Deterministic Capacity of Networks in the Low Power Regime

Anders Høst-Madsen

Abstract

The deterministic capacity of a network is the capacity of a network when nodes are restricted to transmitting reliable information, that is, (asymptotically) deterministic functions of the source messages. In this paper it is shown that the deterministic capacity of a number of networks can be found in the low power regime where $\text{SNR} \to 0$. This is accomplished through a new technique for deriving outer bounds in the limit, without first finding bounds for general SNR.

Index Terms

Relay, Low Power, Wideband, Network Information Theory.

I. INTRODUCTION

Recently there has been a large renewed interest in analyzing capacity of networks, in particular wireless networks. It has been found that the capacity of wireless networks can be increased by using the fact that the wireless signal propagates widely (the multicast advantage) and letting nodes cooperate (cooperative diversity) [1]. Coding methods for networks can generally be divided into two classes: those where relays process the received signal and forwards it, and those where the relays decode (a function of) the original message, and encodes this into a new signal. In the first class (some times denoted estimate-forward) are methods such as amplify-forward and compress-forward [2], [3] that have wide set of generalizations (including mixed schemes such as Theorem 7 in [4]). The second class (sometime denoted regenerative coding) has as its source the original decode-forward strategy of [4]. For a single antenna single relay channel [4]'s original strategy appears to be the only member of this class. However, for multiple antenna relay channels [5], [6] and multi-node networks [7], [8] there are many possible generalizations. What characterizes these methods, as opposed to the first class, is that relays decode the original message, or more generally, a function of the original message, reliably (which also relates it to [9]), and transmits a message which is a possibly different function of the decoded message (e.g., in [8] the parity information). One way to characterize this class is that the transmission is reliable. Relays decode their messages with a vanishing error probability, and base their transmission on deterministic functions\(^1\) of the messages. We will therefore denote this type of coding reliable coding. In contrast, amplify-forward type methods introduces further randomness through the noise at the relays. Informally one could say that amplify-forward type methods introduces errors in their transmission streams, while reliable transmission eliminates errors.

\(^1\)Deterministic in the sense that the functions do no depend on the noise realization in the network. Random encoding can be considered deterministic as nodes can share a common random number generator.
The aim of this paper is to bound the capacity that can be achieved with reliable transmission. Why reliable transmission? One way to think of reliable transmission is as follows: nodes decode transmitted packages, with few errors. They then compute functions of these packages, and transmit these new computed packages. In a traditional multi-hop network, the functions are the identity functions. With network coding, the functions operate only on the contents of the packages. With reliable transmission, the functions operate on both contents of packets and channel coding parts. In all cases, a big advantage is that the layering of the network according to the OSI 7-layer model [10] is maintained, even if some crosslayer interfacing is needed. On the other hand, general network information theory methods may require completely scrapping the separation between the lower 3-4 layers. While that may not seem like a high price to an information theorist, in practical systems it is very disruptive. It might be worth having that disruption; but to argue that we should be able to quantify the gain achieved by non-reliable transmission. Thus, we need to be able to bound the rate achievable by reliable transmission, even if we want to consider non-reliable transmission.

To be able to bound the rate achievable by reliable transmission, we need an abstract definition of reliable transmission, and we need to show that bounds are tight. This paper provides both. We define deterministic capacity, which is an abstraction of reliable transmission, derive outer bounds for deterministic capacity, and show that these bounds are tight for some networks in the low power regime.

The rest of the paper is organized as follows. First we need a precise definition of what is meant by reliable transmission; this is provided in Section II. In Section III some results about the low power regime, \( \text{SNR} \to 0 \), are derived. As the main part, the capacity for reliable transmission for a number of example networks are found in Sections IV and V. The intention is not to provide an encyclopedic listing of every single network and condition under which we currently know the capacity of reliable transmission (e.g., some results could be generalized to \( N \) antennas from two antennas), but to introduce examples that each showcases properties and techniques of reliable transmission, without making the paper overly technical.

We will use the following conventions in the paper. All logarithms are natural logarithms, and all rates are measured in terms of nats. We use \( \epsilon_n \) to denote any sequence satisfying

\[
\lim_{n \to \infty} \epsilon_n = 0
\]

II. DEFINITIONS AND INITIAL REMARKS

We consider a network with \( N \) nodes as in [11] Sec. 15.10; each node may have multiple antennas. We denote the transmitted symbol (which might be a vector) at time \( m \) at node \( i \) by \( X_{ij}[m] \), the transmitted symbol from the \( j \)-th antenna by \( X_{ij}[m] \), and the string of transmitted symbols in the interval \( 1 \ldots k \) by \( X_{ij}[k] = [X_{ij}[1], X_{ij}[2], \ldots, X_{ij}[k]] \). Similarly for the received signals \( Y_{ij}[n] \) and \( Y_{ij}[n] \) at node \( i \). The output alphabet at node \( i \) is \( \mathcal{X}_i \), the input alphabet \( \mathcal{Y}_i \), and the nodes are connected through memoryless channels. For simplicity we at first consider a single flow network, that is, a relay network, with a single source and destination. A generalization to multiple flows can be found in Section V. A (length \( n \)) code for the network is defined as in [11] Sec. 15.10: Node 1, the source has a message \( W \) intended for node \( N \) that it transmits at rate \( R \); we consider the message a uniform random variable over \( \{1, 2, \ldots, 2^{nR}\} \). The encoder at node 1 is a function \( X_1[n] : \{1, 2, \ldots, 2^{nR}\} \to \mathcal{X}_1^n \) and the transmission at node 1 is \( X_1[n](W) \). At node \( i > 1 \) the encoder is a function \( X_i[m] : \mathcal{Y}_i^{m-1} \to \mathcal{X}_i \) that depends only on past received symbols, that is the transmission is \( X_i[m](Y_i[1] \ldots Y_i[m-1]), m \in \{1, \ldots, n\} \). The decoder at node \( N \) is a function \( \hat{W} : \mathcal{X}_N^n \to \{1, 2, \ldots, 2^{nR}\} \), and the performance is measured by the average error probability

\[
P_e = P(\hat{W} \neq W).
\]
Thus far the setup is exactly like in [11] Sec. 15.10. We now deviate by introducing a corresponding deterministic code:

\[ \hat{X}_i[n] : \{1, 2, \ldots, 2^{nR}\} \to \mathcal{X}_i^n \]  \hspace{1cm} (2)

and making the following definition

**Definition 1.** A rate \( R \) is said to be achievable by deterministic codes if there exists a sequence of \( (2^{nR}, n) \) codes \( \{X_i[n], i = 1 \ldots N\} \) and a sequence of corresponding deterministic codes \( \{\hat{X}_i[n], i = 1 \ldots N\} \), so that

\[
\forall i \in \{1, \ldots, N\} \lim_{n \to \infty} P\{\hat{X}_i[n](W) \neq X_i[n](Y_i[n])\} = 0 \hspace{1cm} (3)
\]

\[
\lim_{n \to \infty} P\{W(Y_N[n]) \neq W\} = 0. \hspace{1cm} (4)
\]

The deterministic capacity is the supremum of all rates \( R \) that is achievable by deterministic codes. We say that a sequence of codes \( X_i[n] \) is reliable if there exist a corresponding sequence of deterministic codes \( \hat{X}_i[n] \) so that (3) is satisfied.

Let us consider the definition from the perspective of the original relay paper [4]. The block-Markov decode-forward scheme of [4] Theorem 1] is achievable by deterministic codes (each block is decoded with vanishing error probability as the block length increase), whereas the compress-forward schemes of Theorems 6 and 7 are not, as they forward the unreliable information \( \hat{Y}_1 \). The question we seek to answer is: is the coding scheme of [4, Theorem 1] best possible among “decode-forward type” schemes? The answer is not totally obvious. Could the relay perhaps decode part of the message instead of the full message? Just considering achievable rates the question is not quite well-posed as there are infinitely many possible ways the relay could process the received data. However, Definition 1 can give a definitive and concise answer. The received signals are

\[
Y_2[n] = c_{21}X_1[n] + Z_2[n] \hspace{1cm} (5)
\]

\[
Y_3[n] = c_{31}X_1[n] + c_{32}X_2[n] + Z_3[n], \hspace{1cm} (6)
\]

where \( Z_i[n] \) is iid circular complex Gaussian noise of power \( N_0B \). The nodes are subject to power constraints \( P_1, P_2 \). Suppose \( |c_{21}| \geq |c_{31}| \). Then node 2 can form

\[
Y'_3[n] = \frac{c_{31}}{c_{21}}Y_2[n] + c_{32}X_2[n] + Z'_3[n] = c_{31}X_1[n] + c_{32}X_2[n] + \frac{c_{31}}{c_{21}}Z_2[n] + Z'_3[n], \hspace{1cm} (7)
\]

where \( Z'_3[n] \) is iid circular complex Gaussian noise with power \( \left(1 - \frac{|c_{31}|^2}{|c_{21}|^2}\right) N_0B \). Now consider the two companion signals

\[
\hat{Y}_3[n] = c_{31}X_1[n] + c_{32}X_2[n] + Z_3[n] \hspace{1cm} (8)
\]

\[
\hat{Y}'_3[n] = \frac{c_{31}}{c_{21}}Y_2[n] + c_{32}X_2[n] + Z'_3[n]. \hspace{1cm} (9)
\]

By assumption node 3 can decode \( W \) with small probability of error for large \( n \). Since we consider deterministic capacity, we know that \( Y_3[n] = \hat{Y}_3[n] \) with high probability for large \( n \). A genie-aided node knowing \( \hat{Y}_3[n] \) therefore also can decode \( W \) with small error probability (formally, the genie-aided node’s error probability is bounded by \( \hat{P}_e^{(n)} \leq P_e^{(n)} + P(Y_3[n] \neq \hat{Y}_3[n]) \)). Now, because \( \hat{X}_2[n] \) is a deterministic function of \( W, \hat{Y}_3[n] \) and
\[ \hat{Y}_3'[n] \] have the same distribution. Thus, a genie-aided node knowing \( \hat{Y}_3'[n] \) can also decode \( W \) with small error probability. Finally, since \( \bar{Y}_3'[n] = \hat{Y}_3'[n] \) with high probability, a node knowing \( \bar{Y}_3'[n] \) can also decode \( W \) with small error probability. Thus, node 2 can decode \( W \). A similar argument shows that for \( |c_{21}| < |c_{31}| \) it does not help the destination to know \( \bar{Y}_3[n] \). Therefore, the (reliable) rate is bounded by

\[
R \leq \max \{ I(X_1; Y_2 | X_2), I(X_1; Y_3 | X_2) \} \tag{10}
\]

\[
R \leq I(X_1, X_2; Y_3), \tag{11}
\]

where (11) is the MAC bound. On the other hand, the coding scheme of [4, Theorem 1] achieves this outer bound, and the answer therefore is: yes, [4, Theorem 1] is optimum among all “decode-forward type” coding schemes.

