A General Approach to Random Coding for Multi-Terminal Networks

Stefano Rini, Andrea Goldsmith
Department of Electrical Engineering
Stanford University
Stanford, CA, USA
Email: {stefano,andrea}@wsl.stanford.edu

***Abstract*—We introduce a general framework to derive achievable rate regions based on random coding for any memoryless, single-hop, multi-terminal network. 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 under which the error probability vanishes as the block-length goes to infinity. The proposed graph representation naturally leads to an "automatic rate region derivator" which produces achievable rate regions combining classic random coding techniques such as coded time-sharing, rate-splitting, superposition coding and binning for the general memoryless network under consideration.**

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

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 the 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 [5], [12], [7]. This proof technique was originally developed for the point-to-point channel [19] and has since been extended to multi-user channels by introducing different codebook constructions for different network structures. Coding techniques such as coded time-sharing, rate-spitting, superposition coding, binning, interference decoding, bin-superpose-match coding are some of the strategies developed to obtain achievable rate regions in various multi-terminal, single-hop channels. All these achievable schemes tend to use a combination of "standard" proof techniques which that utilize a few statistical properties of the typical sets of iid random codewords such as size and probability. Given the similarity in the derivation of achievable rate regions based on these various coding strategies, one might expect to be able to obtain a general expression of the achievable rate region for a large class of networks. This is the aim of this work. In particular, we introduce an achievable scheme involving rate-splitting, coded time-sharing, superposition coding and binning valid for a general one-hop channel without feedback or transmitter/receiver 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 in the codebook. This joint distribution is easily represented using a graphical Markov model [17]: more specifically we define a Markov chain graph in which vertexes represents codewords, a set of edges represents superposition coding and another set binning. When an edge connects two vertices, this indicates an encoding operation among the two codewords that is either superposition coding or binning, depending on the type of edge. 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 corresponding achievable rate region. With this approach, we are able to define an "automatic rate region generator" which outputs the best known random coding achievable rate region for any single-hop multi-terminal channel of choice.

Related Work

The key bounding techniques to analyze the error probability of transmission schemes based on random coding using the properties of typical sequences are presented in an unified fashion by Csiszár and Körner [5, Ch. 1.2] and, more recently, by Kramer [12, Ch. 1] and El Gamal and Kim [7, Ch. 2].

Graphs representing codes are employed by Kramer in [11] in the form of "code trees". In [11] a general discrete memoryless network with feedback is considered: in this context a transmission strategy can be efficiently represented using a tree in which each node represents the channel input at a given time and a path to such node a sequence of feedback values which determine, together with the transmitted message, the channel input. The author [11] also introduces the "functional dependence graph", a directed acyclic graph which describes the Markov relationship between the transmitted messages and the channel inputs and outputs at any time instant. Both code trees and functional dependence graphs are used to investigate the role of the directed information in multi-terminal systems with feedback. Despite of the different context, the analysis in [11] has some similarities with our approach in the attempt of compactly represent coding strategies and Markov dependencies.

An systematic approach to the analysis of achievable schemes employing superposition coding is hinted in [7], where tables are utilized to derive the error events for such transmission schemes. Even if a general procedure is not explicitly detailed, [7] suggests an algorithmic derivation of the achievable rate region.

An attempt to generalize the derivation of achievable regions is provided in [10], but no closed-form characterization of the achievable rate is provided.

Our general approach to the derivation of achievable regions based on random coding employs four fundamental techniques: coded time-sharing, rate-splitting, superposition coding and binning[1].

Coded time-sharing was originally introduced by Han and Kobayashi [9] in their derivation of an achievable region for the interference channel. It consists of choosing a specific transmission codebook according to a random but known sequence.

Rate-Splitting was also introduced by Han and Kobayashi in [9] and it consists of splitting a message of one user into multiple sub-messages, each decoded by a different set of receivers. This techniques is useful in the interference channel as it allows the receivers to decode a portion of the interference created by the transmission of the other user.

Superposition coding is a method of stacking a codeword for one user on top of another user's codeword and it was first introduced by Cover [3] for deriving an achievable region for the broadcast channel. In superposition coding a first message is conveyed using a "base" codebook as in the point-to-point channel while a second message is transmitted using an "overlay" codebook that is generated conditionally dependent each codeword in the base codebook.

Binning is applied when an encoder has knowledge of the interference experienced at its intended decoder to pre-code the transmission against such interference. It was originally devised by Sleepian and Wolf [20] for distributed lossless compression and it was later used by Marton [16] to derive an achievable region for the general broadcast channel .

As we shall see, these relatively simple coding strategies can be combined to generate a complex codebook construction to fit a number of communication scenarios.

Contributions:

In the following we:

- **propose a novel chain graph representation for encoding schemes based on coded time-sharing, rate-splitting, superposition coding and binning.** This new formalism provides a clear and unified framework to represent achievable schemes based on random coding arguments as it provides a compact description of the codebook generation as well as encoding and decoding procedures.
- **generalize the encoding and decoding error analysis for random coding based achievable schemes.** The proposed graph representation also describes the distribution of the codewords and is used to bound the error probabilities. This allows the achievable rate region to be determined for any scheme that can be represented with the proposed formalism.
- **compactly represent the achievable rate region.** We consider three class of schemes with increasing complexity and, in each scenario, present a compact representation of the achievable rate region which relies only on only graph theoretic properties of the chain graph representation.

[1]sometimes referred to as Cover random binning [7] or Gel'fand-Pinsker coding [8]

Paper Organization:

The paper is organized as follows: Section II presents a general class of memoryless, one-hop networks. Section III provides a few basic graph theoretic notions that will be used in the reminder of the paper. Section IV introduces the novel chain graph representation of the encoding and decoding operations in a general random-coding-based achievable scheme. Section V derives the rate bounds that define the achievable rate region based on the proposed *chain graph* representation for three different classes of graph, listed in increasing order of complexity. Section VI concludes the paper.

II. NETWORK MODEL

We consider a general one-hop multi-terminal network in which $N_{\rm TX}$ transmitting nodes want to communicate with $N_{\rm RX}$ receiving nodes. The network is assumed to be single-hop and without feedback or cooperation among transmitters or receivers. The transmitting node $k \in [1 \ldots N_{\rm TX}]$ has input X_k to the channel while the receiving node $z \in [1 \ldots N_{\rm RX}]$ accesses the channel output Y_z. The channel is assumed to be memoryless with transition probability

$$P_{\mathbf{Y}|\mathbf{X}} = P_{Y_1 \ldots Y_{N_{\rm RX}}|X_1 \ldots X_{N_{\rm TX}}}. \tag{1}$$

The subset of transmitting nodes $\mathbf{i}$, is interested in reliably communicating the message $W_{\mathbf{i}\to\mathbf{j}}$ to the subset of receiving nodes $\mathbf{j}$ over N channel uses. The message $W_{\mathbf{i}\to\mathbf{j}}$, is uniformly distributed in the interval $[1 \ldots 2^{NR_{\mathbf{i}\to\mathbf{j}}}]$, where N is the block-length and $R_{\mathbf{i}\to\mathbf{j}}$ the transmission rate.

