An Achievable Region for a General Multi-terminal Network and its Chain Graph Representation

Stefano Rini
Lehrstuhl für Nachrichtentechnik ,
Technische Universität München
Arcisstraße 21, 80333 München, Germany,
Email: stefano.rini@tum.de

***Abstract*—Random coding, along with various standard techniques such as coded time-sharing, rate-splitting, superposition coding, and binning, are traditionally used in obtaining achievable rate regions for multi-terminal networks. The error analysis of such an achievable scheme relies heavily on the properties of strongly joint typical sequences and on bounds of the cardinality of typical sets. In this work, we obtain an achievable rate region for a general (i.e. an arbitrary set of messages shared amongst encoding nodes, which transmit to arbitrary decoding nodes) memoryless, single-hop, multi-terminal network without feedback or cooperation by introducing a general framework and notation, and carefully generalizing the derivation of the error analysis. We show that this general inner bound may be obtained from a graph representation that captures the statistical relationship among codewords and allows one to readily obtain the rate bounds that define the achievable rate region. The proposed graph representation naturally leads to the derivation of all the achievable schemes that can be generated by combining classic random coding techniques for any memoryless network used without feedback or cooperation.**

***Index Terms*—achievable region, chain graph, multi-terminal network, superposition, binning, rate-splitting, time-sharing.**

I. Introduction

In random coding, codewords are generated by drawing symbols in an independent, identically distributed (iid) fashion from a prescribed distribution; the performance of the ensemble of codes is then analyzed as a function of the block-length, which is eventually taken to infinity. Thanks to the iid symbols, and a block-length which tends to infinity, it is possible to derive the asymptotic performance of the ensemble of codes using the properties of jointly typical sets [1]. This proving technique was originally developed for the point-to-point channel [2] but is easily extended to multi-user channels by introducing multiple codebooks, one for each message to be transmitted. Multiple codebooks can be further organized in complex schemes that takes into account the structure of the network and the distribution of the messages among the users. Coded time-sharing, rate-spitting, superposition coding, binning, Markov encoding, quantize and forward, are some of the strategies that have been developed for multi-terminal channels. Given that all achievability schemes tend to use a combination of "standard" techniques applied in different fashions (leading to different dependencies amongst codewords), one might expect to be able to derive a general achievability scheme for a large class of networks. The key bounding techniques to analyze the error probability of transmission schemes are presented in an unified fashion by Csiszár and Körner [1, Ch. 1.2] and, more recently, by Kramer [3, Ch. 1] and El Gamal [4, Ch. 1]. In this paper we introduce an achievable scheme involving superposition coding, binning, rate-splitting, and coded time-sharing valid for a general one-hop channel without feedback or cooperation. This achievable scheme is defined by the random variables representing different codewords and by the factorization of the joint distribution among these random variables. We develop a graph representation of the factorization of the joint distribution based on Markov graphs [5], [6]: in particular we define a Markov chain graph with two types of edges, one representing superposition coding and one representing binning.

In information theory, a similar attempt to capture the relationship between random variables is the *functional dependence graph* [3, App. 8]. Despite fundamental similarities, the functional dependence graph is used mainly to represent the Markov relationship between random variables, while, in our case, the Markov chain graph describes a transmission scheme and the relationship between codewords of different users. Another attempt to provide achievable regions for general channels is hinted in [7], but the authors do not provide a rigorous approach to the problem.

By building upon the fundamental results in random coding theory and graph theory, we define a formal representation and a standard notation for a general achievable scheme as well as the derivation of the achievable region. Our ultimate goal is to define a form of "automatic rate region generator" which outputs the best known random coding achievable rate region for any channel of choice.

Paper Organization

Section II presents the class of networks considered in this work and revises the standard random coding techniques that we employ in our general achievable scheme. Section III introduces the novel *chain graph* representation of the encoding operations. Section IV describes the codebook generation, encoding and decoding procedures of the achievable scheme associated to a specific graph representation. Section V derives the rate bounds that define the achievable rate region based on the proposed *chain graph* representation. Section VI concludes the paper.

II. Channel Model and Random Coding Techniques for Achievability

A. Notation

In the following, bold letters indicate subsets of power sets $\mathfrak{P}$, i.e.

$$\mathbf{i} \subset \mathfrak{P}_N = \mathfrak{P}\left([1 \dots N]\right), \tag{1}$$

• index k/z: transmitter/receiver index with X_k, Y_z being the channel input/output at transmitter k and receiver z,
• index $\mathbf{i}/\mathbf{j}$: subset of transmitters/receivers. We also use $\mathbf{l}/\mathbf{m}$ and $\mathbf{v}/\mathbf{t}$,
• index $\mathbf{S}$: set of containing $(\mathbf{i}, \mathbf{j})$ pairs and $\overline{\mathbf{S}}$ its complement.