The definition of deterministic capacity is abstract and very general. The essence is that it allows usage of traditional methods of information theory. Equation (3) essentially says that node \( i \) should be able to decode the function \( \bar{X}_n[i] \). One can then use Fano’s inequality to outer bound the rate region. If there exist (decode-forward type) coding methods achieving this outer bound, this is the deterministic capacity.

The definition of deterministic capacity is abstract and very general. The essence is that it allows arbitrary computations on the data; general capacity allows arbitrary computations on the signals. In this way it is related to the computation rate in [9]. It excludes, purposefully, methods such as amplify-forward [11] and compress-forward [4, 2, 3], but allows a wide class of “decode-forward” type schemes such as partial-decode forward and parity forwarding [6]. It also, perhaps more interesting, includes networking coding [12] for multflow networks. But as we shall see in Section V-A, network coding is not sufficient to achieve deterministic capacity. The class of methods allowed by the definition of deterministic capacity is larger. In fact, one way to think of deterministic capacity is as fundamental limit for network coding at the channel coding layer, such as in [13, 14].

While Definition 1 applies to general channel models, we will restrict attention to Gaussian channels as they are representative of wireless networks. The received signal is subject to additive complex Gaussian noise of power \( B N_0 \), where \( N_0 \) is the noise power spectral density, and \( B \) is the bandwidth. The complex channel gain from node \( i \) to node \( j \) is \( c_{ij} \), and if node \( i \) has more than one antenna \( c_{ijk} \); we also define \( c_{ij} = [c_{ijk} \ldots c_{ijkN}] \). Each node is subject to a power constraint

\[
\frac{1}{n} \sum_{i=1}^{n} |X_i[n]|^2 \leq P_i; \tag{12}
\]

In many cases we will consider a total power constraint \( \sum_{i=1}^{N} P_i \leq P \).

We consider a number of different type of channels with corresponding differing channel state information (CSI):

\(^2\) whereas \( Y_3[n] \) and \( \bar{Y}_3'[n] \) do not necessarily have the same distribution because \( X_3[n] \) and \( Z_3[n] \) could be dependent.

\(^3\) The conditioning on \( X_2 \) enters the same way as the proof of Theorem 15.10.1 in [10]; for details see the proof of Theorem 6.
• **Static channel (complete CSI):** all nodes are assumed to have full channel state information, i.e., to know perfectly all $c_{ijk}$, and capacity is defined for a fixed set of $c_{ijk}$.

• **Fading channel (receiver CSI):** the coefficients $c_{ijk}$ are assumed to be independent random with $E[c_{ijk}] = 0$, $E[|c_{ijk}|^2] < \infty$. Receivers know $c_{ijk}$, but transmitters only know $E[|c_{ijk}|^2]$ (for all $i, j, k$). In this case we calculate *ergodic capacity*. There are two special cases:
  
  - **Scaled fading:** all channel coefficients have the same distribution except for scaling. An example is if all channels are Rayleigh fading with different average gain.
  
  - **Phase fading:** all nodes know all $|c_{ijk}|$, whereas the phase of $c_{ijk}$ is unknown to transmitters, but known at receivers. The phase of $c_{ijk}$ is assumed to vary ergodic during transmission. This can be used to model nodes that do not have synchronized local oscillators.

  Of course, the phase fading case is a special case of the fading case with scaled channels, and all results we currently have apply to at least the scaled channel case. The reason we extract phase fading as a separate case above is that it a good model of many real wireless networks. However, in the rest of the paper we will consider only the complete CSI and the (scaled) fading cases, with all results for the scaled fading case applying to the phase fading case.

We notice that in the fading case, the deterministic condition imposes a rather strict condition on transmission. Essentially, a code is only reliable if it does not depend on the fading realization.

Definition 1 for Gaussian networks applies to general SNR – the example above with the single relay is valid for any SNR. However, it is very hard to generalize this example to larger networks. The reason is that it is difficult to obtain useful outer bounds. The rest of the paper is therefore solely about the low power regime where SNR → 0. We develop a new methodology (a generalization of techniques used in [16]) for finding outer bounds in the low power regime directly, without first finding bounds for general SNR and taking limits. This technique allows us to find the deterministic capacity of a number of channels.

### III. The Low Power Regime

The capacity of the channel depends on the bandwidth as follows [11]: Fix $P$ (in Watts) and let the available bandwidth be $B$ (in Hz). The available power per sample is then $P/(2B)$ and the noise variance per sample $N_0/2$. If we denote by $C(B)$ the capacity (or spectral efficiency, [17]) in nats/s/Hz for a given bandwidth, we can define the following limit (if it exists)

$$ C = \lim_{B \to \infty} BC(B), $$

which is the limit of the capacity in (nats/s) · (Wats/Hz) for infinite bandwidth; we have multiplied with $N_0$ to simplify formulas in the following. If the actual numerical rate in nats/s is of interest, just multiply capacity results with the numerical value of $N_0$. We call the infinite bandwidth limit the low power regime; this has been considered in many papers, with the two papers [17], [18] key papers. Signaling in the low power regime has a number of advantages: robustness to interference, little interference generation, covertness, etc., and is the principle behind UWB. For a point-to-point channel it is also the most energy efficient signaling. For multi-terminal channels it is not clear if this is still true, see e.g., [19].

The low power regime also has the theoretical advantage that $C$ may be calculated without having explicit expressions for $C(B)$ using the techniques in [18] combined with the further results in [20], as we will see in the following.

We will denote rates in the low power regime by sans serif, i.e., if $R \leq C$ we say that the rate $R$ (in (nats/s) · (Wats/Hz)) is achievable. Similarly, if $R \leq C(B)$, we say that the rate $R$ (in nats/s/Hz) is achievable.
We need the following generalizations of results in [20]. This is also a generalization of results in section V.B of [16].

**Lemma 2.** Suppose that for each value of $B$ a random ($N$-vector) random variable $X(B)$ that satisfy $\text{var}[X(B)] \leq P$ is given. Let $Y = c^H X(B) + Z$, where $Z \sim N(0, N_0 B)$. If $c$ is a constant vector

$$
\lim_{B \to \infty} BI(X(B); Y) = \lim_{B \to \infty} \frac{\text{var}[c^H X(B)]}{N_0}.
$$

(14)

In the fading case

$$
\lim_{B \to \infty} BI(X(B); Y|c) = \lim_{B \to \infty} \sum_{i=1}^{N} E[|c_i|^2 \text{var}[X_i(B)]]/N_0.
$$

(15)

**Proof:** The proof follows quite closely the proof of Lemma 1 in [20]. The result is also mentioned under note 4 in [16], but not explicitly proven there; in fact, [16] has a fourth order moment condition, which is not needed with the proof technique of [20]. For completeness we will therefore provide the proof in the fading case. We can assume that $X(B)$ has zero mean, as the mean will not influence the mutual information. Put $\tilde{Y} \sim N\left(0, \sum_{i=1}^{N} E[|c_i|^2 \text{var}[X_i(B)]] + N_0 B\right)$, and write

$$
I(X(B); Y|c) = D\left(P_{Y|X(B)}||P_{\tilde{Y}}|P_{X(B)}, P_c\right) - D\left(P_{Y||P_{\tilde{Y}}|P_c}\right).
$$

(16)

The first term is

$$
D\left(P_{Y|X(B)}||P_{\tilde{Y}}|P_{X(B), P_c}\right) = \int D\left(P_{Y|X=B,c}||P_{\tilde{Y}}\right) dP_{X(B)}dP_c
$$

$$
= \int \log \left(\sum_{i=1}^{N} |c_i|^2 \text{var}[X_i(B)] + N_0 B\right) dP_{X(B)}dP_c
$$

$$
+ \int \sum_{i=1}^{N} \frac{|c_i|^2 \text{var}[X_i(B)]}{N_0 B} \sum_{j=1}^{N} \frac{|c_j|^2 \text{var}[X_j(B)]}{N_0 B} - \frac{1}{N_0 B} dP_{X(B)}dP_c
$$

$$
= \log \left(1 + \sum_{i=1}^{N} E[|c_i|^2 \text{var}[X_i(B)]]/N_0 B\right),
$$