The allocation of multiple the messages at transmitters and receivers is described by the set $\mathbf{V}$, which defines the set of messages $W_{\mathbf{V}} = \{W_{\mathbf{i}\to\mathbf{j}},\ (\mathbf{i},\mathbf{j}) \in \mathbf{V}\}$. A rate vector $R_{\mathbf{V}} = \{R_{\mathbf{i}\to\mathbf{j}},\ (\mathbf{i},\mathbf{j}) \in \mathbf{V}\}$ is said to be achievable if, for all $(\mathbf{i},\mathbf{j}) \in \mathbf{V}$, there exists a sequence of encoding functions

$$X_k^N = X_k^N\left(\{W_{\mathbf{i}\to\mathbf{j}},\ k \in \mathbf{i}\}\right), \quad \forall\ (\mathbf{i},\cdot) \in \mathbf{V}, \tag{2}$$

and a sequence of decoding functions

$$\widehat{W}^z_{\mathbf{i}\to\mathbf{j}} = \widehat{W}^z_{\mathbf{i}\to\mathbf{j}}\Big(Y_z^N\Big), \quad \forall\ z\ \in \mathbf{j},\ (\cdot,\mathbf{j}) \in \mathbf{V}, \tag{3}$$

such that

$$\lim_{N\to\infty} \max_{z,\mathbf{i},\mathbf{j}} \mathbb{P}\left[\widehat{W}^z_{\mathbf{i}\to\mathbf{j}} \neq W_{\mathbf{i}\to\mathbf{j}}\right] = 0. \tag{4}$$

The capacity region $\mathcal{C}(R_{\mathbf{V}})$ is the convex closure of the region of all achievable rates in the vector $R_{\mathbf{V}}$.

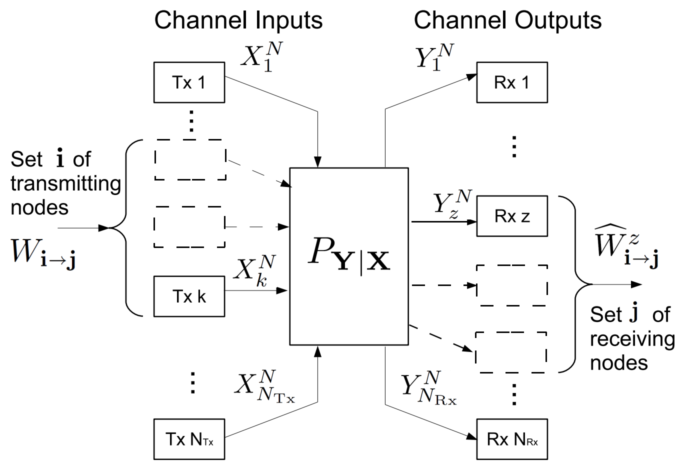

Fig. 1. The general memoryless, one-hop multi-terminal network in Sec. II.

The channel model under consideration is depicted in Fig. 1: on the left side are the N_{TX} transmitting nodes while on the right are the N_{RX} receiving nodes. A message $W_{\mathbf{i}\to\mathbf{j}}$ is encoded by to the set $\mathbf{i}$ of the transmitting nodes and decoded at the set $\mathbf{j}$ of receiving nodes. The channel input X_k at each decoder k is a function of the messages available at this decoder according to (2). Receiver z decodes all the messages $W_{\mathbf{i}\to\mathbf{j}}$ such that $z\in\mathbf{j}$ from the channel output Y_z using the decoding function in (3).

The channel under consideration is a variation to the network model in [4, Ch. 15.10], but allows for messages to be allocated to multiple users while not considering feedback and cooperation.

III. Graphical Markov Models

This following section presents a few basic graph theoretic notions and describes Graphical Markov Models (GMMs). GMMs were introduced by Perl in 1988 [17] and are used to represent the factorization of a multivariate distribution in a graphical manner. A GMM is constructed using the graph $\mathcal{G}(\mathbf{V},\mathbf{E})$ and associating the set nodes in the graph, $\mathbf{V}$, to a set of Random Variables (RVs) and the set edges $\mathbf{E}$ to the conditional dependencies among these RVs. Although conceptually simple, GMMs can be used to represent a highly varied and complex system of multivariate dependencies by means of the global structure of the graph, thereby obtaining efficiency in modeling, inference, and probabilistic calculations [14], [21], [6].

A. Some Graph Theoretic Notions

A *graph* $\mathcal{G}(\mathbf{V},\mathbf{E})$ is defined by a finite set of *vertices* $\mathbf{V}$ and a set of *edges* $\mathbf{E}\subseteq\mathbf{V}\times\mathbf{V}$ i.e. a set of ordered pairs of distinct vertices. An edge $(\alpha,\beta)\in\mathrm{E}$ whose opposite $(\beta,\alpha)\in\mathbf{E}$ is called an *undirected edge*, whereas an edge $(\alpha,\beta)\in\mathbf{E}$ whose opposite $(\beta,\alpha)\notin\mathbf{E}$ is a *directed edge*. Two vertices α and β are *adjacent* in $\mathcal{G}$ if $(\alpha,\beta)\in\mathbf{E}$ or $(\beta,\alpha)\in\mathbf{E}$. If $\mathbf{A}\subseteq\mathbf{V}$ is a subset of the vertex set, it induces a *subgraph* $\mathcal{G}_{\mathbf{A}}=(\mathbf{A},\mathbf{E}_{\mathbf{A}})$, where $\mathbf{E}_{\mathbf{A}}=\mathbf{E}\cap(\mathbf{A}\times\mathbf{A})$. The *parents* of a node $\alpha\in\mathbf{V}$ in $\mathbf{A}$ are those vertices linked to α by a directed edge in $\mathbf{E}_{\mathbf{A}}$, i.e.

$$\mathrm{pa}_{\mathbf{E}_{\mathbf{A}}}(\beta)=\left\{\alpha\in\mathbf{A}|\ (\alpha,\beta)\in\mathbf{E}_{\mathbf{A}},\ (\beta,\alpha)\notin\mathbf{E}_{\mathbf{A}}\right\}.$$

This definition readily extends to subsets of nodes $\mathbf{B}\subseteq\mathbf{A}$:

$$\mathrm{pa}_{\mathbf{E}_{\mathbf{A}}}(\mathbf{B})=\left\{\alpha\in\mathbf{A}|\exists\ \beta\in\mathbf{B},(\alpha,\beta)\in\mathbf{E}_{\mathbf{A}},\ (\beta,\alpha)\notin\mathbf{E}_{\mathbf{A}}\right\}.$$

A path π of length n from α_0 to α_n is a sequence $\pi=\{\alpha_0,\alpha_1,...,\alpha_n\}\subseteq\mathbf{V}$ of distinct vertices such that $(\alpha_{n-1},\alpha_n)\in\mathbf{E}$ for all $i=1...n$. If (α_{n-1},α_n) is directed for at least one of the nodes i, we call the path *directed*. If none of the edges are directed, the path is called *undirected*. A cycle is a path in which $\alpha_0=\alpha_n$. We define the *future* of a node α in G, denoted by $\phi(\alpha)$, as the set of nodes that can be reached by α through a directed path.

Graph are generally classified in three categories:

• If all the edges are undirected, the graph is said to be an *UnDirected Graph* (UDG).
• If all the edges are directed and the graph contains no cycles, the graph is said to be an *Directed Acyclic Graph* (DAG).
• If edges are both directed and undirected and the graph does not contain directed cycles, the graph is called a *Chain Graph* (CG)

B. Graphical Markov Models

The idea of GMMs is to define dependencies between RVs through a graph: each node in the graph represents a RV, an edge between two RVs indicates that the two RVs are conditionally dependent. For simple scenarios, this implies that the distribution of a node depends only on its neighboring nodes while it is conditionally independent from the rest of the graph.

