CCIT Report #669 November 2007

Communication via Decentralized Processing

Amichai Sanderovich*,Shlomo Shamai (Shitz)*,Yossef Steinberg*and Gerhard Kramer†

*Department of Electrical Engineering, Technion, Haifa 32000, Israel

†Bell Laboratories, Alcatel-Lucent, NJ, USA

Email:{amichi@tx,sshlomo@ee,ysteinbe@ee}.technion.ac.il, gkr@bell-labs.com

Abstract

The problem of a nomadic terminal sending information to a remote destination via agents with lossless connections to the destination is investigated. Such a setting suits, e.g. access points of a wireless network where each access point is connected by a wire to a wireline-based network. The Gaussian codebook capacity for the case where the agents do not have any decoding ability is characterized for the Gaussian channel. This restriction is demonstrated to be severe, where allowing the nomadic transmitter to use other signaling improves the rate. For both general and degraded discrete memoryless channels, lower and upper bounds on the capacity are derived. An achievable rate with unrestricted agents, which are capable of decoding, is also given and then used to characterize the capacity for the deterministic channel.

Index Terms

Cooperative reception, decentralized detection, relay channel, wireless networks

I. Introduction

Information theory for networks and especially wireless networks is in the focus of an extensive research activity. This interest is partly due to many recent results on multiple antenna channels, which demonstrate significant gains, especially for fading channels.

Many papers propose and analyze ad-hoc wireless networks in information theoretic terms. Among these, coding schemes which achieve $O(n)$ transport capacity were given in [1]. Multi-hop relaying makes use of several intermediate wireless nodes to assist the communication between two nodes that are far apart. An information theoretic framework for the relay channel was given by Cover and El Gamal in [3] for a single relay node and extended by [4] to several relaying nodes. Relaying techniques can be coarsely divided into compress-and-forward and decode-and-forward, depending on whether the relays attempt to decode the transmitted message or just forward the processed received signal to the destination. By using cooperation, relaying schemes can take advantage of the inherent dependencies for efficient forwarding to the final destination. Such cooperation is commonly used and selected examples are [5],[6], while cooperation between receiving nodes in a degraded broadcast channel is described in [7]. We conclude with an upper bound derived by [8], that suggests that as the number of users in an ad-hoc network goes to infinity, the total rate per user tends to zero. This bound motivates the use of networks that are not solely ad-hoc, but are composed of base stations or access points as well.

The problems of conveying a source which is observed by remote agents to a single destination are built around similar settings, where the source is modelled as an i.i.d. random variable. Several aspects of these problems are analyzed in information theoretic frameworks such as distributed source coding, lossless CEO (Chief Executive Officer) [9], CEO [10] and sensor networks. A small sample from the extensive work that is relevant to our distributed detection setting includes [11],[12] and [9] for distributed source coding. Allowing lossy source encoding, in a distributed scheme, as opposed to non-distributed [13] is still essentially an unsolved problem. An exception is the Gaussian CEO problem [14],[15] which was recently solved using the entropy power inequality by [16],[17]. Multi-terminal lattice approaches are described in [18]. These rate-distortion problems are linked to network models in [19],[20] and [21]. The use of other measures, instead of the plain distortion, is addressed, for example in [22] and [23]. The dissertation of Schein [24] focuses on the characteristics of the problem of communicating via two agents, and several achievable rates are demonstrated there.

Here we consider the problem of reliable communication from a nomadic transmitter to a remote destination, through non-decoding agents which are connected to the final destination via lossless links. The agents have the noisy version (via their respective channels) of the transmitted message, and are able to transmit a predetermined number of bits to the destination without any errors. The destination is reached only via the agents which serve as the noisy channel access points. By nomadic transmitter we mean that the receiving devices can not and will not decode the transmitted signal. Only the remote final destination will decode the transmitted message. Such a setting is of interest to numerous applications. The main motivation, however, is for systems where the agents can not decode because of added noise or interference. We also consider the less restrictive case, where the agents are informed about the transmitter's code, and give several achievable rates, which turn up to be capacity achieving for the deterministic channel.

The rest of the paper is organized as follows: in section II the setting of the problem is given. An achievable rate and an upper bound for the nomadic transmitter are presented in sections III and IV respectively. An achievable rate for the case of cognizant agents is given for both degraded and non-degraded channels in section V, where the capacity is fully characterized for the deterministic channel. The Gaussian channel is presented as an example in section VI which for the case where the agents are unaware of the code used, and where the codebook is Gaussian, also includes a characterization of the rate-region.

In this paper we use capital letters X for random variables, lower case letters x for the realization of these variables, and calligraphic letters $\mathcal{X}$ for their alphabets. Vectors are denoted by bold face letters $\boldsymbol{X}$, $\boldsymbol{x}$, or vector spaces by $\boldsymbol{\mathcal{X}}$, and are of length n, unless otherwise specified. A calligraphic letter denotes a set, e.g. $\mathcal{T} \triangleq \{1,\ldots,T\}$. A complement (denoted by the superscript C) of some subset $\mathcal{S}$ of a set $\mathcal{T}$ refers to the subset $\mathcal{S}^C$ which fulfills: $\mathcal{S} \cup \mathcal{S}^C = \mathcal{T}$ and $\mathcal{S} \cap \mathcal{S}^C = \phi$. Cardinality of any set $\mathcal{T}$ is written as $|\mathcal{T}|$. A subscript, e.g. X_i, denotes the i-th element in the vector $\boldsymbol{X}$ and a superscript X^n denotes the vector $(X_1,\ldots,X_n)$. The notation X_k^m refers to the vector $(X_k,\ldots,X_m)$, and $X_{\mathcal{S}}$ refers to $\{X_i\}_{i\in\mathcal{S}}$. Let $P_{A_1,A_2,\ldots,A_L}(a_1,a_2,\ldots,a_L)$ be the probability function of the random variables $A_1,\ldots,A_L$ which take values in $\mathcal{A}_1,\ldots,\mathcal{A}_L$, respectively.

II. Problem settings

We consider the problem of a single transmission from the transmitter S through T agents, playing the role of decentralized processors, to the final destination D, seen in Fig. 1 for $T = 2$. Our model consists of a nomadic transmitter S, which uses random coding, where the agents do not know the codebook used. An example for this setting is the case where the transmitter uses one code out of a group of possible channel codes, where the receiving agents are not equipped with the corresponding decoder. However, the agents are still needed to allow communication with the destination. What the agents do know, is some characteristics of the group of codes, which include their rate, and that they are capacity achieving over a standard single user Gaussian channel. An example can be a group of codes (Turbo, Convolution, Reed-Solomon etc.), a group of interleavers, code-polynomials, puncturing patterns within a specific code, and also a group of modulation techniques. Such random coding is also used in [25] for a mis-match scenario, while the advantages of random coding were demonstrated in [26] for unknown channels.

The following properties and definitions hold, unless stated otherwise:

1) The channel input is $X \in \mathcal{X}$ for every channel use (output of the transmitter S).
2) The T agents $A_1, \dots, A_T$ receive the outputs of a memoryless broadcast channel without feedback, defined by
$$P_{\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_T|\boldsymbol{X}}(\boldsymbol{y}_1,\dots,\boldsymbol{y}_T|\boldsymbol{x}) = \prod_{i=1}^{n} P_{Y_1,\dots,Y_T|X}(y_{1,i},\dots,y_{T,i}|x_i), \tag{1}$$
where $y_{t,i} \in \mathcal{Y}_t$. Denote $\mathcal{T} \triangleq \{1,\dots,T\}$. The agents have full knowledge of the distribution P_X, induced by the nomadic transmission, and thus also of $P_{\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_T}(\boldsymbol{y}_1,\dots,\boldsymbol{y}_T)$.
3) The bandwidth C_t, in bits per channel use, characterizes the lossless link that connects the agent A_t to the final destination D.
4) The communication rate is denoted by R. The message M to be sent is encoded by a random encoding function $\boldsymbol{X} = \phi_{S,F}(M)$ such that for all messages M, the outputs of the encoding function are randomly and independently chosen according to probability $P_{\boldsymbol{X}}(\boldsymbol{x})$. We index the random encoding function by the random variable F. We define the range of F to be $[1,2,\dots,|\mathcal{X}|^{n2^{nR}}]$, which is the number of ways of mapping 2^{nR} messages to the $|\mathcal{X}|^n$ possible codewords. Then let every f correspond to a unique such mapping, i.e. $F = f$ corresponds to one such mapping. That is, we choose
$$\phi_{S,F} : [1,\dots,2^{nR}] \to \mathcal{X}^n, \tag{2}$$
and the probability of selecting f is:
$$P_F(f) = \prod_{m=1}^{2^{nR}} P_{\boldsymbol{X}}(\phi_{S,f}(m)), \tag{3}$$
where $P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} P_X(x_i)$, for some single letter probability P_X. The agents are not informed about the selected encoding F, but are fully aware of P_X.

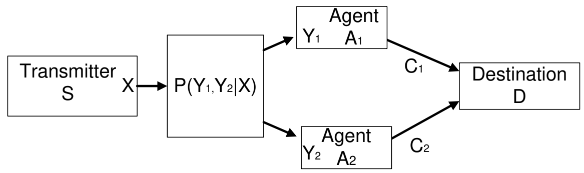

Fig. 1. A system with two agents between the transmitter and the destination.

5) Every agent t, $t = 1, 2, \ldots, T$, encodes its n channel outputs with an encoding function

$$\phi_{At} : \boldsymbol{\mathcal{Y}}_t \rightarrow [1, \ldots, 2^{nC_t}] \tag{4}$$

so that

$$V_t = \phi_{At}\left(\boldsymbol{Y}_t\right). \tag{5}$$

V_t is sent through a lossless link to the final destination.

6) The destination decodes the message M from $V_{\mathcal{T}}$, i.e. we have

$$\hat{M} = \phi_{D,F}(V_{\mathcal{T}}), \tag{6}$$

where $\phi_{D,F} : [1, \ldots, 2^{\sum^{T} nC_t}] \rightarrow [1, \ldots, 2^{nR}]$.

7) The rate R is said to be achievable if for every $\epsilon > 0$, there exists n sufficiently large such that:

$$\frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} \Pr(\hat{M} \neq m | M = m) \leq \epsilon, \tag{7}$$

where $\Pr(\hat{M} \neq m | M = m)$ includes averaging over the channel and the random coding.

Notice that with the knowledge of F, with high probability, $\boldsymbol{X}$ is uniformly distributed over 2^{nR} codewords. However, with knowledge of F we have the following simple lemma

Lemma 1: Without the knowledge of the selected encoding F, the vector $\boldsymbol{X}$ is distributed according to $P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} P_X(x_i)$, and therefore $\boldsymbol{Y}_t$ is distributed as

$$P_{\boldsymbol{Y}_t}(\boldsymbol{y}_t) = \prod_{i=1}^{n} \sum_{x} P_{Y_t|X}(y_{t,i}|x) P_X(x). \tag{8}$$

Proof: See Appendix II. ■

The above setting models the problem where the final destination decodes the message from the transmitter via simple agents, which are not able to decode the transmitted message and use compression of the received signals. When the agents are allowed to decode, as is the case in section V, then obviously randomized encoding is superfluous. However, in order to allow combined approaches, and for the sake of consistency, we use the same settings also for when the agents are informed about the coding used.

III. An achievable rate for nomadic transmitter

We denote the setting of section II as *nomadic transmitter*. The following theorem is a special case of Theorem 3 (proved in Appendix III) applied to the nomadic setting. In fact, by proper modeling, Theorem 3 is also a special case of [31]. But here we give cardinality constraints and the proof is simpler because there is no need for the block Markov superposition encoding.

Theorem 1: Define a positive rate R and a set of auxiliary random variables $U_{\mathcal{T}}$, with bounded cardinalities of $|\mathcal{U}_t| \leq |\mathcal{Y}_t| + 2^{T-1}$ such that

$$R < I(X; U_{\mathcal{T}}) \tag{9}$$

with the constraints

$$\forall \mathcal{S} \subseteq \mathcal{T} : \sum_{t \in \mathcal{S}} C_t > I(U_{\mathcal{S}}; Y_{\mathcal{S}} | U_{\mathcal{S}^C}) \tag{10}$$

and the joint distribution

$$P_{X,U_{\mathcal{T}},Y_{\mathcal{T}}}(x, u_{\mathcal{T}}, y_{\mathcal{T}}) = P_X(x) P_{y_{\mathcal{T}}|X}(y_{\mathcal{T}}|x) \prod_{t=1}^{T} P_{U_t|Y_t}(u_t|y_t). \tag{11}$$

Then R is achievable for nomadic transmitter.

Proof: See Appendix III and Remark 1. ■

We remark that (11) means that

$$U_t \;-\; Y_t \;-\; \{X, U_{\mathcal{T}\setminus t}, Y_{\mathcal{T}\setminus t}\}. \tag{12}$$

forms a Markov chain. The auxiliary random variables $U_{\mathcal{T}}$ are used to compress $Y_{\mathcal{T}}$, and are forwarded to the final destination. The constraints (10) are required so that the final destination can reliably recover $\boldsymbol{U}_{\mathcal{T}}$ from $V_{\mathcal{T}}$.

Corollary 1: The achievable rate of Theorem 1 can be improved by taking into account only errors that involve incorrect X, where the destination is allowed to make errors in $U_{\mathcal{T}}$. Such approach gives an achievable rate which is written with no constraints, albeit we feel is less intuitive. Rate R is achievable if:

$$R < \min_{\mathcal{S} \subseteq \mathcal{T}} \left\{ \sum_{t \in \mathcal{S}} [C_t - I(U_t; Y_t | X)] + I(U_{\mathcal{S}^C}; X) \right\}, \tag{13}$$

where the cardinality and the probability space of the random variables are the same as in Theorem 1.

Proof: See Appendix IV. ■

IV. An upper bound for nomadic transmitter

An upper bound for capacity of the communication problem described in section II is given by the following theorem, which is based on the fact that the agents do not know the selected encoding F. The problem is thus similar to the general CEO problem in the sense that the transmitter source sequence should be reproduced. Meaning that the transmitter is a source which should be reproduced at the destination with a modified distortion. Since the agents are ignorant of the codebook used, there is an inherent gap compared with the case where the agents know the codebook used. The achievable rate when the agents are unrestricted can be upper bounded by the cut-set upper bound [27]. This gap between the achievable rate and the cut-set bound will be demonstrated for the Gaussian channel, in section VI.