(18)

since

$$
\int |c^H X(B)|^2 dP_{X(B)}dP_c = \int c^H E\left[X(B)X(B)^H\right] c dP_c
$$

$$
= \sum_{i=1}^{N} E[|c_i|^2 \text{var}[X_i(B)]].
$$

(21)

$^4\text{var}X(B) = \text{tr}E\left[X(B)X(B)^H\right]
Thus

\[
\lim_{B \to \infty} BD \left( P_{Y|X(B)} \| P_Y | P_{X(B)}, P_c \right)
= \lim_{B \to \infty} \frac{\sum_{i=1}^{N} E[|c_i|^2 \var[X_i(B)]]}{N_0}.
\] (23)

The second term in (16) satisfies \(\lim_{B \to \infty} BD \left( P_Y | P_{Y|X}, P_c \right) = 0\), which can be proven as follows. Fix \(c\). Then

\[
\log \frac{P_Y(y)}{P_{\tilde{Y}}(y)}
= \log \left( \frac{1}{\pi N_0 B} E \left[ \exp \left( - \frac{1}{N_0 B} |y - c^H X|^2 \right) \right] \right)
- \log \left( \frac{1}{\pi \left( \sum_{i=1}^{N} |c_i|^2 \var[X_i(B)] + N_0 B \right)} \exp \left( - \frac{1}{\sum_{i=1}^{N} |c_i|^2 \var[X_i(B)] + N_0 B} |y|^2 \right) \right)
= \log \left( \frac{\sum_{i=1}^{N} |c_i|^2 \var[X_i(B)] + N_0 B}{N_0 B} \right)
- \log E \left[ \exp \left( \frac{1}{\sum_{i=1}^{N} |c_i|^2 \var[X_i(B)] + N_0 B} |y|^2 \right) \right]
- \frac{1}{N_0 B} \left( |y|^2 - 2 \Re \{ y c^H X \} - |c^H X|^2 \right).
\] (24)

Using series expansion we then get

\[
\log \frac{P_Y(y)}{P_{\tilde{Y}}(y)}
= \log \left( \frac{\sum_{i=1}^{N} |c_i|^2 \var[X_i(B)] + N_0 B}{N_0 B} \right)
- \log E \left[ 1 + o \left( \frac{1}{B} \right) + \frac{1}{N_0 B} \left( 2 \Re \{ y c^H X \} - |c^H X|^2 \right) \right]
= \sum_{i=1}^{N} \frac{|c_i|^2 \var[X_i(B)]}{N_0 B} + o \left( \frac{1}{B} \right)
+ o \left( \frac{1}{B} \right) + E \left[ \frac{1}{N_0 B} \left( 2 \Re \{ y c^H X \} - |c^H X|^2 \right) \right],
\] (27)

where (28) uses the Lebesgue bounded convergence theorem [21], [22] to exchange limit and expectation: by

November 20, 2011 DRAFT
assumption $E \left[ |c^H X(B)|^2 \right] \leq E[|c|^2] P$. Then

$$E_c \log \frac{P_{X}(y)}{P_{\tilde{X}}(y)} = \sum_{i=1}^{N} \frac{|c_i|^2 \var[X_i(B)]}{N_0 B} + o \left( \frac{1}{B} \right)$$

$$+ E_{X,c} \left[ \frac{1}{N_0 B} \left( 2 \Re \{ yc^H X \} - |c^H X|^2 \right) \right]$$

$$= o \left( \frac{1}{B} \right), \quad (29)$$

where we have used (22), Lebesgue bounded convergence, and

$$\int \Re \{ yc^H X \} dP_{X(B)} dP_e = 0. \quad (31)$$

Lemma 3. Suppose that for each value of $B$ we are given random variables $U(B), V(B), \text{and a random } (N\text{-vector}) \text{ random variable } X(B)$ that satisfy $\var[X(B)] \leq P$. Define

$$Y_1 = c^H_1 X(B) + Z_1 \quad (32)$$

$$Y_2 = c^H_2 X(B) + Z_2, \quad (33)$$

where $Z_1$ and $Z_2$ are independent, $Z_1, Z_2 \sim \mathcal{N}(0, N_0 B)$. Suppose that $(U(B), V(B)) \rightarrow X(B) \rightarrow Y_1 \text{ and } (U(B), V(B)) \rightarrow X(B) \rightarrow Y_2 \text{ form Markov chains. If } c \text{ is a constant vector then}

$$\lim_{B \rightarrow \infty} BI(U(B); Y_1)$$

$$= \lim_{B \rightarrow \infty} \frac{\var[c^H_1 X(B)]}{N_0}$$

$$- \lim_{B \rightarrow \infty} \frac{\var[c^H_1 X(B)|U(B)]}{N_0}$$

$$\lim_{B \rightarrow \infty} BI(X(B); Y_2|U(B))$$

$$= \lim_{B \rightarrow \infty} \frac{\var[c^H_2 X(B)|U(B)]}{N_0} \quad (34)$$

$$\lim_{B \rightarrow \infty} BI(U(B); Y_1|V(B))$$

$$= \lim_{B \rightarrow \infty} \frac{\var[c^H_1 X(B)|V(B)]}{N_0}$$

$$- \lim_{B \rightarrow \infty} \frac{\var[c^H_1 X(B)|U(B), V(B)]}{N_0}$$

$$\lim_{B \rightarrow \infty} BI(U(B); Y_2|V(B))$$

$$= \lim_{B \rightarrow \infty} \frac{\var[c^H_2 X(B)|V(B)]}{N_0} \quad (35)$$

$$- \lim_{B \rightarrow \infty} \frac{\var[c^H_2 X(B)|U(B), V(B)]}{N_0}$$

$$\lim_{B \rightarrow \infty} BI(U(B); Y_1|U(B), V(B))$$

$$= \lim_{B \rightarrow \infty} \frac{\var[c^H_1 X(B)|U(B), V(B)]}{N_0} \quad (36)$$
In the fading case,

\[
\begin{align*}
\lim_{B \to \infty} BI(U(B); Y_1 | c) &= \lim_{B \to \infty} \sum_{i=1}^{N} \frac{E[|c_{1i}|^2] \text{var}[X_i(B)]}{N_0} \\
&- \lim_{B \to \infty} \sum_{i=1}^{N} \frac{E[|c_{1i}|^2] \text{var}[X_i(B) | U(B) | Y_1]}{N_0} \quad (37)
\end{align*}
\]

\[
\begin{align*}
\lim_{B \to \infty} BI(X(B); Y_2 | U(B), c) &= \lim_{B \to \infty} \sum_{i=1}^{N} \frac{E[|c_{2i}|^2] \text{var}[X_i(B) | U(B) | Y_2]}{N_0} \\
&- \lim_{B \to \infty} \sum_{i=1}^{N} \frac{E[|c_{1i}|^2] \text{var}[X_i(B) | U(B) | V(B) | Y_2]}{N_0} \quad (38)
\end{align*}
\]

\[
\begin{align*}
\lim_{B \to \infty} BI(U(B); Y_1 | V(B), c) &= \lim_{B \to \infty} \sum_{i=1}^{N} \frac{E[|c_{2i}|^2] \text{var}[X_i(B) | V(B)]}{N_0} \\
&- \lim_{B \to \infty} \sum_{i=1}^{N} \frac{E[|c_{1i}|^2] \text{var}[X_i(B) | V(B) | U(B)])}{N_0} \quad (39)
\end{align*}
\]

assuming all limits are defined.

**Proof:** For (35, 38), we use a conditional version of Lemma 2. For (34, 37) we write

\[
I(U(B); Y_1 | c) = I(X(B); Y_1 | U(B)) 
\]

(34, 37) using the Markov chain property. For (36, 39) we write

\[
I(U(B); Y_1 | V(B), c) = I(X(B); Y_1 | V(B)) - I(X(B); Y_1 | U(B), V(B)).
\]

IV. **Deterministic Capacity of Relay Networks in the Low Power Regime**

In this section we will determine the deterministic capacity of a number of relay networks in the low power regime. We will start out with some general definitions and tools.

Consider a network that uses the sequence of reliable codes \(X_i[n]\), that is, there exists a corresponding sequence of deterministic codes \(\hat{X}_i[n]\) satisfying \(3\). Define

\[
W_i = \hat{X}_i[n] 
\]

and notice that for deterministic capacity, node \(i\) must be able to decode at least \(W_i\) with asymptotically zero error probability. Specifically, we have the following statement of Fano’s inequality, proven similarly to Fano’s inequality in [11].

**Lemma 4** (Fano’s inequality). Suppose that the source message \(W \in \{1 \ldots 2^{nR}\}\). If a node uses the sequence of reliable codes \(X_i[n]\) the following inequality holds

\[
1 + Pr\{X_i[n] \neq \hat{X}_i[n]\} nR \geq H\left(\hat{X}_i[n] | Y_i[n]\right). \quad (41)
\]
In the following define

\[
R = \lim_{n \to \infty} \frac{H(W)}{n}
\]
\[
R_i = \lim_{n \to \infty} \frac{H(W_i)}{n}
\]
\[
R_{ij} = \lim_{n \to \infty} \frac{H(W_i | W_j)}{n}
\]
\[
R_{ij} = \lim_{n \to \infty} \frac{H(W_j | W_i)}{n}.
\]

(42)

Conceptually, wireless networks are built of broadcast and MAC channels – since we consider the low power regime, interference is not a factor. Understanding these channels would get us a long way towards determining deterministic capacity. Unfortunately, the general capacity region of broadcast channels is not known. The best outer bounds are those of Marton [23] and Nair and El Gamal [24]. Most of our results are based on modifications of the results in [24]. Specifically, we will use the following generalization of Lemma 3.5 in [24].

