Communication Via Decentralized Processing

Amichai Sanderovich, *Student Member, IEEE*, Shlomo Shamai (Shitz), *Fellow, IEEE*, Yossef Steinberg, *Member, IEEE*, and Gerhard Kramer, *Member, IEEE*

***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, and 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]. Multihop relaying makes use of several intermediate wireless nodes to assist the communication between two nodes that are far apart, e.g., [2]. 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], [5] 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 [2], [6], [7], while cooperation between receiving nodes in a degraded broadcast channel is described in [8]. We conclude with an upper bound derived in [9], 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 also include base stations or access points.

Problems of conveying a source which is observed by remote agents to a single destination are built around similar settings, where the source is modeled as a sequence of independent and identically distributed (i.i.d.) random variables. Such problems are analyzed in information-theoretic frameworks such as distributed source coding, lossless CEO (Chief Executive Officer) [10], CEO [11] and sensor network problems. A small sample from the extensive work that is relevant to our distributed detection setting includes [12], [13], and [10] for distributed source coding. Allowing distributed lossy source encoding, as opposed to centralized encoding [14], is still essentially an unsolved problem. An exception is the Gaussian CEO problem [15], [16] which was recently solved using the entropy power inequality in [17], [18]. Multiterminal lattice approaches are described in [19]. These rate–distortion problems are linked to network models in [20]–[22]. The use of other measures, instead of the standard distortion, is addressed, for example in [23] and [24]. Schein's dissertation [25] focuses on the problem of communicating via two agents, and develops several achievable rates.

Here we consider the problem of reliable communication from a *nomadic* transmitter to a remote destination via nondecoding agents that are connected to the destination via lossless links. These agents have noisy versions of the transmitted signal, and transmit a predetermined number of bits to the destination without any errors. The destination is reached only via the agents that serve as access points. By *nomadic* transmitter we mean that the receiving devices cannot or will not decode the transmitted signal. Such a setting is of interest for numerous applications. The main motivation, however, is for systems where the agents cannot 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 out to be capacity achieving for the deterministic channel.

The rest of the paper is organized as follows: in Section II, we describe the problem. An achievable rate and a capacity 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 nondegraded channels in Section V, where the capacity is fully characterized for the deterministic channel. The Gaussian channel is considered

Manuscript received November 20, 2005; revised October 15, 2007. This work was supported by the EU 7th framework program via the NEWCOM and NEWCOM++ networks of excellence and the REMON Consortium. The work of G. Kramer was supported in part by the Board of Trustees of the University of Illinois Subaward no. 04-217 under National Science Foundation Grant CCR-0325673 and the Army Research Office under ARO Grant W911NF-06-1-0182. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Adelaide, Australia, September 2005.

A. Sanderovich, S. Shamai (Shitz), and Y. Steinberg are with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: amichi@tx.technion.ac.il; sshlomo@ee.technion.ac.il; ysteinbe@ee.technion.ac.il).

G. Kramer is with Bell Laboratories, Alcatel-Lucent, Murray Hill, NJ 07974-0636 USA (e-mail: gkr@bell-labs.com).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.

Color versions of Figures 2 and 3 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2008.924659

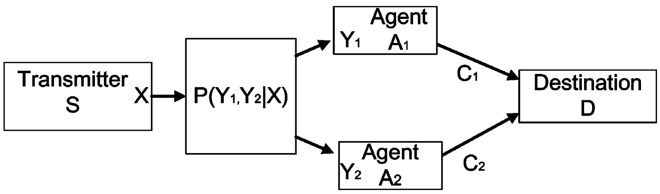

Fig. 1. A system with two agents between the transmitter and the destination.

in Section VI. For the case where the agents are unaware of the code used, and where the codebook is Gaussian, we characterize the capacity region.

In this paper, we use capital letters, e.g., X, for random variables, lower case letters, e.g., x, for the realization of these variables, and calligraphic letters, e.g., $\mathcal{X}$, for their alphabets. Vectors are of length n unless otherwise specified and are denoted by bold-face letters, e.g., $\boldsymbol{X}$, $\boldsymbol{x}$, or vector spaces by calligraphic bold-face letters, e.g., $\boldsymbol{\mathcal{X}}$. 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 $\mathcal{S} \cap \mathcal{S}^C = \phi$. The cardinality of any set $\mathcal{T}$ is written as $|\mathcal{T}|$. A subscript, e.g., X_i, denotes the ith 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 of the event $A_1 = a_1, \ldots, A_L = a_L$.

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, as seen in Fig. 1 for $T = 2$. Suppose the agents do not know the transmitter's codebook. We model this by having the transmitter use one code out of a set of possible codes. The agents know some characteristics of these codes, e.g., their rate and that they are capacity achieving over a standard single-user Gaussian channel. An example can be a set of interleavers and also a set of modulation techniques. Such random coding is also used in [26] for a mismatch scenario. The advantages of random coding were demonstrated in [27] for unknown channels.

The following properties and definitions hold, unless stated otherwise.

1) The channel input (output of the transmitter S) is $\boldsymbol{X} \in \mathcal{X}$.
2) The T agents $A_1, \ldots, A_T$ receive the outputs of a memoryless broadcast channel without feedback, defined by

$$P_{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T|\boldsymbol{X}}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T|\boldsymbol{x}) = \prod_{i=1}^{n} P_{Y_1, \ldots, Y_T|X}(y_{1,i}, \ldots, y_{T,i}|x_i) \quad (1)$$

where $y_{t,i} \in \mathcal{Y}_t$. Denote $\mathcal{T} \triangleq \{1, \ldots, T\}$. The agents have full knowledge of the distribution P_X, induced by the nomadic transmission, and thus also of $P_{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$.

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 $[1, 2, \ldots, |\mathcal{X}|^{n2^{nR}}]$, 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

$$\phi_{S,F} : [1, \ldots, 2^{nR}] \rightarrow \mathcal{X}^n \quad (2)$$

and the probability of selecting f is

$$P_F(f) = \prod_{m=1}^{2^{nR}} P_{\boldsymbol{X}}(\phi_{S,f}(m)) \quad (3)$$

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

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

so that

$$V_t = \phi_{At}(\boldsymbol{Y}_t). \quad (5)$$

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

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

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

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

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, without 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

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

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 for both cases.

III. An Achievable Rate

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 [32]. 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 the nomadic transmitter.

Proof: See Appendix III and Remark 2. □

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 an 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 cardinalities and the probability spaces of the random variables are the same as in Theorem 1.

Proof: See Appendix IV. □

Remark 1: The achievable rate of Theorem 1 can be further improved by considering some common knowledge shared by the agents. For example, such information can be another transmission which was decoded, by all agents, and they can thus compress conditioned on this common information.

IV. A Capacity Upper Bound

An upper bound on the 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. Since the agents are ignorant of the codebook used, there is an inherent loss compared with the case where the agents know the codebook. The achievable rate when the agents are unrestricted can be upper-bounded by the cut-set bound [28]. This gap between the achievable rate and the cut-set bound will be demonstrated for the Gaussian channel in Section VI.

Theorem 2: A 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 $|\mathcal{W}|$, and 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 (RHS) of (10) is smaller than the RHS 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 RHS of (15) is identical to the RHS of (10). This is since

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

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, namely

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

where again, the random variables have the same cardinalities and satisfy 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 final destination.

The next theorem is based on Marton's scheme [29] for 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). Let $\tilde{\mathcal{T}}(\mathcal{S},t) \triangleq \{i : i \in \mathcal{S}$ and $i < t\}$ and W_ϕ be a constant.

Theorem 3: For decoding agents, any rate R satisfying

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

with the constraints as shown in (21) at the bottom of the page, and with the joint distribution

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

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 $w_{\mathcal{T}}$. 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 adding a time-sharing random variable to the rate region of Theorem 3.

Proof: The proof appears in Appendix III and uses compression in addition to Marton's broadcast coding. □

Remark 2: 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 3: 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

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

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 4: The rate (20) can be improved by sending common broadcast messages in addition to the individual broadcast messages to the agents. This is done by extending Theorem 2 in [29] to more than two users and adding compression. Notice that such a construction includes Theorem 3 and Theorem 4 below as special cases. Such a scheme is given in Appendix VII for two agents.