A rigorous formulation of this concept has to take into account global features of the graph, such as paths and cycles, in order to define a joint distribution associated with any graph.

Definition 1. Global G-Markov Property:
Consider a graph $\mathcal{G}(\mathbf{V},\mathbf{E})$, a probability measure P on

$$\mathcal{U}=\bigtimes_{\alpha\in\mathbf{V}}\mathcal{U}_\alpha, \tag{5}$$

is said to be *global G-Markovian* if

$$\alpha\perp\beta\ |[\mathbf{V}\backslash\phi(\alpha)]\backslash\{\alpha,\beta\}\quad[P], \tag{6}$$

for α and β are not adjacent and $\beta\notin\phi(\alpha)$ and if, given four disjoint subsets A,B,C and D, the following holds:

$$\begin{aligned}A\perp B\ |C\cup D\quad[P]\text{ and }A\perp C\ |B\cup D\quad[P]\\ \Longrightarrow\ A\perp B\cup C\ |D\quad[P],\end{aligned} \tag{7}$$

where the notation $A\perp B\ |C\quad[P]$ indicates that A is conditionally independent of B given C under P.

The above formulation of the global Markov property is not the most general, but we refer to this definition for simplicity. Definition 1 can be interpreted as follows: two nodes α and β that are not adjacent and such that β is not in the future of α

are conditionally independent given all the nodes in $\mathbf{V}$ minus the future of α and α and β themselves.

The global Markov properties assures that a the distribution P over the graph $\mathcal{G}(\mathbf{V}, \mathbf{E})$ is well defined, regardless of the connectivity properties of the graph. This definition is particularly important as it makes it possible to establish a notion of equivalence among graphs. Graphs with different sets of edges can result in the same joint distribution among the RVs and in general it is not trivial to determine which graphs describe the same dependency structure.

Definition 2. Markov Equivalence, [1, Th. 3.1]: Two chain graphs $\mathcal{G}_1 = (V, E_1)$ and $\mathcal{G}_2 = (V, E_2)$ are called *Markov equivalent* if, for every product space $\mathcal{X}$ indexed by V, the classes of probability measures that are global G-Markovian on $\mathcal{G}_1$ and $\mathcal{G}_2$ are equivalent.

As the Markov properties of an arbitrary probability distribution can be difficult to establish, a commonly used class of GMMs are those that offer a very convenient factorization of the joint distribution P in (6). UDGs offer a factorization in terms of cliques, subset of vertices such that every two vertices in the subset are connected by an edge. For DAGs we have that P factorizes in terms of parents nodes of a vector, i.e.

$$P = \prod_{\alpha \in V} P_{\alpha|\mathrm{pa}(\alpha)}. \tag{8}$$

In general, a convenient and recursive factorization of P for CGs is not available: the only case in which such a factorization exists is when the chain graph is Markov equivalent to a DAG, in which case the factorization in (8) can be used. In [1] a rigorous theory to establish the equivalence between GMMs is developed.

IV. The Chain Graph Representation of an Achievable Scheme

We now introduce a general achievable scheme employing rate-splitting, coded time-sharing, superposition coding and binning for the general channel model in Sec. II. As we shall see, rate-splitting effectively transforms the problem of achieving a given rate vector into the problem of achieving a different rate vector for a channel model in which more messages are allocated to each user. Each additional message can be though of as an additional virtual user which is introduced in the network is such a way as to preserve the rate of the transmitted messages. By increasing the total number of messages to be encoded, rate-splitting increases the overall number of viable transmission strategies and this can only result in larger achievable rates.

Coded time-sharing is used to convexify the achievable region and is obtained by choosing the transmission codebook according to a random sequence made available to all the terminals in the network. Coded time-sharing also outperforms both time and frequency division multiplexing (TDM/FDM respectively).

The core coding techniques of the proposed general achievable scheme are superposition coding and binning. These two coding steps are represented using two GMMs over the same set of nodes. A first GMM, the *superposition coding graph*, describes how superposition coding is applied when creating the transmission codebook. A second GMM, the *binning graph*, describes how binning is used to select codewords from the transmission codebook so as to appear as if generated according to a different distribution than the codebook distribution.

A. Rate-Splitting

In rate-splitting the message of a user is split into multiple sub-messages which are each then encoded/decoded at different sets of transmitters/receivers. Since encoding capabilities and decoding requirements cannot be altered, a message $W_{\mathbf{i}\to\mathbf{j}}$ can be split into the messages $W_{\mathbf{l}\to\mathbf{m}}$ only when $\mathbf{i} \supseteq \mathbf{l}$ and $\mathbf{j} \subseteq \mathbf{m}$, that is, the new set of messages can only be encoded by a smaller set of transmitters or decoded by a larger set of receivers.

More generally, rate-splitting can be expressed by the matrix by Γ such that

$$R_{\mathbf{V}'} = \Gamma R_{\mathbf{V}}, \tag{9}$$

for $\mathbf{V}' \subseteq \mathbf{V}$, i.e. $\mathbf{V}'$ is the original message allocation and $\mathbf{V}$ the one after rate-splitting, and where

$$\Gamma_{(\mathbf{i},\mathbf{j})\times(\mathbf{l},\mathbf{m})}, \tag{10}$$

indicates the portion of the message $W_{\mathbf{i}\to\mathbf{j}}$, $(\mathbf{i},\mathbf{j}) \in \mathbf{V}'$ in the original allocation that is embedded in the message $W_{\mathbf{l}\to\mathbf{m}}$, $(\mathbf{l},\mathbf{m}) \in \mathbf{V}$ after rate-splitting has been applied.

For rate-splitting to be feasible, the following must hold:

$$\Gamma_{(\mathbf{i},\mathbf{j})\times(\mathbf{l},\mathbf{m})} \neq 0 \implies \mathbf{i} \supseteq \mathbf{l},\ \mathbf{j} \subseteq \mathbf{m}, \tag{11}$$

also, since the rate of each message must be preserved:

$$\sum_{(\mathbf{i},\mathbf{j})} \Gamma_{(\mathbf{i},\mathbf{j})\times(\mathbf{l},\mathbf{m})} = 1. \tag{12}$$

Note that (12) implies that multiple messages in $W_{\mathbf{i}\to\mathbf{j}}$, $(\mathbf{i},\mathbf{j}) \in \mathbf{V}'$ are compounded to form a single message $W_{\mathbf{l}\to\mathbf{m}}$, $(\mathbf{l},\mathbf{m}) \in \mathbf{V}$ after rate-splitting. This is sometimes referred to as *rate-sharing* and is useful in a number of scenarios [15].

Given an achievable rate vector $R_{\mathbf{V}'}$, there are multiples rate-splitting matrices Γ in (10) that can be considered and thus the largest achievable region is obtained by considering the union over all possible rate-splitting matrices, that is

$$R_{\mathbf{V}'} = \mathit{Conv}\left\{ \bigcup_{\Gamma} \Gamma R_{\mathbf{V}} \right\}, \tag{13}$$

where *Conv* indicates the convex hull operation. Since the achievable rate region is expressed in terms of linear inequalities, it is sometimes possible to utilize the Fourier-Motzkin Elimination (FME) [13] to obtain a closed form expression for (13).