**Proposition 5.** Consider a discrete memoryless broadcast channel with two receivers that need to decode the messages \(W_2\) and \(W_3\), which may be dependent. Rates are defined according to (42). The capacity region of the broadcast channel is contained in the convex closure of all \((R, R_2, R_3)\) that satisfy

\[
R_2 \leq I(U_2; Y_2)
\]
\[
R_3 \leq I(U_3; Y_3)
\]
\[
R = R_{23} + R_3 \leq I(U_2; Y_2 | U_3) + I(U_3; Y_3)
\]
\[
R = R_{32} + R_2 \leq I(U_3; Y_3 | U_2) + I(U_2; Y_2)
\]

for some joint distribution \(p(u_2, u_3)p(x|u_2, u_3)\).

The proof follows quite closely that of [24], so we will not provide it here. As in [24] cardinality bounds on the auxiliary random variables \(U_2, U_3\) can be made. We will use Proposition 5 for Gaussian channels. This can be done through discretizing the continuous channel as mentioned in [11].

### A. Two-relay single-antenna channels

In this section we consider a number of generalizations of the one-relay channel of [4] mentioned in Section II to two relays. The destination is not transmitting. The limitation to only two relays is because the results depend on bounds for the broadcast channel such as Proposition 5 that do not have known generalizations to more than two receivers.
Figure 2. Equivnt diamond relay channel.

At first we consider the diamond relay channel in Fig. 1 where the relays cannot communicate with each other. We consider the static channel case, as this is the more interesting case. Without loss of generality we can assume that $|c_{21}| \geq |c_{31}|$. For this channel we consider the following coding scheme, which we will call successive cooperation. We divide the transmission time into three intervals:

1) In the first time interval node 1 transmits $X_1$, so that node 2 can decode the message $W$.
2) In the second time interval node 2 transmits $X_2$ and node 1 transmits $X_1^3 = k_1 X_2$ (where $k_1$ is a complex constant), so that the signals from nodes 1 and 3 add up coherently at the destination, node 4. This is the same coding scheme as [4] for the one-relay channel.
3) Node 3 decodes the message $W$ from $Y_1$ and $Y_2$, the signals received during the first and second time interval.
4) In the third time interval node 3 transmits $X_3$ and nodes 1 and 2 transmit $X_1^3 = k_2 X_3$ and $X_2^3 = k_3 X_3$, so that all the signals add up coherently at the destination.
5) The destination node decodes the message $W$ from the signal received in all three time intervals.

What is curious about this transmission scheme is that node 2 helps node 3 decoding the message, although there is no direct connection! That is of course because node 1 transmits a signal identical to that of node 2. The scheme is described for the low power regime, where time and bandwidth are free. It can be used for general SNR, but should then be combined with block-Markov coding as in [4] to be time and bandwidth efficient.

It is not quite obvious that this should be the optimum decode-forward scheme for this channel, but the following theorem proves this to be the case if $|c_{31}|^2$ is large enough.

**Theorem 6.** Consider the diamond relay channel with a total power constraint. If

$$|c_{31}|^2 \geq \frac{|c_{42}|^2 |c_{43}|^2 (|c_{42}|^2 + |c_{43}|^2) + |c_{41}|^2 (4|c_{42}|^2 + |c_{43}|^2) + |c_{41}|^2 (4|c_{42}|^4 + 2|c_{42}|^2 |c_{43}|^2 + |c_{43}|^4)}{|c_{43}|^2 (|c_{41}|^2 + |c_{43}|^2)^2}$$

(47) successive cooperation achieves the deterministic capacity in the low power regime.

**Proof:** We can conceptually think of this transmission scheme as that of the broadcast channel in Fig. 2. We split nodes into a receiver part (r) and a transmitter part (t). While separate, a causality constraint between the reception at receiver nodes and the transmission at transmitter nodes should still be observed. For the channel in Fig. 2 we
have the bounds

\[ R \leq I(U; Y_4) \]
\[ R_3 \leq I(U_3; Y_3 | X_3) \]
\[ R \leq I(U; Y_4 | U_3, X_3) + I(U_3; Y_3 | X_3) \]
\[ R \leq I(X_1; Y_2 | X_2, X_3) \]

(48) (49) (50) (51)

for some joint distribution \( p(u_3)p(u_3 | u_3)p(x_3 | u_3)p(x_1, x_2 | u) \). First we will prove (51). The first step is to argue that node 2 can decode the full message \( W \) as \( |c_{21}| \geq |c_{31}| \geq |c_{41}| \) (the last inequality is implied by (47)). The argument is essentially the same as for the one-relay channel in Section II, but we will provide it for completeness.

Node 2 can form

\[ Y'_{3}[n] = \frac{c_{31}}{c_{21}} Y_2[n] + Z'_3[n] \]
\[ = c_{31} X_1[n] + \frac{c_{31}}{c_{21}} Z_2[n] + Z'_3[n], \]

(52)

where \( Z'_3[n] \) is iid Gaussian noise with power \( \left(1 - \frac{|c_{31}|^2}{|c_{21}|^2}\right) N_0B \). Since \( Y'_{3}[n] \) has the same distribution as \( Y_3[n] \), node 2 can decode \( W_3 \) with high probability for large \( n \). It then forms

\[ Y'_{4}[n] = \frac{c_{41}}{c_{21}} Y_2[n] + c_{42} X_2[n] + c_{43} \hat{X}_3[n] + Z'_4[n] \]
\[ = c_{41} X_1[n] + c_{42} X_2[n] + c_{43} \hat{X}_3[n] + \frac{c_{41}}{c_{21}} Z_2[n] + Z'_4[n], \]

(53)

where \( Z'_4[n] \) is iid Gaussian noise with power \( \left(1 - \frac{|c_{41}|^2}{|c_{21}|^2}\right) N_0B \) and \( \hat{X}_3[n] \) is the deterministic code corresponding to \( X_3[n] \). Consider the companion signals

\[ \hat{Y}_4[n] = c_{41} \hat{X}_1[n] + c_{42} \hat{X}_2[n] + c_{43} \hat{X}_3[n] + Z_4[n] \]
\[ \hat{Y}'_4[n] = c_{41} \hat{X}_1[n] + c_{42} \hat{X}_2[n] + c_{43} \hat{X}_3[n] + \frac{c_{41}}{c_{21}} Z_2[n] + Z'_4[n]. \]

(54) (55)

A genie provided with \( Y_4[n] \) can decode \( W \), as with high probability \( \hat{Y}_4[n] = Y_4[n] \). Since \( \hat{Y}_4[n] \) and \( \hat{Y}'_4[n] \) have the same distribution\(^5\), a genie provided \( \hat{Y}'_4[n] \) can also decode \( W \) with high probability. Finally, with high probability \( Y'_{4}[n] = \hat{Y}'_4[n] \), and node 2 can therefore decode \( W \) with high probability.

We now get (51) as follows. First define a random variable

\[ E = \begin{cases} 1 & \hat{X}_3[n] \neq X_3[n] | Y'_3[n] \text{ or } \hat{X}_3[n] \neq X_3[n] | Y'_4[n] \\ 0 & \hat{X}_3[n] = X_3[n] | Y'_3[n] = X_3[n] | Y'_4[n] \end{cases} \]

(56)

where \( \hat{X}_3[n] | Y'_3[n] \) denote the estimate node 2 makes of node 3’s transmission. Notice that by the definition of deterministic capacity, Definition I] \( P(E = 1) = \epsilon_n \). Furthermore,

\[ |I(A; B) - I(A; B | E)| \leq H(\epsilon_n) = \epsilon_n \]

(57)

\(^5\)The subtlety here is that \( \hat{X}_3[n] \) and \( Z_2[n] \) are independent due to the deterministic assumption, whereas \( X_3[n] \) and \( Z_2[n] \) are not independent; that is the reason we need three different companion signals.
as

\[ I(A;B) = H(B) - H(B|A) \]
\[ \leq H(B|E) + H(E) - H(B|A,E) \]
\[ I(A;B|E) = H(B|E) - H(B|A,E) \]
\[ \leq H(B) - H(B|A) + H(E) \]

and \( \lim_{p \to 0} H(p) = 0 \). With this we get,

\[ nR \leq n\epsilon_n + I(W; Y_2[n]) \]
\[ \leq (a) \ n\epsilon_n + I(W; Y_2[n]|E = 0)P(E = 0) + I(W; Y_2[n]|E = 1)P(E = 1) + H(\epsilon_n) \]
\[ \leq (b) \ n\epsilon_n + I(W; Y_2[n]|E = 0)P(E = 0) + H(W)P(E = 1) + H(\epsilon_n) \]
\[ \leq (c) \ n\epsilon_n + P(E = 0) \sum_{m=1}^{n} H(Y_2[m]|Y_2[m-1],E = 0) \]
\[ - H(Y_2[m]|Y_2[m-1],W,E = 0) \]
\[ \leq (d) \ n\epsilon_n + P(E = 0) \sum_{m=1}^{n} H(Y_2[m]|X_2[m],X_3[m],E = 0) \]
\[ - H(Y_2[m]|X_1[m],X_2[m],X_3[m],E = 0) \]
\[ \leq (e) \ n\epsilon_n + P(E = 0) \sum_{m=1}^{n} I(X_1[m]; Y_2[m]|X_2[m],X_3[m],E = 0) \]
\[ + P(E = 1) \sum_{m=1}^{n} I(X_1[m]; Y_2[m]|X_2[m],X_3[m],E = 0) \]
\[ = n\epsilon_n + \sum_{m=1}^{n} I(X_1[m]; Y_2[m]|X_2[m],X_3[m],E) \]
\[ \leq (f) n\epsilon_n + \sum_{m=1}^{n} I(X_1[m]; Y_2[m]|X_2[m],X_3[m]) + nH(\epsilon_n) \]
\[ \leq (g) n\epsilon_n + \sum_{m=1}^{n} I(X_1[m]; Y_2[m]|X_2[m],X_3[m]) \].