Corollary 3: For the case of deterministic channels, where $Y_t = g_t(X)$ for some 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 U_t to be constant and $W_t = Y_t$ for all 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 [33]).

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

For the case where the channels $P_{Y_t|X}$ are either stochastically or physically degraded (see [28, Sec. 14.6.2]) we can use superposition coding, which is known to achieve capacity over degraded broadcast channels.

The received signal Y_2 is a physically degraded version of Y_1 if the following forms a Markov chain:

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

Notice that this relation leaves $I(X;Y_1,Y_2) = I(X;Y_1)$. On the other hand, Y_2 is a stochastically degraded version of Y_1 if 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 (28)). Since (27) is not necessarily true we can have $I(X;Y_1,Y_2) \geq I(X;Y_1)$. 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

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

and the joint distribution

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

is achievable, and one can restrict attention to

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

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

Proof: See Appendix IX. □

Remark 5: Theorem 3 does not seem to include Theorem 4 as a special case, as it does not account for a common rate (see Remark 4).

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 3 which is built along the lines of Corollary 1. This gives the rate

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

VI. The Gaussian Channel

The Gaussian channel is defined by $Y_t = X + N_t$, where $\{N_t\}_{t=1}^T$ are independent Gaussian random variables with $\mathrm{E}N_t^2 = \rho_{N_t}$ and $\mathrm{E}N_t = 0$ where E denotes statistical expectation. Let X be zero mean Gaussian with variance $\mathrm{E}X^2 = \rho_X$. Here we use ρ to denote the variance of a random variable.

We use Corollary 1 with continuous alphabets instead of discrete, where this extension relies on standard arguments (see [16], for example). We also use the generalized Markov Lemma for Gaussian variables that appears in [18].

A. Nondecoding Agents

We prove the following result in Appendix X.

Theorem 5: The capacity of the nomadic transmitter for the Gaussian channel and with X chosen to be a Gaussian random variable, is

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

Proof: Use Corollary 1 for the direct part and the upper bound from (109) along with results from [16] for the converse part. □

Note that we restrict the transmitter to use Gaussian codebooks. The parameters $\{r_t\}_{t=1}^T$ in (34) indicate the bandwidth wasted by quantizing the additive noise, which cannot be avoided because of the nomadic transmitter. This bandwidth reduces the bandwidth for forwarding the actual transmission to $(C_t - r_t)$, and, on the other hand, improves the expected signal-to-noise ratio at the destination to $\rho_X \sum_{t\in\mathcal{T}} \frac{1-2^{-2r_t}}{\rho_{N_t}}$. Notice that (34) is concave in r_t, so that it can be efficiently maximized 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 capacity 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 $\gamma = \frac{\rho_X}{\rho_{N_1}} = \frac{\rho_X}{\rho_{N_2}}$, 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: Suboptimality of Gaussian Signaling

The previous section described the capacity of the nomadic transmitter in the Gaussian setting when the transmitter used a Gaussian codebook. However, Gaussian signaling is not necessarily optimal because of the capacity limitations between the agents and the destination. For example, suppose that $C_1 = C_2 = 1$, and we use binary phase-shift keying (BPSK) at the transmitter. The agents know that BPSK was used, and can

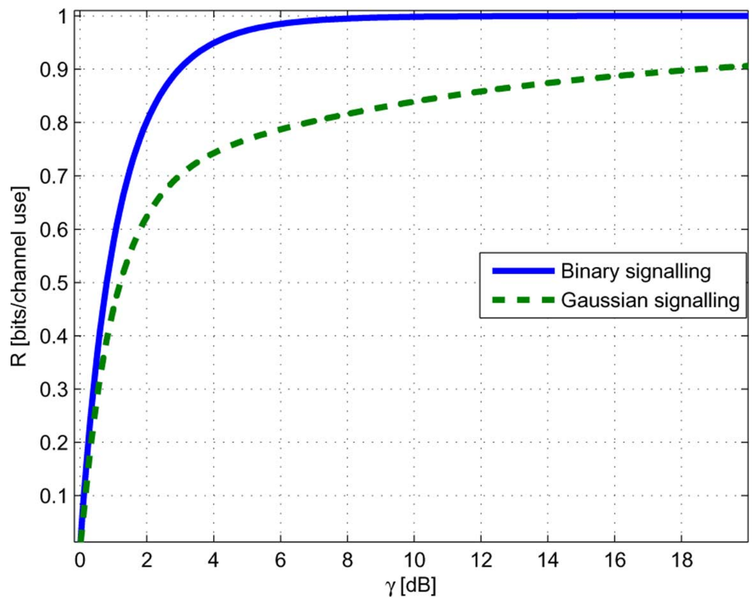

Fig. 2. The achievable rate of a system with two agents, with link bandwidths of $C_1 = C_2 = 1$ and a signal-to-noise ratio of γ. The dotted line designates the use of Gaussian signaling at the transmitter and the solid line designates the use of BPSK.

demodulate every received channel output into one bit ($j = 1, \ldots, n$)

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

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

$$\begin{aligned} R_{\text{bpsk}} &\leq \frac{1}{n} I(\boldsymbol{X}; V_1, V_2) \\ &= I(X(k); V_1(k), V_2(k)) \\ &= G(Q(\sqrt{2\rho_X})) \end{aligned}$$

where

$$\begin{aligned} G(x) \triangleq\, & (1-x)^2 \log_2\left((1-x)^2\right) + x^2 \log_2\left(x^2\right) \\ & - (1 - 2x(1-x)) \log_2\left(\frac{1}{2}(1 - 2x(1-x))\right) \end{aligned}$$

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 Fig. 2. We indeed see that BPSK signaling outperforms Gaussian signaling. This is because demodulation is some form of primitive decoding, which is not possible for the Gaussian signaling.

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 ρ_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

$$R < \sum_{t=1}^{2} R_t + \frac{1}{2} \log_2 \left(1 + \alpha \frac{\rho_X}{\rho_N} \sum_{t=1}^{2} (1 - 2^{-2r_t})\right) \tag{37}$$

where $\{r_t, R_t, \alpha\}$ satisfy the conditions in (38) shown at the bottom of the page.

Using a time sharing random variable can improve the rates for this example. The achievable rate as a function of the bandwidth C, for a signal-to-noise ratio $\frac{\rho_X}{\rho_N} = 10$, is presented in Fig. 3. In this figure, the leftmost dashed line $R = 2C$ and the upper flat dashed line $R = \frac{1}{2}\log_2(1 + 2\frac{\rho_X}{\rho_N})$ are the two cut-set bounds [28], and the lower flat dashed line is the rate of a system without compression $R = \frac{1}{2}\log_2(1 + \frac{\rho_X}{\rho_N})$. The dotted line represents time sharing, which is 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

$$\begin{cases} 0 \leq \alpha \leq 1 \\ t = 1, 2 : 0 \leq R_t \leq C \\ \sum_{t=1}^{2} R_t < \frac{1}{2} \log_2 \left(\frac{\rho_N + \rho_X}{\rho_N + \alpha \rho_X}\right) \\ \forall\, \mathcal{S} \subseteq \{1, 2\} : \sum_{\mathcal{S}} [C - R_t] > \sum_{\mathcal{S}} r_t + \frac{1}{2} \log_2 \left(1 + \alpha \rho_X \sum_{t=1}^{2} \frac{1 - 2^{-2r_t}}{\rho_N}\right) - \frac{1}{2} \log_2 \left(1 + \alpha \frac{\rho_X}{\rho_N} \sum_{\mathcal{S}^C} (1 - 2^{-2r_t})\right). \end{cases} \tag{38}$$