B. Network Model

We consider a general multi-terminal network where N_{TX} transmitting nodes want to communicate with N_{RX} receiving nodes. A given node may only be a transmitting or a receiving node , that is, the network is single-hop and without feedback. The transmitting node k, $k \in [1 \dots N_{\mathrm{TX}}]$, inputs X_k to the channel, while the receiving node z, $z \in [1 \dots N_{\mathrm{RX}}]$, has access to the channel output Y_z. The channel transition probability is indicated with $P_{Y_1 \dots Y_{N_{\mathrm{RX}}} | X_1 \dots X_{N_{\mathrm{TX}}}}$ and the channel is assumed to be memoryless. The subset of transmitting nodes $\mathbf{i}$, $\mathbf{i} \in \mathfrak{P}_{N_{\mathrm{TX}}}$, is interested in sending the message $W_{\mathbf{i}\to\mathbf{j}}$ to the subset of receiving nodes $\mathbf{j} \in \mathfrak{P}_{N_{\mathrm{RX}}}$ over N channel uses. The message $W_{\mathbf{i}\to\mathbf{j}}$, $(\mathbf{i}, \mathbf{j}) \in \mathfrak{P}_{N_{\mathrm{TX}}} \times \mathfrak{P}_{N_{\mathrm{RX}}}$, uniformly distributed in the interval $[0 \dots 2^{N R_{\mathbf{i}\to\mathbf{j}}} - 1]$, where N is the block-length and $R_{\mathbf{i}\to\mathbf{j}}$ the transmission rate.

A rate vector $\mathbf{R} = \{R_{\mathbf{i}\to\mathbf{j}}, \ \forall \ (\mathbf{i}, \mathbf{j}) \in \mathfrak{P}_{N_{\mathrm{TX}}} \times \mathfrak{P}_{N_{\mathrm{RX}}}\}$ is said to be achievable if there exists a sequence of encoding functions

$$X_k^N = X_k^N, \left(\{W_{\mathbf{i}\to\mathbf{j}}, \ \text{s.t.} \ \ (\mathbf{i}, \mathbf{j}) \subset \mathfrak{P}_{N_{\mathrm{TX}}} \times \mathfrak{P}_{N_{\mathrm{RX}}}, k \in \mathbf{i}\}\right),$$

and a sequence of decoding functions

$$\widehat{W}^z_{\mathbf{i}\to\mathbf{j}} = \widehat{W}^z_{\mathbf{i}\to\mathbf{j}}\left(Y_z^N\right) \text{ if } z \in \mathbf{j}, \tag{2}$$

such that

$$\lim_{N\to\infty} \max_{\mathbf{i},\mathbf{j},z} \mathbb{P}\left[\widehat{W}^z_{\mathbf{i}\to\mathbf{j}} \neq W^z_{\mathbf{i}\to\mathbf{j}}\right] = 0.$$

The capacity region $\mathcal{C}(\mathbf{R})$ is the convex closure of the region of all achievable rates in the vector $\mathbf{R}$-pairs. The general network model we consider is a variation to the network model in [8, Ch. 14], but we allow for messages to be distributed to more than one user while not considering feedback.

Fig. 1 shows the channel model considered in this work.

C. Random Coding Techniques for Achievability

We now revise standard random coding techniques used in the literature for achievability in single-hop networks used without feedback/cooperation: coded-time-sharing, rate-splitting, superposition coding, and binning.

• **Coded Time-Sharing** consists of using different transmission strategies according to a random schedule [9] and allows one to achieve the convex closure of the set of achievable points and, in some cases, an even larger region [1, pp. 288-290]. Let q^N denote denote the outcome of N iid draws from the distribution P_Q and let q^N be revealed to all the encoders and decoders in the network. In coded time-sharing the users generate multiple random codebooks conditioned on the value of the sequence q^n and choose their transmission strategy according to this random outcome.

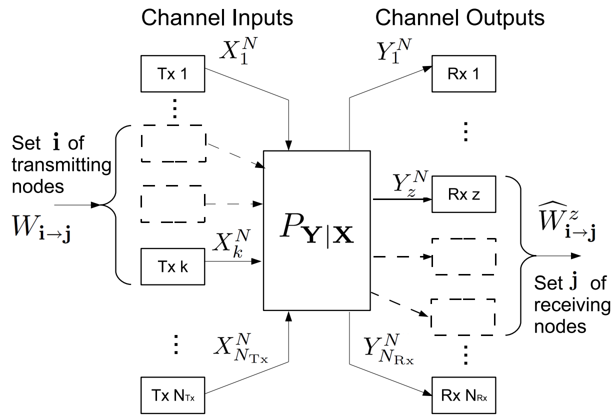

Fig. 1. The general cognitive multi-terminal network.

• **Rate-Splitting** corresponds to dividing a message $W_{\mathbf{i}\to\mathbf{j}}$ into a set of sub-messages $\{W_{\mathbf{i}'\to\mathbf{j}'}\}$ where $\mathbf{j} \subset \mathbf{j}'$ and $\mathbf{i} \supset \mathbf{i}'$, that is, sub-messages that are encoded by a smaller subset of transmitters and decoded at a larger set of receivers than the original message. Rate splitting preserves the rate of the transmitted messages but increases the number of messages to be transmitted over the channel which, in terms, increases the possible transmission strategies [9]. Let $W^{[\mathbf{i}\to\mathbf{j}]}_{\mathbf{l}\to\mathbf{m}}$ indicate that sub-message $W_{\mathbf{l}\to\mathbf{m}}$ obtained by splitting the message $W_{\mathbf{i}\to\mathbf{j}}$, then the message $W_{\mathbf{i}\to\mathbf{j}}$ can be split in a sequence of sub-messages $W^{[\mathbf{i}\to\mathbf{j}]}_{\mathbf{l}\to\mathbf{m}}$ for every $(\mathbf{l}, \mathbf{m})$ such that $\mathbf{j} \subset \mathbf{m}$ and $\mathbf{i} \supset \mathbf{l}$ so that

