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

Stefano Rini,
Department of Electrical Engineering, Stanford University,
Stanford, CA 94305, USA,
Email: stefano@wsl.stanford.edu

***Abstract*—Random coding, along with various standard techniques such as coded time-sharing, superposition coding, rate-splitting and binning, are traditionally used in obtaining achievable rate regions in multi-terminal information theory. The corresponding error analysis relies heavily on the properties of strongly joint typical sequences. In this work, we obtain an achievable rate region for a general (i.e., arbitrary set of messages shared amongst encoding nodes, which transmit to arbitrary decoding nodes) memoryless networks without feedback/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 *chain graph representation* of the encoding operations. This graph representation captures the statistical relationship among codewords and allows 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 classical random coding techniques for any memoryless network used without feedback/cooperation.**

I. Introduction

Random coding was first introduced by Shannon in his seminal 1948 paper [1] to prove the channel coding theorem. In Shannon's words:

> The method of proving [capacity of a discrete channel with noise] is not by exhibiting a coding method having the desired properties, but by showing that such a code must exist in a certain group of codes.

In random coding, codewords are generated by drawing symbols in an independent, identically distributed (i.i.d.) fashion from a prescribed distribution; the performance of the ensemble of codes is analyzed and is a function of the block-length, which is eventually taken to infinity. Thanks to the i.i.d. symbols, and 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. This proof technique was originally developed for the point-to-point channel but is easily extended to multi-user channels by introducing a dependency structure among codewords. Time-sharing, rate-spitting, superposition coding, binning, Markov encoding, compress and forward, decode and forward are some of the strategies that have been developed for multi-terminal channels using the random coding proof technique. 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. Kramer [2] and El Gamal [3] have identified the key bounding techniques, called "covering lemma" and "packing lemma" [3], for bounding the error events of random codes. By building upon these results, 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 define a form of "automatic rate region generator" which outputs the best known random coding achievable rate region for any channel of choice.

A. Main Contributions

In this paper:

1) **We propose a novel *chain graph* representation for encoding schemes based on standard random coding techniques. This representation is for a general, single-hop, memoryless multi-terminal network used without feedback/cooperation where the number of transmitters and receivers, as well as the set of messages amongst transmitters, is arbitrary.** This new formalism provides a clear and unified framework to represent achievability schemes based on random coding arguments; it includes and generalized all known achievable schemes for the class of networks considered.
2) **We derive the achievable rate region based on of the proposed *chain graph* representation** The proposed chain graph representation naturally leads to an algorithm that automatically outputs the achievable rate region for any channel in the class of networks considered.

B. Paper Organization

Section II presents the class of networks considered in this work and revises the standard random coding techniques that are employed in the literature to derive achievable scheme. Section III introduces the novel *chain graph* representation of the encoding operations. Section IV describes the codebook generation, encoding and decoding procedures. Section V derives the rate bounds that define the achievable rate region based on the proposed *chain graph* representation. Section VIII concludes the paper.

II. Channel Model and Random Coding Techniques for Achievability

A. Notation

In order to designate arbitrary subsets of transmitter and receiver nodes we use the following notation:

$$S_i \triangleq \{ k, \ k \in i \}, \quad \forall \ i \subset \left[1 \dots 2^{\max\{N_{\rm RX}, N_{\rm TX}\}} \right] \qquad (1)$$

i.e. is a set containing a specific subset of encoder or decoders. To compactly index the messages from certain transmitting nodes to other receiving nodes, we introduce the notation

$$S_{\boldsymbol{i}} \triangleq \{ (j,k) \in \boldsymbol{i} \}, \quad \forall \ \boldsymbol{i} \subset \left[1 \dots 2^{N_{\rm RX}} \right] \times \left[1 \dots 2^{N_{\rm TX}} \right] \quad (2)$$

We adopt the following conventions for superscripts and subscripts:

- index k/z: transmitters/receivers and channel inputs/outputs
- index i/j: subset S_i/S_j of transmitters/receivers. We will also use l/m and v/t.

B. Network Model

We consider a multi-terminal network where $N_{\rm TX}$ transmitting nodes want to communicate with $N_{\rm RX}$ receiving nodes. A given node may only be a transmitting or a receiving node (and may not alternate between them), that is, the network is single-hop and it is used without feedback/cooperation. The transmitting node k, $k \in [1 : N_{\rm TX}]$, inputs X_k to the channel, while the receiving node z, $z \in [1 : N_{\rm RX}]$, has access to the channel output Y_z. The channel transition probability is indicated with $P_{Y_1,\dots,Y_{N_{\rm RX}}|X_1,\dots,X_{N_{\rm TX}}}$. The channel is assumed to be memoryless.

The subset of transmitting nodes S_i, $i \in [1 : 2^{N_{\rm TX}} - 1]$, is interested in sending the message $W_{i\to j}$ to the subset of receiving nodes S_j, $j \in [1 : 2^{N_{\rm RX}} - 1]$. The total number of messages is $(2^{N_{\rm TX}} - 1)(2^{N_{\rm RX}} - 1)$ and included all form of "degraded message sets"/cognition. The message $W_{i\to j}$, $(i,j) \in [1 : 2^{N_{\rm TX}} - 1] \times [1 : 2^{N_{\rm RX}} - 1]$, is uniformly distributed on the interval $[0 : 2^{N R_{i\to j}} - 1]$, where N is the block-length and $R_{i\to j}$ the transmission rate. The outcome of the Random Variable (RV) $W_{i\to j}$ is denoted with $w_{i\to j}$ and the set of all messages is denoted by $\boldsymbol{w} = [w_{1\to 1}, \dots, w_{2^{N_{\rm TX}}-1 \to 2^{N_{\rm RX}}-1}]^T$. A rate vector $\boldsymbol{R} = [R_{1\to 1}, \dots, R_{2^{N_{\rm TX}}-1 \to 2^{N_{\rm RX}}-1}]^T$ is said to be achievable if there exists a sequence of encoding functions