B. Coded Time-Sharing

Coded time-sharing is attained by generating the multiple transmission codebooks conditionally dependent on the time-sharing RV Q. The codebook used for transmission is then selected according to the time-sharing random sequence q^n obtained with iid draws from Q and available at all transmitters and receivers. Coded time-sharing outperforms time-division and frequency-division multiplexing (TDM/FDM) and is thought to achieve larger rates than the convex closure of the achievable rates [2].

C. Chain Graph Representation of Superposition Coding and Binning

For a given rate-splitting matrix Γ and a time-sharing sequence q^N, we define two chain graph GMMs to describe how superposition coding and binning are performed. The conditions on the message allocations under which a codeword can be superposed over or binned against another are specified by the next two lemmas.

Lemma 1. Conditions for Superposition Coding
The superposition of one $U_{\mathbf{i}\to\mathbf{j}}$ over another $U_{\mathbf{l}\to\mathbf{m}}$ can be performed when the following two conditions 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.*

Moreover, if $U_{\mathbf{i}\to\mathbf{j}}$ is superposed over $U_{\mathbf{l}\to\mathbf{m}}$ and $U_{\mathbf{v}\to\mathbf{t}}$ over $U_{\mathbf{l}\to\mathbf{m}}$, then $U_{\mathbf{i}\to\mathbf{j}}$ is also superposed over $U_{\mathbf{v}\to\mathbf{t}}$.

If $U_{\mathbf{i}\to\mathbf{j}}$ is superposed over $U_{\mathbf{l}\to\mathbf{m}}$, then the reverse cannot hold.

Lemma 2. Conditions for Binning
Binning of the RV $U_{\mathbf{i}\to\mathbf{j}}$ against the RV $U_{\mathbf{l}\to\mathbf{m}}$ can be performed when the following condition holds:

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

Note that if $U_{\mathbf{i}\to\mathbf{j}}$ can be binned against $U_{\mathbf{i}\to\mathbf{m}}$, then $U_{\mathbf{i}\to\mathbf{j}}$ can also be binned against $U_{\mathbf{i}\to\mathbf{m}}$, regardless of the value of $\mathbf{j}$ and $\mathbf{m}$. This is referred to as joint binning.

With Lemma 1 and Lemma 2 we can now formally define the Chain Graph Representation of an Achievable Scheme (CGRAS):

Definition 3. CGRAS
The GMMs $\mathcal{G}(\mathbf{V},\mathbf{S})$ and $\mathcal{G}(\mathbf{V},\mathbf{E})$, together with the rate-splitting matrix Γ and the time-sharing RV Q, represents an achievable scheme for the network model in Sec. II where:

• every vertex $(\mathbf{i},\mathbf{j}) \in \mathbf{V}$ is associated with 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 the rate-splitting matrix Γ.

• $\mathcal{G}(\mathbf{V},\mathbf{S})$ is the *superposition coding graph* and describes how superposition coding is performed. The conditions in Lemma 1 hold for the set of edges $\mathbf{S}$ and thus $\mathcal{G}(\mathbf{V},\mathbf{S})$ is an DAG. The superposition of $U_{\mathbf{v}\to\mathbf{t}}$ over $U_{\mathbf{i}\to\mathbf{j}}$ is also indicated as $U_{\mathbf{i}\to\mathbf{j}} \mapsto U_{\mathbf{v}\to\mathbf{t}}$.

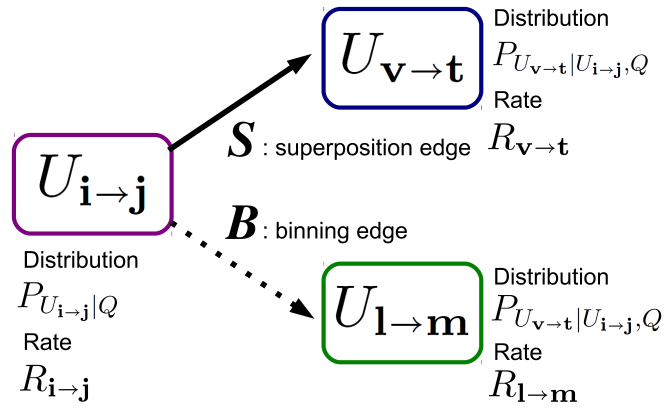

Fig. 2. A schematic representation of the CGRAS is provided in Def (3).

• $\mathcal{G}(\mathbf{V},\mathbf{E})$ for $\mathbf{E} = \mathbf{S} \cup \mathbf{B}$ and $\mathbf{S} \cap \mathbf{B} = \emptyset$, is the *binning coding graph* and describes how binning is performed. The conditions in Lemma 2 hold for the set of edges $\mathbf{B}$. Binning of $U_{\mathbf{l}\to\mathbf{m}}$ against $U_{\mathbf{i}\to\mathbf{j}}$ is also indicated as $U_{\mathbf{i}\to\mathbf{j}} \dashrightarrow U_{\mathbf{l}\to\mathbf{m}}$. Similarly. the joint binning of $U_{\mathbf{l}\to\mathbf{m}}$ and $U_{\mathbf{i}\to\mathbf{j}}$ is indicated as $U_{\mathbf{l}\to\mathbf{m}} \text{ - - - } U_{\mathbf{i}\to\mathbf{j}}$.

A schematic representation of the CGRAS is provided in Fig 2: each node of the graph is associated with a codeword encoding a specific message obtained through rate-splitting. Codewords can be superposed and binned only when the conditions in Lemma 1 and Lemma 2 are satisfied. When $U_{\mathbf{v}\to\mathbf{t}}$ is superposed over $U_{\mathbf{i}\to\mathbf{j}}$, this is indicated by a directed, solid arrow from $U_{\mathbf{i}\to\mathbf{j}}$ to $U_{\mathbf{v}\to\mathbf{t}}$. Similarly, when $U_{\mathbf{l}\to\mathbf{m}}$ is binned against $U_{\mathbf{i}\to\mathbf{j}}$, this is indicated by a directed, dashed arrow from $U_{\mathbf{i}\to\mathbf{j}}$ to $U_{\mathbf{l}\to\mathbf{m}}$.

The superposition coding graph is associated with the *codebook distribution* defined as

$$P_{\text{codebook}} = \prod_{(\mathbf{i},\mathbf{j})\in\mathbf{V}} P_{U_{\mathbf{i}\to\mathbf{j}}|\mathrm{pa}_{\mathbf{S}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}})}, \tag{14}$$

which satisfies the global Markov property in Def. 1 since $\mathcal{G}(\mathbf{V},\mathbf{S})$ is a DAG.

We now impose further assumptions on the binning graph to obtain a convenient factorization of the joint distribution associated with this graph.

Assumption 1. **No directed cycles in the binning graph:** The binning graph $\mathcal{G}(\mathbf{V},\mathbf{E})$ contains no directed cycles. If cycles exist, they must be undirected.

Assumption 2. **Joint binning forms cliques:** Nodes in the binning graph that are connected by an undirected edge form fully connected sets, that is

$$U_{\mathbf{i}\to\mathbf{j}} \text{ - - - } U_{\mathbf{i}\to\mathbf{m}},\ U_{\mathbf{i}\to\mathbf{j}} \text{ - - - } U_{\mathbf{i}\to\mathbf{t}} \Rightarrow U_{\mathbf{i}\to\mathbf{m}} \text{ - - - } U_{\mathbf{i}\to\mathbf{t}}. \tag{15}$$

Moreover, jointly binned codewords have the same parent nodes, that is, if $U_{\mathbf{i}\to\mathbf{j}} \text{ - - - } U_{\mathbf{i}\to\mathbf{t}}$ then

$$\mathrm{pa}_{\mathbf{E}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}}) = \mathrm{pa}_{\mathbf{E}}(\mathrm{U}_{\mathbf{i}\to\mathbf{t}}). \tag{16}$$

Note that the joint binning can be applied only when two nodes are known at the same set of nodes, so the expressions in (15) and (16) are without loss of generality.

