The Sensor Reachback Problem

João Barros     Sergio D. Servetto

Abstract

We consider the problem of reachback communication in sensor networks. In this problem, a large number of sensors are deployed on a field—the goal is for them to measure the state of some physical process unfolding over the field, and cooperate to send back this information to a far receiver for further processing. Formulating the problem as one of communicating multiple correlated sources over an array of independent channels, and with partial cooperation among encoders, we prove a set of coding theorems that give a complete characterization of the reachback capacity, i.e., of the exact conditions on the sources and the channels under which reliable communication with the far receiver is possible. An important conclusion that follows from the results presented in this paper is the fact that, for a large (and arguably most relevant) class of communication networks, separate source and channel coding does provide an optimal system architecture.

J. Barros is with the Institute for Communications Engineering, Munich University of Technology, Munich, Germany. URL: http://www.ei.tum.de/~barros/. S. D. Servetto is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. URL: http://cn.ece.cornell.edu/. Work supported by a scholarship from the Fulbright commission, and by the National Science Foundation, under awards CCR-0238271 (CAREER), CCR-0330059, and ANR-0325556. Parts of this work have been presented at the 2002 IEEE Workshop on Multimedia Signal Processing held in the US Virgin Islands (paper invited to a special session on “Signal Processing for Wireless Networks”) [1], at the 2003 IEEE International Symposium on Information Theory held in Yokohama (Japan) [2], and at the 2003 DIMACS Workshop on Network Information Theory [3].
I. INTRODUCTION

A. Reachback Communication in Wireless Sensor Networks

Wireless sensor networks made up of small, cheap, and mostly unreliable devices equipped with limited sensing, processing and transmission capabilities, have recently sparked a fair amount of interest in communications problems involving multiple correlated sources and large-scale wireless networks. It is envisioned that an important class of applications for such networks involves a dense deployment of a large number of sensors over a fixed area, in which some kind of physical process unfolds—the task of these sensors is then to collect measurements, encode them, and relay them over a noisy channel to some data collection point where this data is to be analyzed, and possibly acted upon. This scenario is illustrated in Fig. 1.

There are several aspects that make this communications problem interesting:

1) Correlated Observations: If we have a large number of nodes sensing a physical process within a confined area, it is reasonable to assume that their measurements are correlated. This correlation may be exploited for efficient encoding/decoding.

2) Cooperation among Nodes: Before transmitting data to the remote receiver, the sensor nodes may establish a conference to exchange information over the wireless medium and increase their efficiency or flexibility through cooperation.

3) Channel Interference: If multiple sensor nodes use the wireless medium at the same time (either for conferencing or reachback), their signals will necessarily interfere with each other. Consequently, reliable communication in a reachback network requires a set of rules that control (or exploit) the interference in the wireless medium.
Based on the assumption of correlated measurements, cooperating sensor nodes and a medium access scheme that eliminates the interference, we formulate the sensor reachback problem as follows. Let $U_1 U_2 \ldots U_M$ be a set of correlated sources drawn i.i.d. from the joint distribution $p(u_1 u_2 \ldots u_M)$. The information generated by the $M$ sources is separately encoded by $M$ encoders, and transmitted to a remote receiver over an array of $M$ independent channels, equivalent to a multiple access channel with $p(y|x_1 x_2 \ldots x_M) = \prod_{i=1}^{M} p(y_i|x_i)$. The encoders are interconnected by an underlying communications network, such that encoder $i$ is able to send messages to encoder $j$ reliably at rates $R_{ij} \leq C_{ij}$, with $i = 1, \ldots, M$, $j = 1, \ldots, M$ and $i \neq j$, before transmitting to the remote receiver. The solution to the problem is a complete characterization of the reachback capacity, i.e., the exact set of conditions under which it is possible to reconstruct the values of $U_1 U_2 \ldots U_M$ at the far receiver with arbitrarily small probability of error. This problem setup is illustrated in Fig. 2 for $M = 2$ sources.

![Fig. 2. A system model for the sensor reachback problem for the case of 2 sources.](image)

**B. Modeling Assumptions**

The previous problem setup forms what we deem to be a reasonable abstraction of the problem of sending the information picked up by a network of cooperating sensor nodes all the way back to a common receiver. In the spirit of George E. P. Box’s maxim that *all models are wrong but some are useful*, we provide next some motivation for the most relevant aspects of the problem formulation and briefly discuss some of the alternatives.

1) **Source Model:** At each time instant the sources generate a random vector $U_1 U_2 \ldots U_M$ drawn i.i.d. from discrete alphabets $\mathcal{U}_1 \times \mathcal{U}_2 \times \cdots \times \mathcal{U}_M$ according to $p(u_1 u_2 \ldots u_M)$. For simplicity, we assume memoryless sources, and thus consider only the spatial correlation of the observed samples and not their temporal dependence (since the latter dependencies could be dealt with by simple extensions of our results to the case of ergodic sources). Furthermore, each sensor node $i$ observes only one component...
$U_i$ and must transmit enough information to enable the remote user to reconstruct the whole vector $U_1 U_2 \ldots U_M$. This assumption is the most natural one to make for scenarios in which data is required at a remote location for fusion and further processing, but the data capture process is distributed, with sensors able to gather local measurements only, and deeply embedded in the environment.

A conceptually different approach would be to assume that all sensor nodes get to observe independently corrupted noisy versions of one and the same source of information $U$, and it is this source (and not the noisy measurements) that needs to be estimated at a remote location. This approach seems better suited for applications involving non-homogeneous sensors, where each one of the sensors gets to observe different characteristics of the same source (e.g., multispectral imaging), and therefore leads to a conceptually very different type of sensing applications from those of interest in this paper. Such an approach leads to the so called CEO problem studied by Berger, Zhang and Viswanathan in [4].

2) An Array of Independent Channels: Our motivation to consider reachback communication over an array of independent channels, instead of a general multiple access channel, is twofold.

From a pure information-theoretic perspective, an array of independent channels is interesting because, as we will see later in this paper, the assumption of independence among channels gives rise to long Markov chains which play a central role in our ability to solve this problem: it is based on those chains that we are able to prove the converse part of our coding theorems, thus obtaining complete results in terms of capacity for the sensor reachback problem. Furthermore, said coding theorems provide solutions for special cases of the multiple access channel with correlated sources, cases for which no general converse is known.

From a more practical point of view, the assumption of independent channels is valid for any network that controls interference by means of a reservation-based medium-access control protocol (e.g., TDMA). Provided individual nodes have enough resources to reach directly the far receiver, this option is perfectly reasonable for sensor networking scenarios in which sensors collect data over extended periods of time, and then an agent collects data at certain time instants (like the plane of Fig. 1). In such a scenario, all sensor nodes must transmit their accumulated measurements simultaneously, and therefore a key assumption in the design of standard random access techniques for multiaccess communication breaks down—the fact that individual nodes will transmit with low probability [5, Ch. 4]. As a result, classical random access would result in too many collisions and hence low throughput.\footnote{Recent work has considered the problem of random multiaccess communication in large-scale sensor networks—see [6]. A collection of papers on multiaccess communication was compiled by Massey for these transactions in 1985—see [7].} Alternatively, instead of
mitigating interference, a MAC protocol could attempt to exploit it, in the form of using cooperation among nodes to generate waveforms that add up constructively at the receiver (cf. [8], [9]). Providing an information theoretic analysis of such cooperation mechanisms would be very desirable, but since it entails dealing with correlated sources and a general multiple access channel, dealing with correlated sources and an array of independent channels appears a very reasonable first step working towards that goal, and also interesting in its own right, since it provides the ultimate performance limits for an important class of sensor networking problems.

3) Communication among Sensors: Before transmitting their data to the remote receiver, the sensors are allowed to exchange messages over a network. Note however that the problem statement only specifies that pairs of sensor nodes are able to communicate reliably below given rates $C_{ij}$, but it does not say anything about how this flow of information actually takes place. The latter would force us to consider classical networking topics like topology formation, routing, and flow control, which would only complicate matters for the goals of this paper. In our context, knowing that there is some network that would allow nodes to exchange information at certain rates is enough to prove our main results.

4) Perfect Reconstruction at the Receiver: In our formulation of the sensor reachback problem, the far receiver is interested in reconstructing the entire field of sensor measurements with arbitrarily small probability of error. This formulation leads us to a natural capacity problem, in the classical sense of Shannon. Alternatively, one could relax the condition of perfect reconstruction, and tolerate some distortion in the reconstruction of the field of measurements at the far receiver. Possible extensions in this direction are discussed in Section VI.

C. Background and Related Work

The sensor reachback problem is closely related to several classical problems in network information theory. To set the stage for the main contributions of this paper, we now review related previous work.

The concept of separate encoding of correlated sources was studied by Slepian and Wolf in their seminal paper [10], where they proved that two correlated sources $(U_1, U_2)$ drawn i.i.d. $\sim p(u_1 u_2)$ can be compressed at rates $(R_1, R_2)$ if and only if

$$R_1 \geq H(U_1|U_2)$$

$$R_2 \geq H(U_2|U_1)$$

$$R_1 + R_2 \geq H(U_1 U_2).$$
Assume now that \((U_1U_2)\) are to be transmitted with arbitrarily small probability of error to a joint receiver over a multiple access channel with transition probability \(p(y|x_1x_2)\). Knowing that the capacity of the multiple access channel with independent sources is given by the convex hull of the set of points \((R_1, R_2)\) satisfying [11, Ch. 14.3]

\[
R_1 < I(X_1; Y|X_2) \\
R_2 < I(X_2; Y|X_1) \\
R_1 + R_2 < I(X_1X_2; Y),
\]

it is not difficult to prove that Slepian-Wolf source coding of \((U_1U_2)\) followed by separate channel coding yields the following sufficient conditions for reliable communication

\[
H(U_1|U_2) < I(X_1; Y|X_2) \\
H(U_2|U_1) < I(X_2; Y|X_1) \\
H(U_1U_2) < I(X_1X_2; Y).
\]

These conditions, which basically state that the Slepian-Wolf region and the capacity region of the multiple access channel have a non-empty intersection, are sufficient but not necessary for reliable communication, as shown by Cover, El Gamal, and Salehi with a simple counterexample in [12]. In that same paper, the authors introduce a class of correlated joint source/channel codes, which enables them to increase the region of achievable rates to

\[
H(U_1|U_2) < I(X_1; Y|X_2U_2) \tag{1} \\
H(U_2|U_1) < I(X_2; Y|X_1U_1) \tag{2} \\
H(U_1U_2) < I(X_1X_2; Y), \tag{3}
\]

for some \(p(u_1u_2x_1x_2y) = p(u_1u_2)\cdot p(x_1|u_1)\cdot p(x_2|u_2)\cdot p(y|x_1x_2)\). Also in [12], the authors generalize this set of sufficient conditions to sources \((U_1U_2)\) with a common part \(W = f(U_1) = g(U_2)\), but they were not able to prove a converse, i.e., they were not able to show that their region is indeed the capacity region of the multiple access channel with correlated sources. Later, it was shown with a carefully constructed example by Dueck in [13] that indeed the region defined by eqns. (1)-(3) is not tight. To this date, the problem still remains open.

Enter cooperation. Assuming independent sources, Willems investigated a different scenario, in which encoders exchange messages over conference links of limited capacity prior to transmission over the
multiple access channel [14]. In this case, the capacity region is given by

\[ R_1 < I(X_1; Y | X_2 Z) + C_{12} \]
\[ R_2 < I(X_2; Y | X_1 Z) + C_{21} \]
\[ R_1 + R_2 < \min\{ I(X_1 X_2; Y | Z) + C_{12} + C_{21}, I(X_1 X_2; Y) \}, \]

for some auxiliary random variable \( Z \) such that \( |Z| \leq \min(|X_1|, |X_2| + 2, |Y| + 3) \), and for a joint distribution \( p(z x_1 x_2 y_1 y_2) = p(z) \cdot p(x_1 | z) \cdot p(x_2 | z) \cdot p(y | x_1 x_2) \).

Clearly, the sensor reachback problem is a close relative of (a) the multiple access channel with correlated sources considered by Cover et al. [12], and (b) of the multiple access channel with partially cooperating encoders solved by Willems [14]. A general problem subsuming these two problems would be a multiple access channel with correlated sources and partially cooperating encoders, which to the best of our knowledge has not been studied before. Our problem is also a special case of this more general problem, since we consider correlated sources, partially cooperating encoders, and a multiple access channel without interference.

D. Main Results

In this paper we prove a number of coding theorems, which give a complete characterization of capacity in all relevant instances of the sensor reachback problem defined in I-A.

First, we study the case of non-cooperating encoders (meaning, the case in which encoders do not exchange any information at all prior to communicating with the far receiver). We show in this case that reliable communication is possible if and only if

\[ H(U(S) | U(S^c)) < \sum_{i \in S} I(X_i; Y_i), \]

for all subsets \( S \subseteq \{1, 2, \ldots, M\} \), where \( S^c \) denotes the complement of \( S \) and \( U(S) = \{U_j : j \in S\} \). Our proof shows that, when the multiple access channel with correlated sources considered in [12] becomes an array of independent channels, distributed source coding (Slepian-Wolf), followed by separate channel coding, is an optimal coding strategy. And although for the general case of [12] only a region of achievable rates is known (without a converse), for our problem setup we are able to give a complete solution, converse included.

\(^2\)Recently, the concept of partially cooperating encoders appeared again in a framework for universal multiterminal source coding proposed by Jaggi and Effros in [15].