Theorem 2: Any reliable communication rate R for the nomadic setting (section II) must satisfy

$$R \leq \max I(X; U_{\mathcal{T}}), \tag{14}$$

where $U_{\mathcal{T}}$ must fulfill the constraints:

$$\forall \mathcal{S} \subseteq \mathcal{T}: \ \sum_{t\in\mathcal{S}} C_t \geq I(U_{\mathcal{S}}; Y_{\mathcal{T}} | U_{\mathcal{S}^C}). \tag{15}$$

The maximization in (14) is over $(X, Y_{\mathcal{T}}, U_{\mathcal{T}}, W)$ which are distributed according to:

$$\begin{cases} P_{X,Y_{\mathcal{T}},W}(x, y_{\mathcal{T}}, w) = P_X(x) P_{Y_{\mathcal{T}}|X}(y_{\mathcal{T}}|x) P_W(w) \\ \forall\ 0 < t \leq T: \ u_t = f_t(w, y_t) \end{cases} \tag{16}$$

for some random variable W and for some deterministic functions $\{f_t\}_{t\in\mathcal{T}}$. The cardinality of W is $|\mathcal{W}|$, where it suffices to use $|\mathcal{W}| \leq |\mathcal{Y}_{\mathcal{T}}| + 2^{T-1}$.

Proof: The theorem is proved in Appendix V. ■

We remark that (16) means that

$$U_t - Y_t - \{X, Y_{\mathcal{T}\setminus t}\} \tag{17}$$

forms a Markov chain. At first look, it seems that the right-hand side of (10) is lower than the r.h.s. of (15), which would result in a contradiction between the necessary conditions of Theorem 2 and the sufficient conditions of Theorem 1. This apparent conflict is resolved by observing the different Markov relations the variables $U_{\mathcal{T}}$ fulfill, where (11) is more restrictive than (16).

Furthermore, when taking the variables $U_{\mathcal{T}}$ in the upper bound such that they fulfill (11), the r.h.s. of (15) is identical to the r.h.s. of (10). This is since

$$I(Y_{\mathcal{T}}; U_{\mathcal{S}} | U_{\mathcal{S}^C}) = I(Y_{\mathcal{S}}; U_{\mathcal{S}} | U_{\mathcal{S}^C}) + I(Y_{\mathcal{S}^C}; U_{\mathcal{S}} | U_{\mathcal{S}^C}, Y_{\mathcal{S}}) = I(Y_{\mathcal{S}}; U_{\mathcal{S}} | U_{\mathcal{S}^C}) \tag{18}$$

where $I(Y_{\mathcal{S}^C}; U_{\mathcal{S}} | U_{\mathcal{S}^C}, Y_{\mathcal{S}}) = I(Y_{\mathcal{S}^C}; U_{\mathcal{S}} | Y_{\mathcal{S}}) = 0$ because of the Markov relations (12).

Corollary 2: Similarly to Corollary 1, we can give an expression with no constraints also for the upper bound. This upper bound is

$$R < \max_{P_W, \{f_t\}} \left\{ \min_{\mathcal{S}\subseteq\mathcal{T}} \left\{ \sum_{t\in\mathcal{S}} [C_t - I(U_t; Y_t | X)] + I(U_{\mathcal{S}^C}; X) \right\} \right\}, \tag{19}$$

where again, the random variables have the same cardinalities and the same Markov chains as in Theorem 2.

Proof: See Appendix VI. ■

V. Agents with code knowledge

In this section we diverge from the nomadic model described in section II. Suppose the agents know the codebook so that the agents and the transmitter can be jointly optimized. This enables to transmit a broadcast message that is decoded by the agents and forwarded to the destination, in addition to the compression operation. Denote this model as *decoding agents*. Such an approach can increase the overall transmission rate.

Obviously, the use of randomized encoding is superfluous here, as the agents are fully informed about the selected coding. Nonetheless, to remain consistent, the same setting as in the nomadic case is used, where the only difference is with the knowledge of F also at the agents.

In the following, we will denote all messages that are decoded at the agents as broadcast messages, although eventually they are always intended for the same final destination.

The next theorem is based on Marton's scheme [28] to the broadcast channel. Denote by M_t the message to be decoded at agent t, and let $M = (M_{\mathcal{T}}, M_{CF})$ (M_{CF} is the message that is decoded only at the final destination).

Theorem 3: For the decoding agents case, any rate R satisfying

$$R < I(X; U_{\mathcal{T}}|W_{\mathcal{T}}) + \sum_{t=1}^{T} R_t, \tag{20}$$

with the constraints

$$\begin{cases} \forall\ 0 < t \leq T:\ \ 0 \leq R_t \\ \forall\ \mathcal{S} \subseteq \mathcal{T}:\ \ \sum_{t\in\mathcal{S}} R_t < \sum_{t\in\mathcal{S}}[I(W_t; Y_t) - I(W_t; W_{\tilde{\mathcal{T}}(\mathcal{S},t)})] \\ \sum_{t\in\mathcal{S}}[C_t - R_t] > I(U_{\mathcal{S}}; Y_{\mathcal{S}}|U_{\mathcal{S}^C}, W_{\mathcal{T}}) \end{cases} \tag{21}$$

on $R_{\mathcal{T}}, U_{\mathcal{T}}, W_{\mathcal{T}}$ (W_0 is a constant). Where $\tilde{\mathcal{T}}(\mathcal{S}, t) \triangleq \{i:\ i \in \mathcal{S} \text{ and } i < t\}$, with the joint distribution

$$P_{X,Y_{\mathcal{T}},W_{\mathcal{T}},U_{\mathcal{T}}}(x, y_{\mathcal{T}}, w_{\mathcal{T}}, u_{\mathcal{T}}) = P_{W_{\mathcal{T}}}(w_{\mathcal{T}})P_{X|W_{\mathcal{T}}}(x|w_{\mathcal{T}})P_{Y_{\mathcal{T}}|X}(y_{\mathcal{T}}|x) \cdot \prod_{t=1}^{T} P_{U_t|Y_t,W_t}(u_t|y_t, w_t), \tag{22}$$

is achievable.

The agent A_t decodes nR_t bits and forwards them to the destination along with $n(C_t - R_t)$ bits used for the compression. This compression is done considering the decoded signal $\boldsymbol{w}_t$. The final destination then decides on the transmitted M_{CF} by using joint typicality for the compressed signals, taking into account $\boldsymbol{w}_{\mathcal{T}}$. The above scheme uses the auxiliary random variables $W_{\mathcal{T}}$ for the messages that will be decoded at the agents, and U_t which depends on W_t, for the compression outcomes that will be decoded at the final destination.

This achievable rate may be further increased by using a convex hull over the rate region which indicated by Theorem 3.

Proof: The proof appears in Appendix III and uses compression in addition to Marton's broadcast (not including common messages). ■

Remark 1: The scheme described in Theorem 1 is obtained as a special case of the above scheme, by taking all $W_{\mathcal{T}}$ to be constants. The cardinality limits in Theorem 1 can be calculated from the limits in Appendix III-F.

Remark 2: The achievable rate in Theorem 3 can be written without $\{R_t\}$ by solving the following linear programming problem: given $P_{X,Y_{\mathcal{T}},W_{\mathcal{T}},U_{\mathcal{T}}}$ which satisfies (22), maximize R from (20), over $R_{\mathcal{T}}$. Using this approach we get that any rate R is achievable if it satisfies:

$$R < I(X; U_{\mathcal{T}}|W_{\mathcal{T}}) + \min_{\mathcal{S}\subseteq\mathcal{T}} \left\{ \sum_{t\in\mathcal{S}} C_t - I(U_{\mathcal{S}}; Y_{\mathcal{S}}|U_{\mathcal{S}^C}, W_{\mathcal{T}}) + \sum_{t\in\mathcal{S}^C}[I(W_t; Y_t) - I(W_t; W_{\tilde{\mathcal{T}}(\mathcal{S}^C,t)})] \right\}, \tag{23}$$

Provided that:

1) $\forall\ \mathcal{S} \subseteq \mathcal{T}$:

$$\sum_{t\in\mathcal{S}} C_t \geq I(U_{\mathcal{S}}; Y_{\mathcal{S}}|U_{\mathcal{S}^C}, W_{\mathcal{T}}) \tag{24}$$

2) $\forall\ \mathcal{S} \subseteq \mathcal{T}$:

$$0 \leq \sum_{t\in\mathcal{S}}[I(W_t;Y_t) - I(W_t;W_{\tilde{\mathcal{T}}(\mathcal{S},t)})]. \tag{25}$$

See the proof in Appendix VIII.

Remark 3: Another improvement upon (20) is done by sending common broadcast messages in addition to the individual broadcast messages to the agents, so that they could divide the bandwidth required for forwarding decoded messages between their links. This is done by extending Theorem 2 in [28] to more than two users and adding compression. Notice that such construction contains Theorem 3 and Theorem 4 (to follow) as special cases. Such scheme is given in Appendix VII for two agents.

Corollary 3: For the case of deterministic channel, where $Y_t = g_t(X)$, for some deterministic functions g_t, the cut-set upper bound is:

$$R \leq \min_{\mathcal{S}\subseteq\mathcal{T}} \left\{ H(Y_{\mathcal{S}^C}) + \sum_{t\in\mathcal{S}} C_t \right\}. \tag{26}$$

This rate is achievable from (23) by taking $\forall\ t:\ U_t = Const, W_t = Y_t$, which fulfills the conditions (24) and (25). So the capacity region is fully characterized for the deterministic channel (This is a special case of the main result in [32]).

For the case where the channels $P_{Y_t|X}$ are either stochastically or physically degraded (see [27] subsection 14.6.2), that is Y_t is better than Y_{t-1} (without loss of generality), we can use superposition coding, which is known to achieve capacity over degraded broadcast channels.

The received signal Y_2 is physically degraded compared to Y_1 if the following Markov chain

$$X - Y_1 - Y_2 \tag{27}$$

is satisfied. Notice that this relation leaves $I(X;Y_1,Y_2) = I(X;Y_1)$. On the other hand, stochastically degraded Y_2 means that the marginal probability $P_{Y_2|X}(y_2|x)$ can be calculated from $P_{Y_1|X}(y_1|x)$ through some $P_{Y_2|Y_1}(y_2|y_1)$ (See equation (28)). Since (27) is not necessarily true for stochastic degradedness, we have that $I(X;Y_1,Y_2) \geq I(X;Y_1)$. So although superposition coding is optimal for the degraded broadcast channel, it is not necessarily optimal for our model.

Theorem 4: For decoding agents with a channel $P_{Y_\mathcal{T}|X}(y_\mathcal{T}|x)$ that satisfies

$$\forall\ 0 < t \leq T:\ P_{Y_{t-1}|X}(y_{t-1}|x) = \sum_{y_t} P_{Y_t|X}(y_t|x)P_{Y_{t-1}|Y_t}(y_{t-1}|y_t), \tag{28}$$

any rate R satisfying (20) with the constraints

$$\begin{cases} \forall\ 0 < t \leq T : 0 \leq R_t \leq I(W_t;Y_t|W^{t-1}) \\ \forall\ \mathcal{S} \subseteq \mathcal{T} : \sum_{t\in\mathcal{S}}[C_t - R_t] > I(U_\mathcal{S};Y_\mathcal{S}|U_\mathcal{S},W_\mathcal{T}). \end{cases} \tag{29}$$

and the joint distribution

$$P_{X|W_\mathcal{T}}(x|w_\mathcal{T})P_{Y_\mathcal{T}|X}(y_\mathcal{T}|x)\prod_{t=1}^{T} P_{U_t|Y_t,W^t}(u_t|y_t,w^t) \cdot \prod_{t=1}^{T} P_{W_t|W^{t-1}}(w_t|w^{t-1}), \tag{30}$$

is achievable, and

$$|\mathcal{W}_t| \leq |\mathcal{X}|^{T-t+1} + \sum_{i=t}^{T} |\mathcal{X}|^{T-i}(2^T + i - t) \quad (31)$$

$$|\mathcal{U}_t| \leq |\mathcal{Y}_t||\mathcal{W}^t| + 2^{T-1}. \quad (32)$$

Theorem 3 does not seem to include Theorem 4 as a special case, as it does not account for a common rate (see Remark 3).

Proof: See Appendix IX. ■

Corollary 4: The rate from Theorem 4 can be expressed with no constraints and no parameters R_t, by solving a linear programming maximization problem, as in Remark 2 which is built along the lines of Corollary 1. This gives the rate

$$R = \min_{\mathcal{S}} \left\{ \sum_{t \in \mathcal{S}} [C_t - I(Y_t; U_t | X)] + I(U_{\mathcal{S}^C}; X) + \sum_{t \in \mathcal{S}^C} I(Y_t; W_t | W^{t-1}) \right\}. \quad (33)$$

VI. The Gaussian channel

In this section we investigate the Gaussian channel under two cases, where the agents are ignorant of the code (sections III and IV) and where they are cognizant of the codebook used by the transmitter (section V). Using the latest results of Oohama for the Gaussian CEO rate-distortion problem [17], a converse for the reliable communication rate is shown for the former case.

We use Corollary 1 with continuous alphabets instead of discrete, where this extension relies on standard arguments (See [15] for example). The Generalized Markov Lemma, for Gaussian variables appears in [17].
The Gaussian channel is defined by $Y_t = X + N_t$, where $\{N_t\}_{t=1}^T$ are independent Gaussian random variables with $\mathrm{E}N_t^2 = P_{N_t}$[1] and $\mathrm{E}N_t = 0$. Let X be zero mean Gaussian with variance $\mathrm{E}X^2 = P_X$. Here we abuse notation and use P_X, P_{N_t} to denote also the variance of the Gaussian random variables X, N_t, in addition to the probability itself, where the meaning will become obvious from the context.

A. Non-decoding agents

We prove the following result in Appendix X.

Theorem 5: The capacity of the nomadic transmitter, for the Gaussian channel, with Gaussian P_X (described above), is

$$R = \max_{\{r_t \geq 0\}} \left\{ \min_{\mathcal{S} \subseteq \mathcal{T}} \left\{ \frac{1}{2} \log_2 \left(1 + P_X \sum_{t \in \mathcal{S}^C} \frac{1 - 2^{-2r_t}}{P_{N_t}} \right) + \sum_{t \in \mathcal{S}} [C_t - r_t] \right\} \right\}. \quad (34)$$

Note that the nomadic assumption here is extended by limiting the transmitter to use Gaussian codebooks. The Theorem is proved using Corollary 1 for the direct part and the upper bound from (113) along with results from [15] for the converse part.