$$X_k^N = X_k^N \Big(\{ \quad W_{i\to j}, \ (i,j) \in [1 : 2^{N_{\rm TX}} - 1] \times [1 : 2^{N_{\rm RX}} - 1] : k \in S_i \} \Big), \quad \forall \ k \in S_i$$

and a sequence of decoding functions

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

for all $(i,j) \in [1 : 2^{N_{\rm TX}} - 1] \times [1 : 2^{N_{\rm RX}} - 1]$ such that

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

The capacity region $\mathcal{C}$ is the convex closure of the region of all achievable rates in the vector $\boldsymbol{R}$-pairs .

$\mathcal{C}(\overline{\boldsymbol{R}})$ denotes the capacity region restricted to the plane $\overline{\boldsymbol{R}} = R_{i_1,j_1} \times R_{i_2,j_2} \times, \dots,$. This corresponds to the capacity region of a sub-channel where some of the rates $R_{i\to j}$ have been set to zero. Fig. 1 shows the channel model considered in this work.

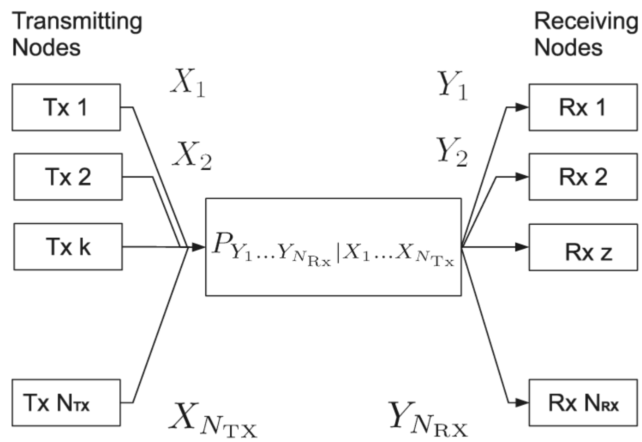

Fig. 1. The general cognitive multi-terminal network.

All memoryless single-hop networks used without feedback/cooperation are included in the class of networks considered in this networks, such as the Multiple Access Channel (MAC) [4], [5], the Broadcast Channel (BC) [6], [7], [8], the InterFerence Channel (IFC) [9], [10], [11], the Cognitive IFC (CIFC) [12], [13], [14], [15].

C. Example: the interference channel with two sources and two destinations

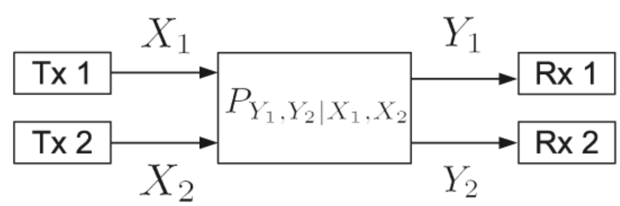

Fig. 2. The general IFC with the most general set of messages amongst transmitters.

An example of the channel included in the class of networks described in Section II is the IFC with two transmitters and two receivers in Fig. 2. The messages to be transmitted between transmitters and receivers are listed in Table I. Table II lists all the classical information theoretical models that are subcases of the general IFC.

D. Random Coding Techniques for Achievability

We revise here standard random coding techniques used in the literature for achievability in single-hop networks used without feedback/cooperation.

• **Coded Time Sharing** consists of using different transmission strategies at different time instants [11] and allows one to achieve the convex closure of the set of achievable points. Let q denote an instance of the time-sharing RV Q with alphabet

TABLE I
THE MESSAGES FOR A GENERAL IFC.

	from Tx1	from Tx2	from both Tx's
to Rx1	$W_{1\to 1}$	$W_{2\to 1}$	$W_{\{1,2\}\to 1}$
to Rx2	$W_{1\to 2}$	$W_{2\to 2}$	$W_{\{1,2\}\to 2}$
to both Rx's	$W_{1\to\{1,2\}}$	$W_{2\to\{1,2\}}$	$W_{\{1,2\}\to\{1,2\}}$

TABLE II
SPECIFIC SUBCASES OF THE GENERAL INTERFERENCE CHANNEL

subcase	channel model
$\mathcal{C}\left(R_{1\to 1}\right)$	point-to-point
$\mathcal{C}\left(R_{1\to 1}, R_{2\to 1}, R_{\{1,2\}\to 1}\right)$	MAC with common message
$\mathcal{C}\left(R_{1\to 1}, R_{1\to 2}, R_{1\to\{1,2\}}\right)$	BC with common message
$\mathcal{C}\left(R_{1\to 1}, R_{2\to 2}, R_{\{1,2\}\to\{1,2\}}\right)$	IFC with common message
$\mathcal{C}\left(R_{1\to 1}, R_{\{1,2\}\to 2}\right)$	CIFC
$\mathcal{C}\left(R_{1\to 1}, R_{\{1,2\}\to\{1,2\}}\right)$	CIFC with degraded message set
$\mathcal{C}\left(R_{1\to\{1,2\}}, R_{2\to\{1,2\}}\right)$	compound MAC