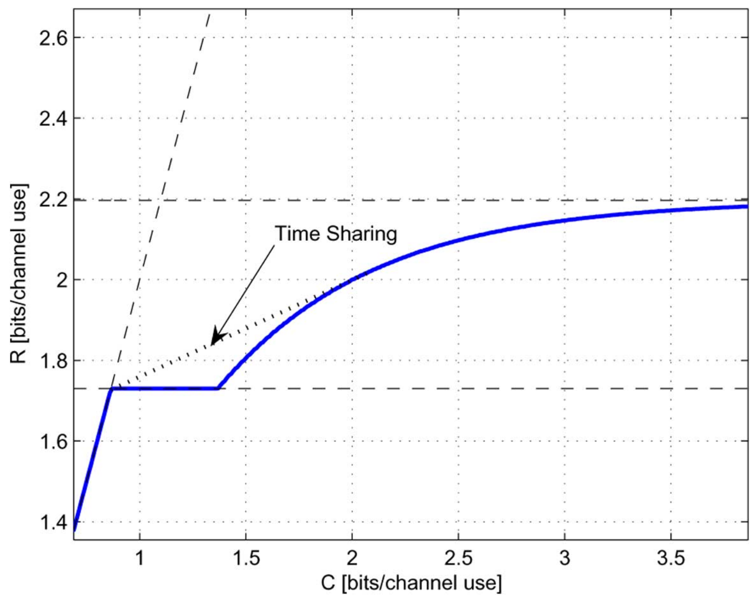

Fig. 3. The achievable rate of a system with two agents, each with link bandwidth of C and a signal-to-noise ratio of 10 dB. The dotted line designates time sharing, and the dashed lines represent the cut-set bounds [28]. The lower flat dashed line is the achievable rate for a system without compression.

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 the achievable rate is 50% of the total bandwidth allocated to the links. It also demonstrates that when the bandwidths of the links are smaller than the sum of capacities of the corresponding broadcast channel $C_1 + C_2 < \frac{1}{2}\log_2(1 + \frac{\rho_X}{\rho_N})$, the achievable rate in the nomadic setting using Gaussian codebooks (35) is strictly smaller than in the fixed transmitter setting (the cut-set bound, $R = C_1 + C_2$).

VII. Conclusion

Communication via distributed 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, 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 [18]. 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.

Appendix I
Definitions and Lemmas

As is commonly done (see [28, Sec. 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 otherwise

$$\mathbf{T}_\epsilon \triangleq \Bigg\{ \boldsymbol{a}_\mathcal{L} : \forall\, \mathcal{S} \subseteq \mathcal{L},\ \forall\, a_\mathcal{S} \in \mathcal{A}_\mathcal{S}, \left| \frac{1}{n} N(a_\mathcal{S}|\boldsymbol{a}_\mathcal{S}) - P_{A_\mathcal{S}}(a_\mathcal{S}) \right| < \frac{\epsilon}{|\mathcal{A}_\mathcal{S}|} \Bigg\} \quad (39)$$

where $N(a_\mathcal{S}|\boldsymbol{a}_\mathcal{S})$ denotes the number of occurrences of the symbol $a_\mathcal{S}$ in the vector $\boldsymbol{a}_\mathcal{S}$. Define the jointly ϵ-typical set $\mathbf{T}_\epsilon$ of $(\boldsymbol{a}_\mathcal{L}, \boldsymbol{w}_\mathcal{L})$ similarly.

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

$$P(\boldsymbol{a}_\mathcal{L} \in \mathbf{T}_\epsilon) \geq 1 - \epsilon. \quad (40)$$

Lemma 3: For some $\mathcal{S} \subseteq \mathcal{L}$ let $\boldsymbol{a}^n_\mathcal{L}$ be generated according to

$$\boldsymbol{a}_\mathcal{L} \sim \prod_{i=1}^{n} \left\{ P_{A_{\mathcal{S}^C}|W_\mathcal{L}}(a_{\mathcal{S}^C,i}|w_{\mathcal{L},i}) \prod_{l\in\mathcal{S}} P_{A_l|W_l}(a_{l,i}|w_{l,i}) \right\} \quad (41)$$

where $\boldsymbol{w}_\mathcal{L}$ is a vector that belongs to $\mathbf{T}_\epsilon$. Then we have

$$\Pr((\boldsymbol{a}_{1,\ldots,L}, \boldsymbol{w}_\mathcal{L}) \in \mathbf{T}_\epsilon) \geq 2^{-n[H(A_{\mathcal{S}^C}|W_\mathcal{L}) - H(A_\mathcal{L}|W_\mathcal{L}) + \sum_{l\in\mathcal{S}} H(A_l|W_l) + \epsilon_1]} \quad (42)$$

$$\Pr((\boldsymbol{a}_{1,\ldots,L}, \boldsymbol{w}_\mathcal{L}) \in \mathbf{T}_\epsilon) \leq 2^{-n[H(A_{\mathcal{S}^C}|W_\mathcal{L}) - H(A_\mathcal{L}|W_\mathcal{L}) + \sum_{l\in\mathcal{S}} H(A_l|W_l) - \epsilon_1]} \quad (43)$$

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

Lemma 4: (Generalized Markov Lemma) Let

$$P_{A_\mathcal{S},W_\mathcal{S},Y_\mathcal{S}}(a_\mathcal{S}, w_\mathcal{S}, y_\mathcal{S}) = P_{W_\mathcal{S},Y_\mathcal{S}}(w_\mathcal{S}, y_\mathcal{S}) \prod_{t\in\mathcal{S}} P_{A_t|W_t,Y_t}(a_t|w_t, y_t). \quad (44)$$

Given $(\boldsymbol{w}_\mathcal{S}, \boldsymbol{y}_\mathcal{S})$ randomly generated 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

$$k_t^* = \phi_t\left(\boldsymbol{y}_t, \boldsymbol{w}_t, \tilde{\boldsymbol{a}}_t^{(1)}, \ldots, \tilde{\boldsymbol{a}}_t^{(N_t)}\right)$$

taking values in $[1 \ldots N_t]$, such that for sufficiently large n

$$\Pr\left(\left(\left\{\boldsymbol{a}_t^{(k_t^*)}\right\}_{t\in\mathcal{S}}, \boldsymbol{w}_\mathcal{S}, \boldsymbol{y}_\mathcal{S}\right) \in \mathbf{T}_\epsilon\right) \geq 1 - \epsilon. \quad (45)$$

Proof: See [28] and [30] for the proofs of Lemmas 2–3. Lemma 4 is a simple extension of Lemma 3.4 in [31]. □

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

We show that when the selected encoding F is unknown, the transmission X^n is a memoryless random process. By definition, memoryless process is 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

$$\begin{aligned}
&P_{X_{t_1},\dots,X_{t_L}}(x_{t_1},\dots,x_{t_L})\\
&\quad= \sum_{f,m:(\phi_{S,f}(m))_{t_{\mathcal{L}}}=x_{t_{\mathcal{L}}}} P_{F,M}(f,m)\\
&\quad= \sum_{m=1}^{2^{nR}} \sum_{f:(\phi_{S,f}(m))_{t_{\mathcal{L}}}=x_{t_{\mathcal{L}}}} P_F(f)P_M(m)\\
&\quad= \sum_{f:(\phi_{S,f}(1))_{t_{\mathcal{L}}}=x_{t_{\mathcal{L}}}} P_F(f)\\
&\quad= \prod_{t=t_1}^{t_L} P_X(x_t) \sum_{x_{t_{\mathcal{L}^C}},\boldsymbol{x}(m=2),\dots,\boldsymbol{x}(m=2^{nR})}\\
&\prod_{j\in t_{\mathcal{L}^C}} P_X(x_j) \prod_{i=1}^{n}\prod_{l=2}^{2^{nR}} P_X(x(m=l)_i) = \prod_{t=t_1}^{t_L} P_X(x_t). \qquad (46)
\end{aligned}$$

This concludes the proof.

APPENDIX III
PROOF OF THEOREM 3

We use ideas from [32] which presents an achievable rate region using the compress-and-forward technique for multiple relays. The difference is that the agents benefit from a fixed noninterfering links to the destination, and thus the interference from simultaneously transmitting relays is avoided. In addition to compress-and-forward, broadcast messages are sent to the agents to be passed on noiselessly to the destination. As before, the network is composed of T agents $t \in \mathcal{T} = \{1,\dots,T\}$, a source transmitter, and a destination. Compared to [32], we do not need the block Markov encoding technique.