\(^3\)This in no way contradicts the results in [12], as explained in Section II-E.
Based on that result, we then proceed to analyze the differences between the two problems (ours and that in [12]), by noting that the crux of the proof in [12] is a class of correlated joint source/channel codes that preserve statistical dependencies of the sources in the transmitted channel codewords. In the context of sensor networks, this property is interesting, since source/channel codes that do not eliminate the source correlation completely are often simpler to implement than distributed source codes, and thus require less processing capabilities at the sensor nodes. For this case, we are able to give a region of achievable rates that is strictly contained in the reachback capacity region above, and we give an exact expression for the rate loss incurred by using correlated codes. In addition, we prove that this is not a trivial region, by showing that it is strictly larger than the one corresponding to independent encoders and decoders. This means that the coding technique proposed for the achievability result does lead to an improvement over the trivial solution based on \( M \) point-to-point problems, which in turn indicates that there is something to gain from exploiting at the decoder the fact that the data streams being uploaded are correlated.

Finally, we provide a complete characterization of reachback capacity for the general sensor reachback problem with an arbitrary number of partially cooperating encoders. In this case, reliable communication is possible if and only if

\[
H(U(S)|U(S^c)) < \sum_{i \in S} I(X_i; Y_i) + I(U(S); Z(S^c)|U(S^c)),
\]

for all subsets \( S \subseteq \{1, 2, \ldots, M\} \), where \( U(S) = \{U_j : j \in S\} \), \( Z(S) = \{Z_{ij} : i \in S \text{ or } j \in S\} \), \( I(Z_{ij}; U_i|U_j) < C_{ij} \) and \( I(Z_{ij}; U_j|U_i) < C_{ji} \). Here, \( Z_{ij} \) denotes auxiliary random variables that represent the information exchanged by the encoders using a network of discrete memoryless independent channels.

**E. A Cautionary Word about Separate Source and Channel Coding in Sensor Networks**

For all instances of the sensor reachback problem considered in this work (with and without cooperation, and for any number of nodes), we are able to prove that natural generalizations of the joint source/channel coding theorem, commonly known as the separation theorem, hold [11, Ch. 8.13]. This observation motivates us to include at the end of the paper an extra section, devoted to discussing the issue of optimality of separate source and channel coding in communication networks, which we argue is a question of “when” and not “if” it holds.

**F. Organization of the Paper**

The rest of the paper is organized as follows. In Section II we formulate the problem of transmitting two correlated sources over independent channels, without cooperation among encoders, and prove a coding
theorem that gives an exact characterization of the conditions for reliable communication to be possible. We also investigate the implications of using correlated codes, and characterize the rate loss resulting in that case. Then, in Section III, we give another characterization of reachback capacity, this time still for two nodes, but now under the assumption that the encoders are allowed to exchange information over network links of limited capacity. Based on this result, we discuss the impact of cooperation on the reachback capacity and again address the issue of encoding constraints. The results obtained for $M = 2$ sources are generalized for arbitrary number of sources $M$ in Section IV, where we also give a Gaussian example and discuss the reachback capacity for a large class of sensor networks with different topologies. In Section V, we revisit the issue of source and channel separation in communication networks, under the light of the results presented in earlier sections. The paper concludes with Section VI.

II. REACHBACK CAPACITY WITH TWO NON-COOPERATING NODES

We begin our study of the sensor reachback problem by providing a solution for the case of $M = 2$ non-cooperating nodes ($C_{12} = C_{21} = 0$). This simple problem setup provides valuable insights into the structure of the problem, allowing us to gain a firm understanding of the main issues involved in reachback communication over independent channels, before discussing the more subtle aspects of cooperation between encoders.

A. Definitions and Problem Statement

Consider two information sources generated by repeated independent drawings of a pair of discrete random variables $U_1$ and $U_2$ from a given joint distribution $p(u_1,u_2)$. We start by providing definitions of independent channels, source/channel block codes, probability of error and reliable communication.

Definition 1: A reachback channel consists of two independent discrete memoryless channels $(X_1, p(y_1|x_1), Y_1)$ and $(X_2, p(y_2|x_2), Y_2)$, with input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, output alphabets $\mathcal{Y}_1$ and $\mathcal{Y}_2$, and transition probability matrices $p(y_1|x_1)$ and $p(y_2|x_2)$.

Note that according to this definition, the reachback channel can be viewed as a multiple access channel with transition probability $p(y|x_1,x_2) = p(y_1|x_1)p(y_2|x_2)$, and $Y = (Y_1 Y_2)$. In other words, without interference, the multiple access channel becomes an array of independent channels.

Definition 2: A source/channel block code consists of an integer $N$, two encoding functions $f_1 : \mathcal{U}_1^N \rightarrow \mathcal{X}_1^N$ and $f_2 : \mathcal{U}_2^N \rightarrow \mathcal{X}_2^N$, and a decoding function $g : \mathcal{Y}_1^N \times \mathcal{Y}_2^N \rightarrow \mathcal{U}_1^N \times \mathcal{U}_2^N$. 

11/17/2003 DRAFT
Definition 3: The probability of error is given by

\[ P_N = p\{(U_1^N U_2^N) \neq g(Y_1^N Y_2^N)\} \]

\[ = \sum_{(u_1^N u_2^N) \in \{U_1^N \times U_2^N\}} p(u_1^N u_2^N) \cdot P\{g(Y_1^N Y_2^N) \neq (u_1^N u_2^N)|(U_1^N U_2^N) = (u_1^N u_2^N)\}, \]

where, for a code assignment \( x_1^N = f_1(u_1^N) \) and \( x_2^N = f(u_2^N) \), the joint probability mass function is given by

\[ p(u_1^N u_2^N y_1^N y_2^N) = \prod_{i=1}^{N} p(u_{1i} u_{2i}) p(y_{1i}|x_{1i}(u_{1i})) p(y_{2i}|x_{2i}(u_{2i})). \]

Definition 4: Reliable communication of the source \((U_1 U_2) \sim p(u_1 u_2)\) over independent channels \((X_1, p(y_1|x_1), Y_1)\) and \((X_2, p(y_2|x_2), Y_2)\) is possible if there exists a sequence of source/channel block codes \(\{x_1^N(u_1^N), x_2^N(u_2^N)\}\), with decoding function \(g(y_1^N, y_2^N)\) such that, as \(N \to \infty\),

\[ P_N = p\{g(Y_1^N Y_2^N) \neq (U_1^N U_2^N)\} \to 0. \]

In the following subsections we make use of the standard notions of jointly \(\epsilon\)-typical sequences and the Asymptotic Equipartion Property (AEP), as defined in [11].

The main goal in this section is to characterize reachback capacity, by giving single-letter information-theoretic conditions for reliable communication.

B. Main Result

The following theorem gives necessary and sufficient conditions for reliable communication under this scenario.

Theorem 1: A source \((U_1 U_2)\) drawn i.i.d. \(\sim p(u_1 u_2)\) can be communicated reliably over two independent channels \((X_1, p(y_1|x_1), Y_1)\) and \((X_2, p(y_2|x_2), Y_2)\), if and only if

\[ H(U_1|U_2) < I(X_1; Y_1) \]

\[ H(U_2|U_1) < I(X_2; Y_2) \]

\[ H(U_1 U_2) < I(X_1; Y_1) + I(X_2; Y_2). \]

C. Proof of Theorem 1

The proof begins with the converse and shows that the conditions of the theorem are necessary conditions for reliable communication to be possible. The forward part of the theorem then follows easily from the region of achievable rates defined by the Slepian-Wolf theorem.
1) Converse Proof: Consider a given code of block length $N$. The joint distribution on $U_1^N \times U_2^N \times X_1^N \times X_2^N \times Y_1^N \times Y_2^N$ is well defined as

$$
p(u_1^N u_2^N x_1^N x_2^N y_1^N y_2^N) = \left( \prod_{i=1}^{N} p(u_i;u_{2i}) \right) p(x_1^N | u_1^N) p(x_2^N | u_2^N) \left( \prod_{j=1}^{N} p(y_1j|x_{1j}) \right) \left( \prod_{k=1}^{N} p(y_2k|x_{2k}) \right)
$$

By Fano’s inequality, we can write:

$$
\frac{1}{N} H(U_1^N U_2^N | Y_1^N Y_2^N) \leq P_N \frac{1}{N} (\log|U_1^N \times U_2^N|) + \frac{1}{N}
$$

$$
= P_N (\log|U_1| + \log|U_2|) + \frac{1}{N} \lambda_N \tag{4}
$$

where $|U_1|$ and $|U_2|$ are the alphabet sizes of $U_1$ and $U_2$, respectively. Notice that if $P_N \to 0$, $\lambda_N$ must also go to zero. Plus, since $H(U_1^N U_2^N | Y_1^N Y_2^N) = H(U_1^N | Y_1^N Y_2^N) + H(U_1^N | U_2^N Y_1^N Y_2^N)$, we must also have $\frac{1}{N} H(U_1^N | Y_1^N Y_2^N) \leq \lambda_N$, and so we can write the following chain of inequalities:

$$
N H(U_1) = H(U_1^N)
$$

$$
= I(U_1^N; Y_1^N Y_2^N) + H(U_1^N | Y_1^N Y_2^N)
$$

$$
\stackrel{(a)}{\leq} I(U_1^N; Y_1^N Y_2^N) + N \lambda_N
$$

$$
\stackrel{(b)}{\leq} I(U_1^N; Y_1^N U_2^N) + N \lambda_N
$$

$$
= I(U_1^N; U_2^N) + I(U_1^N; Y_1^N | U_2^N) + N \lambda_N
$$

$$
\leq I(U_1^N; U_2^N) + I(X_1^N; Y_1^N | U_2^N) + N \lambda_N,
$$

where (a) follows from (4), and (b) and (c) follow from the data processing inequality for the long Markov chain of the form $Y_1^N - X_1^N - U_1^N - U_2^N - X_2^N - Y_2^N$. While the first term can be written as $I(U_1^N; U_2^N) = \sum_{i=1}^{N} I(U_1i; U_2i)$, the second term can be upper bounded by

$$
I(X_1^N; Y_1^N | U_2^N) = H(Y_1^N | U_2^N) - H(Y_1^N | X_1^N U_2^N)
$$

$$
\stackrel{(a)}{=} H(Y_1^N | U_2^N) - H(Y_1^N | X_1^N)
$$

$$
= H(Y_1^N | U_2^N) - \sum_{i=1}^{N} H(Y_{1i} | Y_{1i}^{i-1} X_1^N)
$$

$$
\stackrel{(b)}{=} H(Y_1^N | U_2^N) - \sum_{i=1}^{N} H(Y_{1i} | X_{1i})
$$

$$
\leq H(Y_1^N) - \sum_{i=1}^{N} H(Y_{1i} | X_{1i})
$$
\[
\begin{align*}
&= \sum_{i=1}^{N} H(Y_{1i}) - \sum_{i=1}^{N} H(Y_{1i}|X_{1i}) \\
&= \sum_{i=1}^{N} I(X_{1i}; Y_{1i}),
\end{align*}
\]

since (a) \(Y_1^N\) and \(U_2^N\) are independent given \(X_1^N\), (b) the channel is memoryless, and (c) conditioning reduces entropy. Thus,
\[
H(U_1) \leq \frac{1}{N} \sum_{i=1}^{N} I(U_{1i}; U_{2i}) + \frac{1}{N} \sum_{i=1}^{N} I(X_{1i}; Y_{1i}) + \lambda_N
\]
\[
= I(U_1; U_2) + I(X_1; Y_1) + \lambda_N,
\]
because the \(U_{1i}\)'s and the \(U_{2i}\)'s are i.i.d. Now, going through identical derivations, we get
\[
H(U_2) \leq I(U_1; U_2) + I(X_2; Y_2) + \lambda_N,
\]
and
\[
H(U_1U_2) \leq I(X_1; Y_1) + I(X_2; Y_2) + \lambda_N.
\]
Taking the limit as \(N \to \infty\), \(P_N \to 0\), we have that
\[
\begin{align*}
H(U_1) &\leq I(U_1; U_2) + I(X_1; Y_1) \\
H(U_2) &\leq I(U_1; U_2) + I(X_2; Y_2) \\
H(U_1U_2) &\leq I(X_1; Y_1) + I(X_2; Y_2).
\end{align*}
\]
By subtracting \(I(U_1; U_2)\) on both sides of the first two inequalities, we arrive at the conditions in the theorem, given by
\[
\begin{align*}
H(U_1|U_2) &\leq I(X_1; Y_1) \\
H(U_2|U_1) &\leq I(X_2; Y_2) \\
H(U_1U_2) &\leq I(X_1; Y_1) + I(X_2; Y_2),
\end{align*}
\]
thus concluding the proof of the converse.

2) Achievability Proof: We deal next with the achievable part of Theorem 1. Consider the following coding strategy. First, encoders 1 and 2 compress the input source blocks \(U_1^N\) and \(U_2^N\) separately using Slepian-Wolf codes. Then, the encoders add classical channel coding to transmit the compressed versions
over the two independent channels. Since the Slepian-Wolf theorem guarantees that the rates

\[ R_1 \geq H(U_1|U_2) \]
\[ R_2 \geq H(U_2|U_1) \]
\[ R_1 + R_2 \geq H(U_1U_2) \]

are achievable, and the channel coding theorem guarantees that the probability of error goes to zero for all rates \( R_1 < I(X_1; Y_1) \) and \( R_2 < I(X_2; Y_2) \), reliable communication is possible using this coding strategy if the conditions of the theorem are fulfilled.

D. A Simple Visualization of the Problem

To illustrate the issues involved in this problem consider the rate regions shown in Fig. 3.

![Diagram](image)

Fig. 3. Relationship between the Slepian-Wolf region and the capacity region for two independent channels. In the left figure, as \( H(U_1|U_2) < C_1 \) and \( H(U_2|U_1) < C_2 \) the two regions intersect and therefore reliable communication is possible. The figure on the right shows the case in which \( H(U_2|U_1) > C_2 \) and there is no intersection between the two regions.