The parameters $\{r_t\}_{t=1}^T$ in (34) indicate the bandwidth wasted for quantizing the additive noise, which can not be avoided because of the nomadic transmitter. This bandwidth reduces the available bandwidth that is left for

[1] E stands for statistical expectation

forwarding the actual transmission to $(C_t - r_t)$, and on the other hand, indicates the expected signal to noise ratio at the destination $\frac{1}{2}\log_2\left(1 + P_X \sum_{t\in\mathcal{T}} \frac{1-2^{-2r_t}}{P_{N_t}}\right)$. Notice that (34) is concave in r_t, so that it can be efficiently solved numerically. In addition, when the problem is symmetric (C_t, N_t are equal among agents), then also the optimal $r_t = r^*$ are identical for all the agents, and an explicit expression can be obtained, provided the roots to a polynomial of degree T are found.

Corollary 5: For the case of two equivalent agents, that is $C_1 = C_2 = C$ and $\gamma = \frac{P_X}{P_{N_1}} = \frac{P_X}{P_{N_2}}$, the rate (34) can be written as the following explicit expression:

$$R = \frac{1}{2}\log\left(1 + 2\gamma\left(1 - \frac{\sqrt{\gamma^2 + 2^{4C}(1+2\gamma)} - \gamma}{2^{4C}}\right)\right). \tag{35}$$

Notice that R in (35) equals $\frac{1}{2}\log(1+2\gamma)$ when $C \to \infty$ and $2C$ when $\gamma \to \infty$.

B. Example: sub-optimality of Gaussian signalling

The previous section described the capacity of the nomadic transmitter, in the Gaussian setting when the transmitter used a Gaussian code book. It is important to note that the Gaussian signalling is not necessarily optimal, where sub-optimality can be due to the limited lossless links between the agents and the final destination. For example, we show that for when $C_1 = C_2 = 1$, using binary phase shift keying (BPSK) at the transmitter surpasses using Gaussian signalling, when the agents are still ignorant of the codebook used.

The agents know that BPSK signal was used, and can demodulate every received channel output into one bit $(j = 1, \ldots, n)$:

$$V_t(k) = \begin{cases} 1 & y_t(k) > 0 \\ 0 & y_t(k) \le 0. \end{cases} \tag{36}$$

This scheme is in fact a special case of Theorem 1, where $P_X(X)$ represents the two equiprobable BPSK symbols, and U_t is a deterministic function of Y_t. Notice that this V_t contains n bits, so that $C_1 = C_2 = 1$ suffices to forward it over to the final destination. The final destination can reliably decode the received message provided the transmission rate is no more than (for all $1 \le k \le n$):

$$R_{bpsk} \le \frac{1}{n} I(\boldsymbol{X}; V_1, V_2) = I(X(k); V_1(k), V_2(k)) = G(Q(\sqrt{2P})) \tag{37}$$

where

$$G(x) \triangleq (1-x)^2 \log_2\left((1-x)^2\right) + x^2 \log_2\left(x^2\right) - (1 - 2x(1-x))\log_2\left(\frac{1}{2}(1 - 2x(1-x))\right), \tag{38}$$

and $Q(x) = (2\pi)^{-0.5}\int_x^\infty \exp\left(-\frac{1}{2}z^2\right)dz$. We compare this rate to (35) in figure 2. We indeed see that BPSK signalling outperforms Gaussian signalling. This is because demodulation is some form of primitive decoding, so that it improves upon the Gaussian signalling, which does not allow any decoding (including any sort of demodulation).

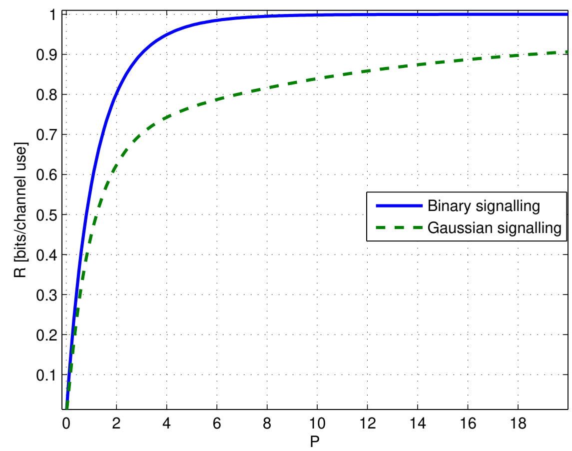

Fig. 2. The achievable rate of a system with two agents, with link bandwidths of $C_1 = C_2 = 1$ and signal to noise ratio of P. The dotted line designates the use of Gaussian signalling at the transmitter and the solid line designates the use of BPSK.

C. Example: agents with decoding capabilities

Consider the symmetric case of a Gaussian channel with statistically equivalent agents (both suffering from an additive Gaussian noise with variance P_N). In addition, both agents are connected via lossless links with equal bandwidth C, to the final destination. The combined approach of broadcast and compression for the degraded channel (theorem 4) is employed, although the optimization considers only Gaussian distributions. The rate R is achievable provided that:

$$R < \sum_{t=1}^{2} R_t + \frac{1}{2}\log_2\left(1 + \alpha\frac{P_X}{P_N}\sum_{t=1}^{2}(1-2^{-2r_t})\right) \tag{39}$$

where $\{r_t, R_t, \alpha\}$ satisfy:

$$\begin{cases} 0 \le \alpha \le 1 \\ t=1,2:\ \ 0 \le R_t \le C \\ \sum_{t=1}^{2} R_t < \frac{1}{2}\log_2(\frac{P_N+P_X}{P_N+\alpha P_X}) \\ \forall \mathcal{S} \subseteq \{1,2\}: \\ \sum_{\mathcal{S}}[C-R_t] > \sum_{\mathcal{S}} r_t + \\ \qquad\qquad \frac{1}{2}\log_2\left(1+\alpha P_X \sum_{t=1}^{2}\frac{1-2^{-2r_t}}{P_N}\right) + \\ \qquad\qquad -\frac{1}{2}\log_2\left(1+\alpha\frac{P_X}{P_N}\sum_{S^C}(1-2^{-2r_t})\right). \end{cases} \tag{40}$$

The convex hull is found to improve rates for this example. The achievable rate as a function of the bandwidth C, for signal to noise ratio $\frac{P_X}{P_N} = 10$, is presented in Fig. 3. In this figure, the left most dashed line $R = 2C$, and the upper flat dashed line $R = \frac{1}{2}\log_2(1 + 2\frac{P_X}{P_N})$ are the two cut-set bounds [27], and the lower flat dashed line is the rate of a system without compression $R = \frac{1}{2}\log_2(1 + \frac{P_X}{P_N})$. The dotted line represents time-sharing, which is

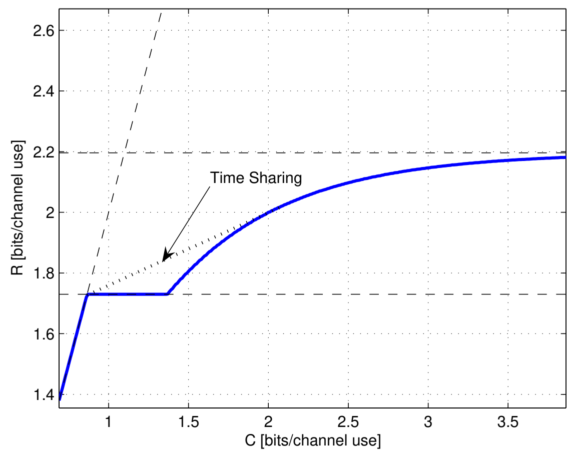

Fig. 3. The achievable rate of a system with two agents, each with link bandwidth of C and signal to noise ratio of 10 dB. The dotted line designates time sharing, and the dashed lines represent the cut-set bounds [27]. The lower flat dashed line is the achievable rate for a system without compression.

useful here. This figure illustrates that if the sum of capacities of the corresponding broadcast channel, (calculated by the signal to noise ratios at the agents) is smaller than the sum of the bandwidths of the links, a compression scheme can significantly improve the performance. A rate of up to 0.2 bits from the cut-set bound is observed with $C = R = 2$, which means 50% excess bandwidth compared to the achievable rate. It also demonstrates that when the bandwidths of the links are smaller then the sum of capacities of the corresponding broadcast channel $C_1 + C_2 < \frac{1}{2}\log_2(1 + \frac{P_X}{P_N})$, the achievable rate in the nomadic setting using Gaussian codebook (equation (35)) is strictly smaller than in the fixed transmitter setting (the cut-set bound, $R = C_1 + C_2$).

VII. Conclusion

Communication via separated agents is considered, focusing on two cases: (1) the agents do not possess any knowledge about the codebook used by the transmitter, and (2) the agents do possess decoding capability. For the first case, the nomadic model was defined and a suitable information theoretic framework, which included random coding, was presented. For this nomadic case, a suitable direct coding theorem based on decentralized compression and the corresponding upper bound were derived. Considering the Gaussian channel, a converse was proved by the entropy power inequality invoking the techniques of [17]. An achievable rate was derived also for the case where the agents are cognizant of the codebook used by the transmitter. These sufficient conditions combined either Marton's or the superposition approaches, with the decentralized compression. For the case of the deterministic channel, the capacity was fully characterized.

ACKNOWLEDGMENT

This research was supported by the EU 6th framework program via the NEWCOM and NEWCOM++ networks of excellence and the REMON Consortium. The authors are indebt to Mr. Michael Peleg for helpful discussions.

APPENDIX I
DEFINITIONS AND LEMMAS

As commonly done (see [27], section 13.6), define the ϵ-typical (strongly typical) set $\mathbf{T}_\epsilon$ of $\boldsymbol{a}_\mathcal{L}$ as the set for which $N(a_\mathcal{S}|\boldsymbol{a}_\mathcal{S}) = 0$ for any $a_\mathcal{S} \in \mathcal{A}_\mathcal{S}$ such that $P_{A_\mathcal{S}}(a_\mathcal{S}) = 0$, and also

$$\mathbf{T}_\epsilon \triangleq \left\{ \boldsymbol{a}_\mathcal{L} : \ \forall \mathcal{S} \subseteq \mathcal{L}, \ \forall a_\mathcal{S} \in \mathcal{A}_\mathcal{S}, \left| \frac{1}{n} N(a_\mathcal{S}|\boldsymbol{a}_\mathcal{S}) - P_{A_\mathcal{S}}(a_\mathcal{S}) \right| < \frac{\epsilon}{|\mathcal{A}_\mathcal{S}|} \right\}, \tag{41}$$

where $N(a_\mathcal{S}|\boldsymbol{a}_\mathcal{S})$ denotes the number of occurrences of the symbol $a_\mathcal{S}$ in the vector $\boldsymbol{a}_\mathcal{S}$.

Lemma 2: For any $\epsilon > 0$, there exist n^* such that for all $n > n^*$ and $\boldsymbol{a}_\mathcal{L} \sim \prod P_{A_\mathcal{L}}$

$$P(\boldsymbol{a}_\mathcal{L} \in \mathbf{T}_\epsilon) \geq 1 - \epsilon. \tag{42}$$

Lemma 3: Define the jointly ϵ-typical set $\mathbf{T}_\epsilon$, as before, by the joint probability

$$P_{A_\mathcal{L}, W_\mathcal{L}}(a_\mathcal{L}, w_\mathcal{L}) \tag{43}$$

For some $\mathcal{S} \subseteq \mathcal{L}$ let $\boldsymbol{a}_\mathcal{L}^n$ be generated according to

$$\boldsymbol{a}_\mathcal{L} \sim \prod_{i=1}^{n} \left\{ P_{A_{\mathcal{S}^C}|W_\mathcal{L}}(a_{\mathcal{S}^C,i}|w_{\mathcal{L},i}) \prod_{l \in \mathcal{S}} P_{A_l|W_l}(a_{l,i}|w_{l,i}) \right\}, \tag{44}$$

where $\boldsymbol{w}_\mathcal{L}$ is a vector that belongs to $\mathbf{T}_\epsilon$. Then the probability that the pair $(\boldsymbol{a}_\mathcal{L}, \boldsymbol{w}_\mathcal{L})$ is in $\mathbf{T}_\epsilon$ is bounded by

$$\Pr((\boldsymbol{a}_{1,\dots,L}, \boldsymbol{w}_\mathcal{L}) \in \mathbf{T}_\epsilon) \geq 2^{-n[H(A_{\mathcal{S}^C}|W_\mathcal{L}) - H(A_\mathcal{L}|W_\mathcal{L}) + \sum_{l \in \mathcal{S}} H(A_l|W_l) + \epsilon_1]} \tag{45}$$

$$\Pr((\boldsymbol{a}_{1,\dots,L}, \boldsymbol{w}_\mathcal{L}) \in \mathbf{T}_\epsilon) \leq 2^{-n[H(A_{\mathcal{S}^C}|W_\mathcal{L}) - H(A_\mathcal{L}|W_\mathcal{L}) + \sum_{l \in \mathcal{S}} H(A_l|W_l) - \epsilon_1]} \tag{46}$$

where $\epsilon_1 \to 0$ as $\epsilon \to 0$.

Lemma 4: Generalized Markov Lemma

Let

$$P_{A_\mathcal{S}, W_\mathcal{S}, Y_\mathcal{S}}(a_\mathcal{S}, w_\mathcal{S}, y_\mathcal{S}) = P_{W_\mathcal{S}, Y_\mathcal{S}}(w_\mathcal{S}, y_\mathcal{S}) \prod_{t \in \mathcal{S}} P_{A_t|W_t, Y_t}(a_t|w_t, y_t). \tag{47}$$

Given randomly generated $\boldsymbol{w}_\mathcal{S} \boldsymbol{y}_\mathcal{S}$ according to $P_{W_\mathcal{S}, Y_\mathcal{S}}$, for every $t \in \mathcal{S}$, randomly and independently generate $N_t \geq 2^{nI(A_t;Y_t|W_t)}$ vectors $\tilde{\boldsymbol{a}}_t$ according to $\prod_{i=1}^{n} P_{A_t|W_t}(\tilde{a}_{t,i}|w_{t,i})$, and index them by $\tilde{\boldsymbol{a}}_t^{(k)}$. Then there exist $|\mathcal{S}|$ functions $k_t^* = \phi_t(\boldsymbol{y}_t, \boldsymbol{w}_t, \tilde{\boldsymbol{a}}_t^{(1)}, \dots, \tilde{\boldsymbol{a}}_t^{(N_t)})$ taking values in $[1 \dots N_t]$, such that for sufficiently large n,

$$\Pr((\{\boldsymbol{a}_t^{(k_t^*)}\}_{t \in \mathcal{S}}, \boldsymbol{w}_\mathcal{S}, \boldsymbol{y}_\mathcal{S}) \in \mathbf{T}_\epsilon) \geq 1 - \epsilon. \tag{48}$$

Proof: See [27] and [29] for the proofs of Lemmas 2-3, while Lemma 4 is a simple extension of Lemma 3.4 (Generalized Markov Lemma) in [30]. ■

In the following, we use only ϵ and remove the distinction between ϵ and ϵ_1, for the sake of brevity.