Under Assumption 1, the binning graph is a chain graph for which a global Markov property can be appropriately defined. Under Assumption 1 and Assumption 2, the binning graph is a chain graph in which any non-cyclical orientation of the undirected edges produces a Markov equivalent DAG $\mathcal{G}(\mathbf{V}, \widetilde{\mathbf{E}})$ for some $\widetilde{\mathbf{E}} \subseteq \mathbf{E}$.

Since undirected binning edges can occur only at nodes known at the same set of encoders, these two assumptions can be imposed without loss of generality. To satisfy Assumption 1, one needs to substitute all the binning steps in the cycle with joint binning, thus obtaining undirected cycles. Similarly, Assumption 2 can be made to hold by adding either superposition coding edges or binning edges between jointly binned RVs and their parents. Since both assumptions can be satisfied by increasing the number of coding step in the CGRAS, we require for the binning graph to satisfy both Assumption 1 and Assumption 2.

Given these two assumptions, the binning graph can be associated with the *encoding distribution* defined as

$$P_{\text{encoding}} = \prod_{(\mathbf{i},\mathbf{j}) \in \mathbf{V}} P_{U_{\mathbf{i}\to\mathbf{j}} | \mathrm{pa}_{\widetilde{\mathbf{E}}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}})}, \tag{17}$$

where $\mathcal{G}(\mathbf{V}, \widetilde{\mathbf{E}})$ is the DAG obtained by a non-cyclical orientation of the undirected edges in $\mathcal{G}(\mathbf{V}, \mathbf{E})$. Since the factorization in (17) holds, the encoding distribution satisfies the global Markov property in Def. 1.

Note that the encoding and decoding distribution in (14) and (17), respectively, have an identical factorization among the RVs except for the RVs connected by a binning edge.

We now detail the codebook generation, encoding and decoding operation which are associated with the CGRAS in Def. 3.

D. Codebook Generation

For a given codebook distribution that factorizes as in (14), the codebook associated with the CGRAS is obtained by recursively applying the following procedure:

- Consider the node $U_{\mathbf{i}\to\mathbf{j}}$ in $\mathcal{G}(\mathbf{V}, \mathbf{S})$ and assume that the codebook of the parent nodes has been generated and indexed by $l_{\mathbf{l}\to\mathbf{m}} \in [1 \ldots 2^{L_{\mathbf{l}\to\mathbf{m}}}]$, i.e.

$$U^N_{\mathbf{l}\to\mathbf{m}}(l_{\mathbf{l}\to\mathbf{m}}), \ \forall \ U_{\mathbf{l}\to\mathbf{m}} \in \mathrm{pa}_{\mathbf{S}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}}), \tag{18}$$

then, for each possible set of base codewords

$$\{U^N_{\mathbf{l}\to\mathbf{m}}(l_{\mathbf{l}\to\mathbf{m}}), \ U_{\mathbf{l}\to\mathbf{m}} \in \mathrm{pa}_{\mathbf{S}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}}), \ \mathrm{l}_{\mathbf{l}\to\mathbf{m}} \in [1 \ldots 2^{\mathrm{L}_{\mathbf{l}\to\mathbf{m}}}]\}, \tag{19}$$

repeat the following:

1) generate $2^{NL_{\mathbf{i}\to\mathbf{j}}}$ codewords, for

$$L_{\mathbf{i}\to\mathbf{j}} = R_{\mathbf{i}\to\mathbf{j}} + \overline{R}_{\mathbf{i}\to\mathbf{j}} \tag{20a}$$

$$\overline{R}_{\mathbf{i}\to\mathbf{j}} = \begin{cases} \geq 0 & \exists \ U_{\mathbf{v}\to\mathbf{t}} \dashrightarrow U_{\mathbf{i}\to\mathbf{j}} \\ 0 & \text{otherwise} \end{cases}, \tag{20b}$$

with iid symbols drawn from the distribution $P^N_{U_{\mathbf{i}\to\mathbf{j}}|\mathrm{pa}_{\mathbf{S}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}})}$ conditioned on the set of base codewords in (19). In the following we refer to $R_{\mathbf{i}\to\mathbf{j}}$ as the *message rate* while to $\overline{R}_{\mathbf{i}\to\mathbf{j}}$ as the *binning rate*.

2) If $\overline{R}_{\mathbf{i}\to\mathbf{j}} \neq 0$, place each codeword $U^N_{\mathbf{i}\to\mathbf{j}}$ in $2^{NR_{\mathbf{i}\to\mathbf{j}}}$ bins of size $2^{N\overline{R}_{\mathbf{i}\to\mathbf{j}}}$ and indexed by $b_{\mathbf{i}\to\mathbf{j}} \in [1...2^{NR_{\mathbf{i}\to\mathbf{j}}}]$. If $\overline{R}_{\mathbf{i}\to\mathbf{j}} = 0$, simply set $b_{\mathbf{i}\to\mathbf{j}} = 1$.

3) Index each codebook of size $2^{NL_{\mathbf{i}\to\mathbf{j}}}$ using the set $\{l_{\mathbf{l}\to\mathbf{m}}, \ \forall \ (\mathbf{l},\mathbf{m}) \text{ s.t. } \ U_{\mathbf{i}\to\mathbf{j}} \mapsto U_{\mathbf{l}\to\mathbf{m}}\}$ so that

$$U^N_{\mathbf{i}\to\mathbf{j}}(l_{\mathbf{i}\to\mathbf{j}}) = U^N_{\mathbf{i}\to\mathbf{j}}\left(w_{\mathbf{i}\to\mathbf{j}}, b_{\mathbf{i}\to\mathbf{j}}, \{l_{\mathbf{l}\to\mathbf{m}}, U_{\mathbf{l}\to\mathbf{m}} \in \mathrm{pa}_{\mathbf{S}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}})\}\right). \tag{21}$$

E. Encoding procedure

Consider an encoding distribution that factorizes as in (17) and differs from the codebook distribution in (14) except for the joint conditional distributions of the RVs involved in binning. In the encoding procedure, the each binning index $b_{\mathbf{i}\to\mathbf{j}}$ in (21) is chosen so that the transmitted codeword appears to be generated according to the encoding distribution in (17) despite of being generated according to the codebook distribution in (14). A set of bin indexes $\{b_{\mathbf{i}\to\mathbf{j}}, \ U_{\mathbf{i}\to\mathbf{j}} \in \mathbf{V}\}$ which satisfies this property can be found if the size of each bin is sufficiently large, that is, if each $\overline{R}_{\mathbf{i}\to\mathbf{j}}$ is sufficiently big.

F. Input generation

The k^{th} encoder produces the channel input X^N_k as a deterministic function of its codebook(s), i.e.

$$X^N_k = X^N_k\left(\{U^N_{\mathbf{i}\to\mathbf{j}}, \ \forall \ (\mathbf{i},\mathbf{j}) \ \text{ s.t. } \ k \in \mathbf{i}\}\right). \tag{22}$$

Restricting the class of encoding functions to deterministic functions instead of random functions can be done without loss of generality [22].

