Submitted to the 2003 IEEE International Symposium on Information Theory.

Reachback Capacity with Non-Interfering Nodes

João Barros   Sergio D. Servetto

***Abstract*— We consider a multiaccess communication problem, in which multiple sources want to send correlated data to a far receiver, over a multiple access channel that is rendered virtually collision free by a suitable MAC protocol. This problem arises, e.g., in the context of sensor networks, where sensors are deployed on a field, and the goal is to send back to a remote location the signals picked up by all nodes. In this paper, we give an exact characterization of the joint source/channel capacity region, i.e., we give conditions on the sources and the channels under which reliable communication is possible in this context. These conditions generalize in a very meaningful way the condition that $\mathcal{H}(X) < C$ (entropy of a source less than channel capacity) for point-to-point channels, and surprisingly, give a separation theorem for this problem as well—we know of no other network information theory problem for which source and channel separation holds.**

I. INTRODUCTION

The capacity of the generic multiple access channel is usually studied under the assumption that messages sent by different users are statistically independent. While this assumption is perfectly reasonable for a large number of communications systems, it becomes too restrictive when studying the fundamental performance limits of a type of communications networks of special interest to us: wireless sensor networks. In a typical sensor scenario, a large number of sensors are deployed over a certain area of interest and collect correlated measurements, all of which are to be transmitted to a central receiver over what we call the *reachback channel*. Therefore, to determine the capacity of the reachback channel in sensor networks, we are required to consider a multiple access channel with correlated inputs. Our present analysis is based on the following assumptions:

- Each source is assigned a separate transmitter and a separate channel.
- Each transmitter uses one encoder, which only observes the outputs of the one source it has been assigned to.
- The receiver has access to all the available channels and is allowed to use the correlation between the sources to achieve the best possible reconstruction of the information sent by each of the transmitters.

These assumptions form what we deem to be a reasonable abstraction of the problem of sending the signals picked up by a large number of sensors all the way back to a common receiver: the nodes pick up correlated information, this information needs to be encoded independently by each node, sent over a multiple access channel for which an ideal MAC protocol (e.g., perfect TDMA/FDMA) ensures virtually collision-free transmission, and decoded at a common node. This model for the communication system is illustrated in Fig. 1.

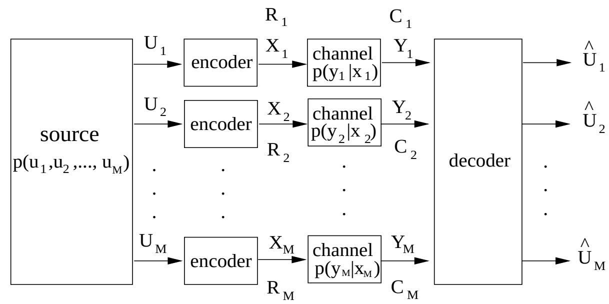

Fig. 1. System model. A set of M correlated sources are independently encoded and sent over M non-interfering channels. The outputs are processed by a joint decoder, which produces an estimate for each of the transmitted symbols.

In this scenario, our main goal is to establish the ultimate information theoretic bounds for reliable communication.

Consider the case illustrated by the block diagram of Fig. 1. The information generated by a set of correlated sources $U_1, U_2, \ldots U_M$ (with joint probability distribution $p(u_1, u_2, \ldots, u_M)$) is to be transmitted over M non-interfering channels. Assume that encoder i (operating at rate R_i) observes only the outputs of source U_i, thus separately generating the codewords X_i. At the receiving end, the decoder knows the joint distribution $p(u_1, u_2, \ldots, u_M)$ and produces the estimates ($\hat{U}_1$, $\hat{U}_2$, ...,$\hat{U}_M$) based on that knowledge and the observed channel outputs $(Y_1, Y_2, \ldots, Y_M)$.

The main question of interest in this work is: *For a given source $p(u_1, u_2, \ldots, u_M)$, and for a multiple access channel $p(y_1, y_2, \ldots, y_M | x_1, x_2, \ldots, x_M) = \prod_{i=1}^{M} p(y_i|x_i)$ (each channel $p(y_i|x_i)$ with capacity C_i), under what conditions is it possible to reproduce the values of $U_1, U_2, \ldots, U_M$ at a remote receiver with arbitrarily small probability of error?*

Clearly, if we were allowed to use a joint encoder, $H(U_1, U_2, \ldots, U_M) < \sum_{i=1}^{M} C_i$ would be a necessary and sufficient condition for reliable communication. Is this also true, when we use a separate encoder for each of the sources? Or do we need to ask for more restrictions?

Another important question is whether or not there is a separation theorem in this context. In constrast to what Shannon established for point-to-point communication [2, Ch. 8.13], in many multi-user problems separating source coding and channel coding does not yield optimal performance. In our case, one could consider compressing the sources first using Slepian-Wolf coding [4], and then adding channel coding to overcome the impairments by the channel. Is this this strategy optimal? If not, how much do we lose in comparison to an optimal strategy which uses joint source and channel coding?

J. Barros is with the Institute for Communications Engineering, Munich University of Technology, Arcisstr. 21, D-80333 München, Germany (email: joao.barros@ei.tum.de). S. D. Servetto is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 (URL: http://people.ece.cornell.edu/servetto/). Work done while the first author was a visiting scientist at Cornell. Work supported in part by a scholarship from the Fulbright Commission, by the National Science Foundation under award CCR-0227676, and by a gift from Lockheed Martin Corp.