$$R'_{\mathbf{i}\to\mathbf{j}} = \sum_{(\mathbf{l},\mathbf{m}) \ \mathbf{j}\subset\mathbf{m}, \ \mathbf{i}\supset\mathbf{l}.} R^{[\mathbf{i}\to\mathbf{j}]}_{\mathbf{l}\to\mathbf{m}} = \sum_{(\mathbf{l},\mathbf{m}), \ \mathbf{j}\subset\mathbf{m}, \ \mathbf{i}\supset\mathbf{l}.} \gamma^{[\mathbf{i}\to\mathbf{j}]}_{\mathbf{l}\to\mathbf{m}} R_{\mathbf{i}\to\mathbf{j}}, \tag{3}$$

with $\gamma^{[\mathbf{i}\to\mathbf{j}]}_{\mathbf{l}\to\mathbf{m}} = \frac{R^{[\mathbf{i}\to\mathbf{j}]}_{\mathbf{l}\to\mathbf{m}}}{R_{\mathbf{i}\to\mathbf{j}}}$ and where $R'_{\mathbf{i}\to\mathbf{j}}$ is the rate of the RV $W_{\mathbf{i}\to\mathbf{j}}$ in the original channel and $R_{\mathbf{i}\to\mathbf{j}}$ in the rate of the RV $W_{\mathbf{i}\to\mathbf{j}}$ in the channel after rate-splitting is applied.

Rate-splitting effectively transforms the problem of achieving a rate vector $\mathbf{R}$ into the problem of achieving the rate vector $\mathbf{R}'$ such that $\mathbf{R}' = \Gamma\mathbf{R}$ with $\Gamma_{(\mathbf{i},\mathbf{j})\times(\mathbf{l},\mathbf{m})} = \gamma^{[\mathbf{i}\to\mathbf{j}]}_{\mathbf{l}\to\mathbf{m}}$.

• **Superposition Coding** can be intuitively thought of as stacking codewords on top of each other [10]. The "base" codewords are decoded first and stripped from the received signal so as to reduce the interference when decoding the "top" codewords. In superposition coding a different top codebook is generated for each base codeword and the codewords in the top codebook have a conditional distribution that depends on the specific base codeword. Let $U^N_{\mathbf{l}\to\mathbf{m}}$ be the codebook with distribution $P^N_{U_{\mathbf{l}\to\mathbf{m}}}$ carrying the message $W'_{\mathbf{l}\to\mathbf{v}}$ and the codeword $U^N_{\mathbf{i}\to\mathbf{j}}$ carrying the message $W'_{\mathbf{i}\to\mathbf{j}}$ be superposed to the codebook $U^N_{\mathbf{l}\to\mathbf{m}}$. A different top codebook is associated to each base codeword $U^N_{\mathbf{l}\to\mathbf{m}} = u^N_{\mathbf{l}\to\mathbf{m}}$ and the codewords

$U_{\mathbf{i}\to\mathbf{j}}^N$ in each of these codebooks are generated according to the distribution $P^N_{U_{\mathbf{i}\to\mathbf{j}}|U_{\mathbf{l}\to\mathbf{m}}=u_{\mathbf{l}\to\mathbf{m}}}$. Superposition of $U_{\mathbf{i}\to\mathbf{j}}^N$ over $U_{\mathbf{l}\to\mathbf{m}}^N$ can be performed when the following hold:

• $\mathbf{l} \subseteq \mathbf{i}$: that is the bottom message is encoded by a larger set of encoders than the top message.

• $\mathbf{m} \subseteq \mathbf{j}$: that is the bottom message is decoded by a larger set of decoders than the top message.

Note that, if $U_{\mathbf{i}\to\mathbf{j}}^N$ is superposed to $U_{\mathbf{l}\to\mathbf{m}}^N$ and $U_{\mathbf{v}\to\mathbf{t}}^N$ is superposed to $U_{\mathbf{l}\to\mathbf{m}}^N$, than $U_{\mathbf{i}\to\mathbf{j}}^N$ is also superposed to $U_{\mathbf{v}\to\mathbf{t}}^N$. In the following $U_{\mathbf{i}\to\mathbf{j}} \overline{<} U_{\mathbf{l}\to\mathbf{m}}$ indicates that the codeword $U_{\mathbf{i}\to\mathbf{j}}^N$ is superposed to the codeword $U_{\mathbf{l}\to\mathbf{m}}^N$.