G. Decoding procedure

Receiver z is required to decode the transmitted messages $W_{\mathbf{i}\to\mathbf{j}}$ for

$$\mathbf{V}^z = \{(\mathbf{i},\mathbf{j}) \in \mathbf{V}, \ \mathbf{j} \in z\}, \tag{23}$$

and it does so by employing a typicality decoder which determines the set of indexes

$$\{w_{\mathbf{i}\to\mathbf{j}}, b_{\mathbf{i}\to\mathbf{j}}, \ (\mathbf{i},\mathbf{j}) \in \mathbf{V}^z\}, \tag{24}$$

such that

$$\left\{Y^N_z, \{U^N_{\mathbf{i}\to\mathbf{j}}(w_{\mathbf{i}\to\mathbf{j}}, b_{\mathbf{i}\to\mathbf{j}}), \ (\mathbf{i},\mathbf{j}) \in \mathbf{V}^z\}\right\} \in \mathcal{T}^n_\epsilon\left(P_{Y_z,\text{encoding}}\right), \tag{25}$$

where

$$P_{Y_z,\text{encoding}} = P_{Y_z|\{U_{\mathbf{i}\to\mathbf{j}}, \ (\mathbf{i},\mathbf{j})\in\mathbf{V}^z\}} \cdot \prod_{(\mathbf{i},\mathbf{j})\in\mathbf{V}^z} P_{U_{\mathbf{i}\to\mathbf{j}}|\mathrm{pa}_{\widetilde{\mathbf{E}}^z}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}})}, \tag{26}$$

for $\widetilde{\mathbf{E}}^z = \widetilde{\mathbf{E}} \cap (\mathbf{V}^z \times \mathbf{V}^z)$. Note that in (26) it is necessary to restrict the parents $U_{\mathbf{i}\to\mathbf{j}}$ to those in $\mathbf{V}^z$ since there could exists a parent of $U_{\mathbf{i}\to\mathbf{j}}$ in the binning graph outside $\mathbf{V}^z$ which is not decoded at receiver z.

A transmission error is committed if any of the receivers decodes (at least) an index incorrectly. It is possible to find a codeword with the desired joint typicality condition if the number of transmitted codewords is sufficiently small, that is if the transmission rate $L_{\mathbf{i}\to\mathbf{j}}$ is low.

V. The Achievable Rate Region of the CGRAS

In this section we derive the achievable region that can be attained with a CGRAS as defined in Def. 3. We do so by considering the following three classes of CGRAS, each more general than the previous: (i) we first consider the CGRAS with only superposition coding, then (ii) the case with superposition coding and binning but no joint binning and finally (iii) the general case. We use the first case to illustrate the decoding error analysis, the second to illustrate the encoding error analysis and the third case to illustrate the effect of joint binning. The decoding error analysis makes use of the packing lemma [7, Sec. 3.2] while the covering lemma [7, Sec. 3.7] is necessary for the encoding error analysis.

A. Superposition Coding

We begin by considering the CGRAS involving only superposition coding. In this case the achievable rate region is expressed as a series of upper bounds on the message rates under which correct decoding occurs with high probability. Each bound relates to the probability that a set of codewords is incorrectly decoded at receiver z and the probability of this event is bounded using the packing lemma [7, Sec. 3.2].

Theorem V.1. Achievable region with superposition coding *Consider any CGRAS employing only superposition coding and let the region $\mathcal{R}$ be defined as*

$$\sum_{(\mathbf{i},\mathbf{j})\in\overline{\mathbf{C}}^z} R_{\mathbf{i}\to\mathbf{j}} \leq \tag{27}$$
$$I(Y_z;\{U_{\mathbf{i}\to\mathbf{j}},\ (\mathbf{i},\mathbf{j})\in\overline{\mathbf{C}}^z\}|\{U_{\mathbf{i}\to\mathbf{j}},\ (\mathbf{i},\mathbf{j})\in\mathbf{C}^z\},Q),$$

for every z and every set $\mathbf{C}^z\subseteq\mathbf{V}^z$ such that

$$\mathrm{pa}_{\mathbf{S}}(\mathbf{C}^z)\subseteq\mathbf{C}^z, \tag{28}$$

then any rate vector $R_{\mathbf{V}}$ which lies in $\mathcal{R}$ is achievable.

Proof: The complete proof is provided in [18]. Each bound in (27) relates to the probability that the codewords in $\mathbf{C}$ are correctly decoded while the ones in $\overline{\mathbf{C}}$ are incorrectly decoded. The channel input is conditionally dependent on the correctly decoded codewords but conditionally independent on the incorrectly decoded ones. A decoding error is committed only if the number of the codewords is high enough for the typicality decoder to find an incorrect codeword that appears jointly typical with the channel output. The probability of this events relates to the mutual information term in the RHS of (27) through the covering lemma. Since the top codeword is chosen conditionally dependent on the bottom codewords, correct decoding is not possible unless all the bottom codewords are also correctly decoded. This condition is expressed by (32).

Note that since all the codewords are chosen from conditionally dependent codebooks, the joint distribution between codewords is the same whether codewords are correctly or incorrectly decoded. As we shall see, this is not the case when the CGRAS also employs binning: in this case incorrect decoding results in a different joint distribution between correctly and incorrectly decoded codewords. ■

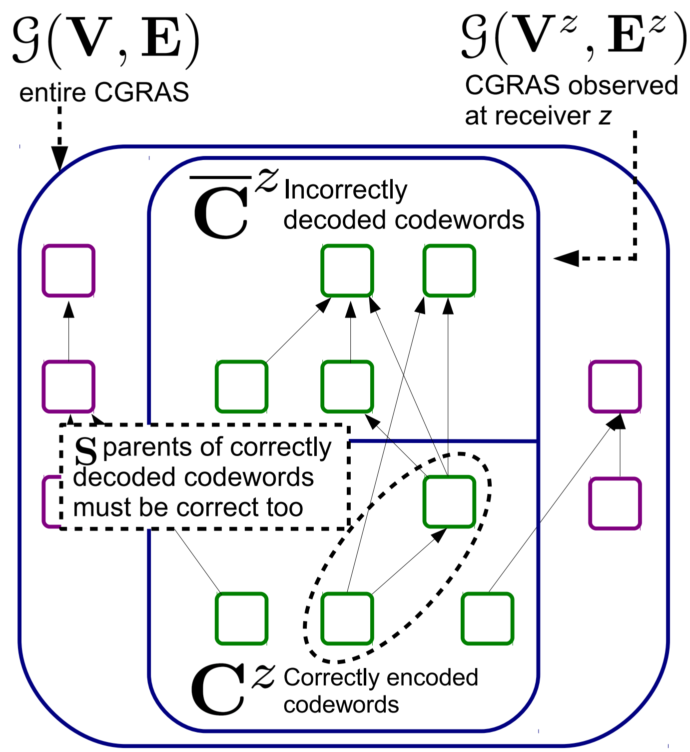

Fig. 3. A graphical representation of Th. V.1.

A graphical representation of Th. V.1 is provided in Fig. 3: the channel output Y_z^N is used at decoder z to decode the set of codeword in $\mathbf{V}^z$ in (23). Since the top codeword is generated conditionally dependent on the bottom codeword, a codeword can be correctly decoded only when all the parents codewords have also been correctly decoded. A lower bound on the message rates can be obtained by considering all the possible error patterns at all the decoders and bounding the probability of each event using the packing lemma.