When the multiple access channel is reduced to two independent channels with capacities \( C_1 \) and \( C_2 \), its capacity region becomes a rectangle with side lengths \( C_1 \) and \( C_2 \) [11, Ch. 14.3]. Also shown is the Slepian-Wolf region of achievable rates for separate encoding of correlated sources, whose limits are defined by \( H(U_1) \), \( H(U_2) \) and \( H(U_1U_2) \). Clearly, \( H(U_1U_2) < C_1 + C_2 \) is a necessary condition for reliable communication as a consequence of Shannon’s joint source and channel coding theorem for point-to-point communication. Assuming that this is the case, consider now the following possibilities:
• $H(U_1) < C_1$ and $H(U_2) < C_2$. The Slepian-Wolf region and the capacity region intersect, so any point $(R_1, R_2)$ in this intersection makes reliable communication possible. Alternatively, we can argue that reliable transmission of $U_1$ and $U_2$ is possible even with independent decoders, therefore a joint decoder will also achieve an error-free reconstruction of the source.

• $H(U_1) > C_1$ and $H(U_2) > C_2$. Since $H(U_1U_2) < C_1 + C_2$ there is always at least one point of intersection between the Slepian-Wolf region and the capacity region, so reliable communication is possible.

• $H(U_1) < C_1$ and $H(U_2) > C_2$ (or vice versa). If $H(U_2|U_1) < C_2$ (or if $H(U_1|U_2) < C_1$) then the two regions will intersect. On the other hand, if $H(U_2|U_1) > C_2$ (or if $H(U_1|U_2) > C_1$), then there are no intersection points, but it is not immediately clear whether reliable communication is possible or not (see Fig. 3), since examples are known in which the intersection between the capacity region of the multiple access channel and the Slepian-Wolf region of the correlated sources is empty and still reliable communication is possible [12].

Theorem 1 gives a definite answer to this last question: in the special case of correlated sources and independent channels an intersection between the capacity region and the Slepian-Wolf rate regions is not only sufficient, but also a necessary condition for reliable communication to be possible. From this observation we conclude that, in the case of independent channels, a two-stage encoder that uses Slepian-Wolf codes to compress the sources to their most efficient representation and then separately adds capacity attaining channel codes, indeed forms an optimal coding strategy—that is, for this reachback network, separation holds.

E. Rate Loss with Correlated Codes

The key ingredient of the achievability proof presented by Cover, El Gamal and Salehi for the multiple access channel with correlated sources is the generation of random codes, whose codewords $X_i^N$ are statistically dependent on the source sequences $U_i^N$ [12]. This property, which is achieved by drawing the codewords according to $\prod_{j=1}^{N} p(x_{ij}|u_{ij})$, implies that $U_i^N$ and $X_i^N$ are jointly typical with high probability. Since the source sequences $U_1^N$ and $U_2^N$ are correlated, the codewords $X_1^N(U_1^N)$ and $X_2^N(U_2^N)$ are also correlated, and so we speak of correlated codes. This class of random codes, which is treated in more general terms in [16], are joint source and channel codes that preserve the given correlation structure of the source sequences, which can then be exploited in the decoder to lower the probability of error.

Since correlated codes yield the best known characterization of achievable rates for the problem of
transmitting correlated sources over the multiple access channel, it is only natural that we ask how this class of codes performs in the sensor reachback problem, for which we know that separate source and channel coding is optimal. This issue is also interesting from a practical point of view, since sensor nodes with limited processing capabilities may be forced to use very simple codes that do not eliminate correlations between measurements prior to transmission [17]. In this case, we are interested in knowing how far the remaining correlation in the codewords can still be used by the receiver to improve the decoding result. The following result gives a characterization of the reachback capacity under this scenario:

**Theorem 2:** A source \( (U_1^N U_2^N) \sim \prod_{i=1}^N p(u_{1i} u_{2i}) \) can be sent with arbitrarily small probability of error over two independent channels \( \{X_1, p(y_1|x_1), Y_1\} \) and \( \{X_2, p(y_2|x_2), Y_2\} \), with correlated codes \( \{X_1^N (U_1^N), X_2^N (U_2^N)\} \) if

\[
\begin{align*}
H(U_1|U_2) &< I(X_1; Y_1|U_2) \\
H(U_2|U_1) &< I(X_2; Y_2|U_1) \\
H(U_1U_2) &< I(X_1X_2; Y_1Y_2),
\end{align*}
\]

for some \( p(u_1 u_2) \cdot p(x_1|u_1) \cdot p(x_2|u_2) \cdot p(y_1|x_1) \cdot p(y_2|x_2) \).

The proof is based on the joint source-channel codes of [12]. We repeat the description of that code construction here to highlight the property that is most relevant to us: the codewords are generated in statistical dependence to the source sequences, and are therefore correlated.

**Proof:** Fix \( p_1(x_1|u_1) \) and \( p_2(x_2|u_2) \). For each \( u_1^N \in U_1^N \), independently generate one \( x_1^N \) sequence according to \( \prod_{i=1}^N p(x_{1i}|u_{1i}) \). Index the \( x_1^N \) sequences by \( x_1^N (u_1^N), u_1^N \in U_1^N \). Similarly, for each \( u_2^N \in U_2^N \), independently generate one \( x_2^N \) sequence according to \( \prod_{i=1}^N p(x_{2i}|u_{2i}) \). Index the \( x_2^N \) sequences by \( x_2^N (u_2^N), u_2^N \in U_2^N \). Notice that each random code is generated according to a conditional probability on the source observed by the corresponding encoder.

**Encoding:** To send sequence \( u_1^N \), transmitter 1 sends the codeword \( x_1^N (u_1^N) \). Similarly, to send sequence \( u_2^N \), transmitter 2 sends codeword \( x_2^N (u_2^N) \).

**Decoding:** Upon observing the received sequences \( y_1^N \) and \( y_2^N \), the decoder declares \( (\hat{u}_1^N \hat{u}_2^N) \) to be the transmitted source sequence pair if \( (\hat{u}_1^N \hat{u}_2^N) \) is the unique pair \( (u_1^N u_2^N) \) such that

\[
(u_1^N, u_2^N, x_1^N (u_1^N), x_2^N (u_2^N), y_1^N, y_2^N) \in A_e,
\]

where \( A_e \) is the appropriate set of jointly \( \epsilon \)-typical sequences according to the definition in [11]. In [12], the code construction described above is used to characterize the conditions for reliable communication of two correlated sources drawn i.i.d. \( \sim p(u_1 u_2) \) over a multiple access channel \( p(y|x_1 x_2) \). The conditions
obtained in that more general setup are given by (1)-(3). Thus, it suffices to specialize this result to the case of independent channels. Let \( y = (y_1y_2) \) and let \( p(y|x_1x_2) = p(y_1y_2|x_1x_2) = p(y_1|x_1)p(y_2|x_2) \). Based on (1) we can write

\[
H(U_1|U_2) \leq I(X_1;Y|U_2X_2)
\]

\[
= I(X_1;Y_1Y_2|U_2X_2)
\]

\[
= H(Y_1Y_2|U_2X_2) - H(Y_1|U_2X_2X_1)
\]

\[
= H(Y_1|U_2X_2) + H(Y_2|U_2X_2Y_1) - H(Y_1|U_2X_2X_1) - H(Y_2|U_2X_2X_1Y_1)
\]

\[
= (a) H(Y_1|U_2) + H(Y_2|X_2) - H(Y_1|U_2X_1) - H(Y_2|X_2)
\]

\[
= H(Y_1|U_2) - H(Y_1|U_2X_1)
\]

\[
= I(X_1;Y_1|U_2),
\]

where (a) follows from the long Markov chain condition \( Y_1 - X_1 - U_1 - U_2 - X_2 - Y_2 \), which implies that \( Y_1 \) is independent of \( X_2 \) given \( U_2 \), \( Y_2 \) is independent of \( U_2 \), \( X_1 \) and \( Y_1 \) given \( X_2 \), and \( Y_1 \) is independent of \( X_2 \) given \( X_1 \). Similarly, for conditions (2) and (3) we get \( H(U_2|U_1) < I(X_2;Y_2|U_1) \) and

\[
H(U_1U_2) < I(X_1X_2;Y_1Y_2)
\]

thus concluding the proof.

It is interesting to observe that in this instance of the problem, we not only have a long Markov chain \( Y^N_1 - X^N_1 - U^N_1 - U^N_2 - X^N_2 - Y^N_2 \) at the block level (due to the functional dependencies introduced by the encoders and the channels), but we also have a single-letter Markov chain \( Y_1 - X_1 - U_1 - U_2 - X_2 - Y_2 \) that comes from using correlated codes and fixing the conditional probability distributions \( p(x_1|u_1) \) and \( p(x_2|u_2) \).

More importantly, the proof shows that the result of Cover, El Gamal and Salehi for the multiple access channel with correlated sources in [12] does not immediately specialize to Theorem 1, when we assume a multiple access channel with conditional probability distribution \( p(y|x_1x_2) = p(y_1y_2|x_1x_2) = p(y_1|x_1)p(y_2|x_2) \). Taking a closer at the first condition in the theorem, we observe that

\[
I(X_1;Y_1|U_2) = H(Y_1|U_2) - H(Y_1|X_1U_2)
\]

\[
= H(Y_1|U_2) - H(Y_1|X_1)
\]

\[
\leq H(Y_1) - H(Y_1|X_1)
\]

\[
= I(X_1;Y_1),
\]

with equality for \( p(u_1u_2x_1x_2) = p(u_1u_2)p(x_1)p(x_2) \). This means that we must choose independent
The crucial difference between the two problems is the presence (or absence) of interference in the channel. Albeit somewhat informally, we can state that correlated codes are advantageous when the transmitted codewords are combined in the channel through interference, which is clearly not the case in our formulation of the sensor reachback problem.

We now give a characterization of the rate loss incurred into by using correlated codes in our problem setup. Comparing the conditions of Theorems 1 and 2, we can describe the gap between the two regions of achievable rates. Focusing on the first of the three conditions, the extent $\delta_1$ of this gap can be written as:

$$
\delta_1 = I(X_1; Y_1) - I(X_1; Y_1|U_2)
= I(X_1; Y_1) - (H(Y_1|U_2) - H(Y_1|X_1, U_2))
= I(X_1; Y_1) - (H(Y_1|U_2) - H(Y_1|X_1))
= I(X_1; Y_1) - (H(Y) - H(Y) + H(Y_1|U_2) - H(Y_1|X_1))
= I(X_1; Y_1) - I(X_1; Y_1) + I(Y_1; U_2)
= I(Y_1; U_2).
$$

Similarly, for the second of the three conditions, we get a gap $\delta_2 = I(Y_2; U_1)$ and for the sum rate condition

$$
\delta_0 = I(X_1; Y_1) + I(X_2; Y_2) - I(X_1X_2; Y_1Y_2)
= I(X_1; Y_1) + I(X_2; Y_2) - (I(X_1; Y_1Y_2) + I(X_2; Y_1|X_1))
= I(X_1; Y_1) + I(X_2; Y_2) - (I(X_1; Y_1) + I(X_1; Y_2|X_1) + I(X_2; Y_2|X_1) + I(X_2; Y_2Y_1|X_1))
= I(X_2; Y_2) - (I(X_1; Y_2Y_1) + I(X_2; Y_2|X_1))
= I(X_2; Y_2) - (H(Y_2|Y_1) - H(Y_2|Y_1X_1) + H(Y_2|X_1) - H(Y_2|X_1X_2))
= I(X_2; Y_2) - (H(Y_2|Y_1) - H(Y_2|X_1X_2))
$$
\[ I(X_2; Y_2) - (H(Y_2) - H(Y_2) + H(Y_2|Y_1) - H(Y_2|X_2)) \]
\[ = I(X_2; Y_2) - (I(X_2; Y_2) - I(Y_1; Y_2)) \]
\[ = I(Y_1; Y_2) \]

Since \( \delta_i \geq 0, i \in \{0, 1, 2\} \) (mutual information is always nonnegative), we conclude that the region of achievable rates given by Theorem 2 is contained in the region defined by Theorem 1. Furthermore, we find that the rate loss terms have a simple, intuitive interpretation: \( \delta_0 \) is the loss in sum rate due to the dependencies between the outputs of different channels, and \( \delta_1 \) (or \( \delta_2 \)) represent the rate loss due to the dependencies between the outputs of channel 1 (or 2) and the source transmitted over channel 2 (or 1).

All these terms become zero if, instead of using correlated codes, we fix \( p(x_1)p(x_2) \) and remove the correlation between the source blocks before transmission over the channels.

Note also that the region defined by Theorem 2 is not trivial, in that it contains more points than those that can be achieved rates by two pairs of independent encoders/decoders. From (5) and (8) it follows that

\[ H(U_1) < I(U_1; U_2) + I(X_1; Y_1|U_2) \]
\[ = I(U_1; U_2) + I(X_1; Y_1) - I(Y_1; U_2), \]

and similarly, using (6) we get

\[ H(U_2) < I(U_1; U_2) + I(X_2; Y_2|U_1) \]
\[ = I(U_1; U_2) + I(X_2; Y_2) - I(Y_2; U_1). \]

Applying the data processing inequality based on the long Markov chain \( Y_1 - X_1 - U_1 - U_2 - X_2 - Y_2 \), we observe that \( I(U_1; U_2) - I(Y_1; U_2) \geq 0 \), and similarly, \( I(U_1; U_2) - I(Y_2; U_1) \geq 0 \), and thus conclude that the region of Theorem 2 is in general larger than the region obtained by solving two independent point-to-point problems. The difference between these two regions is what we can expect to gain from exploiting the correlation between codewords in the decoding step.