• **Binning** allows a transmitter to "pre-cancel" (portions of) the interference known to be experienced at a receiver [11]. This is done by generating a codebook that has a larger number of codewords than the cardinality of the associated message so that the transmitted codeword can be chosen according to both the message and the specific value of the interference. Although in the original setting of [11] the interference is a random process, this technique can be extended to the case where the interference is a different user's codewords. Let $U_{\mathbf{l}\to\mathbf{m}}^N$ be the codeword with distribution $P^N_{U_{\mathbf{l}\to\mathbf{m}}}$ carrying the message $W'_{\mathbf{l}\to\mathbf{v}}$ and the codeword $U_{\mathbf{i}\to\mathbf{j}}^N$ carrying the message $W'_{\mathbf{i}\to\mathbf{j}}$ be binned against $U_{\mathbf{l}\to\mathbf{m}}^N$. The codewords $U_{\mathbf{i}\to\mathbf{j}}^N$ are generated independently from the codewords $U_{\mathbf{l}\to\mathbf{m}}^N$ but chosen so to look as if generated according to the distribution $P^N_{U_{\mathbf{i}\to\mathbf{j}}|U_{\mathbf{l}\to\mathbf{m}}}$. The codeword $U_{\mathbf{i}\to\mathbf{j}}^N$ can be binned against $U_{\mathbf{l}\to\mathbf{m}}^N$ when

• $\mathbf{i} \subseteq \mathbf{l}$: that is, the set of encoders performing binning has knowledge of the interfering codeword

Note that if $U_{\mathbf{i}\to\mathbf{j}}^N$ can be binned against $U_{\mathbf{i}\to\mathbf{m}}^N$, then $U_{\mathbf{i}\to\mathbf{j}}^N$ can also be binned against $U_{\mathbf{i}\to\mathbf{m}}^N$, : this is referred to as "joint binning" [12]. In the following $U_{\mathbf{i}\to\mathbf{j}} \lessdot U_{\mathbf{l}\to\mathbf{m}}$ indicates that the codeword $U_{\mathbf{i}\to\mathbf{j}}^N$ is binned against the codeword $U_{\mathbf{l}\to\mathbf{m}}^N$ while $U_{\mathbf{i}\to\mathbf{j}} \lessdot\gtrdot U_{\mathbf{l}\to\mathbf{m}}$ indicates joint binning.

III. The Chain Graph Representation of a General Achievable Schemes

The elements included in the random coding construction of Sec. II-C may be compactly represented using the following graph $\mathcal{G}(V,E)$:

• every vertex $v = (\mathbf{i},\mathbf{j}) \in V \subset \mathfrak{P}(N_{\rm TX}) \times \mathfrak{P}(N_{\rm RX})$ is associated to the RV $U_{\mathbf{i}\to\mathbf{j}}$ carrying the message $W'_{\mathbf{i}\to\mathbf{j}}$ at rate $R'_{\mathbf{i}\to\mathbf{j}}$ obtained through rate-splitting,

• the set of edges E, connecting the vertices in V, contains two subsets $\boldsymbol{S}$ and $\boldsymbol{B}$ such that $E = \boldsymbol{S} \cup \boldsymbol{B}$ and $\boldsymbol{S} \cap \boldsymbol{B} = \emptyset$

• the vertex $U_{\mathbf{i}\to\mathbf{j}}$ is connected with $U_{\mathbf{l}\to\mathbf{m}}$ by a directed edge of type $\boldsymbol{S}$ (for *superposition*), if $U_{\mathbf{l}\to\mathbf{m}} \overline{<} U_{\mathbf{i}\to\mathbf{j}}$ (solid line).

• the vertex $U_{\mathbf{i}\to\mathbf{j}}$ is connected with $U_{\mathbf{v}\to\mathbf{t}}$ by a directed edge of type $\boldsymbol{B}$ (for *binning*), if $U_{\mathbf{v}\to\mathbf{t}} \lessdot U_{\mathbf{i}\to\mathbf{j}}$ (dotted line).

Moreover, all the RVs are generated according to a marginal distribution that depends on Q.

A schematic representation of the graphical Markov model associated with the graph $\mathcal{G}(V,E)$ is presented in Fig. 2: each vertex $(\mathbf{i},\mathbf{j})$ is associated with the message $W'_{\mathbf{i}\to\mathbf{j}}$, a rate-splitting equation and an auxiliary RV $U_{\mathbf{i}\to\mathbf{j}}$ with distribution $P_{U_{\mathbf{i}\to\mathbf{j}}|Q}$ according to which the codebook is generated. The

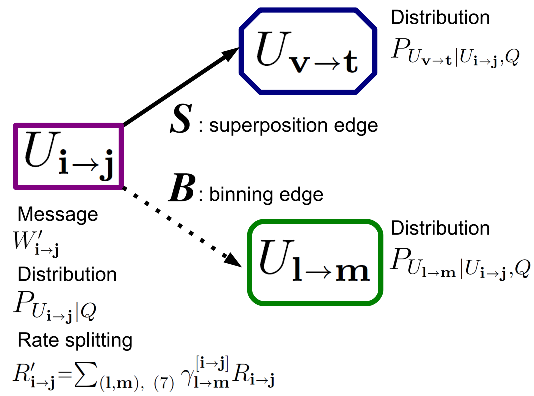

Fig. 2. A schematic representation of the graph in Sec. III .

vertex $(\mathbf{i},\mathbf{j})$ can be connected to a vertex $(\mathbf{l},\mathbf{m})$ by two types of edges: a superposition edge $\boldsymbol{S}$ and a binning edge $\boldsymbol{B}$; both edges indicate the Markov dependency between $U_{\mathbf{i}\to\mathbf{j}}$ and $U_{\mathbf{l}\to\mathbf{m}}$ given Q.