The transmission is as follows: the transmitter sends $\boldsymbol{X}(M)$ 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 destination D, respectively. Agent t 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 just decoded. Agent t uses the compression rate $\hat{R}_t$ 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 $\{\boldsymbol{U}_t\}$ depend on $\hat{M}_{\mathcal{T}}$, bandwidth from the agents to D can be saved by using a Wyner–Ziv lossy distributed source coding. Each agent then uses the remaining bandwidth after sending the broadcast message $(C_t - R_t)$ to send the Wyner–Ziv bin index $s_t \in [1, 2^{n(C_t-R_t)}]$. The 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 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 labeled 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 $P_{\boldsymbol{X}|\boldsymbol{W}_{\mathcal{T}}}(\boldsymbol{x}|\boldsymbol{w}_{\mathcal{T}}) = \prod_i P_{X|W_{\mathcal{T}}}(x_i|w_{\mathcal{T},i})$.
   - 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 drop the index f in the sequel since decoding agents know the chosen f and the achievable rate is valid with high probability for a random f.

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 the collection of $\boldsymbol{w}_{\mathcal{T}}$ such that for any $a_{\mathcal{S}}$ with $P_{W_{\mathcal{S}}}(a_{\mathcal{S}}) = 0$, $N(a_{\mathcal{S}}|\boldsymbol{w}_{\mathcal{S}}) = 0$, and such that

$$\forall\, \mathcal{S} \subseteq \mathcal{T}, \quad \forall\, a_{\mathcal{S}} \in \mathcal{W}_{\mathcal{S}},\quad \left|\frac{1}{n}N(a_{\mathcal{S}}|\boldsymbol{w}_{\mathcal{S}}) - P_{W_{\mathcal{S}}}(a_{\mathcal{S}})\right| < \frac{\epsilon}{|\mathcal{W}_{\mathcal{S}}|}. \qquad (47)$$

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

$$(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_T) \in \mathbf{T}_\epsilon^{BC}. \quad (48)$$

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}. 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 (39).

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

$$(\boldsymbol{y}_t, \hat{\boldsymbol{w}}_t) \in \mathbf{T}_\epsilon^{t,1}. \quad (49)$$

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: Agent t chooses any of the z_t such that

$$(\boldsymbol{u}_t(z_t, \hat{\boldsymbol{w}}_t), \boldsymbol{y}_t, \hat{\boldsymbol{w}}_t) \in \mathbf{T}_\epsilon^{t,2}. \quad (50)$$

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

$$R_t < C_t \quad (51)$$

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

$$\begin{cases} \left(\hat{\boldsymbol{u}}_1(\hat{z}_1, \hat{\boldsymbol{w}}_1(\hat{M}_\mathcal{T})), \ldots, \hat{\boldsymbol{u}}_T(\hat{z}_T, \hat{\boldsymbol{w}}_T(\hat{M}_\mathcal{T})), \hat{\boldsymbol{w}}_\mathcal{T}(\hat{M}_\mathcal{T})\right) \in \mathbf{T}_\epsilon^3 \\ \hat{z}_\mathcal{T} \in S_{s_1} \times \cdots \times S_{s_T} \end{cases} \quad (52)$$

where $\mathbf{T}_\epsilon^3$ is defined in the standard way, as (39). 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

$$\left(\boldsymbol{x}(\hat{M}_\mathcal{T}, \hat{M}_{CF}), \hat{\boldsymbol{u}}_\mathcal{T}(\hat{z}_\mathcal{T}, \hat{\boldsymbol{w}}_\mathcal{T}(\hat{M}_\mathcal{T})), \hat{\boldsymbol{w}}_\mathcal{T}(\hat{M}_\mathcal{T})\right) \in \mathbf{T}_\epsilon^4. \quad (53)$$

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

$$P(\text{error}) = P\left(\cup_{i=1}^{11} E_i\right) \leq \sum_{i=1}^{11} P(E_i). \quad (54)$$

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 $\{\boldsymbol{w}_t\}$ to be larger than zero, we must have

$$\forall\, 1 \leq t \leq T : R_t \leq I(W_t; Y_t) - \delta. \quad (55)$$

For any subset $\mathcal{S} \subseteq \mathcal{T}$, we have

$$\Pr\left(\boldsymbol{w}_\mathcal{T} \notin \mathbf{T}_\epsilon^{BC}\right) \leq 1 - 2^{n[H(W_\mathcal{S}) - \sum_{t\in\mathcal{S}} H(W_t) - 2|\mathcal{S}|\epsilon]} \quad (56)$$

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

$$P \leq \left(\Pr\left(W_\mathcal{T}^n \notin \mathbf{T}_\epsilon^{BC}\right)\right)^{2^{n\left[\sum_\mathcal{S} I(Y_t;W_t) - R_t - \delta\right]}}. \quad (57)$$

It is easy to see that this probability is as small as desired as long as n is sufficiently large and

$$\begin{aligned} \sum_{t\in\mathcal{S}} R_t &< \sum_{t\in\mathcal{S}} [I(Y_t; W_t) - H(W_t) - \delta] + H(W_\mathcal{S}) - 2|\mathcal{S}|\epsilon \\ &= \sum_{t\in\mathcal{S}} I(Y_t; W_t) - I(W_t; W_{\tilde{\mathcal{T}}(\mathcal{S},t)}) - |\mathcal{S}|\delta - 2|\mathcal{S}|\epsilon. \end{aligned} \quad (58)$$

Recall that $\tilde{\mathcal{T}}(\mathcal{S}, t) \triangleq \{i : i \in \mathcal{S} \text{ and } i < t\}$. Let $\epsilon' = \delta + 2\epsilon$.

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

3) E_3 and E_4: 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: 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 from Lemma 4 can be made arbitrarily small for sufficiently large n as long as

$$\hat{R}_t > I(U_t; Y_t | W_t). \quad (59)$$

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

$$\hat{z}_\mathcal{S} \neq z_\mathcal{S} \quad (60)$$

and

$$\hat{z}_{\mathcal{S}^C} = z_{\mathcal{S}^C}. \quad (61)$$

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 with high probability $\hat{\boldsymbol{u}}_\mathcal{T}(\hat{z}_\mathcal{T}, \hat{\boldsymbol{w}}_\mathcal{T}(\hat{M}_\mathcal{T})), \hat{\boldsymbol{w}}_\mathcal{T}(\hat{M}_\mathcal{T})$ belongs to a typical set with 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 a vector belongs to $\mathbf{T}_\epsilon^3$ is upper-bounded by

$$2^{n\left[H(W_\mathcal{T},U_\mathcal{T})-H(W_\mathcal{T})-H(U_{\mathcal{S}^C}|W_\mathcal{T})-\sum_{t\in\mathcal{S}}H(U_t|W_t)+\epsilon\right]}. \quad (62)$$

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

$$2^{n\left[\sum_{t\in\mathcal{S}}\left[\hat{R}_t-C_t+R_t-H(U_t|W_t)\right]\right]}2^{n\left[H(U_\mathcal{S}|W_\mathcal{T},U_{\mathcal{S}^C})+\epsilon+|\mathcal{S}^C|\epsilon'\right]}. \quad (63)$$

This means that as long as

$$\sum_{t\in\mathcal{S}}[C_t-R_t] > \sum_{t\in\mathcal{S}}\left[\hat{R}_t-H(U_t|W_t)\right]+H(U_\mathcal{S}|U_{\mathcal{S}^C},W_\mathcal{T}) \quad (64)$$

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

6) E_{10}, E_{11}: The probability that $\hat{M}_{CF}\neq M_{CF}$ satisfies (53) is upper-bounded by (again Lemma 3)

$$2^{-n[I(X;U_\mathcal{T}|W_\mathcal{T})-\epsilon]}. \quad (65)$$

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

$$R_{CF} < I(X;U_\mathcal{T}|W_\mathcal{T}). \quad (66)$$

Taking (59) and (64) and noticing that $\{U_t\}_{t\in\mathcal{T}}$ are independent given (Y_t, W_t), we can write the constraints as

$$\forall\,\mathcal{S}\in\mathcal{T}: \sum_{t\in\mathcal{S}}[C_t-R_t] > I(U_\mathcal{S};Y_\mathcal{S}|U_{\mathcal{S}^C},W_\mathcal{T}). \quad (67)$$

Notice that (55) is superfluous given (58), and (51) is superfluous given (67). Now (58) and (67) constitutes (21). The achievable rate (20) follows by (66).