Appendix II
Proof of Lemma 1

In this Appendix we show that when the selected encoding F is unknown, the transmission X^n is a memoryless random process, distributed according to: $P_{X_{t_1},\dots,X_{t_L}}(x_{t_1},\dots,x_{t_L}) = \prod_{t=t_1}^{t_L} P_X(x_t)$ for all $t_{\mathcal{L}} \triangleq t_1 < \cdots < t_L$ and all $x_{t_{\mathcal{L}}} \triangleq x_{t_1},\dots,x_{t_L}$. We have that

$$P_{X_{t_1},\dots,X_{t_L}}(x_{t_1},\dots,x_{t_L}) = \sum_{f,m:(\phi_{S,f}(m))_{t_{\mathcal{L}}}=x_{t_{\mathcal{L}}}} P_{F,M}(f,m) = \sum_{m=1}^{2^{nR}} \sum_{f:(\phi_{S,f}(m))_{t_{\mathcal{L}}}=x_{t_{\mathcal{L}}}} P_F(f)P_M(m) =$$
$$\sum_{f:(\phi_{S,f}(1))_{t_{\mathcal{L}}}=x_{t_{\mathcal{L}}}} P_F(f) == \prod_{t=t_1}^{t_L} P_X(x_t) \sum_{x_{t_{\mathcal{L}^C}},\boldsymbol{x}(m=2),\dots,\boldsymbol{x}(m=2^{nR})} \prod_{j\in t_{\mathcal{L}^C}} P_X(x_j) \prod_{i=1}^{n}\prod_{l=2}^{2^{nR}} P_X(x(m=l)_i) = \prod_{t=t_1}^{t_L} P_X(x_t) \tag{49}$$

This concluds the proof.

Appendix III
Proof of Theorem 3

For the proof we use ideas from [31], which presents an achievable rate region using the compress and forward technique for the multiple relays problem. The difference being that here the agents benefit from a fixed non-interfering links to the destination, thus the multiple access communication and the interferences from simultaneously transmitting relays are avoided. In addition to compress-and-forward, broadcast messages are sent to the agents to be passed on noiselessly to the destination. Notice that the agents do not require any knowledge of the codebook used by the transmitter. As before, the network is composed of T agents $t \in \mathcal{T} = \{1,\dots,T\}$, a source transmitter and a final destination. Apposed to [31], we do not need the *block Markov encoding* technique here, because there is no direct link between the transmitter and the final destination.

The transmission goes as follows, the transmitter is sending $\boldsymbol{X}(M)$ where $M \in [1, 2^{nR}]$. Divide M into $M_{\mathcal{T}}$ and M_{CF}, where $M_{\mathcal{T}} \triangleq (M_1,\dots,M_T) \in [1,2^{nR_1}] \times \cdots \times [1,2^{nR_T}]$ and $M_{CF} \in [1,2^{nR_{CF}}]$ are the messages that are decoded at the agents and the message that is decoded only at the final destination D, respectively. The t-agent decodes $\hat{M}_t$ and forwards it to D with nR_t bits. It then compresses the received signal $\boldsymbol{Y}_t$ given the broadcast message that was decoded. The t agent uses compression rate of $\hat{R}_t$ to compress $\boldsymbol{Y}_t$ into $\boldsymbol{U}_t$, indexed by z_t, where $z_t \in [1, 2^{n\hat{R}_t}]$. Since the compressed signals $\{\boldsymbol{U}_t\}$ are dependent with each other and with $\hat{M}_{\mathcal{T}}$, bandwidth from the agents to D can be saved by using the Wyner-Ziv lossy distributed source coding, that is partition the output of the compression into bins, and send only the bin index. Each agent then uses the remaining bandwidth after sending the broadcast message $(C_t - R_t)$ to send the corresponding bin $\{s_t \in [1, 2^{n(C_t-R_t)}]\}$. The final destination receives $\hat{M}_{\mathcal{T}}$ from all the agents and then uses it with $s_1,\dots,s_T$ to decode $\hat{z}_1,\dots,\hat{z}_T$ and then to decode $\hat{M}_{CF} \in [1,2^{nR_{CF}}]$. The detailed proof goes as follows: we first describe the code construction. Next, the processing at transmitter, agents and the decoding at the final destination are given. The conditions (21) result from the described construction so that when $n \to \infty$ the error probability is arbitrary small.

A. Code construction:

Fix $\delta > 0$ and then for every $t = (1, \dots, T)$:

1) For the broadcast transmissions:
   - Randomly generate $2^{n(I(W_t;Y_t)-R_t-\delta)}$ vectors $\boldsymbol{w}_t$, of length n, according to $P_{\boldsymbol{W}_t}(\boldsymbol{w}_t) = \prod_{i=1}^{n} P_{W_t}(w_{t,i})$.
   - Repeat the last step for 2^{nR_t} times, label the resulting $2^{n(I(W_t;Y_t)-R_t-\delta)}$ vectors of each repetition by M_t, where $M_t \in [1, 2^{nR_t}]$. Define $\mathcal{M}_{M_t}$ as the set labelled by M_t.
   - Define the bin $\mathcal{M}_{M_\mathcal{T}} \triangleq \mathcal{M}_{M_1} \times \cdots \times \mathcal{M}_{M_T}$ as the product union of the sets $\{\mathcal{M}_{M_t}\}$.
2) For compress and forward transmission: at the agents

   For all $\{\boldsymbol{w}_t\}$ generated in the previous step,
   - Randomly generate $2^{n[\hat{R}_t-(C_t-R_t)]}$ vectors $\boldsymbol{u}_t$ of length n according to $\prod_i P_{U_t|W_t}(u_{t,i}|w_{t,i})$.
   - Repeat the last step for $s_t = 1, \dots, 2^{n(C_t-R_t)}$, define the resulting set of $\boldsymbol{u}_t$ of each repetition by S_{s_t}.
   - Index all the generated $\boldsymbol{u}_t$ with $z_t \in [1, 2^{n\hat{R}_t}]$. We will interchangeably use the notation S_{s_t} for the set of vectors $\boldsymbol{u}_t$ as well as for the set of the corresponding z_t.
   - Notice that the mapping between the indices z_t and the vectors $\boldsymbol{u}_t$ depends on $\boldsymbol{w}_t$. So we will write $\boldsymbol{u}_t(z_t, \boldsymbol{w}_t)$ to denote $\boldsymbol{u}_t$ which is indexed by z_t for some specific $\boldsymbol{w}_t$ from the previous stage.
3) For compress and forward transmission: at the transmitter

   For every codebook realization f, and every $\boldsymbol{w}_\mathcal{T}$ generated in the first step:
   - Randomly choose $2^{nR_{CF}}$ vectors $\boldsymbol{x}$, of length n, with probability $P_{\mathbf{X}|\boldsymbol{W}_\mathcal{T}}(\boldsymbol{x}|\boldsymbol{w}_\mathcal{T}) = \prod_i P_{X|W_\mathcal{T}}(x_i|w_{\mathcal{T},i})$.
   - Index these vectors by M_{CF} where $M_{CF} \in [1, 2^{nR_{CF}}]$.
   - So we have $2^{n[\sum_\mathcal{T} I(W_t;Y_t)-\delta]}$ different mappings between indices M_{CF} and vectors $\boldsymbol{x}$, where the one used is determined by $\boldsymbol{w}_\mathcal{T}$. We will therefore denote $\boldsymbol{x}(M_{CF}, \boldsymbol{w}_\mathcal{T})$ as the vector indexed by M_{CF} for some $\boldsymbol{w}_\mathcal{T}$ out of the ones chosen on the first step. We leave out the notation of f in the sequel, for the sake of brevity, since for decoding agents, the chosen f is known at the agents, so the achievable rate is valid for every realization of f, with high probability.

B. Encoding:

Let $M = (M_\mathcal{T}, M_{CF})$ be the message to be sent ($M_\mathcal{T}$ is defined at the beginning of this section).

- Define $\mathbf{T}_\epsilon^{BC}$ as:

$$\mathbf{T}_\epsilon^{BC} \triangleq \left\{\boldsymbol{w}_\mathcal{T} :\ \forall \mathcal{S} \subseteq \mathcal{T},\ \forall a_\mathcal{S} \in \mathcal{W}_\mathcal{S}, \left|\frac{1}{n}N(a_\mathcal{S}|\boldsymbol{w}_\mathcal{S}) - P_{W_\mathcal{S}}(a_\mathcal{S})\right| < \frac{\epsilon}{|\mathcal{W}_\mathcal{S}|}\right\}. \tag{50}$$

- Find a T-tuple $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_T)$ in the bin $\mathcal{M}_{M_\mathcal{T}}$ such that

$$(\boldsymbol{w}_1, \dots, \boldsymbol{w}_T) \in\ \mathbf{T}_\epsilon^{BC}. \tag{51}$$

  If no such T-tuple is found, declare error event E_1.
- Define the n functions $\boldsymbol{w}_\mathcal{T}(M_\mathcal{T})$, as the mapping of $M_\mathcal{T}$ into the typical $\boldsymbol{w}_\mathcal{T}$ that was chosen in the last step.

- Transmit to the channel the vector $\boldsymbol{x}$ which is indexed by M_{CF} when distributed by $\prod_i P_{X|W_{\mathcal{T}}}(x_i|w_{\mathcal{T},i}(M_{\mathcal{T}}))$. Denote $\boldsymbol{x}(M) \triangleq \boldsymbol{x}(M_{CF}, \boldsymbol{w}_{\mathcal{T}}(M_{\mathcal{T}}))$.

C. Processing at the agents:

In the following, $\mathbf{T}_\epsilon^{t,1}$ and $\mathbf{T}_\epsilon^{t,2}$ are defined in the standard way, as (41).

1) Decoding: The t agent receives $\boldsymbol{y}_t$ and looks for $\hat{\boldsymbol{w}}_t$ so that

$$(\boldsymbol{y}_t, \hat{\boldsymbol{w}}_t) \in \mathbf{T}_\epsilon^{t,1}. \tag{52}$$

If no such $\hat{\boldsymbol{w}}_t$ exists, declare error event E_2. If there is more than one such $\hat{\boldsymbol{w}}_t$, declare error event E_3. Denote by E_4 the error event where the chosen vector $\hat{\boldsymbol{w}}_t \neq \boldsymbol{w}_t(M_{\mathcal{T}})$.

2) Compression: The t agent chooses any of the z_t such that

$$\big(\boldsymbol{u}_t(z_t, \hat{\boldsymbol{w}}_t), \boldsymbol{y}_t, \hat{\boldsymbol{w}}_t\big) \in \mathbf{T}_\epsilon^{t,2}. \tag{53}$$

The event where no such z_t is found is defined as the error event E_5.

After deciding on z_t the agent transmits s_t, which fulfills $z_t \in S_{s_t}$, and $\hat{M}_t$ to the final destination through the lossless link, where $\hat{M}_t$ corresponds to $\hat{\boldsymbol{w}}_t$.

D. Decoding (at the destination):

The destination retrieves $\hat{M}_{\mathcal{T}}$ and $s_{\mathcal{T}} \triangleq (s_1, \ldots, s_T)$ from the lossless links.

As long as

$$R_t < C_t, \tag{54}$$

the transmitted $s_{\mathcal{T}}$ and $\hat{M}_{\mathcal{T}}$ are properly received, with no errors, since the link is lossless. The destination then finds the set of indices $\hat{z}_{\mathcal{T}} \triangleq \{\hat{z}_1, \ldots, \hat{z}_T\}$ of the compressed vectors $\hat{\boldsymbol{u}}_{\mathcal{T}}^n$ and the decoded vectors $\hat{\boldsymbol{w}}_{\mathcal{T}}$ which satisfy

$$\begin{cases} \big(\hat{\boldsymbol{u}}_1(\hat{z}_1, \hat{\boldsymbol{w}}_1(\hat{M}_{\mathcal{T}})), \ldots, \hat{\boldsymbol{u}}_T(\hat{z}_T, \hat{\boldsymbol{w}}_T(\hat{M}_{\mathcal{T}})), \hat{\boldsymbol{w}}_{\mathcal{T}}(\hat{M}_{\mathcal{T}})\big) \in \mathbf{T}_\epsilon^3 \\ \hat{z}_{\mathcal{T}} \in S_{s_1} \times \cdots \times S_{s_T}. \end{cases} \tag{55}$$

Where $\mathbf{T}_\epsilon^3$ is defined in the standard way, as (41). If there is no such $\hat{z}_{\mathcal{T}}$, the destination declares error E_6 and if there is more than one such $\hat{z}_{\mathcal{T}}$ the destination declares error E_7. The event where $\hat{z}_{\mathcal{T}} \neq z_{\mathcal{T}}$ is defined as E_8.

Finally the destination decides that $\hat{M}_{CF}$ was sent if

$$\big(\boldsymbol{x}(\hat{M}_{\mathcal{T}}, \hat{M}_{CF}), \hat{\boldsymbol{u}}_{\mathcal{T}}(\hat{z}_{\mathcal{T}}, \hat{\boldsymbol{w}}_{\mathcal{T}}(\hat{M}_{\mathcal{T}})), \hat{\boldsymbol{w}}_{\mathcal{T}}(\hat{M}_{\mathcal{T}})\big) \in \mathbf{T}_\epsilon^4. \tag{56}$$

If no such $\hat{M}_{CF}$ is found, declare error E_9, if more than one such $\hat{M}_{CF}$ is found, declare error E_{10}. Further define error E_{11} as the event where $\hat{M}_{CF} \neq M_{CF}$.

Correct decoding means that the destination decides $\hat{M} = M$. An achievable rate R was defined as when the final destination receives the transmitted message with an error probability which is made arbitrarily small for sufficiently large block length n.