A. Related Work

It is important to relate the previous discussion to the more general case of correlated sources over a *generic* multiple access channel, i.e., one in which simultaneous transmissions do actually interfere with each other.

The capacity region for the multiple access channel is given by the convex hull of the set of points (R_1, R_2) satisfying

$$\begin{aligned} R_1 &< I(X_1; Y|X_2), \\ R_2 &< I(X_2; Y|X_1), \\ R_1 + R_2 &< I(X_1, X_2; Y), \end{aligned} \quad (1)$$

and is generally derived under the assumption that the sources are independent with some distribution $p_1(x_1)p_2(x_2)$ [2, Ch. 14.3].

When the sources are no longer independent, computing the capacity region becomes a more complicated matter. For separate encoding of the correlated sources, the Slepian-Wolf theorem provides us with the achievable compression rates. Clearly, if the Slepian-Wolf rate region intersects the capacity region of the multiple access channel reliable communication is possible. However, it is possible to show through a very simple example [1] that this condition is sufficient, but not necessary. The theoretical result which comes closest to a capacity region for this case was published in [1], where it is shown that U and V can be sent with arbitrarily small error to Y if

$$\begin{aligned} H(U|V) &< I(X_1; Y|X_2, V), \\ H(V|U) &< I(X_2; Y|X_1, U), \\ H(U, V) &< I(X_1, X_2; Y), \end{aligned} \quad (2)$$

for some $p(u, v, x_1, x_2, y) = p(u, v) \cdot p(x_1|u) \cdot p(x_2|v) \cdot p(y|x_1, x_2)$. This region of achievable rates can be generalized to sources (U, V) with a common part $W = f(U) = g(V)$, but the authors were not able to prove a converse for the relevant case, i.e., they were not able to show that their region is indeed the capacity region of the multiple access channel with correlated sources [1]. A counterexample was later given to show that the region of achievable rates of [1] is not tight, and so this remains an interesting open problem [3].

It is reasonable to assume that the solution to the case of independent channels we present in this paper yields an outer bound for the capacity region of the general multiple access channel with correlated sources, taking into consideration that the absence of interference between the channels increases the capacity region. To the best of our knowledge however, there are no references to this apparently simpler case (i.e., independent channels).

B. Main Contributions and Organization of the Paper

In this paper we present four main contributions:

1) First, we motivate the problem of transmission of correlated information over independent channels, based on an application of particular interest to us: wireless sensor networks.

2) Then we give a formal statement for the problem of communicating correlated information over independent channels, which is a special case of the general problem of sending correlated sources over multiple access channels considered by Cover et al. [1]. And although for the general case of [1] only a region of achievable rates is known (without a converse), for our model we are able to give an *exact* characterization of the joint source/channel capacity region, converse included.

3) To reach all points on the surface of the capacity region for this problem, in some cases Slepian-Wolf codes are required. In the context of large networks however this may turn out to be impractical, since such distributed codes require knowledge of the joint distribution among all the dependent variables, and this information may be hard to obtain/estimate. Therefore, in the context of wireless sensor networks, one is also interested in finding conditions under which reliable communication is possible, under the assumption that the i-th encoder only knows the marginal distribution $p(u_i)$ of its data, *but does not know the joint distribution* $p(u_1 ... u_N)$ (this case is interesting because the marginal $p(u_i)$ can be estimated from the observed samples if not known beforehand). For this case, we are able to give a region of achievable rates that is strictly contained in the joint source/channel capacity region above, and we give an exact expression for the inefficiency resulting from lack of knowledge of the global statistics. Furthermore, we are able to determine that this region is strictly larger than the region corresponding to independent encoders and decoder—i.e., this is *not* a trivial case, the coding technique proposed does lead to an improvement over the trivial solution based on N point-to-point problems.

4) Finally, unlike the case of most other network source/channel coding problems we are aware of, for this problem we are able to prove that a natural generalization of the joint source/channel coding theorem [2, Ch. 8.13] (commonly known as the *separation* theorem) holds.

The rest of this paper is organized as follows. In Section II we state and prove a theorem that gives an exact characterization of when reliable communication is possible for a scenario in which two sources are sent over two independent channels, assuming the sources have knowledge of the joint statistics. Then, in Section III we present a similar result, now under the assumption that sources only have knowledge of their marginal distribution, but *not* of the joint distribution of all the dependent variables. The region found in Section III is strictly contained in the region found in Section II, and so we study the loss of performance that is incurred into by being forced to operate with this partial knowledge. In Section IV we generalize the previous results to the case of M sources communicating to a single decoder over M independent channels, for arbitrary $M \geq 2$. In Section V, we establish a source/channel separation result for this problem. The paper concludes with Section VI.