$\mathcal{Q}$ and $\boldsymbol{R}_q$ denote the rate achievable under the strategy q. The achievable rate with time-sharing is

$$\boldsymbol{R} = \sum_{q\in\mathcal{Q}} \mathbb{P}[Q=q]\boldsymbol{R}_q.$$

• **Rate Splitting** corresponds to a scheme where a message is split in different sub-messages that are then encoded and decoded separately which make it possible to apply different encoding techniques for each sub-message [11]. The message $W_{i\to j}$ can be split in a sequence of sub-messages $W_{l\to m}^{[i\to j]}$ for every (l,m) such that

$$S_j \subset S_m \text{ and } S_i \supset S_l.$$

That is, $W_{l\to m}^{[i\to j]}$ is encoded by a smaller number of encoders and decoded by a larger set of decoders. The sub-messages $W_{l\to m}^{[i\to j]}$ are uniformly distributed over the interval $[0:2^{NR_{l\to m}^{[i\to j]}}-1]$ so that

$$R'_{i\to j} = \sum_{(l,m)} R_{l\to m}^{[i\to j]} = \sum_{(l,m)} \gamma_{l\to m}^{[i\to j]} R_{i\to j}$$

for

$$\gamma_{l\to m}^{[i\to j]} = \frac{R_{l\to m}^{[i\to j]}}{R_{i\to j}} \qquad \sum_{(l,m)} \gamma_{l\to m}^{[i\to j]} = 1. \tag{3}$$

Rate-splitting effectively transforms the problem of achieving a rate vector $\boldsymbol{R}$ into the problem of achieving the rate vector $\boldsymbol{R}'$ where

$$R'_{l\to m} = \sum_{(i,j)} R_{l\to m}^{[i\to j]} = \sum_{(i,j)} \gamma_{l\to m}^{[i\to j]} R_{i\to j}, \tag{4a}$$

$$\boldsymbol{R}' = \Gamma \boldsymbol{R}, \tag{4b}$$

and where the element in position $(i,j)\times(l,m)$ of the matrix Γ is the coefficient $\gamma_{l\to m}^{[i\to j]}$ in (3). Rate-splitting is useful in the cases where it allows to increase the number of messages in the channel, thus effectively increasing the viable transmission strategies.

• **Superposition Coding** can be intuitively thought of as stacking codewords on top of each other [6]. The "bottom" codewords are decoded first and stripped from the received signal so to reduce the interference when decoding the "top" codewords. Let $U_{i\to j}$ be the RV with distribution $P_{U_{i\to j}}$ carrying the message $W'_{i\to j}$ obtained through the rate-splitting matrix Γ: when the RV $U_{i\to j}$ is superposed to the RV $U_{l\to m}$, the former may depend on the latter, that is $U_{i\to j}$ may be generated according any distribution $P_{U_{i\to j}|U_{l\to m}}$. By introducing dependency among the codewords, superposition coding increases the error performance of the code. To see this, notice that the incorrect decoding of any of two messages relates to an incorrect joint distribution among the corresponding codewords. If the two messages where encoded in independent codewords, the joint distribution among them would not change with a decoding error. Superposition of one RV $U_{i\to j}$ over another RV $U_{l\to m}$ can be performed when the following two conditions hold:

- $S_l \subset S_i$: that is the bottom message is encoded by a larger set of encoders than the top message.
- $S_m \subset S_j$: that is the bottom message is decoded by a larger set of decoders than the top message.

If $U_{i\to j}$ is superposed to $U_{l\to m}$ and $U_{v\to t}$ is superposed to $U_{l\to m}$, than $U_{i\to j}$ is also superposed to $U_{v\to t}$. Similarly, if $U_{i\to j}$ is superposed to $U_{l\to m}$, then the reverse cannot hold. The binary relation *"A is superposed to B"* therefore establishes a partial ordering amongst the encoding RVs denoted by $A \overline{>} B$.

• **Binning** sometimes referred to as Gel'fand-Pinsker coding [16], allows a transmitter to "pre-cancel" (portions of) the interference known to be experienced at a receiver. When the encoding RV $U_{i\to j}$ in binned against $U_{l\to m}$, $U_{i\to j}$ is generated independently from $U_{l\to m}$ but chosen so to look as if generated according to the distribution $P_{U_{i\to j}|U_{l\to m}}$. In order to find a codeword that appears to have the desired marginal distribution, it is necessary to produce more codewords than $2^{NR_{i\to j}}$ and the number of excess codewords depends on the joint distribution $P_{U_{i\to j},U_{l\to m}}$. As for superposition coding, binning introduces correlation between two decoding errors but only if a receiver decodes both messages. When this is not the case, binning is effective in increasing the error performance of the code by introducing a dependency between the desired codeword, $U_{i\to j}$, and the interfering one, $U_{l\to m}$. With the appropriate choice of conditional distribution, the encoder can reduce the effect of the interference at the intended decoder. The RV $U_{i\to j}$ can be binned against the RV $U_{l\to m}$ when

- $S_i \subset S_l$: that is the binning RV must have knowledge of the interfering RV

Note that $U_{i\to j}$ can be binned against $U_{i,m}$ and vice-versa, regardless of the value of j and m. This is commonly referred to as "joint binning" [17]. The relationship *"A can be binned against B"* is transitive and symmetric for the RV known at the same set of decoders. In the following "binning $U_{i\to j}$ against the RV $U_{l\to m}$" is indicated as $U_{i\to j} \lessdot U_{l\to m}$.