F. Cardinality Bounds

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

Consider the functionals on a generic probability Q_{Y_t,W_t}, over $\mathcal{Y}_t$, $\mathcal{W}_t$. 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

$$|\mathcal{U}_t| \le |\mathcal{Y}_t||\mathcal{W}_t|+2^{T-1}. \quad (68)$$

When trying to apply the Support Lemma for the cardinalities of $W_\mathcal{T}$, the structure of the constraints in (58), and specifically the rightmost 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 in Marton's original broadcast technique [29], which to the best of our knowledge has not been done.

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) The decoding is now done in a single stage, in which the destination looks for $\hat{M}_{CF}$, $\hat{z}_\mathcal{T}$, such that (53) is fulfilled and

$$\hat{z}_\mathcal{T}\in S_{s_1}\times\cdots\times S_{s_T}. \quad (69)$$

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

2) Error analysis: E_7—E_{12} are replaced by E_7' and E_8', so we have fewer error events and the achievable rate might be larger. Note that $P(E_7')\to 0$ as $n\to\infty$ 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 $2^{n[R+\sum_{t\in\mathcal{S}}[\hat{R}_t-C_t+R_t]]}$ corresponding vectors $(\boldsymbol{x}(\hat{M}), \boldsymbol{u}_\mathcal{S}(\hat{z}_\mathcal{S}), \boldsymbol{u}_{\mathcal{S}^C}(\hat{z}_{\mathcal{S}^C}))$, and the probability that any of them is 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

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

where the second inequality is due to (59) and because of the Markov chain $U_t - (W_\mathcal{T}, X) - U_{\mathcal{T}\setminus t}$. Notice that for t^* such that $\hat{R}_{t^*}\le C_{t^*}-R_{t^*}$, which means $I(U_{t^*};Y_{t^*}|W_{t^*})\le C_{t^*}-R_{t^*}$, we get full reconstruction of $\boldsymbol{u}_{t^*}$, with no need for binning. In that case, the subsets $\mathcal{S}: t^*\in\mathcal{S}$ are not the minimum, and do not determine the achievable rate. This is since

$$\begin{aligned}&C_{t^*}-R_{t^*}-I(U_{t^*};Y_{t^*}|X,W_{t^*})+I(U_{\mathcal{S}^C};X|W_\mathcal{T})\\&\quad\ge I(U_{t^*};Y_{t^*}|W_{t^*})-I(U_{t^*};Y_{t^*}|X,W_{t^*})\\&\qquad + I(U_{\mathcal{S}^C};X|W_\mathcal{T})\\&\quad= I(U_{t^*};X|W_{t^*})+I(U_{\mathcal{S}^C};X|W_\mathcal{T})\\&\quad\ge I(U_{t^*};X|U_{\mathcal{S}^C},W_{t^*})+I(U_{\mathcal{S}^C};X|W_\mathcal{T})\\&\quad\ge I(U_{\mathcal{S}^C\cup t^*};X|W_\mathcal{T}).\end{aligned} \quad (70)$$

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

$$H(M|V_{\mathcal{T}},F) \leq n\epsilon_n \tag{71}$$

where $n\epsilon_n \to 0$ as $n \to \infty$. Now we have

$$\begin{aligned} nR &\leq H(M) = I(M;V_{\mathcal{T}},F) + H(M|V_{\mathcal{T}},F) && (72)\\ &\leq H(V_{\mathcal{T}}) + H(F|V_{\mathcal{T}}) - H(F|M) && (73)\\ &\quad - H(V_{\mathcal{T}}|M,F) + n\epsilon_n && (74)\\ &= I(V_{\mathcal{T}};M,F) - I(F;V_{\mathcal{T}}) + n\epsilon_n && (75)\\ &\leq I(V_{\mathcal{T}};M,F) + n\epsilon_n && (76)\\ &\leq I(V_{\mathcal{T}};\boldsymbol{X}(M,F)) + n\epsilon_n && (77)\end{aligned}$$

where (75) is since F is independent of M so $H(F|M) = H(F)$ and (77) stems from the Markov chain $\{MF\} - \boldsymbol{X} - V_{\mathcal{T}}$.

Turning to the agents, we bound

$$\begin{aligned} I(V_{\mathcal{T}};\boldsymbol{X}) &= H(\boldsymbol{X}) - H(\boldsymbol{X}|V_{\mathcal{T}}) && (78)\\ &\leq \sum_{i=1}^{n} H(X_i) - H\left(X_i|V_{\mathcal{T}},X^{i-1}\right) && (79)\\ &\leq \sum_{i=1}^{n} H(X_i) - H\left(X_i|V_{\mathcal{T}},Y_{\mathcal{T}}^{i-1},X^{i-1}\right) && (80)\\ &= \sum_{i=1}^{n} H(X_i) - H(X_i|U_{\mathcal{T},i}) && (81)\\ &= \sum_{i=1}^{n} I(X_i;U_{\mathcal{T},i}) && (82)\end{aligned}$$

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

We further have

$$\begin{aligned} &\sum_{i=1}^{n} I(U_{\mathcal{S},i};Y_{\mathcal{T},i}|U_{\mathcal{S}^C,i}) && (83)\\ &\quad = \sum_{i=1}^{n} I\left(V_{\mathcal{S}},Y_{\mathcal{T}}^{i-1},X^{i-1};Y_{\mathcal{T},i}|V_{\mathcal{S}^C},Y_{\mathcal{T}}^{i-1},X^{i-1}\right) && (84)\\ &\quad = \sum_{i=1}^{n} I\left(V_{\mathcal{S}};Y_{\mathcal{T},i}|V_{\mathcal{S}^C},Y_{\mathcal{T}}^{i-1},X^{i-1}\right) && (85)\\ &\quad \leq \sum_{i=1}^{n} I\left(V_{\mathcal{S}};Y_{\mathcal{T},i},X_i|V_{\mathcal{S}^C},Y_{\mathcal{T}}^{i-1},X^{i-1}\right) && (86)\\ &\quad = I(V_{\mathcal{S}};\boldsymbol{Y}_{\mathcal{T}},\boldsymbol{X}|V_{\mathcal{S}^C}) && (87)\\ &\quad = H(V_{\mathcal{S}}|V_{\mathcal{S}^C}) \leq H(V_{\mathcal{S}}) \leq n\sum_{t\in\mathcal{S}} C_t. && (88)\end{aligned}$$

Next, following [15], 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 $[1,n]$. We have

$$\begin{aligned} R &\leq \frac{1}{n}\sum_{i=1}^{n} I(X_i;U_{\mathcal{T},i}) && (89)\\ &= \left[\sum_{i=1}^{n} P(S=i)I(X_i;U_{\mathcal{T},i}|S=i)\right] + I(X^*;S) && (90)\\ &= I(X^*;U_{\mathcal{T},S}|S) + I(X^*;S) && (91)\\ &= I(X^*;U_{\mathcal{T}}^*) && (92)\end{aligned}$$

where (90) follows from Lemma 1: it is known that X^* and S are independent without the key F (X is memoryless). We further have

$$\begin{aligned} \sum_{t\in\mathcal{S}} C_t &\geq \sum_{i=1}^{n} \frac{1}{n} I(U_{\mathcal{S},i};Y_{\mathcal{T},i}|U_{\mathcal{S}^C,i}) && (93)\\ &= \sum_{i=1}^{n} \frac{1}{n} I(U_{\mathcal{S}}^*;Y_{\mathcal{T}}^*|U_{\mathcal{S}^C}^*,S=i) && (94)\\ &= I(U_{\mathcal{S}}^*;Y_{\mathcal{T}}^*|U_{\mathcal{S}^C}^*). && (95)\end{aligned}$$

Define $W^* \triangleq \left(\boldsymbol{Y}_{\mathcal{T},S+1},Y_{\mathcal{T}}^{S-1},X^{S-1},S\right)$, 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

$$\begin{aligned} U_t^* &= \left(V_t,Y_{\mathcal{T}}^{S-1},X^{S-1},S\right)\\ &= (g_t(W^*,Y_t^*),g(W^*)) && (96)\\ f_t(W^*,Y_t^*) &\triangleq (g_t(W^*,Y_t^*),g(W^*)) && (97)\end{aligned}$$

where $g_t(W^*,Y_t^*) = \left(V_t,Y_{\mathcal{T}}^{S-1}\right)$ 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 no longer satisfies (16), so that the upper bound in Theorem 2 is not applicable when the agents are cognizant of the codebook used.