II. A Theorem for Two Correlated Sources and Two Independent Channels

Let us assume that two correlated sources U and V with joint probability distribution $p(u, v)$—known to all encoders—are to be transmitted over two independent channels at rates R_1 and R_2, respectively. We show that arbitrarily small probability of error is possible for some $p(u, v, x1, x2, y1, y2) = p(u, v) \cdot$

$p(x_1|u) \cdot p(x_2|v) \cdot p(y_1|x_1) \cdot p(y_2|x_2)$ if and only if

$$\begin{aligned} R_1 &< I(X_1;Y_1) + I(U;V), \\ R_2 &< I(X_2;Y_2) + I(U;V), \\ R_1 + R_2 &< I(X_1;Y_1) + I(X_2;Y_2), \end{aligned} \tag{3}$$

where X_1 and X_2 are the sent codewords and Y_1 and Y_2 are the outputs of the channels. In other words, in the absence of interference the capacity region of the multiple access channel increases by the full amount of mutual information between the two sources.

Relating this result to the conditions of the Slepian-Wolf problem, we can formulate a more general statement, given by the following theorem:

Theorem 1: A source $(\boldsymbol{U}, \boldsymbol{V}) \sim \prod_i p(u_i, v_i)$ can be sent with arbitrarily small probability of error over two independent channels $\{(\mathcal{X}_1, \mathcal{Y}_1, p(y_1|x_1)\}$ and $\{(\mathcal{X}_2, \mathcal{Y}_2, p(y_2|x_2)\}$, with allowed codes $\{\boldsymbol{x_1}(\boldsymbol{u}), \boldsymbol{x_2}(\boldsymbol{v})\}$ if and only if

$$\begin{aligned} H(U|V) &< I(X_1;Y_1), \\ H(V|U) &< I(X_2;Y_2), \\ H(U,V) &< I(X_1;Y_1) + I(X_2;Y_2). \end{aligned} \tag{4}$$

for some $p(u,v)p(x_1|u)p(x_2|v)p(y_1|x_1)p(y_2|x_2)$.

A. Definitions

Consider two information sources generated by repeated independent drawings of a pair of discrete random variables U and V from a given joint distribution $p(u,v)$. For the communications problem shown in Fig. 1, we will now provide the definitions of independent channels, source/channel block codes, probability of error and reliable transmission of the sources over the channels.

Definition 1: Two independent discrete memoryless channels consist of four alphabets $\mathcal{X}_1$, $\mathcal{X}_2$, $\mathcal{Y}_1$ and $\mathcal{Y}_2$ and a probability transition matrix $p(Y|x_1, x_2) = p(y_1|x_1)p(y_2|x_2)$, for a vector output $Y = (y_1, y_2)$.

Definition 2: A source/channel block code consists of an integer n, two encoding functions $\boldsymbol{x}_1^n : \mathcal{U}^n \to \mathcal{X}_1^n$ and $\boldsymbol{x}_2^n : \mathcal{V}^n \to \mathcal{X}_2^n$, and a decoding function $g : \mathcal{Y}_1^n \times \mathcal{Y}_2^n \to \mathcal{U}^n \times \mathcal{V}^n$.

Definition 3: The probability of error is given by

$$\begin{aligned} P_n &= p\{(\boldsymbol{U}, \boldsymbol{V}) \neq g(\boldsymbol{Y_1}, \boldsymbol{Y_2})\} \\ &= \sum_{(\boldsymbol{u},\boldsymbol{v}) \in \mathcal{U}^n \times \mathcal{V}^n} p(\boldsymbol{u}, \boldsymbol{v}) \\ &\quad \cdot P\{g(\boldsymbol{Y_1}, \boldsymbol{Y_2}) \neq (\boldsymbol{u}, \boldsymbol{v}) | (\boldsymbol{U}, \boldsymbol{V}) = (\boldsymbol{u}, \boldsymbol{v})\} \end{aligned}$$

where, for a code assignment $\{\boldsymbol{x}_1(\boldsymbol{u}), \boldsymbol{x}_2(\boldsymbol{v})\}$, the joint probability mass function is given by

$$p(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{y_1}, \boldsymbol{y_2}) = \prod_{i=1}^{n} p(u_i, v_i) p(y_{1i}|x_{1i}(\boldsymbol{u})) p(y_{2i}|x_{2i}(\boldsymbol{v})). \tag{5}$$

Definition 4: The source $(\boldsymbol{U}, \boldsymbol{V}) \sim \prod p(u_i, v_i)$ can be reliably transmitted over the two independent channels $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{Y}_1 \times \mathcal{Y}_2, p(y_1|x_1)p(y_2|x_2))$ if there exists a sequence of source/channel block codes $\{\boldsymbol{x}_1(\boldsymbol{u}), \boldsymbol{x}_2(\boldsymbol{v})\}$, with decoding function $g(\boldsymbol{y}_1, \boldsymbol{y}_2)$, such that as $n \to \infty$,

$$P_n = p\{g(\boldsymbol{Y_1}, \boldsymbol{Y_2}) \neq (\boldsymbol{U}, \boldsymbol{V})\} \to 0.$$

In the following sections we will use the notions of jointly ϵ-typical sequences and the Asymptotic Equipartion Property (AEP) as described in [2].

B. Proof of Theorem 1

To prove the theorem, we will start by the converse and show that the conditions of the theorem are necessary conditions for reliable communication to be possible. The forward part of the theorem follows easily from the region of achievable rates defined by the Slepian-Wolf theorem.