By combining these four encoding techniques one can design an achievability scheme with specific characteristics. Although the code design detailed above is fully general, some coding choices result in an ill defined construction. Consider for instance the code

$$U_{1\to 1} \lessdot U_{1\to 2} \lessdot U_{1\to\{1,2\}} \lessdot U_{1\to 1}.$$

In this case it is not possible to define a joint probability distribution for $U_{1,1}, U_{1,2}, U_{14}$ because of the cyclic dependency between the RVs. We address this issue in the next section.

III. A Graphical Representation of Achievability Schemes

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

- every vertex $v \in V$ of the $\mathcal{G}$ is associated to a RV $U_{i\to j}$ carrying the message $w'_{i\to j}$ at rate $R'_{i\to j}$ obtained through the rate-splitting Γ from rate vector $\boldsymbol{R}$ as in (3),
- the vertex $U_{i\to j}$ is connected with $U_{l\to m}$ by a directed edge of type ***S*** (for *superposition*), if $U_{i\to j} \overline{\lessdot} U_{l\to m}$ (solid line).
- the vertex $U_{i\to j}$ is connected with $U_{l\to m}$ by a directed edge of type ***B*** (for *binning*), if $U_{i\to j} \lessdot U_{l\to m}$ (dotted line).

Since $\overline{\lessdot}$ is a transitive relationship, it is convenient to omit the edges of type ***S*** that are implied by transitivity. The time-sharing RV Q is not represented in this graph as it is assumed that each RV is generated according to a marginal distribution that depends on Q; the overall region may then be composed of a time-sharing of the regions obtained by the marginal distributions.

The graph representation of the achievable scheme is particularly useful in deriving the joint distribution of the coding RVs $U_{i\to j}$. When trying to determine this distribution, superposition coding and binning effectively result in allowing for any joint distribution among the connected RVs.

Graphs representing conditional dependencies among RVs have been extensively studied in the literature [18] and we can utilize such results to determine the joint distribution of the RVs $U_{i\to j}$. That is, if the achievable scheme employs only superposition coding, $\mathcal{G}$ is an Acyclic DiGraph (ADG) since the **S** relationship is transitive and anti-symmetric. In an ADG the joint distribution of the RVs is obtained as

$$P_{\{U_{i\to j},\ \forall\ (i,j)\}} = \prod_{(i,j)} P_{U_{i\to j}|\ \{U_{l\to m},\ U_{i\to j} \overline{\lessdot} U_{l\to m}\}}$$

If the edges **B** edges are all directed, then the corresponding graph is still an ADG. If $\mathcal{G}$ possesses only undirected edges ,obtained both either two edges of type **B** or one of type **B** and one edge of type and **S**, then it is possible to obtain a (non-unique) joint distribution if the graph is *chordal*, i.e. if every cycle of length $n > 3$ has a *chord* – two non consecutive edges that are neighbors [18]. For the most general case, a necessary condition to obtain a fully joint probability distribution is for the graph to be a *chain graph*, or a mixed graph – a graph with both directed and undirected nodes that contains no directed cycles. A more detailed discussion of the conditions under which the graph representation of an achievable scheme defines a joint probability distribution can be found in [19].

In the following we will assume that all the chain components of the graph $\mathcal{G}$ are complete. That is, if $U_{i\to j}$ and $U_{i,m}$ are connected by an undirected edge and $U_{i,m}$ and $U_{i,t}$ are also connected by an undirected edge, then $U_{i\to j}$ and $U_{i,t}$ are as well. Under this condition the graph representation of the achievable scheme is Markov equivalent to an ADG [18] obtained by choosing the proper orientation of the undirected edges in the original graph. A modified Maximum Cardinality Search (MCS) algorithm can be used to efficiently construct a Markov equivalent ADG from a decomposable chain graph $\mathcal{G}$ [18]. Since undirected edges are generated by ***B*** edges, it is possible to determine a smaller set of edges, ***B-***, so that the graph that has ***B-*** and ***S*** is an ADG. The joint distribution can then be written as

$$P_{\{U_{i\to j}\forall(i,j)\}} = \prod_{(i,j)} P_{U_{i\to j}|\{U_{l\to m},\ U_{i\to j} \overline{\lessdot} U_{l\to m} \text{ or } U_{i\to j} \overleftarrow{\lessdot} U_{l\to m}\}} \tag{5}$$

where $\overleftarrow{\lessdot}$ represents the edges in ***B-*** and are indicated with a dashed line.

A. An example of our graphical representation: the CIFC

We now consider the CIFC, the subcase $C(R_{1,1}, R_{3,2})$ from Section II-C to illustrate the coding procedure and the construction of the graph representing the achievability scheme.

The message w_{11} in rate-split into $w'_{1,1}, w'_{1,3}$ and the message $w_{3,2}$ into $w'_{3,2}, w'_{2,2}, w'_{1,2}$ and $w'_{3,3}$, i.e. $\boldsymbol{R}' = \Gamma\boldsymbol{R}$ for

$$\begin{bmatrix} R'_{1\to 1} \\ R'_{1\to 2} \\ R'_{1\to\{1,2\}} \\ R'_{2\to 2} \\ R'_{\{1,2\}\to 2} \\ R'_{\{1,2\}\to\{1,2\}} \end{bmatrix} = \begin{bmatrix} \gamma_{1\to 1}^{[1\to 1]} & 0 \\ \gamma_{1\to 2}^{[1\to 1]} & \gamma_{1\to 2}^{[\{1,2\}\to 2]} \\ \gamma_{1\to\{1,2\}}^{[1\to 1]} & 0 \\ 0 & \gamma_{\{1,2\}\to 2}^{[\{1,2\}\to 2]} \\ 0 & \gamma_{\{1,2\}\to\{1,2\}}^{[\{1,2\}\to 2]} \end{bmatrix} \cdot \begin{bmatrix} R_{1\to 1} \\ R_{\{1,2\}\to 2} \end{bmatrix}$$