III. Reachback Capacity with \( M = 2 \) Partially Cooperating Nodes

We now extend the results in Section II to allow for partial cooperation between the encoders. Once again we start with the some definitions and a formal statement of the problem. The latter differs from the previous problem statement in the definition of the encoders, and in their ability to establish a conference prior to transmission over the reachback channel.
We start with some discussion about the conferencing mechanism, inspired by [14]. Assume that encoder 1 can send messages to encoder 2 over a channel with capacity $C_{12}$, and vice versa (encoder 2 to encoder 1 over a channel with capacity $C_{21}$). These messages could represent, for example, synchronization information: “In transmission 12 I will send $X_1 = 3$”, “I will transmit zeros in transmissions 22, 24 and 26”, etc. They could also represent quantized versions of the observed source values: “My sample is positive”, “I will send an index between 128 and 132”, etc. The simplest form of conference can be characterized as two simultaneous monologues: encoder 1 sends a block of messages to encoder 2, and encoder 2 sends a block of messages to encoder 1. In [14], Willems presents a more general definition of a conference, which is closer to a dialogue. Let $V_{ik}$ be the message sent by encoder $i$ at the $k$th transmission; encoder 1 sends the first message $V_{11}$, then encoder 2 sends its first message $V_{21}$, after which encoder 1 sends another message $V_{12}$, then encoder 2 sends $V_{22}$, and so on. This type of conference is more general not only because it admits multiple messages to be exchanged between the encoders, but, more interestingly, because it allows the next message $V_{1k}$ (or $V_{2k}$) to be sent by encoder 1 (or encoder 2) to depend on all previously received messages $V_{1}^{k-1}$ (or $V_{2}^{k-1}$). It turns out that both in the capacity problem considered by Willems in [14] and in the sensor reachback problem, two simultaneous monologues are sufficient to achieve all points in the capacity region.

A. Definitions and Problem Statement

A reachback network consists of two sender nodes and one receiver node. Sender 1 is connected to the receiver via a discrete memoryless channel $(X_1, p(y_1|x_1), Y_1)$ and sender 2 via $(X_2, p(y_2|x_2), Y_2)$. Senders 1 and 2 are joined by network links of capacity $C_{12}$ and $C_{21}$ with information being exchanged in opposite directions. This setup was illustrated in Fig. 2.

A conference among encoders is specified by a set of $2K$ functions

$$h_{1k} : U_1^N \times V_{21} \times \ldots \times V_{2(k-1)} \rightarrow V_{1k}$$

$$h_{2k} : U_2^N \times V_{11} \times \ldots \times V_{1(k-1)} \rightarrow V_{2k},$$

such that the conference message $V_{1k} \in V_{1k}$ (or $V_{2k} \in V_{2k}$) of encoder 1 (or encoder 2) at time $k$ depends on the previously received messages $V_{2}^{k-1}$ (or $V_{1}^{k-1}$) and the corresponding source message.
The conference rates\(^4\) are given by
\[
R_{12} = \left(\frac{1}{K}\right) \sum_{k=1}^{K} \log_2 |V_{1k}| \quad \text{and} \quad R_{21} = \left(\frac{1}{K}\right) \sum_{k=1}^{K} \log_2 |V_{2k}|.
\]
A conference is said to be \((C_{12}, C_{21})\)-admissible if and only if
\[
KR_{12} \leq NC_{12} \quad \text{and} \quad KR_{21} \leq NC_{21}.
\]
The encoders are two functions:
\[
f_1 : U_1^N \times V_{21} \times \cdots \times V_{2K} \rightarrow X_1^N
\]
\[
f_2 : U_2^N \times V_{11} \times \cdots \times V_{1K} \rightarrow X_2^N.
\]
These functions map a block of \(N\) source symbols observed by each encoder, and a block of \(K\) messages received from the other encoder, to a block of \(N\) channel inputs. The decoder is a function
\[
g : Y_1^N \times Y_2^N \rightarrow \hat{U}_1^N \times \hat{U}_2^N.
\]
g maps two blocks of channel outputs (one from each channel) into two blocks of reconstructed source sequences.

An \((R_1, R_2, R_{12}, R_{21}, N, K, P_e)\) code for this problem is defined by:
- Two encoders \(f_1, f_2\), with \(|f_1| = 2^{NR_1}\), \(|f_2| = 2^{NR_2}\),
- A decoder \(g\) for the two encoders \(f_1\) and \(f_2\),
- A \((C_{12}, C_{21})\)-admissible conference of length \(K\) and rates \(R_{12}\) and \(R_{21}\),
- \(\Pr(\hat{U}_1^N \hat{U}_2^N \neq U_1^NU_2^N) \leq P_N\).

Finally, we say that reliable communication is possible, meaning that the sources \((U_1U_2)\) can be sent over this network with arbitrarily small probability of error, if, for sufficiently large blocklength \(N\), there exists a \((C_{12}, C_{21})\)-admissible conference of length \(K\) and a \((R_1, R_2, R_{12}, R_{21}, N, K, \epsilon)\) code, for all \(\epsilon > 0\).

The goal of the problem is to characterize the reachback capacity of the network by giving single-letter information-theoretic conditions for reliable communication.

\(^4\)At first glance, it might seem puzzling that the conference rates are defined in terms of the size of the alphabets as in [14], because in our problem the conference messages sent by one encoder and the source values observed by the other encoder are dependent. Note, however, that the definition of the \(2K\) encoding functions that characterize the conference is general enough to admit a random binning mechanism that eliminates said statistical dependence. The present definition of the conference rates is therefore perfectly reasonable.
B. Statement of Main Result

**Theorem 3:** Reliable communication is possible if and only if

\[
H(U_1|U_2) < I(X_1;Y_1) + I(U_1;Z|U_2) \tag{9}
\]
\[
H(U_2|U_1) < I(X_2;Y_2) + I(U_2;Z|U_1) \tag{10}
\]
\[
H(U_1U_2) < I(X_1;Y_1) + I(X_2;Y_2) \tag{11}
\]

for some auxiliary variable \( Z \) such that \( I(U_1;Z|U_2) < C_{12} \), \( I(U_2;Z|U_1) < C_{21} \), \(|Z| \leq |U_1||U_2|\).

C. Achievability Proof based on Cooperative Source Coding

The achievability part of the proof is based on separate source and channel coding. First, we describe the conferencing mechanism, then we give the rate region for distributed source coding with partial cooperation between encoders. The conditions in the theorem then follow from the intersection of this rate region with the capacity region of the channels. The resulting system architecture is illustrated in Fig. 4.

Fig. 4. Coding strategy for the achievability proof of Theorem 3: cooperative Slepian-Wolf source codes followed by classical channel codes.

**Proof:** Partition the set \( U_1 \) in \( M_1 \) cells, indexed by \( v_1 \in \{1, 2, \ldots, M_1\} \), such that \( v_1(u_1) = c_1 \) if \( u_1 \) is inside cell \( c_1 \). Similarly, partition the set \( U_2 \) in \( M_2 \) cells, indexed by \( v_2 \in \{1, 2, \ldots, M_2\} \), such that \( v_2(u_2) = c_2 \) if \( u_2 \) is inside cell \( c_2 \).

Upon observing a block \( u_1^N \) of source outputs, encoder 1 determines \( v_1 \) for each observed value \( u_1 \). Similarly, encoder 2 determines \( v_2 \) for each observed value \( u_2 \) of the source output block \( u_2^N \). Using the conference mechanism encoder 1 can send a block \( v_1^N \) to encoder 2 at rate \( R_{12} < C_{12} \), and encoder 2 can send a block \( v_2^N \) to encoder 1 at rate \( R_{21} < C_{21} \). We will now show that the rates \( R_{12} = H(V_1|U_2) \) and
\( R_{21} = H(V_2|U_1) \) are sufficient for \( V_1^N \) and \( V_2^N \) to be exchanged between the encoders with arbitrarily small probability of error. Since \((U_1U_2)\) are random and \((V_1V_2)\) are functions of \((U_1U_2)\), \((V_1V_2)\) are random as well. The encoders are assumed to have knowledge of the joint distribution \( p(u_1u_2v_1v_2) \), from which they can obtain the marginals \( p(u_1v_2) \) and \( p(u_2v_1) \). Notice that these two distributions can be viewed as two pairs of correlated sources \((U_1V_2)\) and \((U_2V_1)\). Since \( U_2^N \) is known at encoder 2, it follows from the Slepian-Wolf theorem for \((U_2V_1)\) that \( V_1^N \) can be compressed at rates \( R_{12} \geq H(V_1|U_2) \) and still be reconstructed perfectly at encoder 2. Similarly, \( V_2^N \) can be compressed at rates \( R_{21} \geq H(V_2|U_1) \) and still be reconstructed perfectly at encoder 1. Thus, using separate source and channel coding, \( V_1^N \) and \( V_2^N \) can be transmitted over the conference links at rates

\[
R_{12} = H(V_1|U_2) < C_{12} \quad \text{(12)}
\]

\[
R_{21} = H(V_2|U_1) < C_{21}, \quad \text{(13)}
\]

with arbitrarily small probability of error.

Let \( Z = (V_1V_2) \). Since \((V_1V_2)\) are functions of the source random variables \((U_1U_2)\), \( Z \) is also a random variable and a function of \((U_1U_2)\), which in turn means that \( p(u_1u_2z) = p(u_1u_2)p(z|u_1u_2) \) is a well-defined probability distribution. Instead of (12), we can now write

\[
C_{12} > H(V_1|U_2)
= H(V_1V_2|U_2)
= H(V_1V_2|U_2) - H(V_1V_2|U_1U_2)
= I(U_1; V_1V_2|U_2)
= I(U_1; Z|U_2).
\]

Similarly, (13) yields \( C_{21} > I(U_2; Z|U_1) \).

After conferencing, the encoders compress their data using distributed source codes. Let \( U'_1 = (U_1Z) \) and \( U'_2 = (U_2Z) \). Since \( U_1 \) and \( U_2 \) are i.i.d. sources, \( Z = f(U_1U_2) \) is also i.i.d. \( U'_1 \) and \( U'_2 \) can be viewed as two i.i.d. sources \( \sim p(u'_1u'_2) = p(u_1u_2z) \). Then, according to the Slepian-Wolf theorem, the following compression rates are achievable:

\[
R_1 > H(U'_1|U'_2)
\]

\[
R_2 > H(U'_2|U'_1)
\]

\[
R_1 + R_2 > H(U'_1U'_2).
\]
Substituting \( U'_1 = (U_1 Z) \) and \( U'_2 = (U_2 Z) \), we get

\[
R_1 > H(U_1 Z | U_2 Z) = H(U_1 | U_2 Z) \\
R_2 > H(U_2 Z | U_1 Z) = H(U_2 | U_1 Z) \\
R_1 + R_2 > H(U_1 U_2 Z) = H(U_1 U_2).
\]

Adding channel codes separately, we conclude that reliable communication is possible if this rate region intersects the capacity region of the channels. We can write this as

\[
H(U_1 | U_2 Z) < I(X_1; Y_1) \\
H(U_2 | U_1 Z) < I(X_2; Y_2) \\
H(U_1 U_2) < I(X_1; Y_1) + I(X_2; Y_2),
\]

which is equivalent to

\[
H(U_1 U_2) < I(X_1; Y_1) + I(Z; U_1 U_2) \\
H(U_2 | U_1) < I(X_2; Y_2) + I(Z; U_2 | U_1) \\
H(U_1 U_2) < I(X_1; Y_1) + I(X_2; Y_2),
\]

thus concluding the proof of achievability.

Notice that the conference mechanism described in the proof relies on two deterministic partitions, which can be chosen arbitrarily. These two partitions determine \( p(v_1 v_2) \), which can easily be obtained from \( p(u_1 u_2) \) by summing over all \( (u_1 u_2) \) in each pair of partition cells indexed by \( (v_1 v_2) \). Since \( p(v_1 v_2) = p(z) \), the choice of partition determines the auxiliary random variable \( Z \), which together with the source and channel encoders define the operation point in the reachback capacity region. In other words, for an arbitrary choice of \( Z \) (or equivalently of partitions) for which there exists an admissible conference such that \( I(U_1; Z | U_2) < C_{12} \) and \( I(U_2; Z | U_1) < C_{21} \), Theorem 3 gives the conditions for reliable communication, i.e., the exact reachback capacity with partially cooperating encoders. The latter includes all achievable points for an arbitrary choice of \( Z \), and so it is not necessary to specify the partitions any further.

Instead of exchanging conference messages first and then performing separate source and channel coding, one could start by compressing the sources using Slepian-Wolf codes, and then allowing the channel encoders to exchange messages as proposed by Willems in [14]. We address this issue in Appendix A, by giving an alternative achievability proof based on the coding strategy shown in Fig. 5. It turns out that there is nothing to lose from moving the conference mechanism to the channel encoders.
Fig. 5. Coding strategy for the alternative achievability proof of Theorem 3: classical Slepian-Wolf source codes followed by Willems’ cooperative channel codes.

D. Converse of Theorem 3

The converse part of Theorem 3 can be proved, similarly to Theorem 1, using Fano’s inequality and standard techniques. By exploiting two long Markov chains, in this case $Y_2^N - X_2^N - (Z^N U_1^N) - X_1^N - Y_1^N$ and $Y_2^N - X_2^N - (U_2^N Z^N) - X_1^N - Y_1^N$, we can show that the conditions obtained in the previous subsection are not only sufficient but also necessary for reliable communication to be possible. Since the complete proof is rather technical and lengthy (but conceptually straightforward), details are only provided in Appendix B.