Proof: Fix n. Consider a given code of block length n. The joint distribution on $\mathcal{U}^n \times \mathcal{V}^n \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ is well defined as

$$\begin{aligned} & p(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{y_1}, \boldsymbol{y_2}) \\ &= \left(\prod_{i=1}^{n} p(u_i, v_i)\right) p(\boldsymbol{x_1}|\boldsymbol{u}) p(\boldsymbol{x_2}|\boldsymbol{v}) \left(\prod_{j=1}^{n} p(y_j|x_j)\right) \\ &\quad \left(\prod_{k=1}^{n} p(y_k|x_k)\right) \end{aligned}$$

By Fano's inequality, we can write:

$$\begin{aligned} \frac{1}{n} H(\boldsymbol{U}, \boldsymbol{V}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) &\leq P_n \frac{1}{n} (\log \|\mathcal{U}^n \times \mathcal{V}^n\|) + \frac{1}{n} \\ &= \underbrace{P_n (\log \|\mathcal{U}\| + \log \|\mathcal{V}\|) + \frac{1}{n}}_{\lambda_n} \end{aligned}$$

where $\|\mathcal{U}\|$ and $\|\mathcal{V}\|$ are the alphabet sizes of U and V, respectively. Notice that if $P_n \to 0$, λ_n must also converge to zero. Plus, since $H(\boldsymbol{U}, \boldsymbol{V}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) = H(\boldsymbol{U}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) + H(\boldsymbol{U}|\boldsymbol{V}, \boldsymbol{Y_1}, \boldsymbol{Y_2})$, we must also have $\frac{1}{n} H(\boldsymbol{U}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) \leq \lambda_n$, and then we can write the following chain of inequalities:

$$\begin{aligned} nH(U) &= H(\boldsymbol{U}) \\ &= I(\boldsymbol{U}; \boldsymbol{Y_1}, \boldsymbol{Y_2}) + H(\boldsymbol{U}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) \\ &\leq I(\boldsymbol{U}; \boldsymbol{Y_1}, \boldsymbol{Y_2}) + n\lambda_n \\ &\leq I(\boldsymbol{U}; \boldsymbol{Y_1}, \boldsymbol{V}) + n\lambda_n \\ &= I(\boldsymbol{U}; \boldsymbol{V}) + I(\boldsymbol{U}; \boldsymbol{Y_1}|\boldsymbol{V}) + n\lambda_n \\ &\leq I(\boldsymbol{U}; \boldsymbol{V}) + I(\boldsymbol{X_1}; \boldsymbol{Y_1}|\boldsymbol{V}) + n\lambda_n \end{aligned} \tag{6}$$

While the first term on the right side of this last inequality can be written as $I(\boldsymbol{U}; \boldsymbol{V}) = nI(U;V)$, the second term can be upper bounded by

$$\begin{aligned} I(\boldsymbol{X_1}; \boldsymbol{Y_1}|\boldsymbol{V}) &= H(\boldsymbol{Y_1}|\boldsymbol{V}) - H(\boldsymbol{Y_1}|\boldsymbol{X_1}, \boldsymbol{V}) \\ &= H(\boldsymbol{Y_1}|\boldsymbol{V}) - H(\boldsymbol{Y_1}|\boldsymbol{X_1}) \\ &= H(\boldsymbol{Y_1}|\boldsymbol{V}) - \sum_{i=1}^{n} H(Y_{1i}|\boldsymbol{Y^{i-1}}, \boldsymbol{X_1^n}) \\ &= H(\boldsymbol{Y_1}|\boldsymbol{V}) - \sum_{i=1}^{n} H(Y_{1i}|X_{1i}) \\ &\leq H(\boldsymbol{Y_1}) - \sum_{i=1}^{n} H(Y_{1i}|X_{1i}) \\ &= \sum_{i=1}^{n} H(Y_{1i}) - \sum_{i=1}^{n} H(Y_{1i}|X_{1i}) \\ &= \sum_{i=1}^{n} I(X_{1i}; Y_{1i}), \end{aligned} \tag{7}$$

since (a) $\boldsymbol{Y_1}$ and $\boldsymbol{V}$ are independent given X_1, (b) the channel is memoryless, and (c) conditioning reduces entropy. Thus, the inequation in (6) becomes

$$\begin{aligned} H(U) &\leq \frac{1}{n}\sum_{i=1}^{n} I(U_i;V_i) + \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_{1i}) + \lambda_n \\ &= I(U;V) + I(X_1;Y_1) + \lambda_n, \end{aligned}$$

because the U_i's and the V_i's are iid. Now, going through identical derivations, we get

$$H(V) \leq I(U;V) + I(X_2;Y_2) + \lambda_n, \tag{8}$$

and

$$H(U,V) \leq I(X_1;Y_1) + I(X_2;Y_2) + \lambda_n. \tag{9}$$

Taking the limit as $n \to \infty$, $P_n \to 0$, and we have the following converse:

$$\begin{aligned} H(U) &\leq I(U;V) + I(X_1;Y_1) \\ H(V) &\leq I(U;V) + I(X_2;Y_2) \\ H(U,V) &\leq I(X_1;Y_1) + I(X_2;Y_2) \end{aligned} \tag{10}$$

By subtracting $I(U;V)$ on both sides of the first two inequations, we arrive at the conditions in the theorem, given by

$$\begin{aligned} H(U|V) &\leq I(X_1;Y_1) \\ H(V|U) &\leq I(X_2;Y_2) \\ H(U,V) &\leq I(X_1;Y_1) + I(X_2;Y_2) \end{aligned} \tag{11}$$

thus concluding the proof of the converse.