In step (a) we have used (57), in (b) \( n\epsilon_n \) has absorbed \( H(W)P(E = 1) \leq nR\epsilon_n \) and \( H(\epsilon_n) \). Step (c) follows from the fact that \( X_2[m] \) is a function of \( Y_2[m-1] \), and, conditioned on \( E = 0 \), so is \( X_3[m] \). In step (d) we have used that conditioning reduced entropy and in step (e) that mutual information is positive. In step (f) we have used (57) again, and in (g) \( nH(\epsilon_n) \) is absorbed into \( n\epsilon_n \).

The proof of (48),(50) follows closely the proof in [24] with some twists due to the deterministic capacity definition. In the following use the notation \( \overline{X}[m] = [X[m],X[m+1],\ldots,X[n]] \).
The bound (49) follows from

\[ nR_3 \leq n\epsilon_n + I(W_3; Y_3[n]) \]

\[ = n\epsilon_n + \sum_{m=1}^{n} H(Y_3[m]|Y_3[m-1]) - H(Y_3[m]|Y_3[m-1], W_3) \]

\[ \leq n\epsilon_n + \sum_{m=1}^{n} H(Y_3[m]|X_3[m]) - H(Y_3[m]|Y_3[m-1], Y_4[m+1], W_3, X_3[m]). \quad (60) \]

where we have used that \( X_3[m] \) is a function of \( Y_3[m-1] \) and that conditioning reduces entropy. Equation (49) now follows as in [24].
We will now prove the key bound (50). Similarly to (59) we have

\[ nR \]

\[ = H(W) = H(W_3) + H(W|W_3) \]

\[ = H(W_3|Y_3[n]) + I(W_3; Y_3[n]) \]

\[ + H(W|Y_4[n], W_3) + I(W; Y_4[n]|W_3) \]

\[ \leq n\epsilon_n + I(W_3; Y_3[n]) + I(W; Y_4[n]|W_3) \]

\[ \leq n\epsilon_n + I(W_3; Y_3[n]) + I(W; Y_4[n]|W_3, E) + H(\epsilon_n) \]

\[ = n\epsilon_n + I(W_3; Y_3[n]) + I(W; Y_4[n]|W_3, E = 0)P(E = 0) \]

\[ + I(W; Y_4[n]|W_3, E = 1)P(E = 1) + H(\epsilon_n) \]

\[ \overset{(a)}{=} n\epsilon_n + \sum_{m=1}^{n} I(W_3; Y_3[m]|Y_4[m - 1]) \]

\[ + \sum_{m=1}^{n} I(W; Y_4[m]|W_3, Y_4[m + 1], E = 0)P(E = 0) \]

\[ + H(W)\epsilon_n + H(\epsilon_n) \]

\[ \overset{(b)}{=} n\epsilon_n + \sum_{m=1}^{n} I(W_3; Y_3[m]|Y_4[m - 1], X_3[m]) \]

\[ + \sum_{m=1}^{n} I(W; Y_4[m]|W_3, Y_4[m + 1], X_3[m], E = 0)P(E = 0) \]

\[ \leq n\epsilon_n + \sum_{m=1}^{n} I(W_3; Y_3[m]|Y_4[m - 1], X_3[m]) \]

\[ + \sum_{m=1}^{n} I(W; Y_4[m]|W_3, Y_4[m + 1], X_3[m], E = 1)P(E = 1) \]

\[ = n\epsilon_n + \sum_{m=1}^{n} I(W_3; Y_3[m]|Y_4[m - 1], X_3[m]) \]

\[ + \sum_{m=1}^{n} I(W; Y_4[m]|W_3, Y_4[m + 1], X_3[m], E) \]

\[ \overset{(d)}{=} n\epsilon_n + \sum_{m=1}^{n} I(W_3; Y_3[m]|Y_4[m - 1], X_3[m]) \]

\[ + nH(\epsilon_n) + \sum_{m=1}^{n} I(W; Y_4[m]|W_3, Y_4[m + 1], X_3[m]) \]

\[ = n\epsilon_n + \sum_{m=1}^{n} I(W_3; Y_3[m]|Y_4[m - 1], X_3[m]) \]

\[ + \sum_{m=1}^{n} I(W; Y_4[m]|W_3, Y_4[m + 1], X_3[m]) \]
In step (a) we have used that \( I(W; \cdots) \leq H(W) \). In step (b) we have included \( H(W)\epsilon_n + H(\epsilon_n) \) in the \( n\epsilon_n \) term, and in the first summation we have used that \( X_3[m] \) is function of only \( \sum_3[m-1] \), in the second summation that \( \hat{X}_3[m] \) is a function of \( W_3 \) only and that if \( E = 0 \), \( X_3[m] = \hat{X}_3[m] \). In step (c) we have used that mutual information is non-negative, and in step (d) (57) again.

The final equation (61) is identical to equation (2.5) in [24], except for the conditioning on \( X_3[m] \), and the proof of (50) follows the proof of equations (2.5)-(2.8) in [24]. We will briefly validate that the conditioning on \( X_3[m] \) can be maintained through the proof. We have

\[
\sum_{m=1}^{n} I(W_3; Y_3[m]|Y_3[m-1], X_3[m]) \\
\leq \sum_{m=1}^{n} I(W_3, Y_3[m]-1; Y_3[m]|X_3[m]) \\
= \sum_{m=1}^{n} I(W_3, Y_3[m]-1, Y_4[m]+1; Y_3[m]|X_3[m]) \\
- \sum_{m=1}^{n} I(Y_4[m]+1; Y_3[m]|X_3[m], W_3, \sum_3[m-1]) \\
(62)
\]

and

\[
\sum_{m=1}^{n} I(W_3; Y_4[m]|W_3, Y_4[m]+1, X_3[m]) \\
\leq \sum_{m=1}^{n} I(Y_3[m]-1; Y_4[m]|W_3, Y_4[m]+1, X_3[m]) \\
+ \sum_{m=1}^{n} I(W_3; Y_4[m]|W_3, Y_4[m]+1, X_3[m], Y_3[m]-1). \\
(63)
\]

Thus, we obtain an equation similar to (2.8) before step (d) in [24], except for the conditioning on \( X_3[m] \) everywhere. Step (d) in (2.8) in [24] is therefore still valid, and we get (50).

We now apply Lemmas 2-3 to equations (48-51). First, a comment on the Markov chain property required for using the Lemma 3, as this is not explicitly mentioned in [24]. We define the auxiliary random variables as

\[
U_3[m] = W_3, \\
U_5[m] = (W_3, Y_4[m]+1, \sum_3[m-1]) \\
(64)
\]

As \( U_3[m] \) does not depend on the current channel output and the channel is memoryless, we do indeed have a Markov chain \((U[m], U_3[m]) \rightarrow (X_1[m], X_2[m], X_3[m]) \rightarrow (Y_2[m], Y_3[m], Y_4[m])\). This is maintained through the time-sharing step in [24]. In (48-51) there is also conditioning on \( (X_2, X_3) \), but it is easily seen that these do not have to satisfy a Markov property to use Lemma 3.
Applying Lemmas 23 we then obtain
\[
\begin{align*}
R & \leq \lim_{B \to \infty} \text{var}[c_{42}X_2(B) + c_{43}X_3(B) + c_{41}X_1(B)] \\
R_3 & \leq |c_{31}|^2 \lim_{B \to \infty} \text{var}[X_1(B), X_3(B)] - |c_{31}|^2 \lim_{B \to \infty} \text{var}[X_1(B)X_3(B), U_3(B)] \\
R & \leq \lim_{B \to \infty} \text{var}[c_{42}X_2(B) + c_{43}X_3(B) + c_{41}X_1(B)U_3(B), X_3(B)] \\
& \quad + |c_{31}|^2 \lim_{B \to \infty} \text{var}[X_1(B)X_3(B)] - |c_{31}|^2 \lim_{B \to \infty} \text{var}[X_1(B)X_3(B), U_3(B)] \\
R & \leq |c_{21}|^2 \lim_{B \to \infty} \text{var}[X_1(B)X_2(B), X_3(B)]. \tag{65}
\end{align*}
\]

Let
\[
\lim_{B \to \infty} \text{cov}[\mathbf{X}(B)] = \mathbf{K}_x = \mathbf{LL}^{-1} \\
\mathbf{S} = \mathbf{L}^{-1}\mathbf{X}.
\tag{66}
\]

Here \( \mathbf{L} \) is a triangular matrix found through Cholesky factorization. The random variables \( \mathbf{S} \) are asymptotically uncorrelated (but not necessarily independent), and
\[
\begin{align*}
X_3(B) & = l_{33}S_3(B) \\
X_2(B) & = l_{23}S_3(B) + l_{22}S_2(B) \\
X_1(B) & = l_{13}S_3(B) + l_{12}S_2(B) + l_{11}S_1(B). \tag{67}
\end{align*}
\]