Superposition may be performed as long as

$$U_{\{1,2\}\to 2} \overline{\lessdot} U_{\{1,2\}\to\{1,2\}}, \tag{6a}$$

$$U_{2\to 2} \overline{\lessdot} U_{\{1,2\}\to 2}, \tag{6b}$$

$$U_{1\to 2} \overline{\lessdot} U_{\{1,2\}\to 2}, \tag{6c}$$

$$U_{1\to 1} \overline{\lessdot} U_{1\to\{1,2\}}, \tag{6d}$$

and binning may be performed as long as

$$U_{1\to\{1,2\}} \lessdot U_{\{1,2\}\to 2}, \tag{7a}$$

$$U_{1\to 1} \lessdot U_{\{1,2\}\to 2}, \tag{7b}$$

$$U_{1\to 1} \lessdot U_{1\to 2} \lessdot U_{1\to 2} \lessdot U_{1\to 1}. \tag{7c}$$

An achievability scheme may be obtained from any feasible combination of these encoding steps. We consider the achievable scheme obtained by combining all these steps but the $U_{1\to 1} \lessdot U_{1\to\{1,2\}}$ and $U_{1\to 2} \lessdot U_{1\to\{1,2\}}$ in order to avoid cycles. The resulting achievability scheme results in a more general scheme that the scheme of [14] which contains the largest known achievable rate region for the CIFC and achieves the

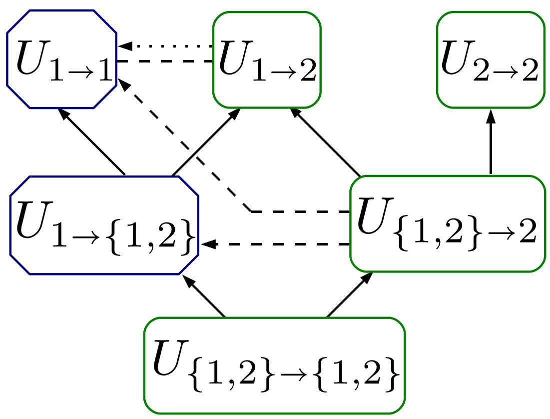

Fig. 3. The graph representation of an achievable scheme in Section III-A for the CIFC.

outer bound for all the channel conditions for which capacity is known. The scheme we propose here has one RV more than the scheme in [14] – $U_{2,2}$. This RV carries a private message between transmitter 2 and receiver 2 that is not known at the cognitive encoder. The graph corresponding to the proposed achievability scheme has one undirected edge. By choosing a direction for this edge we obtain a ADG that is Markov equivalent to the original graph; in this case both directions are can be selected . The joint probability distribution of this graph is

$$\begin{aligned}
&P_{U_{\{1,2\}\to\{1,2\}},U_{\{1,2\}\to2},U_{1\to\{1,2\}},U_{1\to1},U_{1\to2},U_{2\to2}} = \\
&\quad P_{U_{\{1,2\}\to\{1,2\}}} P_{U_{\{1,2\}\to2}|U_{\{1,2\}\to\{1,2\}}} \\
&\quad P_{U_{1\to\{1,2\}}|U_{2\to2},U_{\{1,2\}\to\{1,2\}}} P_{U_{2\to2}|U_{\{1,2\}\to2}\to U_{\{1,2\}\to\{1,2\}}} \\
&\quad P_{U_{1\to2}|U_{1\to\{1,2\}}\to U_{\{1,2\}\to2}\to U_{\{1,2\}\to\{1,2\}}} \\
&\quad P_{U_{1\to1}|U_{1\to2}\to U_{\{1,2\}\to2}\to U_{1\to\{1,2\}}\to U_{\{1,2\}\to\{1,2\}}}.
\end{aligned}$$

The graph representation of this region in Fig. 3: the solid line represent the ***S*** edges, the dashed lines the ***B*** edges and the dotted lines indicate the direction in ***B-*** to be associated with the undirected edges. The solid edges that are implied by the transitivity of the $\overline{<}$ have been omitted from the plot. The blue,rhomboidal nodes carry a rate-split of the $w_{1,1}$ message while the green, square ones of the message $w_{3,2}$.

IV. Code-book Generation, Encoding and Decoding procedures

We now outline the code-book generation, encoding, and decoding operations at each node. The codewords are generated by drawing i.i.d. symbols according to the joint distribution associated with the ***S*** edges of $\mathcal{G}$. They are furthermore chosen so as to appear as if generated according to the distribution associated with edges ***S*** and ***B-***. Finally the channel inputs are produced as function of the codewords known at each transmitter. At each receiver, the decoder looks for a set of codewords that possesses the correct conditional distribution given the received channel outputs. The conditions under which encoding and decoding errors vanish as the block-length goes to infinity are determined in the next sections.