To prove the forward part of the theorem we need to show that the rates, which are in the region defined by the conditions of the theorem are achievable, i.e., there exist codes at rates (R_1, R_2) that result in an arbitrarily small probability of error. As the Slepian-Wolf theorem guarantees the existence of such codes for

$$\begin{aligned} R_1 &\geq H(U|V) \\ R_2 &\geq H(V|U) \\ R_1 + R_2 &\geq H(U,V), \end{aligned} \tag{12}$$

no further proof is necessary for the forward part of the theorem. ∎

III. A Theorem for Two Correlated Sources Without Knowledge of the Joint Statistics

A. Main Result

Theorem 1 assumes that both encoders know the joint distribution of the sources and can thus use Slepian-Wolf codes to guarantee reliable communication. If the a priori knowledge of each encoder is limited to the marginal probability distribution of the source it observes, we get a region of achievable rates, which is given by the following theorem:

Theorem 2: A source $(\boldsymbol{U},\boldsymbol{V}) \sim \prod_i p(u_i,v_i)$ can be sent with arbitrarily small probability of error over two independent channels $\{(\mathcal{X}_1,\mathcal{Y}_1,p(y_1|x_1)\}$ and $\{(\mathcal{X}_2,\mathcal{Y}_2,p(y_2|x_2)\}$, with allowed codes $\{\boldsymbol{x_1}(\boldsymbol{u}),\boldsymbol{x_2}(\boldsymbol{v})\}$ if

$$\begin{aligned} H(U|V) &< I(X_1;Y_1|V), \\ H(V|U) &< I(X_2;Y_2|U), \\ H(U,V) &< I(X_1;Y_1) + I(X_2;Y_2), \end{aligned}$$

for some $p(u,v)p(x_1|u)p(x_2|v)p(y_1|x_1)p(y_2|x_2)$.

B. Proof of Theorem 2

Proof: To prove Theorem 2 we first describe the encoding and decoding schemes, and then proceed with the analysis of the probability of error.

Generation of Random Codes: Fix $p_1(x_1|u)$ and $p_2(x_2|v)$. For each $\boldsymbol{u} \in \mathcal{U}^n$ independently generate one $\boldsymbol{x_1}$ sequence according to $\prod_{i=1}^n p(x_{1i}|u_i)$. Index the $\boldsymbol{x_1}$ sequences by $\boldsymbol{x_1}(\boldsymbol{u})$, $\boldsymbol{u} \in \mathcal{U}^n$. Similarly, For each $\boldsymbol{v} \in \mathcal{V}^n$ independently generate one $\boldsymbol{x_2}$ sequence according to $\prod_{i=1}^n p(x_{2i}|v_i)$. Index the $\boldsymbol{x_2}$ sequences by $\boldsymbol{x_2}(\boldsymbol{v})$, $\boldsymbol{v} \in \mathcal{V}^n$.

Notice that each random code is generated according to a conditional probability on the source observed by the corresponding encoder.

Encoding: To send sequence $\boldsymbol{u}$, transmitter 1 sends the codeword $\boldsymbol{x_1}(\boldsymbol{u})$. Similarly, to send sequence $\boldsymbol{v}$, transmitter 2 sends codeword $\boldsymbol{x_2}(\boldsymbol{v})$.

Decoding: Upon observing the received sequences $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$, the decoder declares $(\hat{\boldsymbol{u}},\hat{\boldsymbol{v}})$ to be the transmitted source sequence pair if $(\hat{\boldsymbol{u}},\hat{\boldsymbol{v}})$ is the unique pair $(\boldsymbol{u},\boldsymbol{v})$ such that

$$(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x_1}(\boldsymbol{u}),\boldsymbol{x_2}(\boldsymbol{v}),\boldsymbol{y_1},\boldsymbol{y_2}) \in A_\epsilon,$$

where A_ϵ is the appropriate set of jointly ϵ-typical sequences according to the definition in [2].

Analysis of the probability of error: Suppose that $(\boldsymbol{u}_0,\boldsymbol{v}_0)$ was the source output pair, then an error occurs if
(i) $(\boldsymbol{u}_0,\boldsymbol{v}_0,\boldsymbol{x_1}(\boldsymbol{u}_0),\boldsymbol{x_2}(\boldsymbol{v}_0),\boldsymbol{y_1},\boldsymbol{y_2}) \notin A_\epsilon$, or
(ii) there exists some $(\boldsymbol{u},\boldsymbol{v}) \neq (\boldsymbol{u}_0,\boldsymbol{v}_0)$ such that $(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x_1}(\boldsymbol{u}),\boldsymbol{x_2}(\boldsymbol{v}),\boldsymbol{y_1},\boldsymbol{y_2}) \in A_\epsilon$. Define the following error events:

$$\begin{aligned} E_{e0} &= \{\exists \boldsymbol{u} \neq \boldsymbol{u}_0 : \\ &\quad (\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{X_1}(\boldsymbol{u}),\boldsymbol{X_2}(\boldsymbol{v}_0),\boldsymbol{Y_1},\boldsymbol{Y_2}) \in A_\epsilon\} \\ E_{0e} &= \{\exists \boldsymbol{v} \neq \boldsymbol{v}_0 : \\ &\quad (\boldsymbol{u}_0,\boldsymbol{v},\boldsymbol{X_1}(\boldsymbol{u}_0),\boldsymbol{X_2}(\boldsymbol{v}),\boldsymbol{Y_1},\boldsymbol{Y_2}) \in A_\epsilon\} \\ E_{ee} &= \{\exists \boldsymbol{u} \neq \boldsymbol{u}_0, \boldsymbol{v} \neq \boldsymbol{v}_0 : \\ &\quad (\boldsymbol{u},\boldsymbol{v},\boldsymbol{X_1}(\boldsymbol{u}),\boldsymbol{X_2}(\boldsymbol{v}_0),\boldsymbol{Y_1},\boldsymbol{Y_2}) \in A_\epsilon\} \end{aligned}$$

By applying the AEP and the union bound to the probability of error P_n, we get

$$\begin{aligned} P_n = \; & P\{(\boldsymbol{u}_0,\boldsymbol{v}_0,\boldsymbol{X_1}(\boldsymbol{u_0}),\boldsymbol{X_2}(\boldsymbol{v_0}),\boldsymbol{Y_1},\boldsymbol{Y_2}) \notin A_\epsilon\} \\ & + P\{\exists(\boldsymbol{u},\boldsymbol{v}) \neq (\boldsymbol{u}_0,\boldsymbol{v}_0) : \\ & (\boldsymbol{u},\boldsymbol{v},\boldsymbol{X_1}(\boldsymbol{u}),\boldsymbol{X_2}(\boldsymbol{v}),\boldsymbol{Y_1},\boldsymbol{Y_2}) \in A_\epsilon, \\ & (\boldsymbol{u}_0,\boldsymbol{v}_0) \in A_\epsilon)\} \\ \leq \; & \epsilon + \sum_{(\boldsymbol{u},\boldsymbol{v}) \in A_\epsilon} p(\boldsymbol{u},\boldsymbol{v}) \cdot \Bigg(\sum_{\substack{\boldsymbol{u} \in A_\epsilon: \\ \boldsymbol{u} \neq \boldsymbol{u}_0}} P\{E_{e0}\} \\ & + \sum_{\substack{\boldsymbol{v} \in A_\epsilon: \\ \boldsymbol{v} \neq \boldsymbol{v}_0}} P\{E_{0e}\} + \sum_{\substack{(\boldsymbol{u},\boldsymbol{v}) \in A_\epsilon: \\ \boldsymbol{u} \neq \boldsymbol{u}_0, \boldsymbol{v} \neq \boldsymbol{v}_0}} P\{E_{ee}\} \Bigg) \end{aligned}$$

where P is the conditional probability given that $(\boldsymbol{u}_0,\boldsymbol{v}_0)$ was sent. For $P(E_{e0})$ we can write

$$\begin{aligned} P(E_{e0}) &= P((\boldsymbol{u},\boldsymbol{v}_0,\boldsymbol{X_1}(\boldsymbol{u}),\boldsymbol{X_2}(\boldsymbol{v}_0),\boldsymbol{Y_1},\boldsymbol{Y_2}) \in A_\epsilon) \\ &= \sum_{\substack{(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x_1}(\boldsymbol{u}),\boldsymbol{x_2}(\boldsymbol{v}), \\ \boldsymbol{y_1},\boldsymbol{y_2}) \in A_\epsilon)}} p(\boldsymbol{u},\boldsymbol{x_1}) p(\boldsymbol{v},\boldsymbol{x_2},\boldsymbol{y_1},\boldsymbol{y_2}) \\ &\leq |A_\epsilon| 2^{-n(H(U,X_1)-\epsilon)} 2^{-n(H(V,X_2,Y_1,Y_2)-\epsilon)} \end{aligned}$$

$$\leq \quad 2^{-n(H(U,X_1)+(H(V,X_2,Y_1,Y_2)-H(U,V,X_1,X_2,Y_1,Y_2)-3\epsilon)} \tag{13}$$

In order to simplify the term in the exponent, consider that since

$$\begin{aligned} & p(u,v,x1,x2,y1,y2) \\ & \quad = \quad p(u,v)p(x_1|u)p(x_2|v)p(y_1|x_1)p(y_2|x_2), \end{aligned}$$

we can write

$$\begin{aligned} H(U,V,X_1,X_2,Y_1,Y_2) = H(U,V) + H(X_1|U) + \\ + H(X_2|V) + H(Y_1|X_1) + H(Y_2|X_2). \end{aligned} \tag{14}$$

Furthermore, we note that the first two terms of the exponent yield

$$\begin{aligned} H(U,X_1) + H(V,X_2,Y_1,Y_2) = H(U) + H(X_1|U) \\ + H(V) + H(X_2|V) + H(Y_1,Y_2|X_2,V). \end{aligned} \tag{15}$$