A. Cardinality Bounds

In this subsection, we develop bounds on the cardinality of the auxiliary variable W through bounds on the variables $U_{\mathcal{T}}$ which fulfill the Markov chain (17). For that, we use the Support Lemma (see, for example, [30, p. 310], and [20]). 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 $\mathcal{P}(\mathcal{X})$, then there exist K elements $Q_k \in \mathcal{P}(\mathcal{X})$ and K nonnegative reals α_k that sum to unity, such that for every $1 \leq j \leq K$

$$\int_{\mathcal{P}(\mathcal{X})} q_j(Q)\mu(dQ) = \sum_{k=1}^{K} \alpha_k q_j(Q_k). \tag{98}$$

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:

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

We remark that $P_{X|Y_{\mathcal{T}}}(x|y_{\mathcal{T}})$ 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 exists a random variable U_t' such that

$$\sum_{u_t} P_{U_t'}(u_t) q_1(P(\bullet|U_t' = u_t)) = H(X) - I(X; U_{\mathcal{T}}) \quad (99)$$

$$\sum_{u_t} P_{U_t'}(u_t) q^c_{\mathcal{S}^C}(P(\bullet|U_t' = u_t)) = H(Y_{\mathcal{T}}|U_{\mathcal{S}^C}) \quad (100)$$

$$\sum_{u_t} P_{U_t'}(u_t) q_{y_{\mathcal{T}}}(P(\bullet|U_t' = u_t)) = P_{Y_{\mathcal{T}}}(y_{\mathcal{T}}) \quad (101)$$

are fulfilled and U_t' has cardinality bounded by

$$|\mathcal{U}_t| \leq |\mathcal{Y}_{\mathcal{T}}| + 2^{T-1}. \quad (102)$$

This is since (99) is one equation, (100) is 2^{T-1}, and (101) is $|\mathcal{Y}_{\mathcal{T}}| - 1$ equations. This is also the cardinality bound on W, namely

$$|\mathcal{W}| \leq |\mathcal{Y}_{\mathcal{T}}| + 2^{T-1}. \quad (103)$$

Appendix VI
Proof of the Upper Bound in Corollary 2

We have

$$\begin{aligned}
\sum_{t\in\mathcal{S}} C_t &\geq \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}; V_{\mathcal{S}}|V_{\mathcal{S}^C}) && (104)\\
&= \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}; V_{\mathcal{T}}) - \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}; V_{\mathcal{S}^C}) && (105)\\
&= \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}, \boldsymbol{X}; V_{\mathcal{T}}) - \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}, \boldsymbol{X}; V_{\mathcal{S}^C}) && (106)\\
&= \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{T}}) - \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C})\\
&\quad + \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{T}}; V_{\mathcal{T}}|\boldsymbol{X}) - \frac{1}{n} I(\boldsymbol{Y}_{\mathcal{S}^C}; V_{\mathcal{S}^C}|\boldsymbol{X}) && (107)\\
&= \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{T}}) - \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C})\\
&\quad + \sum_{t\in\mathcal{S}} \frac{1}{n} I(\boldsymbol{Y}_t; V_t|\boldsymbol{X}) && (108)
\end{aligned}$$

where (106) and (108) are since V_t is a deterministic function of Y_t^n. We use Fano's inequality (77), and find that

$$R \leq \min_{\mathcal{S}\subseteq\mathcal{T}} \left\{ \sum_{t\in\mathcal{S}} \left[C_t - \frac{1}{n} I(\boldsymbol{Y}_t; V_t|\boldsymbol{X}) \right] + \frac{1}{n} I(\boldsymbol{X}; V_{\mathcal{S}^C}) \right\}. \quad (109)$$

Next, from (82) we know that

$$I(\boldsymbol{X}; V_{\mathcal{S}^C}) \leq \sum_{i=1}^n I(X_i; U_{\mathcal{S}^C,i})$$

and

$$I(\boldsymbol{Y}_t; V_t|\boldsymbol{X}) \leq \sum_{i=1}^n I(Y_{t,i}; U_{t,i}|X_i).$$

Again using the time-sharing random variable S here, we get 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 [29]) is exemplified here for two users $T = 2$. The achievable rate in this case is

$$R \leq I(X; U_1, U_2|W_1 W_2 W_C) + R_1 + R_2 \quad (110)$$

where we get (111) at the bottom of the page, and where

$$\begin{aligned}
&P_{W_C,W_1,W_2,X,Y_1,Y_2,U_1,U_2}(w_C, w_1, w_2, x, y_1, y_2, u_1, u_2)\\
&\quad = P_{W_C,W_1,W_2}(w_C, w_1, w_2) P_{X|W_C,W_1,W_2}(x|w_C, w_1, w_2)\\
&\qquad \cdot P_{Y_1,Y_2|X}(y_1, y_2|x) P_{U_1|Y_1,W_1,W_C}(u_1|y_1, w_1, w_C)\\
&\qquad \cdot P_{U_2|Y_2,W_2,W_C}(u_2|y_2, w_2, w_C). && (112)
\end{aligned}$$

Proof Outline: The proof involves generating three vectors $\boldsymbol{w}_1$, $\boldsymbol{w}_1$, $\boldsymbol{w}_C$ i.i.d., where $\boldsymbol{w}_C$ is distributed according to $\prod_{i=1}^n P_{W_C}(w_{C,i})$ and $\boldsymbol{w}_1$ 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 tests. 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 destination. An extension to more than $T = 2$ agents can be done using similar steps.

Appendix VIII
Solving the Linear Programming Presented in Remark 3

Define the raw vector $\boldsymbol{x} = R_{\mathcal{T}}$. We have the following problem:

$$\max \boldsymbol{x} (\underbrace{1, \ldots, 1}_{T})^H \quad (113)$$

where the superscript H denotes transposition, and we have the constraints

$$\boldsymbol{A}\boldsymbol{x} \leq \begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{b}_{BC} \\ \boldsymbol{b}_q \end{pmatrix} \quad (114)$$

$$\begin{cases}
0 \leq R_1 \leq I(W_1, W_C; Y_1)\\
0 \leq R_2 \leq I(W_2, W_C; Y_2)\\
R_1 + R_2 \leq \min\{I(W_C; Y_1), I(W_C; Y_2)\} + I(W_1; Y_1|W_C) + I(W_2; Y_2|W_C) - I(W_1; W_2|W_C)\\
I(U_1; Y_1|U_2, W_1, W_2, W_C) \leq C_1 - R_1\\
I(U_2; Y_2|U_1, W_1, W_2, W_C) \leq C_2 - R_2\\
I(U_1, U_2; Y_1, Y_2|W_1, W_2, W_C) \leq C_1 + C_2 - R_1 - R_2
\end{cases} \quad (111)$$

where $\mathbf{0}$ is a column vector of T zeros, and where

$$\boldsymbol{b}_{BC} = \begin{pmatrix} I(Y_1;W_1) \\ \vdots \\ I(Y_T;W_T) \\ \sum_{t=1,2} I(Y_t;W_t) - I(W_t;W_{t-1}) \\ \vdots \\ \sum_{t=1}^{T} I(Y_t;W_t) - I\left(W_t;W^{t-1}\right) \end{pmatrix} \tag{115}$$

and

$$\boldsymbol{b}_q = \begin{pmatrix} C_1 - I(Y_1;U_1|U_{2,\dots,T},W_\mathcal{T}) \\ \vdots \\ C_T - I(Y_T;U_T|U_{1,\dots,T-1},W_\mathcal{T}) \\ C_1+C_2 - I(Y_{1,2};U_{1,2}|U_{3,\dots,T},W_\mathcal{T}) \\ \vdots \\ \sum_{t=1}^{T} C_t - I(Y_\mathcal{T};U_\mathcal{T}|W_\mathcal{T}) \end{pmatrix}. \tag{116}$$