E. Error analysis

The error probability is upper bounded by:

$$P(error) = P\left(\cup_{i=1}^{11} E_i\right) \leq \sum_{i=1}^{11} P(E_i). \tag{57}$$

Where:

1) E_1: No T-tuple $\boldsymbol{w}_{\mathcal{T}}$ jointly typical is found.
2) E_2: No $\hat{\boldsymbol{w}}_t$ which is jointly typical with $\boldsymbol{y}_t$ exists.
3) E_3: More than one $\hat{\boldsymbol{w}}_t$ is jointly typical with $\boldsymbol{y}_t$.
4) E_4: A different $\hat{\boldsymbol{w}}_t \neq \boldsymbol{w}_t$ is jointly typical with $\boldsymbol{y}_t$.
5) E_5: No $\boldsymbol{u}_t(z_t, \hat{\boldsymbol{w}}_t)$ is jointly typical with $(\boldsymbol{y}_t, \hat{\boldsymbol{w}}_t)$.
6) E_6: There is no $\left(\hat{\boldsymbol{u}}_{\mathcal{T}}(\hat{z}_{\mathcal{T}}, \hat{M}_{\mathcal{T}}), \hat{\boldsymbol{w}}_{\mathcal{T}}(\hat{M}_{\mathcal{T}})\right)$ jointly typical.
7) E_7: There is more than one $\left(\hat{\boldsymbol{u}}_{\mathcal{T}}(\hat{z}_{\mathcal{T}}, \hat{M}_{\mathcal{T}}), \hat{\boldsymbol{w}}_{\mathcal{T}}(\hat{M}_{\mathcal{T}})\right)$ jointly typical.
8) E_8: There is another $\left(\hat{\boldsymbol{u}}_{\mathcal{T}}(\hat{z}_{\mathcal{T}}, \hat{M}_{\mathcal{T}}), \hat{\boldsymbol{w}}_{\mathcal{T}}(\hat{M}_{\mathcal{T}})\right)$ jointly typical.
9) E_9: There is no $\boldsymbol{x}$ jointly typical with $\hat{\boldsymbol{u}}_{\mathcal{T}}, \hat{\boldsymbol{w}}_{\mathcal{T}}$.
10) E_{10}: There is more than one $\boldsymbol{x}$ jointly typical with $\hat{\boldsymbol{u}}_{\mathcal{T}}, \hat{\boldsymbol{w}}_{\mathcal{T}}$.
11) E_{11}: The jointly typical $\boldsymbol{x}(\hat{M}_{CF}, \hat{M}_{\mathcal{T}}) \neq \boldsymbol{x}(M)$.

The error events E_3, E_7, E_{10} are added since they are declared by the system as errors. We will upper bound the probabilities of the individual error events by arbitrarily small ϵ.

1) E_1: Notice that in order for the number of generated vectors to be larger than zero,

$$\forall\ 1 \leq t \leq T:\ \ R_t \leq I(W_t; Y_t) - \delta. \tag{58}$$

For any subset $\mathcal{S} \subseteq \mathcal{T}$, the probability that a randomly generated t-tuple $\boldsymbol{w}_{\mathcal{S}}$ is not jointly typical (so $\boldsymbol{w}_{\mathcal{T}} \notin \mathbf{T}_{\epsilon}^{BC}$) is upper bounded by

$$P_{\boldsymbol{w}_{\mathcal{T}} \notin \mathbf{T}_{\epsilon}^{BC}} \leq 1 - 2^{n[H(W_{\mathcal{S}}) - \sum_{t\in\mathcal{S}} H(W_t) - 2|\mathcal{S}|\epsilon]}, \tag{59}$$

and the probability P that some bin $\mathcal{M}_{M_{\mathcal{S}}}$ does not contain any jointly typical $\mathcal{S}$-tuple is upper bounded by

$$P \leq \left(P_{W_{\mathcal{T}}^n \notin \mathbf{T}_{\epsilon}^{BC}}\right)^{2^{n[\sum_{\mathcal{S}} I(Y_t; W_t) - R_t - \delta]}}. \tag{60}$$

It is easy to see that this probability is small as desired for any n sufficiently large when

$$\begin{aligned}\sum_{t\in\mathcal{S}} R_t &< \sum_{t\in\mathcal{S}} [I(Y_t; W_t) - H(W_t) - \delta] + H(W_{\mathcal{S}}) - 2|\mathcal{S}|\epsilon \\ &= \sum_{t\in\mathcal{S}} I(Y_t; W_t) - I(W_t; W_{\tilde{\mathcal{T}}(\mathcal{S},t)}) - |\mathcal{S}|\delta - 2|\mathcal{S}|\epsilon. \end{aligned} \tag{61}$$

Where $\tilde{\mathcal{T}}(\mathcal{S}, t) \triangleq \{i:\ i \in \mathcal{S}\ and\ i < t\}$. Define for the sequel $\epsilon' = \delta + 2\epsilon$.

2) E_2, E_6, E_9: By Lemmas 2 and 4, the probability of jointly distributed variables not to be ϵ-typical is as small as desired for n sufficiently large.

3) E_3 and E_4: We have that according to Lemma 3, the probability that another $\hat{\boldsymbol{w}}_t$ belongs to $\mathbf{T}_\epsilon^{t,1}$ is upper bounded by $2^{-n[I(W_t;Y_t)-\epsilon]}$. Since there are no more than $2^{n[I(W_t;Y_t)-\delta]}$ such $\hat{\boldsymbol{w}}_t$, the probability of E_3 and E_4 can be made arbitrarily small as n goes to infinity as long as $\delta > \epsilon$.

4) E_5: From Lemma 4, there is no z_t such that $\boldsymbol{u}_t(z_t, \boldsymbol{w}_t)$ is in $\mathbf{T}_\epsilon^{t,2}$ with probability $P(E_5)$, which can be made arbitrary small for sufficiently large n as long as

$$\hat{R}_t > I(U_t; Y_t|W_t). \tag{62}$$

5) E_7 and E_8: Assume that for some $\mathcal{S} \subseteq \mathcal{T}$:

$$\hat{z}_\mathcal{S} \neq z_\mathcal{S}, \tag{63}$$

and

$$\hat{z}_{\mathcal{S}^C} = z_{\mathcal{S}^C}. \tag{64}$$

This means that the compression vectors $\hat{\boldsymbol{u}}_t$ for $t \in \mathcal{S}$, are jointly typical with the corresponding $\hat{\boldsymbol{w}}_t$, with high probability (Lemma 2) as they are generated that way. But they are not necessarily jointly typical with the other $\{\hat{\boldsymbol{u}}_j, \hat{\boldsymbol{w}}_j\}_{j\neq t}$. On the other hand, since $\hat{\boldsymbol{u}}_t = \boldsymbol{u}_t$ for $t \in \mathcal{S}^C$, they are jointly typical together with $\hat{\boldsymbol{w}}_\mathcal{T}$ with high probability, due to Lemma 4. So $\hat{\boldsymbol{u}}_\mathcal{T}(\hat{z}_\mathcal{T}, \hat{\boldsymbol{w}}_\mathcal{T}(\hat{M}_\mathcal{T})), \hat{\boldsymbol{w}}_\mathcal{T}(\hat{M}_\mathcal{T})$ with high probability, belongs to a typical set of the distribution:

$$\prod_{i=1}^{n} \left\{ P_{W_\mathcal{T}}(\hat{w}_{\mathcal{T},i}) P_{U_{\mathcal{S}^C}|W_\mathcal{T}}(\hat{u}_{\mathcal{S}^C,i}|\hat{w}_{\mathcal{T},i}) \prod_{t\in\mathcal{S}} P_{U_t|W_t}(\hat{u}_{t,i}|\hat{w}_{t,i}) \right\}.$$

Thus according to Lemma 3, the probability that such vector belongs to $\mathbf{T}_\epsilon^3$ is upper bounded by

$$2^{n[H(W_\mathcal{T},U_\mathcal{T})-H(W_\mathcal{T})-H(U_{\mathcal{S}^C}|W_\mathcal{T})-\sum_{t\in\mathcal{S}} H(U_t|W_t)+\epsilon]}. \tag{65}$$

Overall there are

$$2^{n[\sum_\mathcal{S}[\hat{R}_t - C_t + R_t]]} - 1$$

such vectors in the set $S_{s_1} \times \cdots \times S_{s_T}$ and the probability of errors E_7 and E_8 is upper bounded by:

$$2^{n[\sum_{t\in\mathcal{S}}[\hat{R}_t - C_t + R_t - H(U_t|W_t)]]} \times 2^{n[H(U_\mathcal{S}|W_\mathcal{T},U_{\mathcal{S}^C})+\epsilon+|\mathcal{S}^C|\epsilon']}. \tag{66}$$

Which means that as long as

$$\sum_{t\in\mathcal{S}}[C_t - R_t] > \sum_{t\in\mathcal{S}} \left[\hat{R}_t - H(U_t|W_t)\right] + H(U_\mathcal{S}|U_{\mathcal{S}^C}, W_\mathcal{T}) \tag{67}$$

the destination will be able to reliably decode $M_\mathcal{T}, z_\mathcal{T}$ for sufficiently large n.

6) E_{10}, E_{11}: The probability that $\hat{M}_{CF} \neq M_{CF}$ satisfies (56) is upper bounded by (again Lemma 3)

$$2^{-n[I(X;U_{\mathcal{T}}|W_{\mathcal{T}})-\epsilon]}. \tag{68}$$

Now summing over the $2^{nR_{CF}}-1$ possible $\hat{M}_{CF}$ and upper bounding, we find that reliable detection of M_{CF} given $M_{\mathcal{T}}$ is possible if

$$R_{CF} < I(X;U_{\mathcal{T}}|W_{\mathcal{T}}). \tag{69}$$

Taking (62) and (67) and noticing that $\{U_t\}_{t\in\mathcal{T}}$ are independent given (Y_t, W_t) we can write the constraints as:

$$\forall \mathcal{S} \in \mathcal{T} : \sum_{t\in\mathcal{S}} [C_t - R_t] > I(U_{\mathcal{S}}; Y_{\mathcal{S}}|U_{\mathcal{S}^C}, W_{\mathcal{T}}). \tag{70}$$

Notice that (58) is surplus given (61) and (54) is surplus given (70). Now (61) and (70) constitutes (21). The achievable rate (20) is through (69).

F. Cardinality Bounds

In this subsection we show the bounds on the cardinality of the auxiliary variables $U_{\mathcal{T}}$. For that, we use the support Lemma, as was done in subsection V-A.

Consider the functionals over generic Q_{Y_t,W_t}. Note that there are 2^{T-1} such functionals from (21), one from (20) and $|\mathcal{Y}_t||\mathcal{W}_t|-1$ from the given probability P_{Y_t,W_t}. This proves that

$$|\mathcal{U}_t| \leq |\mathcal{Y}_t||\mathcal{W}_t| + 2^{T-1}. \tag{71}$$

When trying to apply the support Lemma for the cardinalities of $W_{\mathcal{T}}$, the structure of the constraints in (61), specifically the right-most elements, prevents isolating a single auxiliary variable from the others, and thus also prevents the application of the support Lemma. The difficulty is faced also when trying to limit the cardinalities of the auxiliary variables of Marton's broadcast approach [28], which to the best of our knowledge have not been done. The auxiliaries entanglement is resolved for the degraded channel, considered in Theorem 4 which is proved in Appendix IX.

APPENDIX IV
PROOF OF COROLLARY 1

The scheme which achieves the rate (13) is basically identical to the one used for Theorem 3, with the following differences:

1) In section III-D, now the decoding is done in a single stage, in which the destination looks for $\hat{M}_{CF}, \hat{z}_{\mathcal{T}}$, such that (56) is fulfilled and

$$\hat{z}_{\mathcal{T}} \in S_{s_1} \times \cdots \times S_{s_T}. \tag{72}$$

If there are no such indices, declare error E_7', and if either more than one, or an erroneous $\hat{M}_{CF}$ are found, is denoted by error E_8'. Otherwise declare the received message to be $\hat{M}_{CF}, \hat{M}_{\mathcal{T}}$.

2) Error analysis: E_7-E_{12} are irrelevant anymore, and instead we have E'_7,E'_8. So we consider less error events, and the achievable rate is larger (either strictly or not).

   $P(E'_7) \to 0$ according to Lemma 4.

   As for E'_8: Consider the case where $\hat{M}_{CF} \neq M_{CF}$ and $\hat{z}_{\mathcal{S}} \neq z_{\mathcal{S}}$. There are $2^{n[R+\sum_{t\in\mathcal{S}}[\hat{R}_t - C_t + R_t]]}$ such vectors, and the probability of $(\boldsymbol{x}(\hat{M}), \boldsymbol{u}_{\mathcal{S}}(\hat{z}_{\mathcal{S}}), \boldsymbol{u}_{\mathcal{S}^C}(\hat{z}_{\mathcal{S}^C}))$ to be jointly typical is upper bounded by (Lemma 3) $2^{n[H(X,U_{\mathcal{T}}|W_{\mathcal{T}})-H(X|W_{\mathcal{T}})-H(U_{\mathcal{S}^C}|W_{\mathcal{T}})-\sum_{t\in\mathcal{S}} H(U_t|W_t)+\epsilon]}$. Thus the rate R_{CF} is achievable if:

$$\begin{aligned} R_{CF} &< \sum_{t\in\mathcal{S}}[C_t - R_t - \hat{R}_t + H(U_t|W_t)] - H(U_{\mathcal{S}}|X, W_{\mathcal{T}}) - H(U_{\mathcal{S}^C}|X, U_{\mathcal{S}}, W_{\mathcal{T}}) \\ &< \sum_{t\in\mathcal{S}}[C_t - R_t - I(Y_t; U_t|X, W_t)] + I(U_{\mathcal{S}^C}; X|W_{\mathcal{T}}), \end{aligned}$$

   where the second inequality is due to (62) and because of the Markov chain $U_t - (W_{\mathcal{T}}, X) - U_{\mathcal{T}\setminus t}$. Notice that for t^* such that $\hat{R}_{t^*} \leq C_{t^*} - R_{t^*}$, which means $I(U_{t^*}; Y_{t^*}|W_{t^*}) \leq C_{t^*} - R_{t^*}$, we get full reconstruction of $\boldsymbol{u}_{t^*}$, with no need for binning. In that case, we get that the constraints $\mathcal{S}: \ t^* \in \mathcal{S}$ are superfluous. This is since:

$$\begin{aligned} &C_{t^*} - R_{t^*} - I(U_{t^*}; Y_{t^*}|X, W_{t^*}) + I(U_{\mathcal{S}^C}; X|W_{\mathcal{T}}) \geq I(U_{t^*}; Y_{t^*}|W_{t^*}) - I(U_{t^*}; Y_{t^*}|X, W_{t^*}) + I(U_{\mathcal{S}^C}; X|W_{\mathcal{T}}) = \\ &I(U_{t^*}; X|W_{t^*}) + I(U_{\mathcal{S}^C}; X|W_{\mathcal{T}}) \geq I(U_{t^*}; X|U_{\mathcal{S}^C}, W_{t^*}) + I(U_{\mathcal{S}^C}; X|W_{\mathcal{T}}) \geq I(U_{\mathcal{S}^C \cup t^*}; X|W_{\mathcal{T}}). \quad (73) \end{aligned}$$

APPENDIX V
PROOF OF THE OUTER BOUND OF THEOREM 2

Using Fano's inequality, we get that reliable decoding at the destination is possible only if:

$$H(M|V_{\mathcal{T}}, F) \leq n\epsilon_n, \tag{74}$$

where $n\epsilon_n \to 0$ as $n \to \infty$.