E. Cooperative versus Non-Cooperative Reachback

We now take a closer look at the implications of Theorem 3. The first thing to note is that $Z$ is a variable that can be interpreted as the information exchanged by the two encoders. Therefore, by explicitly solving the channel capacity problems for communication between the two encoders and with the far receiver, a simpler (and more intuitive) version of the theorem is obtained:\(^5\)

\[
\begin{align*}
H(U_1|U_2) &< C_1 + C_{12} \\
H(U_2|U_1) &< C_2 + C_{21} \\
H(U_1 U_2) &< C_1 + C_2,
\end{align*}
\]

where $C_1, C_2, C_{12}, C_{21}$ are the capacities of the corresponding channels. Now, from this simpler version, we can see clearly that there is indeed a strict improvement over the conditions given in Theorem 1 for

\(^5\)We thank Prof. Tom Cover for pointing this out to us.
the case of no cooperation. For example, it is easy to see that if $C_{12} > H(U_1|U_2)$, and $C_{21} > H(U_2|U_1)$, then these conditions reduce to the one for the classical point-to-point problem: $H(U_1U_2) < C_1 + C_2$.

Any point on the surface of the sum-rate face of the region can be achieved by having encoders send their realization to each other using Slepian-Wolf codes—in this way, both can reconstruct both sources, generate a joint encoding, and then split this encoding in whatever way they choose to. Some of these points are clearly not achievable without cooperation among encoders: e.g., $(R_1, R_2) = (0, H(U_1U_2))$ and $H(U_1|U_2) > 0$, even if $H(U_1U_2) < I(X_2; Y_2)$. This is illustrated in Fig. 6.

It is also interesting to observe in Fig. 6 that, contrary to what happens with general multiple access channels, in the case of independent channels considered in this work cooperation does not lead to an increase in the achievable sum-rate. This suggests a routing interpretation for Theorem 3. If, say, encoder 1 has too much data to send and a channel not good enough (that is, $H(U_1|U_2) > C_1$), and encoder 2 has enough idle capacity (that is, $H(U_2|U_1) < C_2$), then encoder 1 can use $Z$ to route some of its data.

---

**Fig. 6.** An example to illustrate the effect of cooperation among encoders on the feasibility of reliable communication between the reachback network and the far receiver. Since the capacity region of the pair of independent channels and the Slepian-Wolf region do not intersect, it follows from Theorem 1 that reliable communication is not possible. However, with cooperation we can enlarge side faces of the capacity region by the capacity of the conference channels until there are points of intersection with the Slepian-Wolf region (the shaded portion of the picture). To achieve such points, it is necessary for the encoder with a bad channel to route some of its data to the joint decoder via the good channel available to the other encoder.
to the far receiver via encoder 2. The total number of bits that can be sent over the reachback network is still bounded by $I(X_1; Y_1) + I(X_2; Y_2)$. But, provided this constraint is not violated, each encoder can also act as a relay for the other encoder, in this way relaxing the conditions on the minimum amount of data required from each encoder. And the total amount of information that can be exchanged among nodes is given by the capacity of the interconnection network between encoders.

Incidentally, note also that the widely accepted view that “according to the Slepian-Wolf theorem separate encoders can achieve the same compression performance of a joint encoder” is only partially accurate. Indeed, the total number of bits required with separate and with joint encoders remains the same. However, for the informal statement above to be an accurate description of things, the joint decoder needs to receive a minimum amount of information from each encoder, so not any point in the sum-rate region $R_1 + R_2 > H(U_1 U_2)$ is achievable without cooperation. The net effect of cooperation is to relax this requirement, to the extent supported by the interconnection network between encoders.

F. Partially Cooperating Nodes with Constrained Encoders

In Section II-E we discussed a reachback scenario in which the sensor nodes send correlated codewords instead of using Slepian-Wolf source coding. As argued there, this constraint is interesting in part because it models the case in which the sensor nodes have limited complexity and are not capable of encoding their data optimally to remove correlations. In this subsection, we address the same issue, now in the presence of partial cooperation among encoders.

Going back to the achievability proof of Theorem 3, we observe that Slepian-Wolf codes are used for two different tasks: (1) to compress the source messages prior to transmission over the reachback channels, and (2) to compress the conference messages prior to transmission over the conference links:

- The first task is identical to the case with non-cooperative encoders, and so it is reasonable to assume that the use of correlated codes will lead to a rate loss relative to the region given by Theorem 3 similar to that shown in Theorem 2.
- The second task imposes one additional requirement on the encoders – to be able to reconstruct the Slepian-Wolf encoded conference messages the encoders must have full knowledge of the joint probability distribution $p(u_1 u_2)$.

From a practical point of view, this requirement could pose some difficulties, since it may be hard for the sensor nodes to obtain or estimate the dependencies among all observed variables. Part of the

6Whereas with the benefit of hindsight this might seem obvious, taking the informal statement above literally is a mistake we made in our initial steps on this work, by assuming that cooperation would not result in an increase of the capacity region.
Reason that makes Slepian-Wolf codes attractive for practical sensor networking applications is the fact that, by not requiring knowledge of typical sets at the encoders, complexity is moved to the decoder. Yet in this conferencing mechanism, we do require knowledge of those typical sets at the encoders, to decode conference messages. We are therefore interested in obtaining a set of conditions for reliable communication in which the encoders do not exploit the joint distribution for conferencing.

The following theorem gives sufficient conditions for reliable communication under these two encoding constraints.

**Theorem 4:** Let \( (U_1, U_2) \) be two correlated sources drawn i.i.d. \( \sim p(u_1 u_2) \), and transmitted over two independent channels \( \{X_1, p(y_1|x_1), Y_1\} \) and \( \{X_2, p(y_2|x_2), Y_2\} \). Assuming that the encoders are connected by communication links of capacities \( C_{12} \) and \( C_{21} \), do not have knowledge of \( p(u_1 u_2) \), and use correlated codes \( \{x_1^n(u_1^n), x_2^n(u_2^n)\} \) then reliable communication is possible if

\[
H(U_1|U_2) < I(X_1; Y_1|U_2 Z) + I(U_1; Z|U_2)
\]

\[
H(U_2|U_1) < I(X_2; Y_2|U_1 Z) + I(U_2; Z|U_1)
\]

\[
H(U_1 U_2) < I(X_1 X_2; Y_1 Y_2),
\]

for some \( p(u_1 u_2) \cdot p(v_1|u_1) \cdot p(v_2|u_2) \cdot p(x_1|u_1) \cdot p(x_2|u_2) \cdot p(y_1|x_1) \cdot p(y_2|x_2) \) and \( Z = (V_1 V_2) \), such that \( H(Z|V_2) < C_{12} \), \( H(Z|V_1) < C_{21} \), and \( |Z| \leq |U_1||U_2| \).

**Proof:** We start with the conferencing mechanism and then obtain the conditions for reliable communication by generalizing Theorem 2.

The conference messages are generated using the same partitions as in the proof of Theorem 3. The pair of conference channels is then equivalent to a two-way channel without interference [18, pp. 351-352], yet with correlated inputs. Since the encoders cannot exploit the joint probability distribution, we assume they compress the conference messages to their marginal entropies and then add channel coding to transmit them reliably over the conference channels. It follows then from the source coding theorem and from the channel coding theorem that a sufficient condition for reliable communication of the conference messages to be possible is \( H(V_1) = H(Z|V_2) < C_{12} \) and \( H(V_2) = H(Z|V_1) < C_{21} \), with \( Z = (V_1 V_2) \).

\[\text{Note that we do not claim separate source and channel coding to be an optimal coding strategy for sending correlated sources over a two-way channel without interference and without knowledge of the joint probability distribution at the encoders. This is because, without proof, we cannot rule out the existence of a universal coding strategy (meaning, without a priori knowledge of source statistics), leading to less restrictive conditions for reliable communication over the conference channels than those stated above.}\]
Now, let $U'_1 = (U_1 Z)$ and $U'_2 = (U_2 Z)$. Since $U_1$ and $U_2$ are i.i.d. sources, $Z = f(U_1 U_2)$ is also i.i.d., and $U'_1$ and $U'_2$ can be viewed as two i.i.d. sources $\sim p(u'_1 u'_2) = p(u_1 u_2 z)$. Then, according to Theorem 2, reliable communication is possible if

$$H(U'_1|U'_2) < I(X_1; Y_1|U'_2)$$
$$H(U'_2|U'_1) < I(X_2; Y_2|U'_1)$$
$$H(U'_1 U'_2) < I(X_1 X_2; Y_1 Y_2).$$

Substituting $U'_1 = (U_1 Z)$ and $U'_2 = (U_2 Z)$, we get

$$H(U_1 U_2) < I(X_1; Y_1|U_2 Z)$$
$$H(U_2 U_1) < I(X_2; Y_2|U_1 Z)$$
$$H(U_1 U_2) < I(X_1 X_2; Y_1 Y_2).$$

The conditions in the theorem then follow from standard identities.

Comparing the expressions in Theorems 2, 3, and 4, we conclude that the rate loss under the given encoding constraints is twofold. First, the use of correlated codes leads to similar rate loss terms as in Section II-E. Secondly, the restriction on the choice of codes for conferencing implies a restriction on the choice of auxiliary variable $Z$, possibly leading to smaller values of $I(U_1; Z|U_2)$ and $I(U_2; Z|U_1)$ and thus to a potential reduction in reachback capacity.

IV. Reachback Capacity with an Arbitrary Number of Nodes ($M \geq 2$)

A. $M \geq 2$ Non-Cooperating Nodes

Having established the reachback capacity region for the case of two non-cooperating nodes, we now generalize this result to the transmission of $M$ correlated sources over $M$ independent channels, for arbitrary $M \geq 2$. The following theorem gives the conditions for reliable communication for this case, illustrated in Fig. 7.

**Theorem 5:** A set of correlated sources $U^M = \{U_1, U_2, \ldots, U_M\}$ can be communicated reliably over independent channels $(X_1, p(y_1|x_1), Y_1) \ldots (X_M, p(y_M|x_M), Y_M)$ if and only if

$$H(U(S)|U(S^c)) < \sum_{i \in S} I(X_i; Y_i),$$

for all subsets $S \subseteq \{1, 2, \ldots, M\}$ and $U(S) = \{U_j : j \in S\}$. 

11/17/2003 DRAFT
**Proof:** The converse can be proved exactly as in Theorem 1, this time dealing 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

$$R(S) > H(U(S)|U(S^c)),$$

(15)

for all $S \subseteq \{1, 2, \ldots, M\}$ where $R(S) = \sum_{i \in S} R_i$ and $U(S) = \{U_j : j \in S\}$. If the conditions in the theorem are fulfilled for all subsets $S \subseteq \{1, 2, \ldots, M\}$, then the Slepian-Wolf region defined by (15) intersects the capacity region given by $R_i < I(X_i; Y_i)$, $i = 1 \ldots M$, which means that the compressed source blocks can be transmitted over the array of $M$ independent channels with arbitrarily small probability of error by adding separate channel codes.

Generalizing Theorem 2 in a similar way, we obtain the following result for correlated codes:

**Theorem 6:** A set of correlated sources $\{U_1, U_2, \ldots, U_M\}$ can be communicated reliably over independent channels $(X_1, p(y_1|x_1), Y_1) \ldots (X_M, p(y_M|x_M), Y_M)$ with correlated codes if

$$H(U(S)|U(S^c)) < \sum_{i \in S} I(X_i; Y_i|U(S^c)),$$

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

---

Fig. 7. A sketch of the regions involved in Theorem 5 for $M = 3$ sources. Once again, when the capacity region of the independent reachback channels does not intersect the Slepian-Wolf rate region (left plot) reliable communication is not possible. If the two regions do intersect (right plot) than all points in the intersection are achievable.
Proof: The proof is similar to the proof of Theorem 2, using the more general version of the theorem by Cover, El Gamal and Salehi for $M > 2$ sources [12].

B. $M \geq 2$ Cooperating Nodes

The sensor reachback problem with $M \geq 2$ cooperating nodes is a network problem in which $M$ encoders observe $M$ correlated sources, then exchange messages over an interconnection network of limited capacity, and finally send the information to a far receiver over an array of $M$ independent channels. This setup is illustrated in Fig. 8.

By generalizing Theorem 3, we get the following conditions for reliable communication:

Theorem 7: A set of correlated sources $U^M = \{U_1 U_2 \ldots U_M\}$ can be communicated reliably with partially cooperating encoders over independent channels $(X_1, p(y_1|x_1), Y_1) \ldots (X_M, p(y_M|x_M), Y_M)$, if and only if there exist random variables $Z_{ij}$, $i = 1, \ldots, M$, $j = 1, \ldots, M$, and $i < j$, such that

$$H(U(S)|U(S^c)) < \sum_{i \in S} I(X_i; Y_i) + I(U(S); Z(S^c)|U(S^c)),$$

for all subsets $S \subseteq \{1, 2, \ldots, M\}$, where $U(S) = \{U_j : j \in S\}$, $Z(S) = \{Z_{ij} : i \in S \text{ or } j \in S\}$, $I(Z_{ij}; U_i|U_j) < C_{ij}$ and $I(Z_{ij}; U_j|U_i) < C_{ji}$.

Proof: The proof is very similar to Theorem 3. We start with the achievability part. First, the encoders establish pairwise conferences over the network. We define $M(M - 1)/2$ auxiliary random
variables $Z_{ij} = (V_{ij} V_{ji})$, $i = 1, \ldots, M$, $j = 1, \ldots, M$, and $i < j$, which are equal to the messages $V_{ij}$ and $V_{ji}$ exchanged by encoders $i$ and $j$ at rates $C_{ij}$ and $C_{ji}$.

