R_2+\sum_{l=3}^{L} R_l)}$*:* All elements in this set have $\tilde{\mathbf{U}}_{*1} = \tilde{\mathbf{U}}_{*1}^0$ (common message is correct, hence this section may take on only one possible value) but the remaining sections $\tilde{\mathbf{U}}_{*c}$ for $c = 2, \cdots L$ may take on any value except the correct ones. Hence, by (44) we may upper bound (we do not subtract the correct values as these do not change the asymptotic rates) this cardinality as follows

$$||\mathcal{U}_{2,0}|| = ||\{\tilde{\mathbf{U}}_{*1}\}|| \cdot ||\{\tilde{\mathbf{U}}_{*2}\}|| \cdots ||\{\tilde{\mathbf{U}}_{*L}\}|| \quad (48)$$
$$< 1 \cdot 2^{n2\rho_2} \cdots 2^{nL\rho_L} \quad (49)$$
$$= 2^{n\sum_{c=2}^{L} c\rho_c} = 2^{n(2R_2+\sum_{l=3}^{L} R_l)}. \quad (50)$$

Proof of $||\mathcal{U}_{2,A}|| \le 2^{n(\sum_{c=||A||+1}^{L}(c-||A||)\rho_c)} = 2^{n(\sum_{l=||A||+1}^{L} R_l)}$*:* All elements in this set have $\tilde{\mathbf{U}}_{*1} = \tilde{\mathbf{U}}_{*1}^0$ (common message is correct, hence this column may take on only one possible value) and $||A||$ out of L equations, indexed by the set A, are also correct. Consider the first $||A||$ equation sections, say $\tilde{\mathbf{U}}_{*c}$ where $c = 1, \cdots, ||A||$. Clearly, there are $||A||$ known rows in each of these $\tilde{\mathbf{U}}_{*c}$'s. Noting $||A|| \ge c$ for such equation sections, we know that all the remaining rows in each of these equation sections are known. As shown in Figure 8, the first $||A||$ message equation sections are fixed and only have 1 possible value.

Consider the remaining $L - ||A||$ equation sections, say $\tilde{\mathbf{U}}_{*c}$ where $c = ||A|| + 1, \cdots, L$. We know that

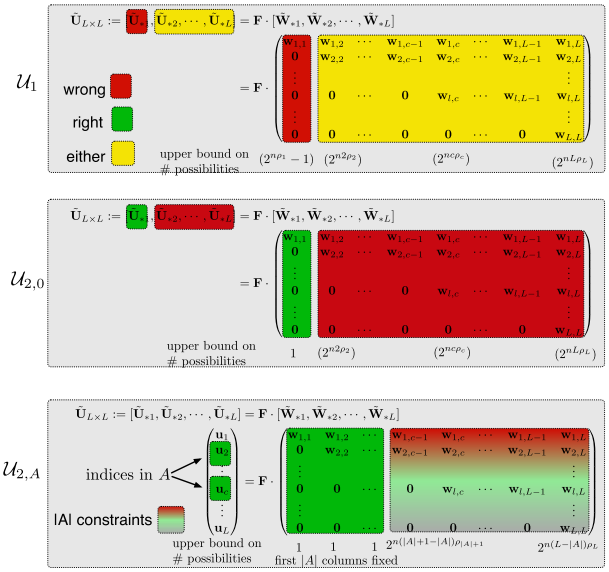


Fig. 8. Illustration of the different sets of equations $\mathcal{U}_1, \mathcal{U}_{2,0}, \mathcal{U}_{2,A}$ for $A \subset \{1, 2, \cdots L\}, 0 < |A| < L$. The number of possibilities (# possibilities) below each column indicates an upper bound on the number of different values this column may take. This is useful for counting the number of elements in each set, as given in Lemma 27 and used in the probability of error analysis of Theorem 20.

$||\{\tilde{\mathbf{U}}_{*c}\}|| \le 2^{n(c-||A||)\rho_c}$. Thus,

$$||\mathcal{U}_{2,A}|| = ||\{\tilde{\mathbf{U}}_{*1}\}|| \times ||\{\tilde{\mathbf{U}}_{*2}\}|| \times \cdots \times ||\{\tilde{\mathbf{U}}_{*L}\}|| \quad (51)$$
$$\le 1 \times \cdots \times 1 \times 2^{n(||A||+1-||A||)\rho_{||A||+1}} \quad (52)$$
$$\times 2^{n(||A||+2-||A||)\rho_{||A||+2}} \times \cdots \times 2^{n(L-||A||)\rho_L} \quad (53)$$
$$= 2^{n(\sum_{c=||A||+1}^{L}(c-||A||)\rho_c)} = 2^{n(\sum_{l=||A||+1}^{L} R_l)}. \quad (54)$$

■

C. Proof of Lemma 28

Proof of Lemma 9: These probability items can be upper bounded depending on the relationship between the observed sequence Y^n and the sequence tuple $(q^n, x_1^n, \cdots, x_L^n)$.