Now we have:

$$\begin{aligned} nR &\leq H(M) = I(M; V_{\mathcal{T}}, F) + H(M|V_{\mathcal{T}}, F) && (75) \\ &\leq H(V_{\mathcal{T}}) + H(F|V_{\mathcal{T}}) - H(F|M) && (76) \\ &\quad -H(V_{\mathcal{T}}|M, F) + n\epsilon_n && (77) \\ &= I(V_{\mathcal{T}}; M, F) - I(F; V_{\mathcal{T}}) + n\epsilon_n && (78) \\ &\leq I(V_{\mathcal{T}}; M, F) + n\epsilon_n && (79) \\ &\leq I(V_{\mathcal{T}}; \boldsymbol{X}(M, F)) + n\epsilon_n && (80) \end{aligned}$$

where (78) is since F is independent with M so $H(F|M) = H(F)$, (80) stems from the Markov chain $MF - \boldsymbol{X} - V_{\mathcal{T}}$.

Turning to the agents, we upper bound

$$\frac{1}{n} I(V_{\mathcal{T}}; \boldsymbol{X}) \tag{81}$$

over the encoding functions of the agents, when the agents do not know F:

$$I(V_{\mathcal{T}};\boldsymbol{X}) = H(\boldsymbol{X}) - H(\boldsymbol{X}|V_{\mathcal{T}}) \tag{82}$$

$$\leq \sum_{i=1}^{n} H(X_i) - H(X_i|V_{\mathcal{T}}, X^{i-1}) \tag{83}$$

$$\leq \sum_{i=1}^{n} H(X_i) - H(X_i|V_{\mathcal{T}}, Y_{\mathcal{T}}^{i-1}, X^{i-1}) \tag{84}$$

$$= \sum_{i=1}^{n} H(X_i) - H(X_i|U_{\mathcal{T},i}) \tag{85}$$

$$= \sum_{i=1}^{n} I(X_i;U_{\mathcal{T},i}). \tag{86}$$

When $U_{\mathcal{S},i} \triangleq (V_{\mathcal{S}}, Y_{\mathcal{T}}^{i-1}, X^{i-1})$ for any $\mathcal{S} \subseteq \mathcal{T}$.

Following the definition, the random variables $U_{\mathcal{S}}$ must fulfill the following constraints:

$$\sum_{i=1}^{n} I(U_{\mathcal{S},i};Y_{\mathcal{T},i}|U_{\mathcal{S}^C,i}) \tag{87}$$

$$= \sum_{i=1}^{n} I(V_{\mathcal{S}}, Y_{\mathcal{T}}^{i-1}, X^{i-1};Y_{\mathcal{T},i}|V_{\mathcal{S}^C}, Y_{\mathcal{T}}^{i-1}, X^{i-1}) \tag{88}$$

$$= \sum_{i=1}^{n} I(V_{\mathcal{S}};Y_{\mathcal{T},i}|V_{\mathcal{S}^C}, Y_{\mathcal{T}}^{i-1}, X^{i-1}) \tag{89}$$

$$\leq \sum_{i=1}^{n} I(V_{\mathcal{S}};Y_{\mathcal{T},i}, X_i|V_{\mathcal{S}^C}, Y_{\mathcal{T}}^{i-1}, X^{i-1}) \tag{90}$$

$$= I(V_{\mathcal{S}};\boldsymbol{Y}_{\mathcal{T}}, \boldsymbol{X}|V_{\mathcal{S}^C}) \tag{91}$$

$$\leq H(V_{\mathcal{S}}|V_{\mathcal{S}^C}) \leq H(V_{\mathcal{S}}) \leq n\sum_{t\in\mathcal{S}} C_t. \tag{92}$$

Next, following [14] define $U_t^* \triangleq (U_{t,s}, s)$, $X^* \triangleq X_s$, $Y^* \triangleq Y_s$ where s is a random variable, uniformly distributed over $[1,n]$. So we have that

$$R \leq \frac{1}{n}\sum_{i=1}^{n} I(X_i;U_{\mathcal{T},i}) \tag{93}$$

$$= \left[\sum_{i=1}^{n} P(s=i)I(X_i;U_{\mathcal{T},i}|s=i)\right] + I(X^*;s) \tag{94}$$

$$= I(X^*;U_{\mathcal{T},s}|s) + I(X^*;s) \tag{95}$$

$$= I(X^*;U_{\mathcal{T}}^*) \tag{96}$$

where (94) follows from Lemma 1: it is known that X^* and s are independent without the key F (X is memoryless). We further have

$$\sum_{t\in\mathcal{S}} C_t \geq \sum_{i=1}^{n} \frac{1}{n} I(U_{\mathcal{S},i};Y_{\mathcal{T},i}|U_{\mathcal{S}^C,i}) \tag{97}$$

$$= \sum_{i=1}^{n} \frac{1}{n} I(U_{\mathcal{S}}^*;Y_{\mathcal{T}}^*|U_{\mathcal{S}^C}^*, s=i) \tag{98}$$

$$= \quad I(U^*_{\mathcal{S}}; Y^*_{\mathcal{T}} | U^*_{\mathcal{S}^C}). \tag{99}$$

Define $W^* \triangleq (\boldsymbol{Y}_{\mathcal{T},s+1}, Y_{\mathcal{T}}^{s-1}, X^{s-1}, s)$, so that by considering Lemma 1, X^* and $Y^*_{\mathcal{T}}$ are independent with W^* when not conditioned on F. The auxiliary variables U^*_t can then be represented as:

$$U^*_t = (V_t, Y_{\mathcal{T}}^{s-1}, X^{s-1}, s) = (g_t(W^*, Y^*_t), g(W^*)) \tag{100}$$

$$f_t(W^*, Y^*_t) \triangleq (g_t(W^*, Y^*_t), g(W^*)) \tag{101}$$

where $g_t(W^*, Y^*_t) = (V_t, Y_{\mathcal{T}}^{s-1})$ and $g(W^*) = (X^{s-1}, s)$. This shows that the probability space is indeed (16). This is possible only because of the nomadic transmitter. In the case when F is known to the agents, the probability space (16) no longer holds, so that the upper bound in theorem 2 isn't applicable when the agents are cognizant of the codebook used.

A. *Cardinality Bounds*

In this subsection we show the bounds on the cardinality of the auxiliary variable W, through bounds on the variables $U_{\mathcal{T}}$, which fulfill the Markov chain, as in eq. (17). For that, we use the support Lemma, (see for example [29] pp. 310, and [19]). According to this Lemma, if there are K functionals $\{q_k\}_{k=1}^K$ on a set $\mathcal{P}(\mathcal{X})$ of probability distributions over the alphabet $\mathcal{X}$, and given any probability measure μ on the Borel σ-algebra of $\mathcal{P}(\mathcal{X})$, there exist K elements $Q_k \in \mathcal{P}(\mathcal{X})$ and K non-negative reals α_k that sum to unity, such that for every $1 \leq j \leq K$:

$$\int_{\mathcal{P}(\mathcal{X})} q_j(Q)\mu(dQ) = \sum_{k=1}^{K} \alpha_k q_j(Q_k). \tag{102}$$

In order to use the Lemma, we define a generic distribution $Q_{Y_{\mathcal{T}},U_{\mathcal{T}\setminus t}}(y_{\mathcal{T}}, u_{\mathcal{T}\setminus t})$ over $\mathcal{Y}_{\mathcal{T}} \times \mathcal{U}_{\mathcal{T}\setminus t}$, which fulfills (16). First write the following functionals as a function of $Q_{Y_{\mathcal{T}},U_{\mathcal{T}\setminus t}}$. Notice that the cardinalities of $U_{\mathcal{T}\setminus t}$ are intact:

$$\begin{aligned}
q_1(Q) = & - \sum_{x, y_{\mathcal{T}}, u_{\mathcal{T}\setminus t}} P_{X|Y_{\mathcal{T}}}(x|y_{\mathcal{T}}) Q_{Y_{\mathcal{T}},U_{\mathcal{T}\setminus t}}(y_{\mathcal{T}}, u_{\mathcal{T}\setminus t}) \cdot \\
& \log \left(\frac{\sum_{y'_{\mathcal{T}}} P_{X|Y_{\mathcal{T}}}(x|y'_{\mathcal{T}}) Q_{Y_{\mathcal{T}},U_{\mathcal{T}\setminus t}}(y'_{\mathcal{T}}, u_{\mathcal{T}\setminus t})}{\sum_{y'_{\mathcal{T}}} Q_{Y_{\mathcal{T}},U_{\mathcal{T}\setminus t}}(y'_{\mathcal{T}}, u_{\mathcal{T}\setminus t})} \right) \\
q^{c}_{\mathcal{S}^C}(Q) = & - \sum_{y_{\mathcal{T}}, u_{\mathcal{T}\setminus t}} Q_{Y_{\mathcal{T}},U_{\mathcal{T}\setminus t}}(y_{\mathcal{T}}, u_{\mathcal{T}\setminus t}) \cdot \\
& \log \left(\frac{\sum_{u'_{\mathcal{S}}} Q_{Y_{\mathcal{T}},U_{\mathcal{T}\setminus t}}(y_{\mathcal{T}}, u'_{\mathcal{S}} u_{\mathcal{S}^C\setminus t})}{\sum_{y'_{\mathcal{T}}, u'_{\mathcal{S}}} Q_{Y_{\mathcal{T}},U_{\mathcal{T}\setminus t}}(y'_{\mathcal{T}}, u'_{\mathcal{S}} u_{\mathcal{S}^C\setminus t})} \right) \\
q_{y_{\mathcal{T}}}(Q) = & \sum_{u'_{\mathcal{T}\setminus t}} Q_{Y_{\mathcal{T}},U_{\mathcal{T}\setminus t}}(y_{\mathcal{T}}, u_{\mathcal{T}\setminus t}).
\end{aligned}$$

Where $P_{X|Y_{\mathcal{T}}}(x|y_{\mathcal{T}})$ is given by (16), and $\mathcal{S}^C$ is such that $\mathcal{S}^C \subseteq \mathcal{T}$ and $t \in \mathcal{S}^C$.

Now applying the support Lemma, we find that there exist a random variable U_t', such that

$$\sum_{u_t} P_{U_t'}(u_t) q_1(P(\bullet|U_t'=u_t)) = H(X) - I(X;U_{\mathcal{T}}) \tag{103}$$

$$\sum_{u_t} P_{U_t'}(u_t) q_{\mathcal{S}^C}^c(P(\bullet|U_t'=u_t)) = H(Y_{\mathcal{T}}|U_{\mathcal{S}^C}) \tag{104}$$

$$\sum_{u_t} P_{U_t'}(u_t) q_{y_{\mathcal{T}}}(P(\bullet|U_t'=u_t)) = P_{Y_{\mathcal{T}}}(y_{\mathcal{T}}), \tag{105}$$

are fulfilled and U_t' has cardinality of:

$$|\mathcal{U}_t| \leq |\mathcal{Y}_{\mathcal{T}}| + 2^{T-1}. \tag{106}$$

This is since (103) is one equation, (104) is 2^{T-1} equations, and (105) is $|\mathcal{Y}_{\mathcal{T}}|-1$ equations. Note that this is also the cardinality of W:

$$|\mathcal{W}| \leq |\mathcal{Y}_{\mathcal{T}}| + 2^{T-1}. \tag{107}$$

Concluding the proof.

Appendix VI
Proof of the upper bound in Corollary 2

Notice that the following holds:

$$\begin{aligned}
\sum_{t\in\mathcal{S}} C_t &\geq \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}; V_{\mathcal{S}}|V_{\mathcal{S}^C}) &(108)\\
&= \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}; V_{\mathcal{T}}) - \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}; V_{\mathcal{S}^C}) &(109)\\
&= \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}, \boldsymbol{X}; V_{\mathcal{T}}) - \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}, \boldsymbol{X}; V_{\mathcal{S}^C}) &(110)\\
&= \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{T}}) - \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C})\\
&\quad + \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}; V_{\mathcal{T}}|\boldsymbol{X}) - \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{S}^C}; V_{\mathcal{S}^C}|\boldsymbol{X}) &(111)\\
&= \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{T}}) - \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C})\\
&\quad + \sum_{t\in\mathcal{S}} \frac{1}{n} I(\boldsymbol{Y}_t; V_t|\boldsymbol{X}). &(112)
\end{aligned}$$

where (110) and (112) are since V_t is a deterministic function of Y_t^n. Again, we use (**??**), so that

$$R \leq \min_{\mathcal{S}\subseteq\mathcal{T}} \left\{ \sum_{t\in\mathcal{S}} [C_t - \frac{1}{n} I(\boldsymbol{Y}_t; V_t|\boldsymbol{X})] + \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C}) \right\} \tag{113}$$

Next, from (86) we know that $I(\boldsymbol{X}; V_{\mathcal{S}^C}) \leq \sum_{i=1}^n I(X_i; U_{\mathcal{S}^C,i})$, and $I(\boldsymbol{Y}_t; V_t|\boldsymbol{X}) \leq \sum_{i=1}^n I(Y_{t,i}; U_{t,i}|X_i)$. Repeating the usage of s here, we get to the desired upper bound (19).