A. Code-book generation

Given a specific coding strategy, specified by a rate-splitting matrix Γ and a chain graph, each message $w'_{i\to j}$ is encoded by $2^{NL_{i\to j}}$ codewords $U^N_{i\to j}$, with

$$L_{i\to j} = R'_{i\to j} + \sum_{\substack{(l,m):\\ U_{i\to j} \lessdot U_{l\to m}}} R^{[l\to m]}_{i\to j} + \sum_{\substack{(l,m):\\ U_{i\to j} \overline{<} U_{l\to m}}} L_{l\to m}, \tag{8}$$

with i.i.d. symbols drawn according to the distribution $P_{U_{i\to j}|\{U_{l\to m},\ U_{l\to m}\overline{<}U_{i\to j}\}}$ to achieve the joint distribution

$$P_{(\text{code-book})} = \prod_{(i,j)} P_{U_{i\to j}|\{U_{l\to m}:\ U_{l\to m}\overline{<}U_{i\to j}\}} \tag{9}$$

among the codewords. The codewords are placed in $(2^{N_{\mathrm{TX}}}-2)(2^{N_{\mathrm{RX}}}-2)$ bins, indexed by $b^{[l\to m]}_{i\to j}$. The size of the bin $b^{[l\to m]}_{i\to j}$ is determined by the relationship between $U_{i\to j}$ and $U_{l\to m}$:

- $b^{[l\to m]}_{i\to j}$ indices the encoded message $w'_{i\to j}$ and the size of the bin is $R'_{i\to j}$,
- if $U_{i\to j}\overline{<}U_{l\to m}$, $b^{[l\to m]}_{i\to j}$ is the superposition coding index and the size of the bin is $L_{l\to m}$,
- if $U_{i\to j} \lessdot U_{l\to m}$, $b^{[l\to m]}_{i\to j}$ is the binning index and the size of the bin is $R^{[l\to m]}_{i\to j}$,
- if $U_{i\to j}$ and $U_{l\to m}$ are not connected, then the size of the bin $b^{[l\to m]}_{i\to j}$ is zero.

B. Encoding procedure

In the encoding procedure the bin indices $b^{[l\to m]}_{i\to j}$ of the codewords $U^N_{i\to j}$ are jointly chosen so that the codewords appear to have been generated with i.i.d. symbols drawn from the distribution

$$P_{(\text{encoding})} = \prod_{(i,j)} P_{U_{i\to j}|\{U_{l\to m},\ U_{i\to j}\overline{<}U_{l\to m}\ \text{or}\ U_{i\to j}\overleftarrow{\lessdot}U_{l\to m}\}} = (5) \tag{10}$$

instead of the code-book distribution (9). We may find such jointly typical codewords if the number of bins $b^{[l\to m]}_{i\to j}$ is sufficiently large.

Finally, node k's encoder produces the channel input X^N_k as a deterministic function of its code-book(s), i.e.

$$X^N_k = X^N_k\left(\{U^N_{i\to j},\ \forall (i,j):\ k\in S_{i\to j}\}\right).$$

In [19] we have shown that there is no loss of generality in considering X^N_k to be a deterministic function of the codewords $U^N_{i\to j}$ instead of a RV.

C. Decoding procedure

To decode the transmitted messages, each receiver z looks for a set of bin indices $b^{[l\to m]}_{i\to j}$ for $z\in S_j$, such that the set $\{Y^N_z, \{U^N_{i\to j}\ :\ z\in S_j\}\}$ looks as if generated i.i.d. according to the distribution $P_{Y_z,\{U_{i\to j}:\ z\in S_j\}}$ An error is committed if any of the receivers decodes (at least) a bin index incorrectly. Note that the decoders, although interested only in the decoding of the messages $w'_{i\to j}$, are decoding all the indices $b_{i\to j}^{[l\to m]}$ that index the codeword $U_{i\to j}^N$. The decoding of the bin indices other than $b_{i\to j}^{[l\to m]} = w'_{i\to j}$ reduces the probability of error as it requires the decoded messages to be distributed according to the joint imposed by the encoding procedure.

V. Derivation of the Rate Bounds

In this section we derive the achievable rate region of the proposed inner bound by using properties of jointly typical sequences. As the codewords are generated by symbols drawn i.i.d. from the appropriate distributions, their empirical distribution as $N \to \infty$ approaches the true generating one. This enables one to bound the probability of encountering codewords that appear to be generated according to a different distribution than the true generating distribution. In particular we make use of two general bounds:

- **Mutual covering lemma** [20], [3] bounds the minimum number of independent i.i.d. sequences $U_{i\to j}^N$, $U_{l\to m}^N$ that need to be generated in order to find two that look as if generated according to a joint distribution $P_{U_{i\to j},U_{l\to m}}^N$,
- **Packing lemma** [10], [3] bounds the maximum number of independent sequences $U_{i\to j}^N$, $U_{l\to m}^N$ that may be generated so that no two of them look as if generated according to a joint distribution $P_{U_{i\to j},U_{l\to m}}^N$.

Both bound are based on standard properties of jointly typical sequences [21] and are common tools used to derive achievable rate regions: our contribution lies in the generalization of the error analysis corresponding to our encoding and decoding scheme. Without loss of generality, the error analysis may be bounded as

$$
\begin{aligned}
&\mathbb{P}[\text{error}] \leq \\
&\quad \mathbb{P}[\text{encoding NOT successful}] \\
&\quad + \mathbb{P}[\text{decoding NOT successful}|\text{encoding IS successful}].
\end{aligned}
$$

In the following sections we provide bounds on the rates $L_{i\to j}$ and $R_{i\to j}^{[l\to m]}$ such that the probability of encoding and decoding error vanishes as $N \to \infty$.