The graph representation of the achievable scheme is particularly useful in deriving the joint distribution of the codewords $U_{\mathbf{i}\to\mathbf{j}}^N$ in a general scheme in Sec. II. When trying to determine this distribution, superposition coding and binning effectively result in allowing for any joint distribution among the connected RVs. In the following we detail how the graph $\mathcal{G}(V,E)$ can be used to describe the dependency structure of the codewords of an achievable scheme for a general multi-terminal network. Graphs representing conditional dependencies among RVs have been extensively studied in the literature in the field of graphical Markov models [5]. In order for the graph representation of achievable scheme to correspond to a feasible distribution, it is necessary to impose certain restrictions on its structure. Moreover, it is convenient to consider graph representations where there exists a convenient factorization of the joint distribution of the RVs in the graph. For this reason we consider a decomposable chain graph: graph with both directed and undirected edges that is equivalent to some Asymmetric Directed Graphs (ADGs).

Assumption 1. RVs that are jointly binned form fully connected sets, that is

$$U_{\mathbf{i}\to\mathbf{j}} \lessdot\gtrdot U_{\mathbf{i}\to\mathbf{m}},\ U_{\mathbf{i}\to\mathbf{m}} \lessdot\gtrdot U_{\mathbf{i}\to\mathbf{j}} \Rightarrow U_{\mathbf{i}\to\mathbf{m}} \lessdot\gtrdot U_{\mathbf{i}\to\mathbf{t}}. \quad (4)$$

Assumption 2. RVs that are jointly binned, have the same parent nodes, that is, if $U_{\mathbf{i}\to\mathbf{j}} \lessdot\gtrdot U_{\mathbf{i}\to\mathbf{t}}$ and $U_{\mathbf{i}\to\mathbf{j}} \lessdot U_{\mathbf{l}\to\mathbf{m}}$ or $U_{\mathbf{i}\to\mathbf{j}} \overline{<} U_{\mathbf{l}\to\mathbf{m}}$, than $U_{\mathbf{i}\to\mathbf{t}} \lessdot U_{\mathbf{l}\to\mathbf{m}}$ or $U_{\mathbf{i}\to\mathbf{t}} \overline{<} U_{\mathbf{l}\to\mathbf{m}}$.

Assumption 3. RVs known at same set of decoders $\mathbf{i}$ do not form directed cycles in the graph representation. If a cycle exists it must be undirected.

Given any achievable scheme, one can always obtain an achievable scheme that satisfies Assumptions 1, 2 and 3 by adding an additional binning steps to the original scheme. These additional encoding operations can only enlarge the achievable region, since the original scheme can be reobtained with the appropriate choice of distribution imposed by the binning step.

Theorem III.1. *If Assumptions 1, 3, and 2, hold, the graph representation of a general achievable scheme corresponds to chain graph which is equivalent to some ADG for which the joint distribution can be written as*

$$P(V) = \prod_{U_k \in V} P_{U_k|\mathrm{pa(U_k)}}, \tag{5}$$

where $\mathrm{pa(U_k)}$ *indicates the parent nodes of* U_k*, that is the set of all the nodes connected to* U_k *by a direct edge.*

Proof: The complete proof is provided in [13]. ■

We refer to a graph representation of an achievable scheme in Th. III.1 as a Chain Graph Representation of an Achievable Scheme (CGRAS).

The edges of the equivalent ADG are generated by orienting undirected edge and therefore the equivalent ADG can be written as $\mathcal{G}(V, \boldsymbol{B}\text{-} \cup \boldsymbol{S})$ for some $\boldsymbol{B}\text{-} \subset \boldsymbol{B}$. The notation $\overset{\leftarrow}{\lessdot}$ is used in the following to indicate the edges in $\boldsymbol{B}$-. With this notation, the joint distribution of a CGRAS can then be factorized as:

$$P_{\{U_{\mathbf{i}\to\mathbf{j}}, \ \forall \ (\mathbf{i},\mathbf{j})\}} = \prod_{(\mathbf{i},\mathbf{j})} P_{U_{\mathbf{i}\to\mathbf{j}}|\{U_{\mathbf{l}\to\mathbf{m}}, \ U_{\mathbf{i}\to\mathbf{j}} \overline{\lessdot} U_{\mathbf{l}\to\mathbf{m}}, \ U_{\mathbf{i}\to\mathbf{j}} \overset{\leftarrow}{\lessdot} U_{\mathbf{l}\to\mathbf{m}}\}}. \tag{6}$$

IV. Codebook Generation, Encoding and Decoding procedures for a CGRAS

We now outline the codebook generation, encoding, and decoding operations for a general transmission scheme associated with a CGRAS.