The bound (114) includes the constraints on $R_\mathcal{T}$ in (21). The matrix $\boldsymbol{A}$ is

$$\boldsymbol{A} = \begin{pmatrix} -1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & -1 \\ 1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & 1 \\ 11 & \dots & 0 \\ & \vdots & \\ 11 & \dots & 11 \\ 1 & \dots & 0 \\ & \ddots & \\ 0 & \dots & 1 \\ 11 & \dots & 00 \\ & \vdots & \\ 11 & \dots & 11 \end{pmatrix}. \tag{117}$$

Since there is at least one feasible $\boldsymbol{x}$, namely, $R_t = 0$, for all $t \in \mathcal{T}$, the maximization (113) is equal to the solution of the following dual problem:

$$\min \boldsymbol{\lambda} \left(\mathbf{0}^H \boldsymbol{b}_{BC}^H \boldsymbol{b}_q^H\right) \tag{118}$$

such that

$$\begin{cases} \boldsymbol{\lambda A} = (1,\dots,1)^H \\ \forall\, 0 < k : (\boldsymbol{\lambda})_k \geq 0. \end{cases} \tag{119}$$

Now since

$$\begin{aligned} &\forall\, \mathcal{S}_1, \quad \mathcal{S}_2 \subseteq \mathcal{T}: \\ &\quad \sum_{t\in\mathcal{S}_1\cup\mathcal{S}_2} [I(W_t;Y_t) - I(W_t;W_{\tilde{\mathcal{T}}}(\mathcal{S}_1\cup\mathcal{S}_2,t))] \\ &\qquad \leq \sum_{t\in\mathcal{S}_1} [I(W_t;Y_t) - I(W_t;W_{\tilde{\mathcal{T}}}(\mathcal{S}_1,t))] \\ &\qquad\quad + \sum_{t\in\mathcal{S}_2} [I(W_t;Y_t) - I(W_t;W_{\tilde{\mathcal{T}}}(\mathcal{S}_2,t))] \end{aligned} \tag{120}$$

and

$$\begin{aligned} &\forall\, \mathcal{S}_1, \quad \mathcal{S}_2 \subseteq \mathcal{T},\ \mathcal{S}_1\cap\mathcal{S}_2 = \phi: \\ &\quad \sum_{t\in\mathcal{S}_1\cup\mathcal{S}_2} [C_t] - I(U_{\mathcal{S}_1\cup\mathcal{S}_2};Y_{\mathcal{S}_1\cup\mathcal{S}_2}|U_{\mathcal{S}_1^C\cap\mathcal{S}_2^C},W_\mathcal{T}) \\ &\qquad \leq \sum_{t\in\mathcal{S}_1}[C_t] - I(U_{\mathcal{S}_1};Y_{\mathcal{S}_1}|U_{\mathcal{S}_1^C},W_\mathcal{T}) \\ &\qquad\quad + \sum_{t\in\mathcal{S}_2}[C_t] - I(U_{\mathcal{S}_2};Y_{\mathcal{S}_2}|U_{\mathcal{S}_2^C},W_\mathcal{T}) \end{aligned} \tag{121}$$

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

Appendix IX
Proof of Theorem 4

The proof of Theorem 4 is very similar to the proof of Theorem 3 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 ([29, Theorem 2]). Such a generalization appears in Appendix VII for $T = 2$.

In this appendix, we prove the achievable rate for the case when the channels are known to be degraded. In this case, it is beneficial 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 [28]).

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

1) Appendix III-A, item 1): for the broadcast transmission, we now repeat the following step for $t = 1,\dots,T$:
   - 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 $\prod_{i=1}^{n} P_{W_t|W^{t-1}}\left(w_{t,i}|w_i^{t-1}\right)$.
   - Label the resulting vectors by M_t, where $M_t \in [1, 2^{nR_t}]$.
2) Appendix III-A, item 2): we now generate $2^{n[\hat{R}_t - C_t + R_t]}$ vectors $\boldsymbol{u}_t$ according to $\prod_{i=1}^{n} P_{U_t|W^t}\left(u_{t,i}|w_i^t\right)$ (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 $\boldsymbol{u}_t\left(z_t, M^t\right)$.
3) Appendix 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) Appendix III-C-I: now the agent t finds M^t such that

$$\left(\boldsymbol{w}^t(M^t), \boldsymbol{y}_t\right) \in \mathbf{T}_\epsilon^{t,4} \tag{122}$$

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

$$\left(\boldsymbol{u}_t\left(z_t, M^t\right), \boldsymbol{y}_t, \boldsymbol{w}^t\right) \in \mathbf{T}_\epsilon^{t,5} \tag{123}$$

   where $\mathbf{T}_\epsilon^{t,5}$ is defined in the usual way.

6) Appendix III-D: Considering the change of labeling, the destination here performs the same steps as the destination in Appendix III.
7) Appendix III-E: 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 due to 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 Lemmas 2 and 4 (generalized Markov Lemma, [31, Lemma 3.4]).
   - 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 if

$$\forall\, 1 \le t \le T : \sum_{i=1}^{t} R_i < I\left(W^t; Y_t\right) = \sum_{i=1}^{t} I\left(W_i; Y_i | W^{i-1}\right). \quad (124)$$

   This condition is fulfilled if

$$\forall\, 1 \le t \le T : R_t < I\left(W_t; Y_t | W^{t-1}\right). \quad (125)$$

   - 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

$$\hat{R}_t > I\left(U_t; Y_t | W^t\right). \quad (126)$$

8) Unlike Appendix III-F, which tries to bound the cardinality of the auxiliary variables, we can now use the Support Lemma ([30, p. 310]), as in [20], to limit the cardinalities of both $\mathcal{W}_t$ and $\mathcal{U}_t$. We start by rewriting the rate (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 $|\mathcal{W}_T| \le |\mathcal{X}| + 2^T$. 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 $\left(\{W_i\}_{t+1}^{T}, X\right)$. Overall, the cardinality of W_t can be limited by

$$|\mathcal{W}_t| \le |\mathcal{X}| \prod_{i=t+1}^{T} |\mathcal{W}_i| - 1 + 2^T - 1 + T - t + 2 = |\mathcal{X}|^{T-t+1} + \sum_{i=t}^{T} |\mathcal{X}|^{T-i}(2^T + i - t). \quad (127)$$

   Next, we limit the cardinality of $U_{\mathcal{T}}$ when provided with the auxiliaries $W_{\mathcal{T}}$ (which have bounded cardinalities). For this, we can repeat what was done in subsection 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

$$|\mathcal{U}_t| \le |\mathcal{Y}_t||\mathcal{W}^t| + 2^T. \quad (128)$$

Considering these differences, the constrains on $\{R_t\}$, and thus also the cardinality limits, are the main differences between Theorems 3 and 4. So by replacing (58) with (125), 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 Section VI-A.