VI. Encoding Errors

For the probability of encoding error to vanish as the block-length increases it is necessary to choose large enough binning rates $R_{i\to j}^{[l\to m]}$ so that it is possible to jointly find a set of bin indices $b_{i\to j}^{[l\to m]}$ for which the codeword $U_{i\to j}$ appears generated according to the conditional distribution in (10) although generated according to the conditional distribution in (9). The encoding error probability then depends on all the possible combinations in which an encoding error may be committed, that is on all the possible ways of failing to find the appropriate bin indices or, equivalently, jointly typical codewords. We index the encoding errors using the notation in (2) and defining $U_{i\to j}(S_{\boldsymbol{i}_{i\to j}})$ to be the RV $U_{i\to j}$ that possess the joint distribution $P_{U_{i\to j},U_{l\to m}}$ in (10) for all $(l,m) \in S_{\boldsymbol{i}_{i\to j}}$. The elements in $S_{\boldsymbol{i}_{i\to j}}$ correspond to the bin indexes of the codeword $U_{i\to j}^N$ that can be successfully chosen in the encoding procedure. Among all the combination of encoding errors, the index $b_{i\to j}^{[l\to m]}$ where $U_{i\to j} \overline{<} U_{l\to m}$ can always be successfully be determined since the codewords are generated according to the desired marginal distribution; the same is true when $U_{i\to j}$ and $U_{l\to m}$ are not connected. Also $b_{i\to j}^{[i\to j]} = w'_{i\to j}$ by definition. For these reasons we only need to consider the cases where

$$
\begin{aligned}
&S_{\boldsymbol{i}_{i\to j}} : \ (i,j) \in S_{\boldsymbol{i}_{i\to j}}, \ (l,m) \in S_{\boldsymbol{i}_{i\to j}} \\
&\quad \forall \ (l,m) : \ U_{i\to j} \overline{<} U_{l\to m} \text{ or } U_{i\to j} \ \not\overline{\lessdot} U_{l\to m}.
\end{aligned} \tag{11}
$$

where $\not\overline{\lessdot}$ indicates the RVs are not connected. We also expect the rate bounds to depend on the quantity in (12) which intuitively measures the distance between the distributions used to generate the code-books and those seen after encoding.

Theorem VI.1. Encoding errors: *the encoding procedure is successful with high probability as $N \to \infty$ if*

$$
\sum_{i\to j} \left(\sum_{(l,m)\notin S_{\boldsymbol{i}_{i\to j}}} R_{i\to j}^{[l\to m]} + \boldsymbol{I}_{S_{\boldsymbol{i}_{i\to j}}} \right) \geq \boldsymbol{I}_{(\text{code}-\text{book})}^{(\text{encoding})} \tag{13}
$$

for

$$
\begin{aligned}
&\boldsymbol{I}_{S_{\boldsymbol{i}_{i\to j}}} \triangleq \\
&I \Big(U_{i\to j}; \{U_{l\to m}, \ U_{i\to j} \overleftarrow{\lessdot} U_{l\to m}, \ (l,m) \ \notin \ S_{\boldsymbol{i}_{i\to j}}\} | \\
&\quad \{U_{l\to m}, \ U_{i\to j} \overline{<} U_{l\to m}\} \cup \\
&\qquad \{U_{l\to m}, U_{i\to j} \overleftarrow{\lessdot} U_{l\to m} \ , \ (l,m) \ \in \ S_{\boldsymbol{i}_{i\to j}}\} \Big)
\end{aligned} \tag{14}
$$

for all $S_{\boldsymbol{i}_{i\to j}}$ for which (11) *holds and such that*

$$
\begin{aligned}
&\{U_{i\to j}(S_{\boldsymbol{i}_{i\to j}}) \not\perp U_{l\to m}(S_{\boldsymbol{i}_{(l,m)}}) | \\
&\quad \{U_{v\to t}(S_{\boldsymbol{i}_{(v,t)}}) : \ (v,t) \neq (i,j), (v,t) \neq (l,m)\}\},
\end{aligned} \tag{15}
$$

where $\perp$ indicates the independence among RVs.

Proof: The detailed proof is provided in [19]. Here we provide an intuitive interpretation of the result. $S_{\boldsymbol{i}}$ is the set of all the possible errors in the finding the bins that impose the desired conditional distribution. The condition in (12) determines the general distance between the distribution of codewords in the code-book and the encoding distribution. The terms $S_{\boldsymbol{i}_{i\to j}}$ contains the set of all the bin index associated with $U_{i\to j} \lessdot U_{l\to m}$ that can be correctly determined. We may reduce the number of possible encoding error events by noticing that if $U_{i\to j} \lessdot U_{l\to m}$ but the two RVs are conditionally independent in a specific error event, then the probability of this error event is dominated by the event where at least one bin index correct $b_{i\to j}^{[l\to m]}$ and $b_{l\to m}^{[i\to j]}$ and where the two RV are conditionally dependent. ∎

VII. Decoding Errors

The analysis of the decoding errors is similar to the encoding counterpart in that we need to consider all the possible decoding error events. Each decoder z is decoding the messages $w'_{i\to j}, \ j \in S_j$, so, for a fixed z we need to consider all the combinations

$$
S_{\boldsymbol{i}_{i\to j}} : \ z \in S_j.
$$

$$\mathrm{E}\left[\log \frac{P_{(\text{encoding})}}{P_{(\text{code}-\text{book})}}\right] = \sum_{(i,j)} I(U_{i\to j}; \{U_{l\to m},\ U_{i\to j} \overleftarrow{\lessdot} U_{l\to m}\} | \{U_{l\to m},\ U_{i\to j} \overline{<} U_{l\to m}\}) \triangleq \boldsymbol{I}^{(\text{encoding})}_{(\text{code}-\text{book})} \tag{12}$$