Codebook Generation: The codebook of a CGRAS is generated according to the distribution imposed by the superposition coding edges. Consider the node $U_{\mathbf{i}\to\mathbf{j}}$ and assume that the codebook of the parent nodes has already been generated and indexed by $l_{\mathbf{l}\to\mathbf{m}} \in [1 \ldots 2^{L_{\mathbf{l}\to\mathbf{m}}}]$, then, for each possible set

$$\{ \ l_{\mathbf{l}\to\mathbf{m}}, \ \forall \ (\mathbf{l},\mathbf{m}) \ U_{\mathbf{i}\to\mathbf{j}} \overline{\lessdot} U_{\mathbf{l}\to\mathbf{m}}\},$$

repeat the following:

- generate $2^{NL_{\mathbf{i}\to\mathbf{j}}}$ codewords, for $L_{\mathbf{i}\to\mathbf{j}} = R'_{\mathbf{i}\to\mathbf{j}} + \overline{R}_{\mathbf{i}\to\mathbf{j}}$ with iid symbols drawn from the distribution $P_{U_{\mathbf{i}\to\mathbf{j}}|\{U_{\mathbf{l}\to\mathbf{m}}, \ U_{\mathbf{l}\to\mathbf{m}} \overline{\lessdot} U_{\mathbf{i}\to\mathbf{j}}\}}$.
- Place the codewords in $2^{NR'_{\mathbf{i}\to\mathbf{j}}}$ bins of size $2^{N\overline{R}'_{\mathbf{i}\to\mathbf{j}}}$ each.
- Index each codebook of size $2^{NL_{\mathbf{i}\to\mathbf{j}}}$ using the set $\{ \ l_{\mathbf{l}\to\mathbf{m}}, \ \forall \ (\mathbf{l},\mathbf{m}) \ U_{\mathbf{i}\to\mathbf{j}} \overline{\lessdot} U_{\mathbf{l}\to\mathbf{m}}\}$ so that

$$U^N_{\mathbf{i}\to\mathbf{j}}(l_{\mathbf{i}\to\mathbf{j}}) = U^N_{\mathbf{i}\to\mathbf{j}}\left(w'_{\mathbf{i}\to\mathbf{j}}, b_{\mathbf{i}\to\mathbf{j}}, \{ \ l_{\mathbf{l}\to\mathbf{m}}, \ \forall \ (\mathbf{l},\mathbf{m}) \ U_{\mathbf{i}\to\mathbf{j}} \overline{\lessdot} U_{\mathbf{l}\to\mathbf{m}}\}\right), \tag{7}$$

where $w'_{\mathbf{i}\to\mathbf{j}}$ is chosen according to the transmitted message. Since graph $\mathcal{G}(V, \boldsymbol{S})$ is an ADG, we can apply the above step starting from the nodes that have no parents to all the nodes in the graph.

Encoding Procedure: The binning index $b_{\mathbf{i}\to\mathbf{j}}$ in each codeword is chosen so that the codewords in (7) appear to have been generated with iid symbols drawn from the distribution in (6). If a codeword cannot be determined, then a value of $b_{\mathbf{i}\to\mathbf{j}}$ is chosen randomly. We may find a jointly typical codeword if the number of bins $b_{\mathbf{i}\to\mathbf{j}}$ is sufficiently large, that is, if $\overline{R}_{\mathbf{i}\to\mathbf{j}}$ is sufficiently large. Finally the encoder k produces the channel input X^N_k as a deterministic function of its codebook(s).

Decoding Procedure Receiver z looks for a set of bin indices $w'_{\mathbf{i}\to\mathbf{j}}$ and $b_{\mathbf{i}\to\mathbf{j}}$ for $z \in \mathbf{j}$, such that the set $\left\{Y^N_z, \left\{U^N_{\mathbf{i}\to\mathbf{j}} : \ z \ \in \ \mathbf{j}\right\}\right\}$ looks as if it generated an iid according to the distribution in (6). If the decoder cannot find such codeword or it finds more than one, it picks one tuple at random.

V. Derivation of the Rate Bounds

In this section we derive the achievable rate region of the transmission strategy described in Sec. IV by bounding the encoding and decoding error probability.

A. Encoding Errors

For the probability of encoding error to vanish as the block-length increases it is necessary to choose a large enough binning rate $\overline{R}_{\mathbf{i}\to\mathbf{j}}$ so the encoders can jointly find a set of bin index $\overline{b}_{\mathbf{i}\to\mathbf{j}}$ for which the codeword $U^N_{\mathbf{i}\to\mathbf{j}}$ appears to be generated according to the distribution imposed by binning although it is generated according to the distribution imposed by superposition coding. Define

$$\mathbf{S_B} = \left\{(\mathbf{i},\mathbf{j}), U_{\mathbf{i}\to\mathbf{j}} \lessdot U_{\mathbf{l}\to\mathbf{m}} \text{ for some } (\mathbf{l},\mathbf{m})\right\}, \tag{8}$$

i.e. $\mathbf{S_B}$ is the set of all the indexes whose codeword possess a bin index; we intuitively expect the condition for a successful encoding to depend on the distance between the codewords' distribution at generation and after encoding:

$$\mathbf{I}^{\text{encoding}}_{\text{codebook}} = E[\log P_{\text{encoding}}] - E[\log P_{\text{codebook}}] = \tag{9}$$
$$\sum_{(\mathbf{i},\mathbf{j}) \in \mathbf{S_B}} I(U_{\mathbf{i}\to\mathbf{j}}; \{U_{\mathbf{l}\to\mathbf{m}}, \ U_{\mathbf{i}\to\mathbf{j}} \overset{\leftarrow}{\lessdot} U_{\mathbf{l}\to\mathbf{m}}\} | \{U_{\mathbf{l}\to\mathbf{m}}, \ U_{\mathbf{i}\to\mathbf{j}} \overline{\lessdot} U_{\mathbf{l}\to\mathbf{m}}\} | Q).$$

The encoding error analysis is obtained using Markov inequality and the mutual covering lemma [3], [4].