After conferencing the encoders compress the source blocks and the received/sent message blocks using Slepian-Wolf codes. Let $U'_k = (U_k Z(k))$ with $k = 1, \ldots, M$ and $Z(k) = \{Z_{ij} : i = k \text{ or } j = k\}$, such that $p(u'_1 \ldots u'_M) = p(u_1 \ldots u_M z_{12} \ldots z_{M-1,M})$. The $M$-source version of the Slepian-Wolf theorem [11, Theorem 14.4.2] guarantees that the rates

$$R(S) > H(U'(S)|U'(S^c))$$

for all subsets $S \subseteq \{1, 2, \ldots, M\}$ with $R(S) = \sum_{s \in S} R_s$, $U'(S) = \{U_k : k \in S\}$, $U(S) = \{U_j : j \in S\}$, and $Z(S) = \{Z_{ij} : i \in S \text{ or } j \in S\}$, are achievable. If conditions

$$H(U(S)|U(S^c)) < \sum_{i \in S} I(X_i; Y_i) + I(U(S); Z(S^c)|U(S^c)),$$

are fulfilled for all subsets $S \subseteq \{1, 2, \ldots, M\}$, then the Slepian-Wolf region defined by (17) intersects the capacity region given by $R_i < I(X_i; Y_i)$, $i = 1, 2, \ldots, M$, which is true whenever the conditions in the theorem are fulfilled, then the compressed source blocks can be transmitted over the array of $M$ independent channels with arbitrarily small probability of error by adding separate channel codes.

The converse part is similar to the converse proof of Theorem 3 with $2^M - 1$ inequalities. Details are provided in Appendix C.

C. Examples

With a few concrete examples, we illustrate the usefulness of Theorems 5 and 7.

1) Reachback Communication over Gaussian Channels with Orthogonal Multiple Access: The capacity of the Gaussian multiple access channel with $M$ independent sources is given by

$$\sum_{i=1}^m R_i \leq \frac{1}{2} \log \left( 1 + \frac{\sum_{i=1}^m P_i}{N} \right),$$

for all $m \subseteq \{1, 2, \ldots, M\}$, where $N$ and $P_i$ are the noise power and the power of the $i$-th user respectively [11, pp. 378-379]. If we use orthogonal accessing (e.g., TDMA), and assign different time
slots to each of the transmitters, then the Gaussian multiple access channel is reduced to an array of $M$

independent single-user Gaussian channels with capacity

$$C_i = \tau_i \cdot \frac{1}{2} \log \left( 1 + \frac{P_i}{N\tau_i} \right), \quad 1 \leq i \leq M,$$

where $\tau_i$ is the time fraction allocated to source user $i$.

Applying Theorem 5, we obtain the reachback capacity of the Gaussian channel with orthogonal

accessing\(^8\). Reliable communication is possible if and only if

$$H(U(S)|U(S^c)) \leq \sum_{i \in S} \tau_i \cdot \frac{1}{2} \log \left( 1 + \frac{P_i}{N\tau_i} \right),$$

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

2) Reachback Networks with Different Topologies: Theorem 7 can be applied to a great variety of

network topologies, depending on the capacities of the links between nodes. Three such examples are

shown in Fig. 9.

![Diagram](image)

Fig. 9. Three sensor network topologies for which Theorem 7 gives the reachback capacity. Dark circles denote the sensor nodes and a light circle represents the remote receiver. Conference channels and reachback channels are depicted as dashed and solid arrows, respectively, with the arrow head indicating the direction of the flow of information. Conferencing is done before transmission over the reachback channels.

Three interesting cases are illustrated in Fig. 9:

- In (a) the sensor nodes build a linear array and conference communication occurs along the vertical axis. To obtain the reachback capacity we must set all capacities to zero, except those corresponding to the reachback channels $C_1 \ldots C_4$ and those corresponding to the active conference links $C_{12}, C_{21}$.

\(^8\)The generalization of Theorem 5 for channels with real-valued output alphabets can be easily obtained using the techniques in [11, Section 9.2 and Chapter 10].
$C_{23}, C_{32}, C_{34}, C_{43}$. The auxiliary random variables become $Z_{12}, Z_{23}$ and $Z_{34}$ and the application of Theorem 7 follows easily.

- In (b) not all sensor nodes are connected to the remote location, and hence the information needs to be relayed to those nodes which do have a reachback channel available. Here, we must consider the capacities $C_5, \ldots, C_7$ for reachback, and $C_{15}, C_{25}, C_{26}, C_{36}, C_{37}, C_{47}$ for conferencing, such that the auxiliary random variables become $Z_{15}, Z_{25}, Z_{26}, Z_{36}, Z_{37}, Z_{47}$.

- Finally, (c) shows a tree topology, in which each parent node has a reachback channel available ($C_5, \ldots, C_3$), and the children can relay their information over conference channels with capacities $C_{42}, C_{52}, C_{63}, C_{73}$. To compute the reachback capacity we must define the auxiliary random variables $Z_{24}, Z_{25}, Z_{36}, Z_{37}$. Once again, the application of Theorem 7 is straightforward.

In all cases, it is interesting to see how reliable communication is possible if and only if the network of interconnections among encoders has enough capacity to redistribute sensed data, so as to match the amount of data to upload to the capacity of the channels from each node to the far receiver, and further expanding on the routing interpretation developed in Section III-E.

V. ON THE SEPARATION OF SOURCE AND CHANNEL CODING IN COMMUNICATION NETWORKS

A. Separation is Optimal for the Sensor Reachback Problem

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 [11, Ch. 8.13]. From this result, commonly known as the separation principle, 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.

In the context of the sensor reachback problem with non-cooperating nodes, the proof of Theorem 5 gives necessary and sufficient conditions for reliable communication that, after solving the capacity problems, can be written as:

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

for all subsets $S \subseteq \{1, 2, \ldots, M\}$. These conditions show that a generalization of the previous statement also holds for the transmission of multiple correlated sources $(U_1, U_2, \ldots, U_M)$ over independent channels of capacities $(C_1, C_2, \ldots, C_M)$ – there is nothing to lose by compressing the sources to their most efficient
representation (Slepian-Wolf coding) and separately adding channel codes. Furthermore, for the sensor reachback problem with partially cooperating encoders, Theorem 3 also shows that an optimal coding strategy for this problem consists once again of a cascade of a cooperative version of Slepian-Wolf source codes, followed by classical channel codes.

Therefore, these results identify an important class of non-trivial communication networks and information theory problems, in which the classical notion of separation between sources and channels holds.

B. Is Separation Relevant for Networks?

Based on the observations above it is only natural that we revisit the issue of optimality of separate source and channel coding in communication networks. This question is certainly not trivial, and we are not yet in a position to provide a definite answer. However, we feel it is only appropriate to discuss some of the intuition we derive about this most relevant issue from the results presented in this paper.

We observe first that both in the converse proof of the separation theorem, as well as in the converse proofs for the different instances of the sensor reachback problem addressed in this paper, the key ingredient that renders separation optimal is the data processing inequality [11, Ch. 2.8]. Application of this inequality requires Markov dependencies among random variables used to model sources and channel inputs/outputs. And as shown in this work, this property arises not only in point-to-point problems, but also in various non-trivial networks. Now, it is well known that this Markov property does not hold for a general multiple access channel with correlated sources, as established by the simple example of a binary adder channel and two binary sources with joint probability \((1/3, 1/3, 0, 1/3)\) in [12], and this has been the basis so far for arguing that separation does not hold in networks. However, after looking at all the evidence available, concluding from that simple example that the separation principle is not useful in the context of communication networks does appear to us to be too hasty a step:

- **Separation holds in other networks.** The team of Effros, Médard, Koetter, et al., showed that separation is optimal for a large class of networks [19], the crux being that all operations are carried out over a common finite field. A most remarkable aspect of their result is that, with this simple and natural restriction, it is shown in [19] that separation is optimal even for the example of the binary adder channel used in [12] to motivate the need for joint source/channel codes in networks. Also, in [20], Merhav and Shamai give an example of a point-to-point problem with side information for which separation holds. Also, in an unpublished manuscript that we recently became aware of, Yeung had also established some initial separation results in some simple networks [21].
So, for one network example where separation fails, there are multiple other network examples where separation holds, and one could easily argue more relevant, too.

- **The performance gap between a separation based approach and an unconstrained approach is not known in general.** In those cases for which separate source and channel coding has been found to be a suboptimal strategy, it seems pertinent to establish the extent to which separation leads to a loss of performance. To the best of our knowledge, no conclusive piece of evidence has been provided to establish beyond a reasonable doubt the need for joint source/channel codes in a network setup.9

- **Reservation-based accessing schemes are common practice, and practical distributed source codes exist.**Because of their simplicity and low complexity, multiple access schemes that deal with the interference issue by dividing the medium into independent channels are widely used in many communication networks, and are of particular appeal for highly resource constrained nodes in sensor networks [24]. One important advantage is the reduction of the very challenging multi-user channel coding problem (e.g., [25] and [26]), to multiple instances of the point-to-point problem, which is well understood both in theory and in practice. Similarly, wireless sensor networks have led a considerable amount of research in the area of distributed source coding, yielding practical schemes that come close to the theoretical limits obtained by Slepian and Wolf [10], and Wyner and Ziv [27] (e.g. [28], [29], [30], [31], [32]). It is therefore interesting to know the ultimate performance limits of communications systems with distributed source coding and separate channel coding even in networks where the separation principle does not hold.

The separation of source and channel coding is one of the cornerstones of digital communications. By representing information in terms of bits, Shannon provided an architecture for point-to-point communication systems in which the task of data compression and the task of channel noise mitigation are carried out by separate modules of the system, without any performance degradation due to this split of tasks. In the context of communication networks, we argue that joint source/channel codes are not always the only viable approach and that, whenever possible we should take advantage of over 50 years of experience in the design of communication systems based on separate source and channel coding.

Even in networks for which the separation principle does not hold, separate design still gives us the

---

9Gastpar and Vetterli have presented some preliminary work along these lines [22], related to a sensor networking application derived from Berger’s CEO problem [4], and built on top of Berger’s results on uncoded transmission [23, pg. 162]. Now, while that preliminary result does hint at a potentially large gap in the extreme case of no cooperation at all, in another extreme case of the same setup (full cooperation), the gap vanishes. Hence, it does seem to us that the work of [22] needs to be further developed (e.g., to address partial cooperation), before valid inferences can be made about the utility of separation in that setup.
practical advantages of a system with multiple reusable components. In a real application, performance is a most important factor, but is not the only one: for example, the ability to quickly assemble working systems out of off-the-shelf components might justify some performance loss in highly dynamic environments such as a battlefield, or a marketplace. Therefore, even for networks in which separation turns out to be suboptimal, it is still of great interest to know what are the performance limits when enforcing separation constraints.

We end this section quoting Ahlswede and Han on the issue of separation in networks, from an early paper on multiterminal source coding [16]:

Another way of coming closer to a real communication situation with our models consists of enforcing the separation principle (in spite of its suboptimality in an ideal situation) and investigating what can be done (also optimally) if source and channel coding are carried out separately.

VI. CONCLUSIONS

A. Summary of Contributions

In this paper we have considered the sensor reachback problem. We formulated this problem as one of communication of multiple correlated sources over an array of independent channels, with partial cooperation among encoders. We defined the notion of reachback capacity, and gave exact characterizations for this capacity in a number of scenarios (nodes cooperating or not, constraints on the encoders, and numbers of nodes in the network). Having found in all cases that a natural network generalization of the classical joint source/channel coding theorem holds, we revisited the issue of source/channel separation in networks, where we argued that it may be too soon to dismiss separation as irrelevant in networks.

B. Future Work

After having established capacity theorems for the sensor reachback problem, there is a question that comes up naturally: what if, in a given scenario, it turns out to be impossible to match the rates of an array of Slepian-Wolf encoders to the capacities of the channels? In that case, the best we can hope for is to reconstruct an approximation to the original source message. Now, in the point-to-point setup, this is what happens when we have a source $U$ with entropy $H(U) > C$, where $C$ is the capacity of the channel over which it needs to be sent—and in this case, the answer is given by rate-distortion theory [23]. Therefore, it is only natural to consider a rate-distortion version of the sensor reachback problem – in
the case of non-cooperating encoders we encounter none other than the classical Multiterminal Source Coding problem [33].

It follows from Theorem 1 that the cascade of Slepian Wolf random binning and channel codes is an architecture capable of achieving capacity for a reachback network. Therefore, by relying informally on the idea that capacity and rate-distortion are dual problems (at least in the point-to-point problem), we are currently exploring whether a cascade of rate-distortion codes and random binning of blocks of quantization indices could result in an optimal architecture for the multiterminal source coding problem. This architecture is illustrated in Fig. 10. For this two-stage process (classical quantization followed by random binning), we have determined an achievable rate-distortion region in [34]. Proving that this bound is tight is part of our ongoing work.

ACKNOWLEDGEMENTS

The authors most gratefully acknowledge discussions with Prof. Toby Berger, whose encouragement, guidance and friendship has made work on information theory problems so much more enjoyable. They also wish to acknowledge support provided by Prof. Joachim Hagenauer, without which the authors would not have been able to work together. Finally, the authors also wish to acknowledge fruitful discussions with Zaher Dawy and Michael Tüchler (LNT/TUM), and with Ron Dabora, An-swol Hu, Christina Peraki (Cornell/ECE) and Megan Owen (Cornell/Applied Math).
APPENDIX

A. An Alternative Achievability Proof for Theorem 3 based on Cooperative Channel Coding

In [14], Willems obtained the capacity region of the multiple access channel with partially cooperating encoders. He did this by introducing a class of channel codes with cooperation, that build on a construction due to Slepian and Wolf for a multiple access channel with two encoders, two independent sources and a third common source observed by both encoders [35]. We now show that in the sensor reachback problem there is no performance loss associated with an alternative architecture to the one presented in the main text (illustrated in Fig. 4). In this new architecture, we use classical Slepian-Wolf codes to remove the correlation between the sources, and then apply Willems’ cooperative channel coding approach (as illustrated in Fig. 5). Besides the historical interest (we developed this proof first), this alternative proof also serves the purpose of showing that there is nothing to lose in terms of performance by moving cooperation from the source coders to the channel coders.