1) Consider $\Pr(\mathcal{E}_1)$. $\tilde{\mathbf{U}}_{L\times L}$ here has the property that $\mathbf{U}_{*1} \neq \mathbf{U}_{*1}^0$ (common message is incorrect). Note that $\mathbf{U}_{*1}$ 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 equation section implies that the observed y^n is independent of the true $(q^n, x_1^n, \cdots, x_L^n)$. Thus,

$$\Pr(\mathcal{E}_1) \le 2^{-n\cdot(I(Q,X_1,\cdots,X_L;Y)-\epsilon)} \overset{(a)}{\le} 2^{-n\cdot(I(X_1,\cdots,X_L;Y)-\epsilon)}.$$

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

2) Consider $\Pr(\mathcal{E}_{2,0})$. $\tilde{\mathbf{U}}_{L\times L}$ here has the property that $\mathbf{U}_{*1} = \mathbf{U}_{*1}^0$ (common equation section is correct) but the remaining equation sections are all wrong. This correct first equation section serves as the index for sequence $q^n(\mathbf{U}_{*1})$, on which codewords $X_m^n(\mathbf{u}_m)$ are conditioned. Thus the correctness of the first section, and incorrectness of the overall equations $\mathbf{u}_m$ which

determine codewords $X_m^n(\mathbf{u}_m)$ (for all $m = 1, 2, \cdots L$) imply that the observed Y^n is independent of the true $(x_1^n, \cdots, x_L^n)$ given the correct q^n. Thus,

$$\Pr\left(\mathcal{E}_{2,0}\right) \leq 2^{-n\cdot(I(X_1,\cdots,X_L;Y|Q)-\epsilon)}.$$

3) Consider $\Pr\left(\mathcal{E}_{2,A}\right)$ for a given subset A. Set A here is one such that $A \subset \{1, 2, \cdots L\}, 0 < ||A|| < L$. $\tilde{\mathbf{U}}_{L\times L}$ here has the property that: 1) its first equation section which determines $q^n(\mathbf{U}_{*1})$ is correct; 2) for $a \in A$, $\mathbf{u}_a$ are correct, thus Y^n is indeed dependent on $X_a^n(\mathbf{u}_a)$ for $a \in A$; 3) the remaining $\mathbf{u}_{a'}$ for $a' \in A^C$ are incorrect, meaning the observed sequence Y^n is independent of $X_{a'}^n(\mathbf{u}_{a'})$ for $a' \in A^C$. Thus,

$$\Pr\left(\mathcal{E}_{2,A}\right) \leq 2^{-n\cdot\left(I(X_{A^C};Y|X_A,Q)-\epsilon\right)}.$$

■

ACKNOWLEDGMENT

The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NSF.

REFERENCES

[1] Y. Song, N. Devroye, and B. Nazer, "Inverse compute-and-forward: Extracting messages from simultaneously transmitted equations," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, Aug. 2011, pp. 415–419.

[2] Y. Chen, Y. Song, and N. Devroye, "The capacity region of three user Gaussian inverse-compute-and-forward channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 1476–1480.

[3] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[4] S.-N. Hong and G. Caire, "Two-unicast two-hop interference network: Finite-field model," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sep. 2013, pp. 1–5.

[5] S.-N. Hong and G. Caire, "Structured lattice codes for 2×2×2 MIMO interference channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2013, pp. 2229–2233.

[6] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7661–7685, Dec. 2014.

[7] S.-N. Hong and G. Caire, "Reverse compute and forward: A low-complexity architecture for downlink distributed antenna systems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 1147–1151.

[8] S.-N. Hong and G. Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227–5243, Sep. 2013.

[9] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1037–1076, 1973.

[10] M. A. Wigger, "Cooperation on the multiple-access channel," Ph.D. dissertation, ETH Zurich, Univ. in Zürich, Switzerland, 2008.

[11] F. M. Willems, "Information theoretical results for multiple access channels," Ph.D. dissertation, K. U. Leuven, Leuven, Belgium, 1982.

[12] T. S. Han, "The capacity region of general multiple-access channel with certain correlated sources," *Inf. Control*, vol. 40, no. 1, pp. 37–60, 1979.

[13] D. Gunduz and O. Simeone, "On the capacity region of a multiple access channel with common messages," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 470–474.

[14] T. Cover, A. E. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 648–657, Nov. 1980.

Yanying Chen is currently a data scientist at Fair Isaac Company (FICO). She obtained M.Sc and Ph.D. in Electrical Engineering from University of Illinois at Chicago in 2015 and B.Sc. in Electrical Engineering from Tianjin University in 2009. Her background is on information theory, image and video processing, machine learning, statistics and computer science. She is interested in any form of mathematics.

Yiwei Song is currently a software engineer at A9. He obtained Ph.D. in Electrical Engineering from University of Illinois at Chicago in 2013 and B.Eng from Nanjing University of Posts and Telecommunications in 2009. He is interested in machine learning and information theory.

Natasha Devroye is an Associate Professor in the Department of Electrical and Computer Engineering at the University of Illinois at Chicago (UIC), which she joined in January 2009. From July 2007 until July 2008 she was a Lecturer at Harvard University. Dr. Devroye obtained her Ph.D in Engineering Sciences from the School of Engineering and Applied Sciences at Harvard University in 2007, and a Honors B. Eng in Electrical Engineering from McGill University in 2001. Dr. Devroye was a recipient of an NSF CAREER award in 2011 and was named UIC's Researcher of the Year in the "Rising Star" category in 2012. She has been an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE JOURNAL OF SELECTED AREAS IN COMMUNICATIONS, and is currently an Associate Editor for the IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. Her research focuses on multi-user information theory and applications to cognitive and software-defined radio, radar, relay and two-way communication networks.