Carrying out the subtraction of (14) from (15), after some simplifications we get

$$\begin{aligned} & H(U) + H(V) - H(U,V) + H(Y_1,Y_2|X_2,V) \\ & -H(Y_1|X_1) - H(Y_2|X_2) \\ = \ & I(U;V) + H(Y_1|V) + H(Y_2|X_2,V) \\ & -H(Y_1|X_1,V) - H(Y_2|X_2) \\ = \ & I(U;V) + I(X_1;Y_1|V), \end{aligned}$$

since (i) Y_1 and Y_2 are conditionally independent given X_2 and V, (ii) Y_1 is conditionally independent of X_2 and V given X_1, (iii) Y_2 and V are conditionally independent given X_2, and (iv) Y_1 and $X2$ are conditionally independent given V.
Thus, we can write

$$P(E_{e0}) \leq 2^{-n(I(U;V)+I(X_1;Y_1|V)-3\epsilon)}. \tag{16}$$

Similarly,

$$P(E_{0e}) \leq 2^{-n(I(U;V)+I(X_2;Y_2|U)-3\epsilon)} \tag{17}$$

and

$$P(E_{ee}) \leq 2^{-n(I(X_1;Y_1)+I(X_2;Y_2)-4\epsilon)}. \tag{18}$$

Consequently, taking into consideration the cardinalities of the sets under the sums, the upper bound in (13) becomes

$$\begin{aligned} P_n \quad \leq \quad & \epsilon + 2^{nH(U)}2^{-n(I(U;V)+I(X_1;Y_1|X_2,V)-3\epsilon)} \\ & +2^{nH(V)}2^{-n(I(U;V)+I(X_2;Y_2|X_1,U)-3\epsilon)} \\ & +2^{nH(U,V)}2^{-n(I(X_1;Y_1)+I(X_2;Y_2)-4\epsilon)}, \end{aligned} \tag{19}$$

Since $\epsilon > 0$ is arbitrary, each term tends to zero if the sources fulfil the following conditions:

$$H(U) \quad < \quad I(U;V) + I(X_1;Y_1|V) \tag{20}$$

$$H(V) \quad < \quad I(U;V) + I(X_2;Y_2|U) \tag{21}$$

$$H(U,V) \quad < \quad I(X_1;Y_1) + I(X_2;Y_2) \tag{22}$$

Subtracting $I(U;V)$ on both sides of the first two inequations in (20) and (21), we arrive at the conditions given by the theorem. ■

If we compare the conditions of theorem 1 and theorem 2, it becomes immediately clear that there is a gap between the regions of achievables rates with and without a priori knowledge of the joint distribution. Focusing on the first of the three conditions, the extent δ_1 of this gap can be written as:

$$\delta_1 \quad = \quad I(X_1;Y_1) - I(X_1;Y_1|V) = I(Y_1;V). \tag{23}$$

Similarly, for the second of the three conditions, we get a gap $\delta_2 = I(Y_2;U)$. Since mutual information is always nonnegative, $\delta_1 \geq 0$ and $\delta_2 \geq 0$, which proves that (as expected) the region of achievable rates given by theorem 2 is contained by the region defined by the conditions in theorem 1.

Furthermore, the gap between the two regions represents the rate loss, which occurs, when the encoders have no knowledge of the joint distribution and are thus unable to use distributed codes. If we rewrite the conditions $R_1 < I(U;V) + I(X_1;Y_1|V)$ as $R_1 < I(X_1;Y_1) + \Delta_1$ and $R_2 < I(U;V) + I(X_2;Y_2|U)$ as $R_2 < I(X_2;Y_2) + \Delta_2$, we conclude that $\Delta_1 = I(U;V) - I(Y_1;V)$ and $\Delta_2 = I(U;V) - I(Y_2;V)$ represent the increase in capacity due to the correlation between the sources.

IV. A Theorem for $M \geq 2$ Correlated Sources and $M \geq 2$ Independent Channels

Having established the joint source/channel capacity region for the case two correlated sources, we now generalize this result to the transmission of M correlated sources over M independent channels, for arbitrary $M \geq 2$.

Theorem 3: A set of correlated sources $\{U_1, U_2, \ldots, U_M\}$ can be communicated reliably over independent channels $(\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_M, \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_M, p(y_1|x_1)p(y_2|x_2)\ldots p(y_M|x_M))$ if and only if

$$H(\boldsymbol{U}(S)|\boldsymbol{U}(S^c)) < \sum_{i \in S} I(\boldsymbol{X}_i;\boldsymbol{Y}_i), \tag{24}$$

for all subsets $S \subseteq \{1,2,\ldots,M\}$.

Proof: The converse can be proved exactly as in theorem 1 with $2^M - 1$ inequalities. To prove the forward part of the theorem consider the Slepian-Wolf region of achievable rates for multiple sources, given by

$$R(S) > H(U(S)|U(S^c)) \tag{25}$$

for all $S \subseteq \{1,2,\ldots,M\}$ where $R(S) = \sum_{i \in S} R_i$ and $U(S) = \{U_j : j \in S\}$. Since all the rate tuples can be achieved using separate encoders and a common decoder, all the rates satisfying the inequation in (24) are achievable, thus concluding the proof of the theorem. ■

For better understanding, let us now rewrite the condition in (24), as

$$H(\boldsymbol{U}(S)) < \sum_{i \in S} I(\boldsymbol{X}_i;\boldsymbol{Y}_i) + I(\boldsymbol{U}(S);\boldsymbol{U}(S^c)). \tag{26}$$

Clearly, the capacity region increases by the amount of mutual information between the set S and the complement S^c. The resulting capacity region is illustrated in Fig. 2, for the case of three sources.