We can then write (65) as
\[
\begin{align*}
R & \leq |c_{42}l_{23} + c_{43}l_{33} + c_{41}l_{13}|^2 \lim_{B \to \infty} \text{var}[S_3] \\
& \quad + |c_{42}l_{22} + c_{41}l_{12}|^2 \lim_{B \to \infty} \text{var}[S_2] + |c_{41}l_{11}|^2 \lim_{B \to \infty} \text{var}[S_1] \\
R_3 & \leq |c_{31}|^2 |l_{11}|^2 \lim_{B \to \infty} (\text{var}[S_1] - \text{var}[S_1|U_3]) + |c_{31}|^2 |l_{21}|^2 \lim_{B \to \infty} (\text{var}[S_2] - \text{var}[S_2|U_3]) \\
R & \leq |c_{42}l_{23} + c_{43}l_{33} + c_{41}l_{13}|^2 \lim_{B \to \infty} \text{var}[S_3|U_3] \\
& \quad + |c_{42}l_{22} + c_{41}l_{12}|^2 \lim_{B \to \infty} \text{var}[S_2|U_3] + |c_{41}l_{11}|^2 \lim_{B \to \infty} \text{var}[S_1|U_3] \\
& \quad + |c_{31}|^2 |l_{11}|^2 \lim_{B \to \infty} (\text{var}[S_1] - \text{var}[S_1|U_3]) + |c_{31}|^2 |l_{21}|^2 \lim_{B \to \infty} (\text{var}[S_2] - \text{var}[S_2|U_3]) \\
R & \leq |c_{21}|^2 \lim_{B \to \infty} \text{var}[S_1]. \tag{68}
\end{align*}
\]

Here we have used \( \text{var}[S_1|S_3] \leq \text{var}[S_1] \) and similar and \( \text{var}[S_1|S_3, U_3] = \text{var}[S_1|U_3] \) and similar, as \( S_3 \) is a function of \( U_3 \). Since \( |c_{31}| \geq |c_{41}| \) the bounds are maximized if we set \( \text{var}[S_1|U_3] = 0 \). Furthermore \( \lim_{B \to \infty} \text{var}[S_i] = 1 \), and we put
\[
\alpha = \lim_{B \to \infty} \text{var}[S_2] - \lim_{B \to \infty} \text{var}[S_2|U_3], \tag{69}
\]
where $0 \leq \alpha \leq 1$. We can then write (65) as

$$R \leq |c_{42}l_{23} + c_{43}l_{33} + c_{41}l_{13}|^2 + |c_{42}l_{22} + c_{41}l_{12}|^2 + |c_{41}l_{11}|^2 \quad (70)$$

$$R_3 \leq |c_{31}|^2|l_{11}|^2 + |c_{31}|^2|l_{12}|^2 \alpha \quad (71)$$

$$R \leq |c_{42}l_{22} + c_{41}l_{12}|^2(1 - \alpha) + |c_{31}|^2|l_{11}|^2 + |c_{31}|^2|l_{12}|^2 \alpha \quad (72)$$

$$R \leq |c_{21}|^2|l_{11}|^2. \quad (73)$$

To get an actual outer bound, these bounds must be maximized over the parameters $l_{ij}$ and $\alpha$ and $R_3$. The simplest approach is to fix $R$ and minimize the total power consumption. Even solving this optimization problem is quite complex, so we will only outline the solution. We will first show that this optimum solution is reached for either $\alpha = 0$ or $\alpha = 1$. For this it is convenient to view it as a vector problem. Let

$$\|l_1\|^2 = |l_{21}|^2 + |l_{22}|^2$$
$$\|l_2\|^2 = |l_{31}|^2 + |l_{32}|^2 + |l_{33}|^2$$
$$\|c_{42}\|^2 = |c_{41}|^2 + |c_{42}|^2$$
$$\|c_{43}\|^2 = |c_{41}|^2 + |c_{42}|^2 + |c_{43}|^2$$

$$v_1 = \angle(c_{41}, c_{42})$$
$$v_2 = \angle(l_{12}, l_{22}). \quad (74)$$

We can then write the two inequalities (70) and (72) for $R$ as

$$R \leq \|c_{43}\|^2\|l_3\|^2 + \|c_{42}\|^2\|l_2\|^2 \cos^2(v_2 - v_1) + |c_{41}l_{11}|^2$$
$$R \leq \|c_{42}\|^2\|l_2\|^2(1 - \alpha) \cos^2(v_2 - v_1) + |c_{31}|^2\|l_2\|^2 \cos^2(v_1) + |c_{31}l_{11}|^2. \quad (75)$$

The optimum solution satisfy both of these inequalities with equality, and the total power is

$$P = \|l_1\|^2 + \|l_2\|^2 + \|l_1\|^2$$
$$= \frac{\|c_{43}\|^2(R - |c_{31}l_{11}|^2) + \alpha|c_{31}|^2(R - |c_{41}l_{11}|^2)}{\|c_{43}\|^2(\alpha|c_{31}|^2 \cos^2 v_1 + \|c_{42}\|^2(1 - \alpha) \cos^2(v_2 - v_1))}$$

$$\frac{\partial P}{\partial \alpha} = \frac{(R - |c_{31}l_{11}|^2)(\|c_{43}\|^2 - \|c_{42}\|^2 \cos^2(v_2 - v_1)) \((\|c_{42}\|^2 \cos^2(v_2 - v_1) - |c_{31}|^2 \cos^2 v_1))}{\|c_{43}\|^2(\alpha|c_{31}|^2 \cos^2 v_1 + \|c_{42}\|^2(1 - \alpha) \cos^2(v_2 - v_1))^2}. \quad (77)$$

We notice that the sign of this derivative does not depend on $\alpha$. Thus, for any $v_1$, the optimum solution is on one of the boundaries, $\alpha = 0$ or $\alpha = 1$.

We next find the optimum solutions on the boundaries. For $\alpha = 1$ we have

$$|l_{11}|^2 = \frac{R}{|c_{21}|^2}$$
$$|l_{12}|^2 = \left(\frac{1}{|c_{31}|^2 - \frac{1}{|c_{21}|^2}}\right)R, \quad (78)$$

so that (72) is satisfied. The solution for $l_{22}$ is more complicated. To satisfy (70) we can either increase $l_{22}$ or $\|l_3\|$. We can think of this as a kind of water filling. Initially, it pays off to increase $l_{22}$ from zero, but at a certain point
it is more efficient to start increase $\|I_3\|$ from zero. This point can be found by comparing derivatives, and some straightforward but lengthy calculation show that the optimum solution for $l_{22}$ and $l_3$ is

$$|l_{22}|^2 = \frac{|c_{41}|^2 |c_{42}|^2 |l_{12}|^2}{(|c_{41}|^2 + |c_{43}|^2)^2}$$

$$= \frac{|c_{41}|^2 |c_{42}|^2}{(|c_{41}|^2 + |c_{43}|^2)^2} \left( \frac{1}{|c_{31}|^2} - \frac{1}{|c_{21}|^2} \right) R$$

$$\|I_3\|^2 = \frac{R - |c_{41}|^2 |l_{12}|^2 - |c_{42}|^2 |l_{22}|^2 - 2|c_{41}| |l_{12}| |c_{42}| |l_{22}| - |c_{41}|^2 |l_{11}|^2}{|c_{41}|^2 + |c_{42}|^2 + |c_{43}|^2}.$$ (79)

The solution for $\alpha = 0$ is more straightforward, as all vectors taking inner products should be parallel. Therefore

$$|l_{11}|^2 = \frac{R}{|c_{21}|^2}$$

$$\|I_2\|^2 = \frac{R - |c_{31}|^2 |l_{11}|^2}{|c_{41}|^2 + |c_{42}|^2}$$

$$\|I_3\|^2 = \frac{R - (|c_{41}|^2 + |c_{42}|^2) |l_2|^2 - |c_{41}|^2 |l_{11}|^2}{|c_{41}|^2 + |c_{42}|^2 + |c_{43}|^2}.$$ (80)

We now compare the total power consumption for these two solutions. If the condition (47) is satisfied, the power consumption with $\alpha = 1$ is smaller than the power consumption with $\alpha = 0$. Again, the condition is found through lengthy but straightforward algebra. The final step is now to realize that any solution with $\alpha = 1$ can be achieved with successive cooperation (whereas this is not possible for $\alpha = 0$). This is perhaps easiest to see by considering (70) with $\alpha = 1$,

$$R \leq |c_{42}| l_{23} + c_{43} l_{33} + c_{41} l_{13}|^2$$

$$+ |c_{42}| l_{22} + c_{41} l_{12}|^2 + |c_{41}| l_{11}|^2.$$ (81)

$$R_3 \leq |c_{31}|^2 |l_{11}|^2 + |c_{31}|^2 |l_{12}|^2.$$ (82)

$$R \leq |c_{31}|^2 |l_{11}|^2 + |c_{31}|^2 |l_{12}|^2.$$ (83)

$$R \leq |c_{21}|^2 |l_{11}|^2.$$ (84)

The inequality (84) can be interpreted so that relay 2 should be able to decode the message from the first interval, (82) and (83) states that relay should be able to decode the whole message from transmissions in intervals 1 and 2, and (81) is exactly the condition that the destination can decode the message based on transmission from all relays and the source.

The inequality (47) is a technical condition for the proof to be valid. It implies the more logical condition $|c_{31}|^2 \geq |c_{41}|^2 + |c_{42}|^2$, but this condition is not sufficient. However, both conditions state that the power required to make relay 3 understand the message (and transmit it) should be sufficiently small compared to the power required for relay 2 and the source alone to transmit the message. On the flip side, it seems intuitively clear that if $|c_{31}|^2$ is small, it is better simply not to use relay 3. Surprisingly, the bounds we have are not strong enough to show that.