Theorem V.1. Encoding Errors Analysis using the Mutual Covering Lemma *For any CGRAS, encoding is successful with high probability as* $N \to \infty$ *if, for any subset* $\mathbf{S} \subset \mathbf{S_B}$ *such that*

$$(\mathbf{i},\mathbf{j}) \in \mathbf{S} \Rightarrow (\mathbf{l},\mathbf{m}) \in \mathbf{S}, \ \forall \ (\mathbf{l},\mathbf{m}) \ \text{ s.t. } U_{\mathbf{l}\to\mathbf{m}} \overline{\lessdot} U_{\mathbf{i}\to\mathbf{j}}, \tag{10}$$

for $\mathbf{S_B}$ *defined in* (8)*, the following holds:*

$$\sum_{(\mathbf{i},\mathbf{j}) \ \in \ \overline{\mathbf{S}}} \overline{R}_{\mathbf{i}\to\mathbf{j}} \geq \mathbf{I}^{\text{encoding}}_{\text{codebook}} - \mathbf{I}^{\text{M.C.L.}}_{\mathbf{S}}, \tag{11}$$

where $\mathbf{I}^{\text{M.C.L.}}_{\mathbf{S}}$ *if defined in* (12) *(* M.C.L. *standing for "Mutual Covering Lemma").*

Proof: The complete proof is provided in [13]. ■

$$\mathbf{I}_{\mathbf{S}}^{\text{M.C.L.}} = \sum_{(\mathbf{i},\mathbf{j})\ \in\ \mathbf{S}} I(U_{\mathbf{i}\to\mathbf{j}}; \{U_{\mathbf{l}\to\mathbf{m}},\ U_{\mathbf{i}\to\mathbf{j}} \overleftarrow{\lessdot} U_{\mathbf{l}\to\mathbf{m}}, (\mathbf{l},\mathbf{m}) \in \mathbf{S}\} | \{U_{\mathbf{l}\to\mathbf{m}},\ U_{\mathbf{i}\to\mathbf{j}} \overline{\lessdot} U_{\mathbf{l}\to\mathbf{m}},\ U_{\mathbf{l}\to\mathbf{m}} \lessdot U_{\mathbf{i}\to\mathbf{j}}, (\mathbf{l},\mathbf{m}) \in \overline{\mathbf{S}}\}, Q), \quad (12)$$

$$\begin{aligned} \mathbf{I}_{\mathbf{S}}^{z,\text{P.L.}} &= I(Y_z; \{U_{\mathbf{i}\to\mathbf{j}}, (\mathbf{i},\mathbf{j}) \in \mathbf{S}\} | \{U_{\mathbf{l}\to\mathbf{m}}, (\mathbf{l},\mathbf{m}) \in \overline{\mathbf{S}}\}, Q) \\ &\quad + \sum_{(\mathbf{i},\mathbf{j})\ \in\ \mathbf{S}} I(U_{\mathbf{i}\to\mathbf{j}}; \{U_{\mathbf{l}\to\mathbf{m}}, U_{\mathbf{i}\to\mathbf{j}} \overleftarrow{\lessdot} U_{\mathbf{l}\to\mathbf{m}},\ (\mathbf{l},\mathbf{m}) \in \mathbf{S}\} | \{U_{\mathbf{l}\to\mathbf{m}},\ U_{\mathbf{i}\to\mathbf{j}} \overline{\lessdot} U_{\mathbf{l}\to\mathbf{m}}\ \in\ \overline{\mathbf{S}}\}, Q), \end{aligned} \quad (13)$$

B. Decoding Errors

For the decoding error probability to vanish as the block-length increases, it is necessary to choose a small enough codebook rate $L_{\mathbf{i}\to\mathbf{j}}$ so that the codewords $U_{\mathbf{i}\to\mathbf{j}}^N$ are sufficiently "spaced apart" in the typical set to allow successful decoding. The bounds on the codebook rates are obtained using the packing lemma [3], [4] which bounds the maximum number of codewords that can be employed at transmission and still allow the decoder to recover the transmitted codeword. Decoding relies on the conditional typicality of the channel output given the transmitted codeword and on the typicality relationship imposed on the codewords by the encoding procedure. The probability of error at each decoder is linked to the specific codewords that the receiver is attempting to decode; for this reason it is convenient to define the set $\mathbf{S}_{\mathbf{D}}^z$ as:

$$\mathbf{S}_{\mathbf{D}}^z = \{(\mathbf{i},\mathbf{j}) \ \text{ s.t. } \ z \in \mathbf{j}\}, \quad (14)$$

that is, $\mathbf{S}_{\mathbf{D}}^z$ is the set of indexes whose codeword is decoded at receiver z. Intuitively we expect the condition for successful decoding to depend on the distance between the joint distribution of the codewords and the channel output at the encoding and the distribution after an incorrect decoding, that is:

$$\begin{aligned} &\mathbf{I}_{\text{encoding}}^{\text{decoding}} = E[\log P_{Y_z,\text{encoding}^z}] - E[\log P_{Y_z} P_{\text{codebook}}^z] \\ &= I(Y_z; \{U_{\mathbf{i}\to\mathbf{j}},\ (\mathbf{i},\mathbf{j}) \in \mathbf{S}_{\mathbf{D}}^z\} | Q) + \\ &\sum_{(\mathbf{i},\mathbf{j})\in\mathbf{S}_{\mathbf{D}}^z} I(U_{\mathbf{i}\to\mathbf{j}}; \{U_{\mathbf{l}\to\mathbf{m}},\ U_{\mathbf{i}\to\mathbf{j}} \overleftarrow{\lessdot} U_{\mathbf{l}\to\mathbf{m}}\} | \{U_{\mathbf{l}\to\mathbf{m}},\ U_{\mathbf{i}\to\mathbf{j}} \overline{\lessdot} U_{\mathbf{l}\to\mathbf{m}}\}, Q), \end{aligned}$$

where

$$P_{Y_z,\text{encoding}^z} = P_{Y_z|\{U_{\mathbf{i}\to\mathbf{j}}\in\mathbf{S}_{\mathbf{D}}^z\},Q} P_{\text{encoding}}^z \quad (15a)$$

$$P_{\text{encoding}}^z = \sum_{(\mathbf{i},\mathbf{j})\notin\mathbf{S}_{\mathbf{D}}^z} P_{\text{encoding}} \quad (15b)$$

$$P_{\text{codebook}}^z = \sum_{(\mathbf{i},\mathbf{j})\notin\mathbf{S}_{\mathbf{D}}^z} P_{\text{codebook}}, \quad (15c)$$

that is, P_{encoding}^z and P_{codebook}^z are the encoding and the codebook distribution, for the set of codewords decoded at receiver z, $\{U_{\mathbf{i}\to\mathbf{j}},\ (\mathbf{i},\mathbf{j}) \in \mathbf{S}_{\mathbf{D}}^z\}$.

Theorem V.2. Decoding error analysis using the Packing Lemma

For any CGRAS, decoding is successful with high probability as $N \to \infty$ if , for any receiver z and for any subset $\mathbf{S} \subset \mathbf{S}_{\mathbf{D}}^z$ such that condition (10) *, for $\mathbf{S}_{\mathbf{D}}^z$ defined in* (14)*, the following holds:*

$$\sum_{(\mathbf{i},\mathbf{j})\in\mathbf{S}} L_{\mathbf{i}\to\mathbf{j}} \leq \mathbf{I}_{\mathbf{S}}^{\text{J.D}}, \quad (16)$$

where $\mathbf{I}_{\mathbf{S}}^{z,\ P.L.}$ is defined in (13)*. (where* P.L *stands for* packing lemma*).*

Proof: The complete proof is provided in [13]. ∎

VI. Conclusion

In this paper we present a new general achievable rate region valid for a general class of multi-terminal networks. This achievable scheme employs rate-splitting, superposition coding, and binning, and generalizes a number of inner bounds and techniques that have been proposed in the literature. This achievable scheme may be represented using a graph representation that allows for a quick comparison between transmission strategies and simplifies the derivation of the corresponding achievable rate regions. This paper attempts to establish a general tool to derive achievable rate regions for multi-terminal networks which contains all standard random coding techniques. A subject of ongoing research is whether there exists a combination of encoding strategies that yields the largest achievable region among all possible transmission strategies (within the proposed framework).

Acknowledgment

The author would like to thank Prof. Andrea Goldsmith,Prof. Daniela Tuninetti, Prof. Natasha Devroye and Prof. Gerhard Kramer for the insightful discussions and their helpful comments.

References

[1] I. Csiszar and J. Körner, *Information theory: Coding theorems for discrete memoryless channels*. Budapest: Akademiai Kiado, 1981.

[2] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[3] G. Kramer, *Topics in Multi-User Information Theory*, ser. Foundations and Trends in Communications and Information Theory. Vol. 4: No 45, pp 265-444, 2008.

[4] A. El Gamal and Y.-H. Kim, "Lecture notes on network information theory," *preprint at arXiv:1001.3404*, 2010.

[5] S. Andersson, D. Madigan, and M. Perlman, "On the markov equivalence of chain graphs, undirected graphs, and acyclic digraphs," *Scandinavian Journal of Statistics*, vol. 24, no. 1, pp. 81–102, 1997.

[6] A. Dawid and S. Lauritzen, "Hyper markov laws in the statistical analysis of decomposable graphical models," *The Annals of Statistics*, vol. 21, no. 3, pp. 1272–1317, 1993.

[7] R. Khosravi-Farsani and F. Marvasti, "Interference networks with general message sets: A random coding scheme," *Arxiv preprint arXiv:1107.1839*, 2011.

[8] T. Cover and J. Thomas, *Elements of information theory*. Wiley, 1991.

[9] T. Han and K. Kobayashi, "A New Achievable Rate Region for the Interference Channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan 1981.

[10] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 2–14, 1972.

[11] S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of control and information theory*, vol. 9, no. 1, pp. 19–31, 1980.

[12] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.

[13] S. Rini, "An achievable region for a general multi-terminal network and the corresponding chain graph representation," *Arxiv preprint arXiv:1112.1497*, 2011.