Proof: Each source encoder takes an input block \( U_i^N, i \in \{1, 2\} \) and outputs a bin index \( W_i \) from the alphabet \( W_i = \{1, 2, ..., 2^{N R_i}\} \).

Prior to transmission the channel encoders exchange messages over the conference channels. The conference messages are obtained as follows. First, we partition the set of messages \( W_1 = \{1, 2, ..., 2^{N R_1}\} \) in \( 2^{NR_{12}} \) cells, indexed by \( i_1 \in \{1, 2, ..., 2^{NR_{12}}\} \), such that \( i_1(w_1) = c_1 \) if \( w_1 \) is inside cell \( c_1 \). Similarly, we partition the set of messages \( W_2 = \{1, 2, ..., 2^{N R_2}\} \) in \( 2^{NR_{21}} \) cells, indexed by \( i_2 \in \{1, 2, ..., 2^{NR_{21}}\} \), such that \( i_2(w_2) = c_2 \) if \( w_2 \) is inside cell \( c_2 \). All messages inside each cell \( c_1 \) are indexed by \( j_1 \in \{1, 2, ..., 2^{N(R_1-R_{12})}\} \), and all messages inside each cell \( c_2 \) are indexed by \( j_2 \in \{1, 2, ..., 2^{N(R_2-R_{21})}\} \).

During the conference, encoder 1 sends index \( i_1 \) to encoder 2, and encoder 2 sends index \( i_2 \) to encoder 1.

Since \( U_1^N \) and \( U_2^N \) are random variables, \( W_1, W_2, I_1 \) and \( I_2 \) are also random.

The conditions for reliable communication under this conference scenario are given in [14] and can be written as

\[
R_1 - R_{12} \leq I(X_1; Y_1 Y_2 | X_2 Z) \\
R_2 - R_{21} \leq I(X_2; Y_1 Y_2 | X_1 Z) \\
R_1 + R_2 - R_{12} - R_{21} \leq I(X_1 X_2; Y_1 Y_2 | Z) \\
R_1 + R_2 \leq I(X_1 X_2; Y_1 Y_2),
\]

where \( Z \) is an auxiliary random variable such that \( Z = (I_1 I_2) \), \( p(w_1 w_2 z) = p(w_1)p(w_2)p(z|w_1 w_2) \).
Since \( R_1 = \frac{1}{N} H(W_1) \), and \( R_1 - R_{12} = \frac{1}{N} H(W_1|Z) \), we have that \( R_{12} = \frac{1}{N}(H(W_1) - H(W_1|Z)) = \frac{1}{N} I(W_1;Z) \). Similarly, \( R_{21} = \frac{1}{N} I(W_2;Z) \). Using these identities and the fact that the channels are independent, we get

\[
R_1 \leq I(X_1;Y_1|Z) + \frac{1}{N} I(W_1;Z) \quad (18)
\]
\[
R_2 \leq I(X_2;Y_2|Z) + \frac{1}{N} I(W_2;Z) \quad (19)
\]
\[
R_1 + R_2 \leq I(X_1;Y_1|Z) + I(X_2;Y_2|Z) + \frac{1}{N} I(W_1W_2;Z) \quad (20)
\]
\[
R_1 + R_2 \leq I(X_1;Y_1) + I(X_2;Y_2), \quad (21)
\]

where we used the fact that \( W_1 \) and \( W_2 \) are independent and therefore \( I(W_1;Z) + I(W_2;Z) = I(W_1W_2;Z) \).

As in the other proof, we know that reliable communication is possible if the capacity region given by (18)-(21) intersects the Slepian-Wolf rate region for \((U_1U_2)\). This is the case if and only if

\[
H(U_1|U_2) \leq I(X_1;Y_1|Z) + \frac{1}{N} I(W_1;Z) \quad (22)
\]
\[
H(U_2|U_1) \leq I(X_2;Y_2|Z) + \frac{1}{N} I(W_2;Z) \quad (23)
\]
\[
H(U_1U_2) \leq I(X_1;Y_1|Z) + I(X_2;Y_2|Z) + \frac{1}{N} I(W_1W_2;Z) \quad (24)
\]
\[
H(U_1U_2) \leq I(X_1;Y_1) + I(X_2;Y_2). \quad (25)
\]

We now develop the sum rate condition (24). First, we note that, since \( W_1W_2 \) are a function of \( U_1^N U_2^N \) (encoding property), and \( U_1^N U_2^N \) are a function of \( W_1W_2 \) (decoding property), both Markov chains \( W_1W_2 - U_1^N U_2^N - Z \) and \( U_1^N U_2^N - W_1W_2 - Z \) hold, and so it follows from the data processing inequality that \( I(W_1W_2;Z) = I(U_1^N U_2^N;Z) \). Noting that \( \frac{1}{N} I(U_1^N U_2^N;Z) = I(U_1U_2;Z) \) (the sources are i.i.d.), we can rewrite (24) as

\[
H(U_1U_2) \leq I(X_1;Y_1|Z) + I(X_2;Y_2|Z) + I(U_1U_2;Z). \quad (26)
\]

To develop the side conditions (22) and (23) accordingly, we use a simple time-sharing argument. Assume the two Slepian-Wolf source encoders operate at rates \( R_1 = H(U) \) and \( R_2 = H(U_2|U_1) \), such that \( H(W_1) = H(U_1^N) \), \( H(U_2^N|W_1) = 0 \) and, consequently,

\[
I(W_1;Z) = I(U_1^N W_1;Z) = I(U_1^N;Z) + I(W_1;Z|U_1^N) = I(U_1^N;Z).
\]

Substituting \( I(W_1;Z) = I(U_1^N;Z) \) and \( R_1 = H(U) \) in condition (18), we get

\[
H(U) + \epsilon = I(X_1;Y_1|Z) + I(U_1;Z), \quad (27)
\]
for some $\epsilon > 0$ arbitrarily small. Since (27) follows from (18), and (24) follows from (20), the source/channel coding theorem by Slepian and Wolf [35] guarantees that there exists a code satisfying both (24) and (27). We can now combine these two conditions by subtracting $H(U_1)$ from both sides of the modified condition (26), so that

$$H(U_2|U_1) \leq I(X_1; Y_1|Z) + I(X_2; Y_2|Z) + I(U_1U_2; Z) - H(U_1) \leq I(X_2; Y_2|Z) + I(U_2; Z|U_1) + I(X_1; Y_1|Z) + I(U_1; Z) - H(U) \leq I(X_2; Y_2|Z) + I(U_2; Z|U_1) + \epsilon,$$

where $\epsilon$ can be made arbitrarily small to yield the second condition in the theorem. The first condition can be obtained by a symmetric argument, with Slepian-Wolf encoders operating at rates $R_1 = H(U_2)$ and $R_2 = H(U_1|U_2)$, so that

$$H(U_2) + \epsilon = I(X_2; Y_2|Z) + I(U_2; Z),$$

and

$$H(U_1|U_2) \leq I(X_1; Y_1|Z) + I(U_1; Z|U_2) + \epsilon.$$  

(29)

By time-sharing between the code construction for $(R_1, R_2) = (H(U_1), H(U_2|U_1))$ and $(R_1, R_2) = (H(U_1|U_2), H(U_2))$, we conclude that conditions (29), (28), (26) and (25) are sufficient for reliable communication.

Looking at the first two terms of the right-hand side of (26), we can write

$$I(X_1; Y_1|Z) + I(X_2; Y_2|Z) = H(Y_1|Z) - H(Y_1|X_1Z) + H(Y_2|Z) - H(Y_2|X_2Z)$$

$$= H(Y_1|Z) - H(Y_1|X_1) + H(Y_2|Z) - H(Y_2|X_2)$$

$$\leq H(Y_1) - H(Y_1|X_1) + H(Y_2) - H(Y_2|X_2)$$

$$= I(X_1; Y_1) + I(X_2; Y_2),$$

with equality for $p(x_1 x_2 z) = p(x_1)p(x_2)p(z)$. Since this choice of $Z$ maximizes the right-handside of (26), we can modify this condition to

$$H(U_1U_2) \leq I(X_1; Y_1) + I(X_2; Y_2) + I(U_1U_2; Z).$$

We conclude that (26) is always satisfied when (25) is satisfied, and so we omit the former.
B. Proof of Converse for Theorem 3

1) Preliminaries: To develop the converse we start with Fano’s inequality. If there is a suitable $(R_1, R_2, R_{12}, R_{21}, N, K, P_e)$-code, then we must have

$$H(U_1^N U_2^N | \hat{U}_1^N \hat{U}_2^N) \leq P_e \log (|U_1^N \times U_2^N|) + h(P_e),$$

where $h(P_e)$ is the binary entropy function. For convenience, define also

$$\delta(P_e) = (P_e \log (|U_1^N \times U_2^N|) + h(P_e)) / N.$$

It follows from eqn. (30) that

$$H(U_1^N U_2^N | Y_1^N Y_2^N) = H(U_1^N U_2^N | Y_1^N Y_2^N g(Y_1^N Y_2^N))$$

$$= H(U_1^N U_2^N | Y_1^N Y_2^N \hat{U}_1^N \hat{U}_2^N)$$

$$\leq H(U_1^N U_2^N | \hat{U}_1^N \hat{U}_2^N)$$

$$\leq N \delta(P_e),$$

and therefore,

$$H(U_1^N U_2^N | Y_1^N Y_2^N V_1^K V_2^K) \leq H(U_1^N U_2^N | Y_1^N Y_2^N) \leq N \delta(P_e),$$

and also

$$H(U_1^N | Y_1^N Y_2^N V_1^K V_2^K) \leq N \delta(P_e)$$

$$H(U_2^N | Y_1^N Y_2^N V_1^K V_2^K) \leq N \delta(P_e)$$

$$H(U_1^N | Y_1^N Y_2^N V_1^K V_2^K U_2^N) \leq N \delta(P_e)$$

$$H(U_2^N | Y_1^N Y_2^N V_1^K V_2^K U_2^N) \leq N \delta(P_e).$$

According to the problem statement, we have two long Markov chains in place: $Y_2^N - X_2^N -(V_1^K V_2^K U_1^N) - X_1^N - Y_1^N$ and $Y_2^N - X_2^N -(U_2^N V_1^K V_2^K) - X_1^N - Y_1^N$. These chains (informally referred to as the long chains in this section) will prove quite useful in our derivations.

2) The Side Faces: Necessity of Equations (9) and (10): We start by bounding $H(U_1^N)$:

$$H(U_1^N) = I(U_1^N ; Y_1^N Y_2^N V_1^K V_2^K U_2^N) + H(U_1^N | Y_1^N Y_2^N V_1^K V_2^K U_2^N)$$

$$\leq I(U_1^N ; Y_1^N Y_2^N V_1^K V_2^K U_2^N) + N \delta(P_e)$$

$$= I(U_1^N ; U_2^N) + I(U_1^N ; Y_1^N Y_2^N V_1^K V_2^K | U_2^N) + N \delta(P_e)$$

$$= I(U_1^N ; U_2^N) + I(U_1^N ; V_1^K V_2^K | U_2^N) + I(U_1^N ; Y_1^N Y_2^N | U_2^N V_1^K V_2^K) + N \delta(P_e)$$
Now, $I(U_1^N; U_2^N)$ and $N\delta(P_e)$ stay the same, and we start with $I(U_1^N; V_1^KV_2^K|U_2^N)$:

$$I(U_1^N; V_1^KV_2^K|U_2^N) = \sum_{n=1}^{N} I(U_{1n}; V_1^KV_2^K|U_2^NU_1^{n-1})$$

$$= \sum_{n=1}^{N} H(U_{1n}|U_2^NU_1^{n-1}) - H(U_{1n}|V_1^KV_2^KU_2^NU_1^{n-1})$$

$$= \sum_{n=1}^{N} H(U_{1n}|U_2n) - H(U_{1n}|V_1^KV_2^KU_2n)$$

$$= \sum_{n=1}^{N} I(U_{1n}; V_1^KV_2^K|U_2n)$$

$$= \sum_{n=1}^{N} I(U_{1n}; Z_n|U_2n),$$

where we set $Z_n = V_1^KV_2^K$. Now, we simplify $I(U_1^N; Y_1^NY_2^N|U_2^NV_1^KV_2^K)$:

$$I(U_1^N; Y_1^NY_2^N|U_2^NV_1^KV_2^K)$$

$$= \sum_{n=1}^{N} I(U_1^N; Y_1nY_2n|U_2^NV_1^KV_2^KY_1^{n-1}Y_2^{n-1})$$

$$= (a) \sum_{n=1}^{N} I(U_1^N; X_{1n}; Y_1nY_2n|U_2^NV_1^KV_2^KY_1^{n-1}Y_2^{n-1}X_{2n})$$