The second example of a two-relay channel is a channel with nodes placed on a line, Fig. 5. We consider the scaled fading model only, and assume that the average gain between nodes is a decreasing function of distance.
The key implications of this is that the following conditions are satisfied

\[
E[|c_{ij}|^2] \leq E[|c_{ik}|^2], \quad j \leq k \\
E[|c_{ij}|^2] \leq E[|c_{kj}|^2], \quad k \leq i.
\]  

The results apply to all channels that satisfies this condition. In this case we have an obvious decode-forward strategy: sequential decode-forward (or multi-hop): the first relay (node 2) decodes the full message based on the transmission from the source. It then re-transmits the message. Node 3 decodes the message based on both transmission from nodes 1 and 2, and finally node 4 decodes based on transmission from all previous nodes. With this scheme, a rate \( R \) is achievable if

\[
R \leq \sum_{i=1}^{j-1} E[|c_{ji}|^2] P_i, \quad j = 2, \ldots, 4.
\]  

This is optimum in the scaled fading case, as shown by the following Theorem.

**Theorem 7.** For scaled fading linear relay channels with four nodes that satisfy condition (85), a sequential decode-forward scheme achieves the deterministic capacity.

**Proof:** We only have to prove that nodes 2 and 3 must be able to decode the full message \( W \) as then (86) must be satisfied. We will first argue that node 2 can decode the message based on a degraded channel argument. The first step is to argue that node 2 can predict what node 3 transmits. Consider the received signals at nodes 2 and 3,

\[
\vdots
\]  

Let

\[
k_1 = \frac{E[|c_{31}|^2]}{E[|c_{21}|^2]}.
\]  

and notice that \( k_1 c_{21} \) has the same distribution as \( c_{31} \) due to the scaling assumption. Node 2 constructs the following
Since we are discussing deterministic capacity, there exist some signals and $Z_k[1]$ so that we can always expand for example $P(n)$ so that

$$P(A \cap B \cap C) = P(C|A \cap B) P(A \cap B) \geq 1 - \epsilon.$$ (91)

Notice that, conditioned on $\left\{ X_3[i-1](Y_3'[i-2]) = \hat{X}_3[i-1](W) \right\} \cap A \cap B$, $Y_3'[i-1]$ and $Y_3[i-1]$ have the same distribution (this is argued recursively based on the equations (89)). Since the code $X_3$ used is the same, then by the deterministic assumption there exists some $n_2$ so that for $n \geq n_2$ $P(C|A \cap B) > \sqrt{1-\epsilon}$.

Thus, conditioned on $A \cap B \cap C$, node 2 can predict what node 3 is transmitting. It can then construct

$$Y_3'[i] = k_2 Y_2[i] + c_{21}'[i] X_2[i] + c_{31}'[i] X_1[i],$$ (94)

where $k_2 = \frac{E[|c_{31}'|^2]}{E[|c_{21}'|^2]}$, $c_{21}'[i]$ and $c_{31}'[i]$ are random sequences with the same distribution as the actual channel coefficients, and $Z_3'[i]$ is iid Gaussian $\mathcal{N}(0, (1-k_2)N_0B)$. Since, again conditioned on $A \cap B \cap C$, $Y_4'[n]$ and $Y_4'[n]$ have the
same distribution, and node 4 by assumption can decode $W$, node 2 can also decode $W$; formally,

$$P\left(\hat{W}(Y'_4[n]) \neq W\right) = P\left(\hat{W}(Y'_4[n]) \neq W|A \cap B \cap C\right) P(A \cap B \cap C)$$

$$+ P\left(\hat{W}(Y'_4[n]) \neq W |(A \cap B \cap C)^c\right) P\left((A \cap B \cap C)^c\right);$$

in the first term the first factor is $\epsilon_n$; in the second term, the second factor is $\epsilon_n$.

The argument that node 3 can also decode is quite different, and we will use the broadcast bound of Proposition 5. Nodes 3 and 4 decode messages $W_3$ and $W_4$. As in the proof of Theorem 6, we split node 3 into a transmitter and receiver part: the transmitter part knows $W_3$ and transmits to node 4 and helps it decode $W_4$, while the receiver part must be able decode $W_3$. From Proposition 5 we directly have the bounds

$$R_3 \leq I(U_3;Y_3)$$

$$R_4 \leq I(U_4;Y_4)$$

$$R \leq I(U_3;Y_3|U_4) + I(U_4;Y_4)$$

$$R \leq I(U_4;Y_4|U_3) + I(U_5;Y_3).$$

In fact, we just need the last of these bounds. Using Lemmas 23 we get the bound

$$R \leq E[|c_41|^2] \lim_{B \to \infty} \text{var}[X_1|U_3] + E[|c_42|^2] \lim_{B \to \infty} \text{var}[X_2|U_3] + E[|c_43|^2] \lim_{B \to \infty} \text{var}[X_3|U_3]$$

$$+ E[|c_31|^2] \lim_{B \to \infty} (\text{var}[X_3] - \text{var}[X_1|U_3]) + E[|c_32|^2] \lim_{B \to \infty} (\text{var}[X_2] - \text{var}[X_2|U_3]).$$

Here $\text{var}[X_3|U_3] = 0$, as $X_3$ is a function of $W_3$ and thus $U_3$. Since $E[|c_31|^2] > E[|c_41|^2]$ and $E[|c_32|^2] > E[|c_42|^2]$, $\text{var}[X_1|U_3] = \text{var}[X_2|U_3] = 0$, is optimum, and we get

$$R \leq E[|c_31|^2]\text{var}[X_1] + E[|c_32|^2]\text{var}[X_2].$$

This shows that also node 3 must be able to decode the message.

**B. MIMO Relay channels**

In this section we consider the two MIMO (multiple input multiple output) relay channels in Figs. 45. The new property of these channels compared to the two-relay channels in the previous section, is that message splitting is required to achieve the deterministic capacity. For the channel 4 the message is split into a part transmitted through the relay and one transmitted directly. For the channel 5 the message is split into three parts: two private messages understood by each relay, and a common message understood by both relays. Furthermore, the separate...
sub-messages are transmitted through superposition. The outer bounds show that this is optimum.

Another way to look at these channels is that they are an extension of the previous two-relay channels to three- and four-relay channels. We can think of the two antennas of the MIMO as separate relays connected to the destination so that they can both decode the whole message $W$. From this point of view, it shows that message splitting is required in larger relay networks.

A final point of note is that the channel in Fig. 4 is the only channel where we can find the deterministic capacity for all channel coefficients for both the static and fading cases\footnote{In this section we do not need to assume scaled fading.} while the channel in Fig. 5 has the new element of a MAC with correlated messages.

We need the following Lemma

**Lemma 8.** For any random variables $X$ and $Y$ with first and second order moments

$$\text{cov}[X|Y] \prec \text{cov}[X] - \frac{\text{cov}[X,Y]\text{cov}[X,Y]^H}{\text{var}[Y]},$$

(102)

**Proof:** We can assume that $X$ and $Y$ are zero mean. First, notice that


(103)

Second, the Cauchy-Schwartz inequality gives


$$\leq \sqrt{E[|Y|^2]}\sqrt{E[v^H E[X|Y] E[X|Y]^H v]},$$

(104)

So,

$$v^H E[X Y^*] E[Y X^H] v \leq \text{var}[Y] v^H E[E[X|Y] E[X|Y]^H] v,$$

(105)


We have the following result

**Theorem 9.** The deterministic capacity of the relay channel in Fig. 4 in the low power regime in the static channel
case is given by maximizing

\[ R \leq (|c_{311}|^2 + |c_{312}|^2) P_{31} + (|c_{211}|^2 + |c_{212}|^2) \cos^2(\alpha - \theta) P_{21} \]

(106)

\[ R \leq (|c_{311}|^2 + |c_{312}|^2) P_{31} + (|c_{311}|^2 + |c_{312}|^2) \cos^2(\theta) P_{21} + \left( \sqrt{P_{31} (|c_{311}|^2 + |c_{312}|^2) + |c_{32}| \sqrt{P_2} \right)^2 \]

(107)

with respect to \( P_{21}, P_{31}, P_{32}, \) and \( \theta, \) subject to \( P_{21} + P_{31} + P_{32} \leq P_1. \) Here \( \alpha = \arccos \left( \frac{c_{21}^H c_{31}}{|c_{21}||c_{31}|} \right). \) In the fading case, the deterministic capacity is given by

\[ C = \min \left\{ \max \{ E[|c_{311}|^2] + E[|c_{312}|^2], E[|c_{211}|^2] + E[|c_{212}|^2] \} P_1, (E[|c_{311}|^2] + E[|c_{312}|^2]) P_1 + E[|c_{32}|^2] P_2 \right\}. \]

(108)

**Proof:** The rate is bounded by

\[ R \leq I(X_1; Y_3|U_2, X_2) + I(U_2; Y_2|X_2) \]

(109)

\[ R \leq I(X_1, X_2; Y_3) \]

(110)

for some distribution \( p(u_2)p(x_1, x_2|u_2). \) The bound \[109\] is essentially the same as \[30\], and is proven the same way. The bound \[110\] is simply the MAC bound into the destination.