B. Superposition Coding and One-Way Binning

We now consider the case of a CGRAS employing both superposition coding and binning but no joint binning. This is a more general scenario than the previous section and thus requires considering the probability of an encoding error which results in a lower bound on the binning rates.

The binning graph of a CGRAS with no joint binning is a DAG as no cycle occurs according to Assumption 1. In this context, it is clear that Assumption 1 not only provides the condition for the encoding distribution to be well defined but also for the encoding procedure to be feasible. If Assumption

1 did not hold, than the choice of the bin index for a codeword in the cycle would depend on the value of the bin index itself, which is not feasible.

Theorem V.2. Achievable region with superposition coding and one-way binning
Consider any CGRAS employing only superposition coding, binning but not joint binning and let the region $\mathcal{R}_{\overline{R}}$ *be defined as*

$$\sum_{(\mathbf{i},\mathbf{j})\in\mathbf{C}} \overline{R}_{\mathbf{i}\to\mathbf{j}} \geq \sum_{(\mathbf{i},\mathbf{j})\in\mathbf{V}} I(U_{\mathbf{i}\to\mathbf{j}}; \mathrm{pa}_{\mathbf{B}}\{\mathrm{U}_{\mathbf{i}\to\mathbf{j}}\}|\mathrm{pa}_{\mathbf{S}}\{\mathrm{U}_{\mathbf{i}\to\mathbf{j}}\}|\mathrm{Q}) \tag{29a}$$

$$- \sum_{(\mathbf{i},\mathbf{j})\in\overline{\mathbf{C}}} I(U_{\mathbf{i}\to\mathbf{j}}; \mathrm{pa}_{\mathbf{B}}\{\mathrm{U}_{\mathbf{i}\to\mathbf{j}}\}|\mathrm{pa}_{\mathbf{S}}\{\mathrm{U}_{\mathbf{i}\to\mathbf{j}}\}|\mathrm{Q}), \tag{29b}$$

for all the subsets $\mathbf{C} \subseteq \mathbf{V}$ *such that*

$$\mathrm{pa}_{\mathbf{S}}(\mathbf{C}) \subseteq \mathbf{C}, \tag{30}$$

and the region $\mathcal{R}_L$ *as*

$$\sum_{(\mathbf{i},\mathbf{j})\in\overline{\mathbf{C}}^z} L_{\mathbf{i}\to\mathbf{j}} \leq$$
$$I(Y_z; \{U_{\mathbf{i}\to\mathbf{j}},\ (\mathbf{i},\mathbf{j}) \in \overline{\mathbf{C}}^z\}|\{U_{\mathbf{i}\to\mathbf{j}},\ (\mathbf{i},\mathbf{j}) \in \mathbf{C}^z\}, Q)$$
$$+ \sum_{(\mathbf{i},\mathbf{j})\in\overline{\mathbf{C}}^z} I(U_{\mathbf{i}\to\mathbf{j}}; \mathrm{pa}_{\mathbf{B}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}})|\mathrm{pa}_{\mathbf{S}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}}), \mathrm{Q}), \tag{31a}$$

for every z *and every set* $\mathbf{C}^z \subseteq \mathbf{V}^z$ *such that*

$$\mathrm{pa}_{\mathbf{S}}(\mathbf{C}^z) \subseteq \mathbf{C}^z, \tag{32}$$

then any rate vector $R_{\mathbf{V}}$ *which lies in* $\mathcal{R}_{\overline{R}} \cap \mathcal{R}_L$ *is achievable.*

Proof: The complete proof is provided in [18]. The bounds in (30) are obtained by bounding the encoding error probability: each term corresponds to the event that the bin indexes in $\mathbf{C}$ have been correctly determined while the bin indexes in $\overline{\mathbf{C}}$ are not. As for the decoding error analysis, correct encoding is possible only when the parent codewords in the superposition coding graph have also been correctly encoded. Each bound is obtained applying Tchebychev's inequality to error events in the spirit of [12, Sec. 7.10]. The bounds (30) are obtained by bounding the decoding error probability at each receiver z. These bounds are obtained in a similar manner than in Th. V.1 with the difference of the additional term in (31a): this term corresponds to a "decoding boost" provided by binning. Incorrectly decoded codewords that are binned against each other are generated conditionally independently; for the typicality decoder to commit an error, then, there must exist a codeword which looks as if generated conditionally dependent with the correctly decoded codewords. This makes it harder for the typicality decoder to commit an error. Unfortunately this boost comes at the cost of having to decode the binning index as well as the message index. ■

The novel ingredient in Th. V-B with respect to Th. V.1 is the encoding error analysis which results in the lower bound on the binning rates in (29). During encoding, the binning indexes are chosen so that the transmitted codewords look as if generated according to the encoding distribution although actually distributed according to the codebook distribution. Encoding can be successful only when the parents codewords in the superposition coding graph are also correctly encoded. Since the codebook of the top codeword is chosen conditionally dependent on the chosen base codeword, the base codeword must be correctly encoded for the top codeword to have a desired joint typicality with the rest of the graph.

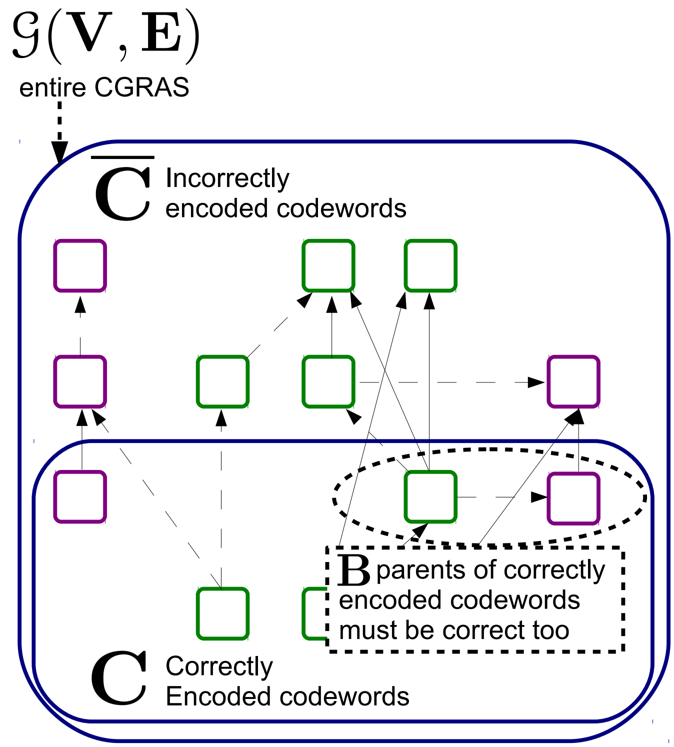

Fig. 4. A graphical representation of Th. V.2.

C. Superposition Coding and Joint Binning

At last we consider the general CGRAS as defined in Def. 3 which includes the case considered in Sec. V-A and in Sec. V-B. For the general case the binning graph is no longer an DAG but rather a CG which requires a more careful analysis of the encoding and decoding error probabilities. This analysis is simplified under the conditions in Ass. 2 which guarantees that any non cyclic orientation of the undirected edges in the binning graph produces a DAG Markov equivalent to the original CG.