Appendix VII
Using common message with two agents

The use of a common message which is decoded by several agents (as outlined by Marton [28]) is exemplified here for two users $T = 2$. The achievable rate in this case is:

$$R \leq I(X; U_1, U_2 | W_1 W_2 W_C) + R_1 + R_2. \tag{114}$$

Where

$$\begin{cases} 0 \leq R_1 \leq I(W_1, W_C; Y_1) \\ 0 \leq R_2 \leq I(W_2, W_C; Y_2) \\ R_1 + R_2 \leq \min\{I(W_C; Y_1), I(W_C; Y_2)\} + I(W_1; Y_1 | W_C) \\ \qquad\qquad + I(W_2; Y_2 | W_C) - I(W_1; W_2 | W_C) \\ I(U_1; Y_1 | U_2, W_1, W_2, W_C) \leq C_1 - R_1 \\ I(U_2; Y_2 | U_1, W_1, W_2, W_C) \leq C_2 - R_2 \\ I(U_1, U_2; Y_1, Y_2 | W_1, W_2, W_C) \leq C_1 + C_2 - R_1 - R_2. \end{cases} \tag{115}$$

and where $P_{W_C,W_1,W_2,X,Y_1,Y_2,U_1,U_2}(w_C, w_1, w_2, x, y_1, y_2, u_1, u_2)$ is the product

$$\begin{aligned} P_{W_C,W_1,W_2,X,Y_1,Y_2,U_1,U_2}(w_C, w_1, w_2, x, y_1, y_2, u_1, u_2) &= P_{W_C,W_1,W_2}(w_C, w_1, w_2) P_{X|W_C,W_1,W_2}(x | w_C, w_1, w_2) \\ &\cdot P_{Y_1,Y_2|X}(y_1, y_2 | x) P_{U_1|Y_1,W_1,W_C}(u_1 | y_1, w_1, w_C) \cdot P_{U_2|Y_2,W_2,W_C}(u_2 | y_2, w_2, w_C). \end{aligned} \tag{116}$$

Proof outline: The proof involves generating three vectors $\boldsymbol{w}_1, \boldsymbol{w}_1, \boldsymbol{w}_C$ i.i.d. according to P_{W_C,W_1,W_2}, where $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are distributed according to $\prod_{i=1}^n P_{W_1|W_C}(w_{1,i} | w_{C,i})$ and $\prod_{i=1}^n P_{W_2|W_C}(w_{2,i} | w_{C,i})$, respectively. The vector $\boldsymbol{w}_C$ is decoded at both agents, using typicality considerations. The random variables W_1, W_2 may be dependent given W_C, so for the transmitted signal to be typical for sufficiently large n, these vectors should be further quantized. This way, the agents receive both common and individual messages. After decoding the messages, the agents compress the received signals $\boldsymbol{y}_1, \boldsymbol{y}_2$ conditioned on the decoded $\boldsymbol{w}_1, \boldsymbol{w}_1, \boldsymbol{w}_C$. Then they forward the decoded messages and the compression information to the final destination. Notice that an extension to more than $T = 2$ follows the same line, however since it includes tedious notations, such as common messages to any subset of the agents, it is not included here.

Appendix VIII
Solving the linear programming presented in Remark 2

The solution of the dual of the linear programming problem is presented next.

Define the raw vector $\boldsymbol{x} = R_{\mathcal{T}}$, where bold symbols represent vectors or matrices. We have the following problem:

$$\max \boldsymbol{x} (\underbrace{1, \ldots, 1}_{T})^H \tag{117}$$

(where ^{H} is for the transpose operation) under the constraints:

$$\boldsymbol{A}\boldsymbol{x} \leq \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{b}_{BC} \\ \boldsymbol{b}_q \end{pmatrix}. \tag{118}$$

Where: $\boldsymbol{0}$ is a column vector of T zeros,

$$\boldsymbol{b}_{BC} = \begin{pmatrix} I(Y_1; W_1) \\ \vdots \\ I(Y_T; W_T) \\ \sum_{t=1,2} I(Y_t; W_t) - I(W_t; W_{t-1}) \\ \vdots \\ \sum_{t=1}^{T} I(Y_t; W_t) - I(W_t; W^{t-1}) \end{pmatrix} \tag{119}$$

and

$$\boldsymbol{b}_q = \begin{pmatrix} C_1 - I(Y_1; U_1 | U_{2,\dots,T}, W_{\mathcal{T}}) \\ \vdots \\ C_T - I(Y_T; U_T | U_{1,\dots,T-1}, W_{\mathcal{T}}) \\ C_1 + C_2 - I(Y_{1,2}; U_{1,2} | U_{3,\dots,T}, W_{\mathcal{T}}) \\ \vdots \\ \sum_{t=1}^{T} C_t - I(Y_{\mathcal{T}}; U_{\mathcal{T}} | W_{\mathcal{T}}) \end{pmatrix}. \tag{120}$$

Where these correspond to the first, second and third sets, respectively, of constraints on $R_{\mathcal{T}}$ in (21). Then $\boldsymbol{A}$ is:

$$\boldsymbol{A} = \begin{pmatrix} -1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & -1 \\ 1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & 1 \\ 11 & \dots & 0 \\ & \vdots & \\ 11 & \dots & 11 \\ 1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & 1 \\ 11 & \dots & 00 \\ & \vdots & \\ 11 & \dots & 11 \end{pmatrix}. \tag{121}$$

Since there is at least one feasible $\boldsymbol{x}$, which is $\sum_{t\in\mathcal{T}} R_t = 0$, the maximization of (117) is equal to the solution of the following dual problem:

$$\min \boldsymbol{\lambda}(\mathbf{0}^H \boldsymbol{b}_{BC}^H \boldsymbol{b}_q^H), \tag{122}$$

such that

$$\begin{cases} \boldsymbol{\lambda}\boldsymbol{A} = (1,\ldots,1)^H \\ \forall\ 0 < k : (\boldsymbol{\lambda})_k \geq 0. \end{cases} \tag{123}$$

Now since:

$$\begin{aligned} &\forall\ \mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{T} : \\ &\sum_{t\in\mathcal{S}_1\cup\mathcal{S}_2} [I(W_t;Y_t) - I(W_t;W_{\tilde{\mathcal{T}}}(\mathcal{S}_1\cup\mathcal{S}_2,t))] \leq \sum_{t\in\mathcal{S}_1} [I(W_t;Y_t) - I(W_t;W_{\tilde{\mathcal{T}}}(\mathcal{S}_1,t))] + \sum_{t\in\mathcal{S}_2} [I(W_t;Y_t) - I(W_t;W_{\tilde{\mathcal{T}}}(\mathcal{S}_2,t))], \end{aligned} \tag{124}$$

and

$$\begin{aligned} &\forall\ \mathcal{S}_1, \mathcal{S}_2 \subseteq \mathcal{T}, \mathcal{S}_1 \cap \mathcal{S}_2 = \phi : \\ &\sum_{t\in\mathcal{S}_1\cup\mathcal{S}_2} [C_t] - I(U_{\mathcal{S}_1\cup\mathcal{S}_2}; Y_{\mathcal{S}_1\cup\mathcal{S}_2} | U_{\mathcal{S}_1^C\cap\mathcal{S}_2^C}, W_{\mathcal{T}}) \leq \sum_{t\in\mathcal{S}_1} [C_t] - I(U_{\mathcal{S}_1}; Y_{\mathcal{S}_1} | U_{\mathcal{S}_1^C}, W_{\mathcal{T}}) + \sum_{t\in\mathcal{S}_2} [C_t] - I(U_{\mathcal{S}_2}; Y_{\mathcal{S}_2} | U_{\mathcal{S}_2^C}, W_{\mathcal{T}}) \end{aligned} \tag{125}$$

we get to (23), where the constraints (24) and (25) stem from the requirement that there will be at least one feasible $\boldsymbol{x}$.

APPENDIX IX
PROOF OF THEOREM 4

The proof for Theorem 4 is very similar to the one used for the general channel from Theorem 3, which appears in Appendix III. In fact, both theorems 3 and 4 are special cases of a generalization of Theorem 3, which includes the transmission of common messages to the agents, so that some subset of them can decode the same information ([28] Theorem 2). Such generalization appears in Appendix VII, considering $T = 2$.

In this Appendix we prove the achievable rate for the case when the channels are known to be degraded. In this case, it is beneficiary to use many common messages, which enables savings of link bandwidth. We use superposition coding for the messages to the agents, since it is known to be optimal for that case (degraded broadcast channel [27]).

Since the proof is very similar to the one provided in Appendix III, we outline only the differences:

1) On III-A, item 1), for the broadcast transmission, we now repeat the following step for $t = 1,\ldots,T$:

   - For every $\boldsymbol{w}^{t-1}$ (total of $2^{n\sum_{i=1}^{t-1} R_i}$) generated in the previous step, generate 2^{nR_t} vectors $\boldsymbol{w}_t$ according to $\prod_{i=1}^n P_{W_t|W^{t-1}}(w_{t,i}|w_i^{t-1})$.

- Label the resulting vectors of each step by M_t, where $M_t \in [1, 2^{nR_t}]$. Define $\mathcal{M}_{M_t}$ as the set labelled by M_t.
- Define the bin $\mathcal{M}_{M_\mathcal{T}} \triangleq \mathcal{M}_{M_1} \times \cdots \times \mathcal{M}_{M_T}$ as the product union of the sets $\{\mathcal{M}_{M_t}\}$.

2) On III-A, item 2), we now generate $2^{n[\hat{R}_t - C_t + R_t]}$ vectors $\boldsymbol{u}_t$ according to $\prod_{i=1}^n P_{U_t|W^t}(u_{t,i}|w_i^t)$ (instead of $\prod_{i=1}^n P_{U_t|W_t}(u_{t,i}|w_{t,i})$ in Appendix III). So here, the mapping of z_t to $\boldsymbol{u}_t$ depends on $\boldsymbol{w}^t$ (not just on $\boldsymbol{w}_t$) and is denoted by $\boldsymbol{u}_t(z_t, \boldsymbol{w}^t)$, and we have $2^{n\sum_{i=1}^t R_i}$ different mappings. Notice that M^t is in one-to-one correspondence with $\boldsymbol{w}^t$, so that we can write $\boldsymbol{u}_t(z_t, M^t)$.
3) On III-B, the transmitter sends $\boldsymbol{x}(M)$ to the channel. Here there is no need to find typical $\boldsymbol{w}_\mathcal{T}$ before transmitting to the channel.
4) On subsection III-C.1 (decoding at the agents), the t agent finds M^t such that

$$(\boldsymbol{w}^t(M^t), \boldsymbol{y}_t) \in \mathbf{T}_\epsilon^{t,4}, \tag{126}$$

instead of M_t in Appendix III. The typical set $\mathbf{T}_\epsilon^{t,4}$ is defined in the usual way.
5) On subsection III-C.2, for the compression, the agent looks for z_t such that

$$\left(\boldsymbol{u}_t(z_t, M^t), \boldsymbol{y}_t, \boldsymbol{w}^t\right) \in \mathbf{T}_\epsilon^{t,5}. \tag{127}$$

Where $\mathbf{T}_\epsilon^{t,5}$ is defined in the usual way.
6) On subsection III-D: Considering the change of labeling, the destination here performs the same as the destination presented in Appendix III.
7) In subsection III-E, the error analysis, there are several differences in the definitions and the calculation of the probability of some error events:

- E_1 is no longer declared by the transmitter.
  E_2-E_4 Consider the typicality of $(\boldsymbol{w}^t, \boldsymbol{y}_t)$ instead of $(\boldsymbol{w}_t, \boldsymbol{y}_t)$.
  E_5-E_8 are changed by the dependence of $\boldsymbol{u}_t$ in $\boldsymbol{w}^t$ instead of $\boldsymbol{w}_t$. This means a change in the variables which are included in the definition of the typical set.
- Error probabilities for E_2, E_6, E_9: As in Appendix III, the probabilities of these events are bounded by ϵ due to Lemma 2 and Lemma 4, (generalized Markov Lemma, Lemma 3.4 in [30]).
- Error probabilities for E_3, E_4: According to Lemma 3, the probability of another typical vector $\hat{\boldsymbol{w}}^t$ is upper bounded by $2^{-n[I(W^t;Y_t)-\epsilon]}$. Since there are $2^{n\sum_{i=1}^t R_i} - 1$ such vectors, the error probability can be made arbitrarily small conditioned that

$$\forall\ 1 \leq t \leq T:\ \sum_{i=1}^t R_i < I(W^t; Y_t) = \sum_{i=1}^t I(W_i; Y_i | W^{i-1}). \tag{128}$$

  This condition is fulfilled if:

$$\forall\ 1 \leq t \leq T:\ R_t < I(W_t; Y_t | W^{t-1}). \tag{129}$$

- Error probability for event E_5: repeating what was done in Appendix III and considering that $\boldsymbol{u}_t$ was generated according to $P_{U_t|W^t,Y_t}$ and not $P_{U_t|W_t,Y_t}$, the probability of E_5 is as small as desired as long as:

$$\hat{R}_t > I(U_t;Y_t|W^t). \tag{130}$$

8) Unlike section III-F, which tries to bound the cardinality of the auxiliary variables, we can now use the support Lemma ([29] p. 310), as in [19], to limit the cardinalities of both $\mathcal{W}_t$ and $\mathcal{U}_t$. We start by rewriting the rate equation (22) and the constraints (29) as functionals of some generic $Q_{X,U_\mathcal{T},W^{T-1}}$. This way, we get $2+2^T-1$ functionals on $Q_{X,U_\mathcal{T},W^{T-1}}$, calculated from $I(X;U_\mathcal{T}|W_\mathcal{T})$ from (22), $H(Y_\mathcal{T}|W_\mathcal{T})$ from the first set of (29) and finally $I(Y_\mathcal{S};U_\mathcal{S}|U_{\mathcal{S}^C},W_\mathcal{T})$ for all $\mathcal{S}\neq\phi$ (2^T-1 such sets) from the second set of (29). In addition, the marginal of Q, with respect to X must be equal to the given P_X. So in total, there are $|\mathcal{X}|-1+2+2^T-1$ functionals on Q, and as a result, the cardinality of W_T can be limited by $|\mathcal{W}_T|\leq|\mathcal{X}|+2^T$. We can apply this technique repeatedly, for the other W_t, where $t<T$. For any such $t<T$, there are $T-t$ functionals as a consequence of limiting $\{R_i\}_{i=t}^T$, in (29) in addition to the marginal distributions with respect to $(\{W_i\}_{t+1}^T,X)$. Over all, the cardinality of W_t can be limited by:

$$|\mathcal{W}_t|\leq|\mathcal{X}|\prod_{i=t+1}^{T}|\mathcal{W}_i|-1+2^T-1+T-t+2=|\mathcal{X}|^{T-t+1}+\sum_{i=t}^{T}|\mathcal{X}|^{T-i}(2^T+i-t). \tag{131}$$

Next, we limit the cardinality of $U_\mathcal{T}$, when provided with the auxiliaries $W_\mathcal{T}$ (which have bounded cardinality). For this, we can repeat what was done in III-F, with the difference that here we look at $P_{Y_t,W^t|U_t}$. So we can limit the cardinality of the auxiliary variables $U_\mathcal{T}$ by:

$$|\mathcal{U}_t|\leq|\mathcal{Y}_t||\mathcal{W}^t|+2^T. \tag{132}$$

Considering these differences, the constrains on $\{R_t\}$, and thus also the cardinality limits, are the main changes between Theorem 3 and 4. So by replacing (61) with (129), one gets to (29).