$$= \sum_{n=1}^{N} H(Y_1nY_2n|U_2^NV_1^KV_2^KY_1^{n-1}Y_2^{n-1}X_{2n})$$

$$- H(Y_1nY_2n|U_2^NV_1^KV_2^KY_1^{n-1}Y_2^{n-1}X_{2n}U_1^NX_{1n})$$

$$\leq \sum_{n=1}^{N} H(Y_1nY_2n|U_2^NV_1^KV_2^KX_{2n})$$

$$- H(Y_1nY_2n|U_2^NV_1^KV_2^KY_1^{n-1}Y_2^{n-1}X_{2n}U_1^NX_{1n})$$

$$= (b) \sum_{n=1}^{N} H(Y_1nY_2n|U_2^NV_1^KV_2^KX_{2n}) - H(Y_1nY_2n|U_2^NV_1^KV_2^KX_{2n}U_1^NX_{1n})$$

$$= \sum_{n=1}^{N} I(U_1^N; X_{1n}; Y_1nY_2n|U_2^NV_1^KV_2^KX_{2n})$$

$$= \sum_{n=1}^{N} I(U_1^N; X_{1n}; Y_1n|U_2^NV_1^KV_2^KX_{2n}) + I(U_1^N; Y_2n|U_2^NV_1^KV_2^KX_{2n}Y_{1n})$$

$$= 0$$
or equivalently, 

\[
\sum_{n=1}^{N} I(U_1^N X_{1n}; Y_{1n} | U_2^N V_1^K V_2^K X_{2n})
\]

\[
= \sum_{n=1}^{N} H(Y_{1n} | U_2^N V_1^K V_2^K X_{2n}) - H(Y_{1n} | U_2^N V_1^K V_2^K X_{2n} U_1^N X_{1n})
\]

\[
= \sum_{n=1}^{N} H(Y_{1n} | U_2^N V_1^K V_2^K X_{2n}) - H(Y_{1n} | V_1^K V_2^K X_{1n})
\]

\[
\leq \sum_{n=1}^{N} H(Y_{1n}) - H(Y_{1n} | V_1^K V_2^K X_{1n})
\]

\[
= \sum_{n=1}^{N} I(Y_{1n}; X_{1n})
\]

where: (a) follows from the fact that \((U_1^N, V_1^K V_2^K) \sim X_1^N \sim X_{1n}\) and \((U_2^N, V_1^K) \sim X_2^N \sim X_{2n}\); (b) follows from the fact that the channels are DMCs, so given \(X_1, X_2\), \(U_1, U_2\) are independent of anything else, so we can drop conditioning terms without changing the entropy; (c) follows from the long chains; and (d) follows from the fact that \((V_1^K, V_2^K)\) are independent of \(Y_{1n}\) given \(X_{1n}\).

Combining all of the above, we get that

\[
H(U_1^N) \leq I(U_1^N; U_2^N) + \sum_{n=1}^{N} I(X_{1n}; Y_{1n}) + \sum_{n=1}^{N} I(U_1^N; Z_{1n} | U_2^N) + N\delta(P_e),
\]

or equivalently,

\[
\frac{1}{N} H(U_1^N | U_2^N) \leq \frac{1}{N} \sum_{n=1}^{N} I(X_{1n}; Y_{1n}) + \frac{1}{N} \sum_{n=1}^{N} I(U_1^N; Z_{1n} | U_2^N) + \delta(P_e).
\]

Symmetric arguments yield

\[
\frac{1}{N} H(U_2^N | U_1^N) \leq \frac{1}{N} \sum_{n=1}^{N} I(X_{2n}; Y_{2n}) + \frac{1}{N} \sum_{n=1}^{N} I(U_2^N; Z_{2n} | U_1^N) + \delta(P_e).
\]

3) The Sum-Rate Face: Necessity of Equation (11)

Again, we start by bounding \(H(U_1^N U_2^N)\):

\[
H(U_1^N U_2^N) = I(U_1^N U_2^N; Y_1^N Y_2^N) + H(U_1^N U_2^N | Y_1^N Y_2^N)
\]

\[
\leq I(U_1^N U_2^N; Y_1^N Y_2^N) + N\delta(P_e).
\]
Now we need to simplify \( I(U_1^N U_2^N; Y_1^N Y_2^N) \). Here we make use of the fact that, from the long chains, it follows that \( U_1^N U_2^N - X_1^N X_2^N - X_{1n} X_{2n} - Y_{1n} Y_{2n} \) also forms a Markov chain. So,

\[
I(U_1^N U_2^N; Y_1^N Y_2^N) \nonumber
\]

\[
= \sum_{n=1}^{N} I(U_1^N U_2^N; Y_{1n} Y_{2n} | Y_1^{n-1} Y_2^{n-1}) 
\]

\[
\leq \sum_{n=1}^{N} I(U_1^N U_2^N; Y_{1n} Y_{2n} | Y_1^{n-1} Y_2^{n-1}) + I(X_{1n} X_{2n}; Y_{1n} Y_{2n} | Y_1^{n-1} Y_2^{n-1} U_1^N U_2^N) 
\]

\[
= \sum_{n=1}^{N} I(U_1^N U_2^N X_{1n} X_{2n}; Y_{1n} Y_{2n} | Y_1^{n-1} Y_2^{n-1}) 
\]

\[
= \sum_{n=1}^{N} I(X_{1n} X_{2n}; Y_{1n} Y_{2n} | Y_1^{n-1} Y_2^{n-1}) + I(U_1^N U_2^N X_{1n} X_{2n}; Y_{1n} Y_{2n} | Y_1^{n-1} Y_2^{n-1} X_{1n} X_{2n}) \nonumber
\]

\[
= \sum_{n=1}^{N} I(X_{1n} X_{2n}; Y_{1n} Y_{2n} | Y_1^{n-1} Y_2^{n-1}) \nonumber
\]

\[
= \sum_{n=1}^{N} H(Y_{1n} Y_{2n} | Y_1^{n-1} Y_2^{n-1}) - H(Y_{1n} Y_{2n} | Y_1^{n-1} Y_2^{n-1} X_{1n} X_{2n}) \nonumber
\]

\[
= \sum_{n=1}^{N} H(Y_{1n} Y_{2n}) - H(Y_{1n} Y_{2n} | X_{1n} X_{2n}) \nonumber
\]

\[
\leq \sum_{n=1}^{N} H(Y_{1n} Y_{2n}) + H(Y_{2n}) - H(Y_{1n} Y_{2n} | X_{1n} X_{2n}) \nonumber
\]

\[
= \sum_{n=1}^{N} H(Y_{1n}) + H(Y_{2n}) - H(Y_{1n} | X_{1n}) - H(Y_{2n} | X_{2n}) \nonumber
\]

\[
= \sum_{n=1}^{N} I(X_{1n}; Y_{1n}) + \sum_{n=1}^{N} I(X_{2n}; Y_{2n}). \quad (32) \nonumber
\]

where: (a) follows from the chain above, and (b) follows from the DMC property.

Therefore, we get that

\[
\frac{1}{N} H(U_1^N U_2^N) \leq \frac{1}{N} \sum_{n=1}^{N} I(X_{1n}; Y_{1n}) + \frac{1}{N} \sum_{n=1}^{N} I(X_{2n}; Y_{2n}) + \delta(P_e). \nonumber
\]
4) Conditions on the Conference Rates: We now obtain necessary conditions for an admissible conference in terms of the auxiliary random variable $Z$:

\[
NC_{12} \geq \sum_{k=1}^{K} \log |V_{1k}|
\]

\[
\geq H(V_1^K)
\]

\[
\geq H(V_1^K | U_2^N)
\]

\[
= H(V_1^K V_2^K | U_2^N)
\]

\[
= H(V_1^K V_2^K | U_2^N) - H(V_1^K V_2^K | U_1^N U_2^N)
\]

\[
= I(V_1^K V_2^K; U_1^N | U_2^N)
\]

\[
= \sum_{n=1}^{N} I(V_1^K V_2^K; U_{1n} | U_1^{n-1} U_2^N)
\]

\[
= \sum_{n=1}^{N} H(U_{1n} | U_1^{n-1} U_2^N) - H(U_{1n} | V_1^K V_2^K U_1^{n-1} U_2^N)
\]

\[
= \sum_{n=1}^{N} H(U_{1n} | U_2^n) - H(U_{1n} | V_1^K V_2^K U_2^n)
\]

\[
= \sum_{n=1}^{N} I(U_{1n}; V_1^K V_2^K | U_2^n)
\]

\[
= \sum_{n=1}^{N} I(U_{1n}; Z_n | U_2^n),
\]

where the first inequality is due to the admissibility condition for conferences, and the rest are standard information theoretic manipulations. Taking the second inequality, a similar argument yields

\[
NC_{21} \geq \sum_{n=1}^{N} I(U_{2n}; Z_n | U_{1n}).
\]

Thus, the conditions on the conference rates become

\[
\frac{1}{N} \sum_{n=1}^{N} I(U_{1n}; Z_n | U_{2n}) \leq C_{12} \quad \text{and} \quad \frac{1}{N} \sum_{n=1}^{N} I(U_{2n}; Z_n | U_{1n}) \leq C_{21}
\]
5) **Final Remarks:** So far, we have established that

\[
\frac{1}{N} H(U_1^N | U_2^N) \leq \frac{1}{N} \sum_{n=1}^{N} I(X_{1n}; Y_{1n}) + \frac{1}{N} \sum_{n=1}^{N} I(U_{1n}; Z_n | U_{2n}) + \delta(P_e)
\]

\[
\frac{1}{N} H(U_2^N | U_1^N) \leq \frac{1}{N} \sum_{n=1}^{N} I(X_{2n}; Y_{2n}) + \frac{1}{N} \sum_{n=1}^{N} I(U_{2n}; Z_n | U_{1n}) + \delta(P_e)
\]

\[
\frac{1}{N} H(U_1^N U_2^N) \leq \frac{1}{N} \sum_{n=1}^{N} I(X_{1n}; Y_{1n}) + \frac{1}{N} \sum_{n=1}^{N} I(X_{2n}; Y_{2n}) + \delta(P_e),
\]

and also that

\[
\frac{1}{N} \sum_{n=1}^{N} I(U_{1n}; Z_n | U_{2n}) \leq C_{12} \quad \text{and} \quad \frac{1}{N} \sum_{n=1}^{N} I(U_{2n}; Z_n | U_{1n}) \leq C_{21}.
\]

Now, from here to the exact form of the conditions in the theorem there is a very short way. First, note that using the standard technique of introducing time-sharing variables (see, e.g., [11, pg. 435]), we can replace the averages above by variables with the exact same distribution as prescribed by Theorem 3. Note also that by its own definition, \( \delta(P_e) \to 0 \) as \( P_e \to 0 \). Finally, note from the achievability proof that \( |Z| \leq |\mathcal{U}_1| \cdot |\mathcal{U}_2| < \infty \) (since \( Z \) is made up of partitions of \( \mathcal{U}_1 \) and \( \mathcal{U}_2 \)). This concludes the proof of Theorem 3. \( \blacksquare \)

**C. Converse Proof for Theorem 7**

The proof uses the same arguments as the converse proof of Theorem 3, therefore we include here only the main steps.

1) **Preliminaries:**

Assume that there exist codes with parameters \((R_1, \ldots, R_M, R_{12}, \ldots, R_{M-1,M}, N, K, P_e)\). Let \( Z_{ij} = (V^K_{ij} V^K_{ji}) \), where \( V^K_{ij} \) denotes the block of messages sent by encoder \( i \) to encoder \( j \). Based on Fano’s inequality, we can write:

\[
H(U_1^N \ldots U_M^N | \hat{U}_1^N \ldots \hat{U}_M^N) \leq P_e \log (|\mathcal{U}_1^N \times \cdots \times \mathcal{U}_M^N|) + h(P_e),
\]

where \( h(P_e) \) is the binary entropy function. Define

\[
\delta(P_e) = \left( P_e \log (|\mathcal{U}_1^N \times \cdots \times \mathcal{U}_M^N|) + h(P_e) \right) / N.
\]

It follows from eqn. (33) that

\[
H(U_1^N \ldots U_M^N Y_1^N \ldots Y_M^N) \leq N \delta(P_e),
\]

11/17/2003
and consequently,

\[
H(U^N(S)|Y_1^N \ldots Y_M^N U^N(S^c) Z(S)) \leq N\delta(P_e)
\]  

(34)

for all subsets \(S \subseteq \{1, 2, \ldots, M\}\) with \(U^N(S) = \{U_j^N : j \in S\}\), and \(Z(S) = \{Z_{ij} : i \in S \text{ or } j \in S\}\).

2) Main Arguments:

The following inequality is true:

\[
H(U^N(S)) = I(U^N(S); Y_1^N \ldots Y_M^N Z(S) U^N(S^c)) + H(U^N(S)|Y_1^N \ldots Y_M^N Z(S) U^N(S^c))
\]

\[
\leq I(U^N(S); U^N(S^c)) + I(U^N(S); Z(S)|U^N(S^c)) + I(U^N(S); Y_1^N \ldots Y_M^N|Z(S) U^N(S^c)) + N\delta.
\]  

(35)

We can now develop each of the mutual information terms on the right-hand side of this inequality to obtain single-letter expressions. The first term can be subtracted on both sides, yielding \(H(U^N(S)|U^N(S^c))\) on the left-hand side of (35), which can be shown to be equal to \(NH(U(S)|U(S^c))\) by arguing that the sources are memoryless and using a standard time-sharing argument. Similarly, using standard information-theoretic identities and inequalities to develop the second term we get

\[
I(U^N(S); Z(S)|U^N(S^c)) \leq NI(U(S); Z(S)|U(S^c)).
\]

Finally, for the third term we obtain

\[
I(U^N(S); Y_1^N \ldots Y_M^N|Z(S) U^N(S^c)) \leq N \sum_{i \in S} I(X_i; Y_i)
\]

repeating the steps of (31) and (32), and using the aforementioned time-sharing argument. 

\[\square\]

REFERENCES