Theorem V.3. Achievable region with superposition coding and joint binning
Consider any CGRAS and let the region $\mathcal{R}_{\overline{R}}$ *be defined as*

$$\sum_{(\mathbf{i},\mathbf{j})\in\overline{\mathbf{C}}} \overline{R}_{\mathbf{i}\to\mathbf{j}} \geq \sum_{(\mathbf{i},\mathbf{j})\in\mathbf{V}} I(U_{\mathbf{i}\to\mathbf{j}}; \mathrm{pa}_{\mathbf{B}^-}\{\mathrm{U}_{\mathbf{i}\to\mathbf{j}}\}|\mathrm{pa}_{\mathbf{S}}\{\mathrm{U}_{\mathbf{i}\to\mathbf{j}}\}, \mathrm{Q})$$
$$- \sum_{(\mathbf{i},\mathbf{j})\in\mathbf{C}} I(U_{\mathbf{i}\to\mathbf{j}}; A_{\mathbf{i}\to\mathbf{j}}|\mathrm{pa}_{\mathbf{S}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}}), \mathrm{Q}), \tag{33}$$

with

$$A_{\mathbf{i}\to\mathbf{j}} = \mathrm{pa}_{\mathbf{B}^-}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}}) \cup \mathrm{pa}_{\mathbf{B}\cap(\overline{\mathbf{C}}\times\overline{\mathbf{C}})}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}}) \tag{34}$$

for any $\mathbf{C} \subseteq \mathbf{V}$ *such that*

$$\mathrm{pa}(\mathbf{C}) \subseteq \mathbf{C}, \tag{35}$$

and the region $\mathcal{R}_{\mathbf{L}}$ *as*

$$\begin{aligned}
&\sum_{(\mathbf{i},\mathbf{j})\in\overline{\mathbf{C}}^z} L_{\mathbf{i}\to\mathbf{j}} \leq \\
&\quad I(Y_z; \{U_{\mathbf{i}\to\mathbf{j}},\ (\mathbf{i},\mathbf{j}) \in \overline{\mathbf{C}}^z\} | \{U_{\mathbf{i}\to\mathbf{j}},\ (\mathbf{i},\mathbf{j}) \in \mathbf{C}^z\}, Q) \\
&\qquad + \sum_{(\mathbf{i},\mathbf{j})\in\overline{\mathbf{C}}^z} I(U_{\mathbf{i}\to\mathbf{j}}; A^z_{\mathbf{i}\to\mathbf{j}} | \mathrm{pa}_{\mathbf{S}}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}}), \mathrm{Q}),
\end{aligned} \tag{36}$$

with

$$A^z_{\mathbf{i}\to\mathbf{j}} = \mathrm{pa}_{\mathbf{B}^-}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}}) \cup \mathrm{pa}_{\mathbf{B}\cap(\overline{\mathbf{C}}^z\times\overline{\mathbf{C}}^z)}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}}) \tag{37}$$

for all z *and all the sets* $\mathbf{C}^z \subseteq \mathbf{V}^z$ *such that*

$$\mathrm{pa}(\mathbf{C}^z) \subseteq \mathbf{C}^z, \tag{38}$$

then any rate vector $R_{\mathbf{V}}$ *which lies in* $\mathcal{R}_{\overline{R}} \cap \mathcal{R}_{\mathbf{L}}$ *is achievable.*

Proof: The complete proof is provided in [18]. The decoding and encoding error analysis from Th. V.1 and Th. V.2 must be updated in order to consider the effect of joint binning. For the encoding error analysis, joint binning of two RVs is successful only when it is possible to jointly determine the bins at both codewords. If the encoding of either bin fails, then the codewords appear as if generated conditionally independent. This intuitively motivates the last mutual information term of (33): the elements in $\mathrm{pa}_{\mathbf{B}\cap(\overline{\mathbf{C}}\times\overline{\mathbf{C}})}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}})$ are those elements for which one RV in the joint binning as been correctly decoded but the other is not.

Joint binning has a similar effect on the last mutual information term in (36). Consider again two RVs which are jointly binned against each other, then the correct joint typicality between the two decoded codewords occurs only when both bin indexes are correctly decoded. The term $\mathrm{pa}_{\mathbf{B}\cap(\overline{\mathbf{C}}^z\times\overline{\mathbf{C}}^z)}(\mathrm{U}_{\mathbf{i}\to\mathbf{j}})$ corresponds to the jointly binned random variables for which only one of the two bin indexes have bee correctly decoded. ■

From the statements of Th. V.2 and Th. V.3 it can be verified that joint binning strictly improves over binning as it results in a smaller RHS in (30) and a larger RHS in (36) which corresponds to smaller binning rates and larger message rates respectively. Despite this, it is yet not clear whether joint binning improves on the convex closure of the achievable regions obtained by the possible orientation of the undirected edges.

VI. Conclusion

This paper presents a general achievable scheme valid for a wide class of single-hop multi-terminal networks. This achievable scheme employs coded time-sharing, rate-splitting, superposition coding, and binning and generalizes a number of inner bounds proposed in the literature. A compact representation of the transmission strategies is offered by graphical Markov models which are also useful in deriving the achievable rate region.

Extensions of this work will consider multi-hop networks and the generalization of transmission techniques such Markov encoding, amplify-and-forward and quantize-and-forward to this general channel model.

References

[1] S.A. Andersson, D. Madigan, and M.D. Perlman. On the markov equivalence of chain graphs, undirected graphs, and acyclic digraphs. *Scandinavian Journal of Statistics*, 24(1):81–102, 1997.

[2] H.F. Chong, M. Motani, H.K. Garg, and A. El Gamal. On The Han–Kobayashi Region for theInterference Channel. *IEEE Trans. Inf. Theory*, 54(7):3188–3195, 2008.

[3] T.M. Cover. Comments on broadcast channels. *IEEE Trans. Inf. Theory*, 44(6):2524–2530, Oct 1998.

[4] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, 1991.

[5] Imre Csiszar and Janos Körner. *Information theory: Coding theorems for discrete memoryless channels*. Budapest: Akademiai Kiado, 1981.

[6] AP Dawid, RG Cowell, SL Lauritzen, and DJ Spiegelhalter. Probabilistic networks and expert systems. *Quality*, 1999.

[7] A. El Gamal and Y.H. Kim. *Network information theory*. Cambridge University Press, 2011.

[8] S.I. Gel'fand and M.S. Pinsker. Coding for channel with random parameters. *Problems of control and information theory*, 9(1):19–31, 1980.

[9] Te Han and K. Kobayashi. A New Achievable Rate Region for the Interference Channel. *IEEE Trans. Inf. Theory*, 27(1):49–60, Jan 1981.

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

[11] G. Kramer. Capacity results for the discrete memoryless network. *Information Theory, IEEE Transactions on*, 49(1):4 – 21, jan 2003.

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

[13] S. Lall. Advanced topics in computation for control. *Lecture Notes for Engr210b at Stanford University, Stanford, CA*, 2005.

[14] S.L. Lauritzen. *Graphical models*, volume 17. Oxford University Press, USA, 1996.

[15] Y. Liang and G. Kramer. Rate regions for relay broadcast channels. *IEEE Trans. Inf. Theory*, 53(10):3517–3535, 2007.

[16] K. Marton. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory*, 25(3):306–311, 1979.

[17] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[18] S. Rini and Goldsmith A. A general approach to random coding for multi-terminal networks. *In Preparation*, 2013.

[19] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[20] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, 19(4):471–480, 1973.

[21] J. Whittaker. *Graphical models in applied multivariate statistics*, volume 16. Wiley New York, 1990.

[22] F. Willems and E. Van der Meulen. The discrete memoryless multiple-access channel with cribbing encoders. *IEEE Trans. Inf. Theory*, 31(3):313–327, 1985.