APPENDIX X
PROOF OF THEOREM 5

First recall the definitions of the Gaussian channel (Y_t,X_t,N_t) from subsection VI-A.

Proof:

A. Direct part

Define the auxiliary random variables U_t as

$$U_t=Y_t+W_t, \tag{133}$$

where $\{W_t\}$ are Gaussian i.i.d. random variables, independent of Y_t (no connection with W in the previous appendices), with zero mean and variance

$$P_{W_t}=\frac{2^{-2I(U_t;Y_t|X)}P_{N_t}}{1-2^{-2I(U_t;Y_t|X)}}. \tag{134}$$

Equation (133) can also be written as

$$U_t = X + D_t, \tag{135}$$

where $D_t = W_t + N_t$, and therefore also

$$P_{D_t} \triangleq \mathrm{E}D_t^2 = P_{N_t} + P_{W_t}. \tag{136}$$

Define $r_t \triangleq I(Y_t; U_t|X)$, so that P_{D_t} can be expressed as

$$\frac{1}{P_{D_t}} = \frac{1 - 2^{-2r_t}}{P_{N_t}}. \tag{137}$$

The terms $\{r_t\}$ can take any positive value, and then $\{P_{D_t}\}$ are determined accordingly (this space $r_{\mathcal{T}} \in \{\mathbb{R}^+\}^T$ is limited, as seen in the next lines, by the available bandwidths). The last equality can be used to explicitly express the maximum mutual information (through maximal ratio combining) in terms of $\{r_t\}$ between X and some subset $U_{\mathcal{S}}$,

$$\begin{aligned} I(X; U_{\mathcal{S}}) &= \frac{1}{2}\log_2\left(1 + \frac{P_X}{(\sum_{t\in\mathcal{S}} P_{D_t})^{-1}}\right) && (138)\\ &= \frac{1}{2}\log_2\left(1 + P_X \sum_{t\in\mathcal{S}} P_{D_t}\right) && (139)\\ &= \frac{1}{2}\log_2\left(1 + P_X \sum_{t\in\mathcal{S}} \frac{1 - 2^{-2r_t}}{P_{N_t}}\right). && (140)\end{aligned}$$

Now we can apply Corollary 1 and then prove the direct part of Theorem 5. Although Corollary 1 considered only discrete channels, and the Gaussian channel is not discrete, the extension is based on standard techniques also used in Oohama [15], who showed the validity of a generalized Markov Lemma for continuous random variables.

B. Upper bound for Gaussian P_X

The upper bound is based on (113), rather than on the single letter expression from Corollary 2, which is too loose for this case. We redefine

$$r_t \triangleq \frac{1}{n} I(Y_t^n; V_t|X^n), \tag{141}$$

and then use the following lemma, which is due to Oohama [15] to upper bound $\frac{1}{n}I(V_{\mathcal{S}}; \boldsymbol{X})$.

Lemma 5:

$$\frac{1}{n}I(V_{\mathcal{S}}; \boldsymbol{X}) \leq \frac{1}{2}\log_2\left(1 + P_X \sum_{t\in\mathcal{S}} \frac{1 - 2^{-2r_t}}{P_{N_t}}\right). \tag{142}$$

Lemma 5 together with (113) and (141) completes the proof. ■

Next we give also the proof of Lemma 5, from [15], which is based on the entropy power inequality.

Proof:

Define the minimum mean square error estimator of $\boldsymbol{X}$ from $\boldsymbol{Y}_{\mathcal{S}}$ by $\hat{\boldsymbol{X}} = \sum_{t\in\mathcal{S}} \frac{P_{\hat{N}_t}}{P_{N_t}} \boldsymbol{Y}_t$. Then we have:

$$\boldsymbol{X} = \hat{\boldsymbol{X}} + \hat{\boldsymbol{N}} \tag{143}$$

where $\hat{\boldsymbol{N}}$ is independent with the estimator $\hat{\boldsymbol{X}}$ or with $\boldsymbol{Y}_{\mathcal{S}}$ and is distributed i.i.d. Gaussian with zero mean and variance $P_{\hat{N}} = (\frac{1}{P_X} + \sum_{t\in\mathcal{S}} \frac{1}{P_{N_t}})^{-1}$. We can use the entropy power inequality:

$$2^{\frac{2}{n}h(\boldsymbol{X}|V_{\mathcal{S}})} \geq 2^{\frac{2}{n}h(\hat{\boldsymbol{X}}|V_{\mathcal{S}})} + 2\pi P_{\hat{N}}. \tag{144}$$

Define $\lambda \triangleq \frac{1}{n}I(\boldsymbol{X};V_{\mathcal{S}})$ and notice that $h(\boldsymbol{X}|\hat{\boldsymbol{X}},V_{\mathcal{S}}) = h(\boldsymbol{X}|\hat{\boldsymbol{X}})$, since $\hat{\boldsymbol{X}}$ is best estimator of $\boldsymbol{X}$ out of $\boldsymbol{Y}_{\mathcal{S}}$ and the Markov chain: $\boldsymbol{X} - \boldsymbol{Y}_{\mathcal{S}} - V_{\mathcal{S}}$. Then we can rewrite (144) as

$$2^{-2\lambda+\frac{2}{n}h(\boldsymbol{X})} \geq 2^{\frac{2}{n}[h(\hat{\boldsymbol{X}}|\boldsymbol{X},V_{\mathcal{S}})+I(\boldsymbol{X};\hat{\boldsymbol{X}})]-2\lambda} + 2\pi P_{\hat{N}}. \tag{145}$$

Next, we can apply the entropy power inequality again, to lower bound $h(\hat{\boldsymbol{X}}|\boldsymbol{X},V_{\mathcal{S}})$:

$$\begin{aligned} 2^{\frac{2}{n}h(\hat{\boldsymbol{X}}|\boldsymbol{X},V_{\mathcal{S}})} &\geq \sum_{t\in\mathcal{S}} 2^{\frac{2}{n}h\left(\frac{P_{\hat{N}_t}}{P_{N_t}}\boldsymbol{Y}_t|\boldsymbol{X},V_{\mathcal{S}}\right)} = \\ &\sum_{t\in\mathcal{S}}\left(\frac{P_{\hat{N}_t}}{P_{N_t}}\right)^2 2^{\frac{2}{n}h(\boldsymbol{Y}_t|\boldsymbol{X},V_{\mathcal{S}})} = \sum_{t\in\mathcal{S}}\left(\frac{P_{\hat{N}_t}}{P_{N_t}}\right)^2 2^{-2r_t+\frac{2}{n}h(\boldsymbol{Y}_t|\boldsymbol{X})} \\ &= \sum_{t\in\mathcal{S}}\left(\frac{P_{\hat{N}_t}}{P_{N_t}}\right)^2 P_{N_t}2^{-2r_t}. \end{aligned} \tag{146}$$

So that overall we have:

$$2^{-2\lambda}P_X \geq \left(\sum_{t\in\mathcal{S}}\left(\frac{P_{\hat{N}_t}}{P_{N_t}}\right)^2 P_{N_t}2^{-2r_t}\right)\frac{P_X}{P_{\hat{N}}}2^{-2\lambda} + P_{\hat{N}}. \tag{147}$$

Or put differently:

$$2^{2\lambda} \leq \frac{P_X}{P_{\hat{N}}} - P_X\sum_{t\in\mathcal{S}}\frac{2^{-2r_t}}{P_{\hat{N}_t}}. \tag{148}$$

Which concludes the proof. ■

Notice that the proof considered nomadic transmitter with P_X which is restricted to be Gaussian. As noticed in subsection VI-B, removing this limitation can improve the achievable rate.

REFERENCES

[1] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: an achievable rate region," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1877–1894, Aug. 2003.

[2] C. Ng and A. Goldsmith, "Transmitter cooperation in ad-hoc wireless networks: Does dirty-paper coding beat relaying?" in *the proceedings of IEEE INFORMATION THEORY WORKSHOP (ITW2004)*, San Antonio, Texas, October 2004, pp. 277–282.

[3] T. M. Cover and A. A. El-Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Jan 1979.

[4] G. Kramer, M. Gastpar, and P. Gupta, "Information-theoretic multi-hopping for relay networks," in *International Zürich seminar on communication*, Swizerland, Feb. 2004, pp. 192–195.

[5] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec 2004.

[6] E. Erkip and M. Yuksel, "Diversity gains and clustering in wireless relaying," in *Proc IEEE Int Symp Info Theory ISIT2004*, Chicago, IL, June 2004, p. 400.

[7] R. Dabora and S. D. Servetto, "Broadcast channels with cooperating receivers: a downlink for sensor reachback problem," in *Proc IEEE Int Symp Info Theory ISIT2004*, Chicago,IL, June 2004, p. 176.

[8] O. Lévêque and E. Telatar, “Information theoretic upper bounds on the capacity of large extended ad hoc wireless networks,” *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 858–865, Mar. 2005.

[9] S. I. Gel’fand and M. Pinsker, “Source coding with incomplete side information,” *Prob. Control. Inform. Theory*, vol. 15, no. 2, pp. 45–57, 1979.

[10] T. Berger, Z. Zhang, and H. Viswanathan, “The CEO problem,” *IEEE Trans. Inform. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[11] A. El-Gamal and T. Cover, “Achievable rates for multiple descriptions,” *IEEE Trans. Inform. Theory*, vol. 6, pp. 8151–8157, November 1982.

[12] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.

[13] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan 1976.

[14] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, “An upper bound on the sum rate distortion function and its corresponding rate allocation schemes for the CEO problem,” *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 977–987, Aug. 2004.

[15] Y. Oohama, “The rate-distortion function for the quadratic Gaussian CEO problem,” *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[16] V. Prabhakaran, D. Tse, and K. Ramchandran, “Rate region of the quadratic Gaussian CEO problem,” in *Proc. of IEEE Int. Symp. Info. Theory (ISIT2004)*, Chicago, IL, June 2004, p. 119.

[17] Y. Oohama, “Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder,” *IEEE Trans. Inform. Theory*, vol. 51, no. 7, pp. 2577–2593, July 2005.

[18] R. Zamir, S. Shamai, and U. Erez, “Nested linear/lattice codes for structured multiterminal binning,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[19] Y. Steinberg and N. Merhav, “On successive refinement for the Wyner-Ziv problem,” *IEEE Trans. Inform. Theory*, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.

[20] J. Barros and S. D. Servetto, “A note on cooperative multiterminal source coding,” in *the Proceedings of the 38th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2004.

[21] A. D. Murugan, P. K. Gopala, and H. El-Gamal, “Correlated sources over wireless channels: cooperative source-channel coding,” *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 988–999, Aug. 2004.

[22] A. Orlitsky and R. Roche, “Coding for computing,” *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[23] G. Kramer and S. A. Savari, “Communicating probability distributions,” *IEEE Trans. Inform. Theory*, vol. 53, no. 2, pp. 518–525, Feb. 2007.

[24] B. E. Schein, “Distributed coordination in network information theory,” Ph.D. dissertation, MIT, October 2001.

[25] S. Shamai and A. Lapidoth, “Fading channels: How perfect need ”perfect side information” be?” *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[26] I. G. Stiglitz, “Coding for a class of unknown channels,” *IEEE Trans. Inform. Theory*, vol. IT-12, no. 2, pp. 189–195, April 1966.

[27] T. M. Cover and J. A. Thomas, *Elements of Information theory*. John Wiley & Sons, Inc., 1991.

[28] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,” *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 306–311, May 1979.

[29] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[30] T. S. Han and K. Kobayashi, “A unified achievable rate region for a general class of multiterminal source coding systems,” *IEEE Trans. Inform. Theory*, vol. IT-26, no. 3, pp. 277–288, May 1980.

[31] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” Bell Labs, Lucent, Murray Hill, NJ,” Technical Memorandum, Feb. 2004.

[32] M. R. Aref “Information Flow in Relay Networks,” Ph.D. thesis, Stanford University, Stanford, CA, October 1980.

Amichai Sanderovich (S'03) Received his B.Sc. (summa cum laude) and M.Sc. (with distinction) from the Technion - Israel Institute of Technology, Haifa, Israel, in 1999 and 2004, respectively.
He is working towards the Ph.D. degree in the department of Electrical Engineering, at the Technion. His research areas include information theory and communication, from the basic concepts and fundamental limits to the practical approaches. He is interested in the information theoretic performance limits of distributed detection, collaborative communication, cellular schemes, ad-hoc networks, sensor networks relay schemes and network coding. Additionally fading channels and (Multiple Input Multiple Output) MIMO systems, and in particular, the combination of networks and a MIMO physical channel are of special interest. Practical modern iterative codes are also of interest, with special interest in LDPC codes and their ultimate performance, and in the combination of sub-optimal detectors.

Shlomo Shamai (Shitz) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975-1985 he was with the Communications Research Labs in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now the William Fondiller Professor of Telecommunications. His research interests encompasses a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is an IEEE Fellow and a member of the Union Radio Scientifique Internationale (URSI). He is the recipient of the 1999 van der Pol Gold Medal of URSI, and a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 joint IT/COM societies paper award, and the 2007 IEEE Information Theory Society Paper Award. He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the SHANNON THEORY OF THE IEEE TRANSACTIONS ON INFORMATION THEORY, and also serves on the Board of Governors of the Information Theory Society.

Gerhard Kramer (S'91-M'94) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Manitoba, Winnipeg, MB, Canada, in 1991 and 1992, respectively, and the Dr. sc. techn. (Doktor der Technischen Wissenschaften) degree from the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, in 1998.

From July 1998 to March 2000, he was with Endora Tech AG, Basel, Switzerland, as a communications engineering consultant. Since May 2000 he has been with Bell Laboratories, Alcatel-Lucent, Murray Hill, NJ.

Dr. Kramer is currently serving as an Associate Editor for Shannon Theory for the IEEE Transactions on Information Theory. He served as Guest Editor-in-Chief for an IEEE Transactions on Information Theory Special Issue on Relaying and Cooperation in Communication Networks in 2006-2007, and as Publications Editor for the same Transactions during 2004-2005. Dr. Kramer is serving as Co-Chair of the Technical Program Committee for the 2008 IEEE International Symposium on Information Theory. He is a co-recipient of the IEEE Communications Society 2005 Stephen O. Rice Prize paper award, a Bell Labs President's Gold Award in 2003, and a recipient of an ETH Medal in 1998