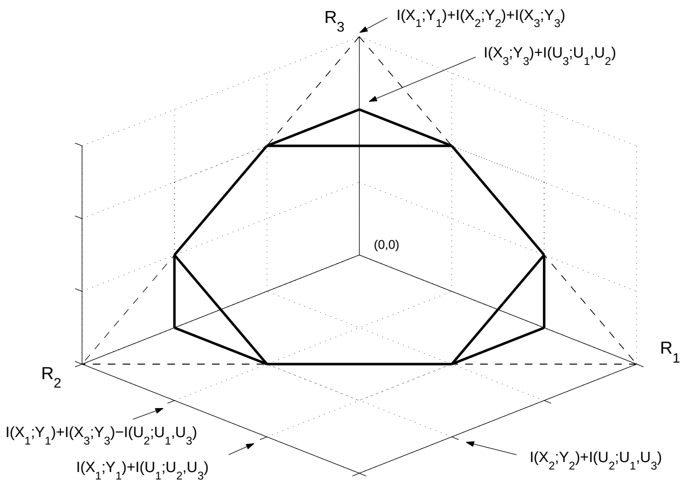

Fig. 2. The region of achievable rates for three independent channels with correlated inputs. If the sources providing the inputs to the three independent channels were independent, the capacity region would be a parallelipiped with side lengths $I(X_1;Y_1)$, $I(X_2;Y_2)$ and $I(X_3,Y_3)$. As the correlation between the sources increases, the edges of this parallelipiped increase accordingly by the amount of mutual information between the sources. At some point the parallelipiped intersects the plane defined by the equation $R_1 + R_2 + R_3 <= I(X_1;Y_1) + I(X_2;Y_2) + I(X_3;Y_3)$. Since this plane represents a necessary condition for reliable communication, the "slice" outside this boundery is not achievable, and so the resulting capacity region becomes a polyhedron.

V. On the Separation Theorem

In the context of point to point communication, given a source U (from a finite alphabet and satisfying the AEP) and a channel of capacity C, it is well known that the condition $H(U) < C$ is both necessary and sufficient for sending the source over the channel with arbitrarily small probability of error [2]. From this condition it follows that there is nothing to lose in using a two-stage encoder, which first compresses the source to its most efficient representation (at a rate close to $H(U)$) and then separately adds channel codes, which can deal with the errors caused by the channel.

A generalization of this statement to multiple sources $(U_1, U_2, \ldots, U_M)$ and multiple channels (of capacities $C_1, C_2, \ldots, C_M$) would imply that there is nothing to lose by compressing the sources to their most efficient representation (Slepian-Wolf coding) and separately adding channel codes if

$$H(\boldsymbol{U}(S)|\boldsymbol{U}(S^c)) < \sum_{i \in S} C_i, \tag{27}$$

for all $S \subseteq \{1, 2, ..., M\}$.

Comparing these inequalities to condition (24) in Theorem 3, we immediately conclude that the separation principle described in the previous paragraph also holds for the transmission of multiple sources over multiple independent channels. This is somewhat surprising, as it is not the general case for the large majority of problems in network information theory.

VI. Conclusions

We have studied the problem of reliable communication of correlated sources over multiple non-interfering channels as an abstraction of the reachback problem in large scale wireless sensor networks. For this case, we were able to give an exact characterization of the conditions under which reliable communication is possible. This is particularly interesting, because our problem is a special case of a problem studied by Cover, El Gamal and Salehi in [1], for which they were not able to find a general converse—our special case does have one such converse. Furthermore, we showed how these conditions we obtained are a natural generalization of the classical "$H < C$" (entropy less than channel capacity) conditions for point-to-point channels, thus establishing a new source/channel separation theorem for this type of networks.

From a practical point of view, our results do provide the ultimate bounds for the performance of a class of communication systems which we deem very relevant. One example is the case of sensors which encode their data only based on the marginal distribution (without requiring full knowledge of the entire joint distribution of all the dependent variables): this case is relevant because, in a large network, it may be difficult to estimate the joint statistics—however, each node is able to estimate the marginals based on data it collects. Another is the case of separate encoders but a single joint decoder: by moving complexity of the signal processing algorithms out of the sensors to the central receiver, very cheap and very simple sensors are feasible.

In terms of future work, there are two natural generalizations to the problem considered in this paper that are of particular interest to us: a rate-distortion version of this problem, and the impact of partial cooperation between the encoders. Both problems are the subject of our current research work on these topics.

References

[1] T. M. Cover, A. A. El Gamal, and M. Salehi. Multiple Access Channels with Arbitrarily Correlated Sources. *IEEE Trans. Inform. Theory*, IT-26(6):648–657, 1980.

[2] T. M. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., 1991.

[3] G. Dueck. A Note on the Multiple Access Channel with Correlated Sources. *IEEE Trans. Inform. Theory*, IT-27(2):232–235, 1981.

[4] D. Slepian and J. K. Wolf. Noiseless Coding of Correlated Information Sources. *IEEE Trans. Inform. Theory*, IT-19(4):471–480, 1973.