If $U_{i\to j}$ is superposed or binned against $U_{l\to m}$ that has been incorrectly decoded, then the decoding of $U_{i\to j}$ will fail with high probability as well. For this reason we only need to consider the combinations for which

$$S_{\boldsymbol{i}_{i\to j}} : (l,m) \in S_{\boldsymbol{i}_{i\to j}} \implies (v,t) \in S_{\boldsymbol{i}_{i\to j}},$$
$$\forall\ (v,t):\ U_{i\to j} \overline{<} U_{v\to t} \ \text{ or } \ U_{i\to j} \lessdot U_{v\to t} \tag{16}$$

Theorem VII.1. Decoding errors: *if the following condition holds,*

$$\sum_{(i,j)\ \notin\ S_{\boldsymbol{i}_{i\to j}}} \left(L_{i\to j} - \boldsymbol{I}_{S_{\boldsymbol{i}_{i\to j}}}\right) \le$$
$$I\left(Y_z; \{U_{i\to j},\ (i,j) \notin S_{\boldsymbol{i}}\} | \{U_{l\to m},\ (l,m) \in S_{\boldsymbol{i}}\}\right) \tag{17}$$

for $\boldsymbol{I}_{S_{\boldsymbol{i}_{i\to j}}}$ *defined in* (14)*, for all* $z \in S_j$ *and for all* $S_{\boldsymbol{i}_{i\to j}}$ *for which* (16) *holds, decoding is successful with high probability.*

Proof: Again we provide only an intuitive argument and provide the full details in [19]. As in the encoding error analysis of Th. VI.1, the term $\boldsymbol{I}_{S_{\boldsymbol{i}_{i\to j}}}$ accounts for the joint distribution among the codewords that have been incorrectly decoded. Intuitively, the more dependency is imposed among the codewords, the less likely the decoder is to decode the incorrect codeword. The RHS of (17) is the measure of the amount of information about the incorrectly decoded RVs contained in the channel output z given the correctly decoded RVs.

■

Theorem VII.2. Achievable region *A rate vector* $\boldsymbol{R}$ *is achievable for a channel in Section II if there exists a rate-splitting matrix* Γ *and an achievable scheme such that* (13), (17) *and* (8) *are satisfied for some choice of* $U_{i\to j}$ *according to the distribution* (5).

Th. VII.2 defines the conditions under which a certain rate point is achievable by imposing that the rate bounds of Th. VI.1 and Th. VII.1 are not violated.

VIII. 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 graphical representation that allows for a quick comparison between transmission strategies and a simplified derivation of the resulting achievable rate region. This paper attempts to establish a general tool to derive achievable rate regions for multi-terminal networks which contain 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). It is commonly believed that superposition coding enlarges the achievable rate region relative to code-books derived from conditionally independent codewords. The same conjecture holds for binning but is less clear that this is the case when considering the union over all the possible distribution of the binning RVs. We believe that the general formulation for achievability schemes and corresponding regions proposed here is a power tool to answer resolve these conjecture, with the ultimate goal of determining optimal coding strategies for general channels.

Acknowledgment

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

References

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

[2] G. Kramer, *Topics in Multi-User Information Theory*, ser. Foundations and Trends in Communications and Information Theory. Vol. 4: No 4?5, pp 265-444, 2008.

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

[4] R. Ahlswede, “Multi-way communication channels,” in *Proc. 2nd Int. Symp. Information Theory*, 1971, pp. 103–135.

[5] H. Liao, “Multiple Access Channels,” University of Hawaii at Manoa, Tech. Rep., 1972.

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

[7] P. Bergmans, “Random coding theorem for broadcast channels with degraded components,” *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 197–207, 1973.

[8] R. Gallager, “Capacity and coding for degraded broadcast channels,” *Problems in the Transmission of Information*, vol. 10, no. 3, pp. 3–14, 1974.

[9] C. Shannon, “Two-way communication channels,” in *Proc. 4th Berkeley Symp. Math. Stat. Prob*, vol. 1. IEEE Press, 1961, pp. 351–384.

[10] R. Ahlswede, “The capacity region of a channel with two senders and two receivers,” *The annals of probability*, vol. 2, no. 5, pp. 805–814, 1974.

[11] 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.

[12] N. Devroye, P. Mitran, and V. Tarokh, “Cognitive Multiple Access Networks,” in *Proc. IEEE International Symposium on Information Theory (ISIT), Adelaide, Australia*, 2005, pp. 57–61.

[13] I. Maric, R. Yates, and G. Kramer, “The Capacity Region of the Strong Interference Channel with Common Information,” in *Proc. Asilomar Conferenece on Signal, Systems and Computers*, Nov. 2005, pp. 1737–1741.

[14] S. Rini, D. Tuninetti, and N. Devroye, “New Inner and Outer Bounds for the Discrete Memoryless Cognitive Channel and some Capacity Results,” *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4087 – 4109, Jul. 2011.

[15] Y. Liang, A. Somekh-Baruch, H. Poor, S. Shamai, and S. Verdú, “Cognitive interference channels with confidential messages,” *Arxiv preprint arXiv:0710.2018*, 2007.

[16] 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.
[17] K. Marton, "A Coding Theorem for the Discrete Memoryless Broadcast Channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 3, pp. 306–311, May 1979.
[18] 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.
[19] S. Rini, "An achievable region for a general multi-terminal network," *Information Theory, IEEE Transactions on*, 2011, in preparation.
[20] A. El Gamal and E. Van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel (Corresp.)," *Information Theory, IEEE Transactions on*, vol. 27, no. 1, pp. 120 – 122, jan 1981.
[21] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley-Interscience, New York, 1991.