A. Direct Part of Proof

Define the auxiliary random variables U_t as

$$U_t = Y_t + W_t \quad (129)$$

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

$$\rho_{W_t} = \frac{2^{-2I(U_t;Y_t|X)}\rho_{N_t}}{1 - 2^{-2I(U_t;Y_t|X)}}. \quad (130)$$

Equation (129) can also be written as

$$U_t = X + D_t \quad (131)$$

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

$$\rho_{D_t} \triangleq \mathrm{E}D_t^2 = \rho_{N_t} + \rho_{W_t}. \quad (132)$$

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

$$\frac{1}{\rho_{D_t}} = \frac{1 - 2^{-2r_t}}{\rho_{N_t}}. \quad (133)$$

The terms $\{r_t\}$ can take any positive value, and then $\{\rho_{D_t}\}$ 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 $U_{\mathcal{S}}$

$$I(X; U_{\mathcal{S}}) = \frac{1}{2}\log_2\left(1 + \frac{\rho_X}{\left(\sum_{t\in\mathcal{S}} \rho_{D_t}\right)^{-1}}\right) \quad (134)$$

$$= \frac{1}{2}\log_2\left(1 + \rho_X \sum_{t\in\mathcal{S}} \rho_{D_t}\right) \quad (135)$$

$$= \frac{1}{2}\log_2\left(1 + \rho_X \sum_{t\in\mathcal{S}} \frac{1 - 2^{-2r_t}}{\rho_{N_t}}\right). \quad (136)$$

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 by Oohama [16], 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 (109), rather than on the single-letter expression from Corollary 2, which is too loose for this case. We redefine

$$r_t \triangleq \frac{1}{n} I\left(Y_t^n; V_t | X^n\right) \tag{137}$$

and then use the following lemma, which is due to Oohama [16] to upper-bound $\frac{1}{n} I(V_{\mathcal{S}}; \boldsymbol{X})$.

Lemma 5:

$$\frac{1}{n} I(V_{\mathcal{S}}; \boldsymbol{X}) \leq \frac{1}{2} \log_2 \left(1 + \rho_x \sum_{t \in \mathcal{S}} \frac{1 - 2^{-2r_t}}{\rho_{N_t}}\right). \tag{138}$$

Lemma 5 together with (109) and (137) completes the proof. □

Next we give also the proof of Lemma 5, from [16], 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 $\hat{\boldsymbol{X}} = \sum_{t \in \mathcal{S}} \frac{\rho_{\hat{N}}}{\rho_{N_t}} \boldsymbol{Y}_t$. Then we have

$$\boldsymbol{X} = \hat{\boldsymbol{X}} + \hat{\boldsymbol{N}} \tag{139}$$

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 $\rho_{\hat{N}} = \left(\frac{1}{\rho_X} + \sum_{t \in \mathcal{S}} \frac{1}{\rho_{N_t}}\right)^{-1}$. We can use the entropy power inequality

$$2^{\frac{2}{n} h(\boldsymbol{x}|V_{\mathcal{S}})} \geq 2^{\frac{2}{n} h(\hat{\boldsymbol{X}}|V_{\mathcal{S}})} + 2\pi\rho_{\hat{N}}. \tag{140}$$

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 the best estimator of $\boldsymbol{X}$ out of $\boldsymbol{Y}_{\mathcal{S}}$ and we have the Markov chain: $\boldsymbol{X} - \boldsymbol{Y}_{\mathcal{S}} - V_{\mathcal{S}}$. Then we can rewrite (140) as

$$2^{-2\lambda + \frac{2}{n} h(\boldsymbol{X})} \geq 2^{\frac{2}{n}[h(\hat{\boldsymbol{X}}|\boldsymbol{X}, V_{\mathcal{S}}) + I(\boldsymbol{X}; \hat{\boldsymbol{X}})] - 2\lambda} + 2\pi\rho_{\hat{N}}. \tag{141}$$

Next, we can apply the entropy power inequality again, to lower-bound

$$\begin{aligned} 2^{\frac{2}{n} h(\hat{\boldsymbol{X}}|\boldsymbol{X}, V_{\mathcal{S}})} &\geq \sum_{t \in \mathcal{S}} 2^{\frac{2}{n} h\left(\frac{\rho_{\hat{N}}}{\rho_{N_t}} \boldsymbol{Y}_t | \boldsymbol{X}, V_{\mathcal{S}}\right)} \\ &= \sum_{t \in \mathcal{S}} \left(\frac{\rho_{\hat{N}}}{\rho_{N_t}}\right)^2 2^{\frac{2}{n} h(\boldsymbol{Y}_t|\boldsymbol{X}, V_{\mathcal{S}})} \\ &= \sum_{t \in \mathcal{S}} \left(\frac{\rho_{\hat{N}}}{\rho_{N_t}}\right)^2 2^{-2r_t + \frac{2}{n} h(\boldsymbol{Y}_t|\boldsymbol{X})} \\ &= \sum_{t \in \mathcal{S}} \left(\frac{\rho_{\hat{N}}}{\rho_{N_t}}\right)^2 \rho_{N_t} 2^{-2r_t}. \end{aligned} \tag{142}$$

So overall we have

$$2^{-2\lambda} \rho_X \geq \left(\sum_{t \in \mathcal{S}} \left(\frac{\rho_{\hat{N}}}{\rho_{N_t}}\right)^2 \rho_{N_t} 2^{-2r_t}\right) \frac{\rho_X}{\rho_{\hat{N}}} 2^{-2\lambda} + \rho_{\hat{N}} \tag{143}$$

or alternatively

$$2^{2\lambda} \leq \frac{\rho_X}{\rho_{\hat{N}}} - \rho_X \sum_{t \in \mathcal{S}} \frac{2^{-2r_t}}{\rho_{N_t}}. \tag{144}$$

□

Acknowledgment

The authors are indebted to Michael Peleg for helpful discussions.

References

[1] P. Gupta and P. R. Kumar, "Towards an information theory of large networks: An achievable rate region," *IEEE Trans. Inf. 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 *Proc. of IEEE Information Theory Workshop (ITW2004)*, San Antonio, TX, Oct. 2004, pp. 277–282.

[3] T. M. Cover and A. A. El-Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 1, pp. 572–584, Jan. 1979.

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

[5] M. Gastpar, G. Kramer, and P. Gupta, "The multiple-relay channel: Coding and antenna-clustering capacity," in *Proc IEEE Int. Symp. Information Theory (ISIT2002)*, Lausanne, Switzerland, Jun./Jul. 2002, p. 136.

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

[7] E. Erkip and M. Yuksel, "Diversity gains and clustering in wireless relaying," in *Proc IEEE Int. Symp. Information Theory (ISIT2004)*, Chicago, IL, Jun./Jul. 2004, p. 400.

[8] R. Dabora and S. D. Servetto, "Broadcast channels with cooperating receivers: A downlink for sensor reachback problem," in *Proc IEEE Int. Symp. Information Theory (ISIT2004)*, Chicago, IL, Jun./Jul. 2004, p. 176.

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

[10] S. I. Gel'fand and M. Pinsker, "Source coding with incomplete side information," *Probl. Control. Inf. Theory*, vol. 15, no. 2, pp. 45–57, 1979.

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

[12] A. El-Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 851–857, Nov. 1982.

[13] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

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

[15] 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.

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

[17] V. Prabhakaran, D. N. C. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. IEEE Int. Symp. Information Theory (ISIT2004)*, Chicago, IL, Jun./Jul. 2004, p. 119.

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

[19] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

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

[21] J. Barros and S. D. Servetto, "A note on cooperative multiterminal source coding," in *Proc. 38th Annu. Conf. Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2004.

[22] 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.

[23] A. Orlitsky and R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[24] G. Kramer and S. A. Savari, "Communicating probability distributions," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 518–525, Feb. 2007.

[25] B. E. Schein, "Distributed Coordination in Network Information Theory," Ph.D. dissertation, MIT, Cambridge, MA, 2001.

[26] S. Shamai (Shitz) and A. Lapidoth, "Fading channels: How perfect need perfect side information be?," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[27] I. G. Stiglitz, "Coding for a class of unknown channels," *IEEE Trans. Inf. Theory*, vol. IT-12, no. 2, pp. 189–195, Apr. 1966.

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

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

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

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

[32] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[33] M. R. Aref, "Information Flow in Relay Networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1980.