We will prove the theorem for the static channel case. The proof in the fading case is a simpler case that we will omit. Using lemma [3] we get the low power limit of \[109, 110\] as

\[ R \leq \lim_{B \to \infty} \text{var}[c_{31}^H X_1(B)|U_2(B), X_2(B)] + \lim_{B \to \infty} \text{var}[c_{21}^H X_1(B)|X_2(B)] - \lim_{B \to \infty} \text{var}[c_{21}^H X_1(B)|U_2(B), X_2(B)] \]

(111)

\[ R \leq \lim_{B \to \infty} \text{var}[c_{31}^H X_1(B) + c_{32} X_2(B)]. \]

(112)

Let \( u \) be a unit vector in the direction of \( \lim_{B \to \infty} \text{cov}[X_1(B), X_2(B)], \) and define

\[ \beta \sqrt{P_1} u = \lim_{B \to \infty} \frac{\text{cov}[X_1(B), X_2(B)]}{\sqrt{\text{var}[X_2(B)]}} \]

(113)

\[ X = \lim_{B \to \infty} \text{cov}[X_1(B)] \]

(114)

\[ A = \lim_{B \to \infty} \text{cov}[X_1(B), X_2(B), U_2(B)] \]

(115)

\[ B = X - \beta^2 P_1 uu^H - A. \]

(116)

Using Lemma [8] we then obtain the following outer bound to the low power rate

\[ R \leq c_{31}^H A c_{31}^H + c_{21}^H B c_{21} \]

(117)

\[ R \leq c_{31}^H (A + B + \beta^2 P_1 uu^H) c_{31} + |c_{32}|^2 P_2 + 2R \{ \beta c_{32} c_{31}^H u \} \sqrt{P_1 P_2} \]

(118)
subject to
\[
\text{tr}A + \text{tr}B + \beta^2 P_1 \leq P_1 \\
A, B \succeq 0 \\
\beta \leq 1.
\]

On the other hand, if \(A, B,\) and \(\beta\) satisfy \((119-121),\) then \((117-118)\) constitute an upper bound on the rate.

It is clear that \((117-118)\) is maximized for \(A = P_{31} c_{31} c_{31}^H \) for some positive constant \(P_{31}.\) Now notice that if the angle between \(c_{21}\) and \(c_{31}\) is acute (if not, we can just use \(-c_{21}\)), the bounds are maximized when the off-diagonal elements of \(B\) are maximized, i.e., if \(B\) has rank one. Thus, we can put \(B = P_{31} vv^H,\) where \(v\) is a unit vector rotated an angle \(\theta\) from \(c_{21}\) in the real plane spanned by \(c_{21}, c_{31},\) as any component outside this plane will not contribute to the bounds. Finally, the bounds are maximized for \(u = \frac{c_{31}}{\|c_{31}\|}.\) It is now a straightforward calculation to get the bounds \((106)\) and \((107).\)

For the achievable rate we split the message \(W\) into two independent parts \(W_d\) and \(W_r.\) The message \(W_d\) is transmitted directly to the destination using power \(P_{31}\) and a rate
\[
R_d = (|c_{311}|^2 + |c_{312}|^2) P_{31}. \quad (122)
\]

The message \(W_r\) is transmitted through the relay using block Markov encoding with a rate
\[
R_r = \min \left\{(|c_{211}|^2 + |c_{212}|^2) \cos^2(\alpha - \theta)P_{21}, \right.
\left. (|c_{311}|^2 + |c_{312}|^2) \cos^2(\theta)P_{21} \right.
\left. + \left(\sqrt{P_{b1}} (|c_{311}|^2 + |c_{312}|^2) + |c_{32}| \sqrt{P_{b2}}\right)^2 \right\}. \quad (123)
\]

Adding up these rates achieves the upper bound.

We now turn to the relay channel in Fig. 5. From the two relays to the destination we have a MAC channel. As opposed to the usual MAC channel, we have messages that can have arbitrary dependency. The usual MAC outer bound can be generalized as follows, with the difference being that \(X_2\) and \(X_3\) can no longer be assumed independent

**Proposition 10.** Consider a discrete memoryless multiple access channel with dependent messages. Transmitters 2 and 3 have the message \(W_2\) and \(W_3,\) which may be dependent, that they need to transmit to the receiver. The capacity region is contained in the convex closure of all rates satisfying
\[
R_{23} \leq I(X_2; Y_4|X_3) \quad (124)
\]
\[
R_{32} \leq I(X_3; Y_4|X_2) \quad (125)
\]
\[
R \leq I(X_3, X_3; Y_4) \quad (126)
\]
for some joint distribution \(p(x_2, x_3).\) Here the rates are defined according to \((42).\)

**Proof:** The proof closely follows the usual MAC proof in \([11,\text{Section 15.3.4}],\) just replacing \(H(W_2)\) with
\( H(W_2|W_3) \) and \( H(W_3) \) with \( H(W_3|W_2) \). For example, the inequality (124) is obtained as follows

\[
\begin{align*}
nR_{23} & \leq H(W_2|W_3) \\
& \leq I(W_2; Y_4[n]|W_3) + n\epsilon_n \\
& = H(Y_4[n]|W_3) - H(Y_4[n]|W_2, W_3) + n\epsilon_n \\
& \leq (a) \quad H(Y_4[n]|X_2[n]) - H(Y_4[n]|W_2, W_3, X_2[n], X_3[n]) + n\epsilon_n \\
& \leq (b) \quad H(Y_4[n]|X_2[n]) - H(Y_4[n]|X_2[n], X_3[n]) + n\epsilon_n,
\end{align*}
\]

where (a) follows from the data processing inequality and the fact that conditioning decreases entropy, and (b) follows from the fact that \( Y_4 \) depends on only \( X_2 \) and \( X_3 \). The equation (127) is identical to [11, eq. (15.108)].

Since the above bound, as the usual Gaussian MAC bound, is maximized by the Gaussian distribution, we get directly the bound in the low power regime

**Corollary 11.** The capacity region of the MAC channel in the low power regime in the static channel case is contained in the convex closure of all all \((R, R_{23}, R_{32})\) that satisfies

\[
\begin{align*}
R_{23} & \leq |c_{42}|^2 P_2 (1 - \rho^2) \\
R_{32} & \leq |c_{43}|^2 P_3 (1 - \rho^2) \\
R & \leq |c_{42}|^2 P_2 + |c_{43}|^2 P_3 + 2\rho |c_{42}| |c_{43}| \sqrt{P_2 P_3}
\end{align*}
\]

for some \( \rho \in [0, 1] \) in the synchronous case.

In the fading case the rates satisfy

\[
\begin{align*}
R_{23} & \leq E[|c_{42}|^2] P_2 \\
R_{32} & \leq E[|c_{43}|^2] P_3 \\
R & \leq E[|c_{42}|^2] P_2 + E[|c_{43}|^2] P_3.
\end{align*}
\]

**Theorem 12.** In the fading case, the deterministic capacity of the relay channel in Fig. 5 is given by transmitting a common message to the two relays in addition to two private messages.

**Proof:** We will prove that upper bound for the broadcast part of the channel can be achieved with a common/private message transmission scheme. Since this is clearly also true for the MAC part, this will be sufficient to prove the Theorem. We will prove the theorem for the case when antenna 1 and antenna 2 of node 1 have separate power constraints, which we will denote \( P_1 \) and \( P_2 \). The result then clearly applies to the case when there is a sum power constraint, but it also applies to the case when the two antennas are actually on separate nodes.

For the broadcast part of the channel, the achievable rate by common/private message transmission is given by

\[
\begin{align*}
R_c &= \min \left\{ E \left[ \log \left( 1 + \frac{|c_{211}|^2 P_{1c} + |c_{212}|^2 P_{2c}}{N_0 B} \right) \right] , E \left[ \log \left( 1 + \frac{|c_{311}|^2 P_{1c} + |c_{312}|^2 P_{2c}}{N_0 B} \right) \right] \right\} \\
R_2 &= R_c + E \left[ \log \left( 1 + \frac{|c_{211}|^2 P_{12} + |c_{212}|^2 P_{22}}{N_0 B} \right) \right] \\
R_3 &= R_c + E \left[ \log \left( 1 + \frac{|c_{311}|^2 P_{13} + |c_{312}|^2 P_{23}}{N_0 B} \right) \right] \\
R &= R_c + E \left[ \log \left( 1 + \frac{|c_{211}|^2 P_{12} + |c_{212}|^2 P_{22} + |c_{311}|^2 P_{13} + |c_{312}|^2 P_{23}}{N_0 B} \right) \right] .
\end{align*}
\]
Taking the limit $B \to \infty$ we get

\[
R_c = \min \{ E[|c_{111}|^2]P_{1c} + E[|c_{212}|^2]P_{2c}, E[|c_{311}|^2]P_{1c} + E[|c_{312}|^2]P_{2c} \}
\]

\[
R_2 = R_c + E[|c_{111}|^2]P_{12} + E[|c_{212}|^2]P_{22}
\]

\[
R_3 = R_c + E[|c_{311}|^2]P_{13} + E[|c_{312}|^2]P_{23}
\]

\[
R = R_c + E[|c_{111}|^2]P_{12} + E[|c_{212}|^2]P_{22} + E[|c_{311}|^2]P_{13} + E[|c_{312}|^2]P_{23}
\]

(135)

with the constraints

\[
P_{1c} + P_{21} + P_{31} \leq P_1
\]

\[
P_{2c} + P_{22} + P_{32} \leq P_2.
\]

(136)

For the upper bound we apply Lemma 3 to the bounds of Proposition 5.

\[
R_2 \leq E[|c_{111}|^2] \lim_{B \to \infty} \text{var}[X_{11}(B)] - \text{var}[X_{11}(B)|U_2(B)] + E[|c_{212}|^2] \lim_{B \to \infty} \text{var}[X_{11}(B)] - \text{var}[X_{11}(B)|U_2(B)]
\]

(137)

\[
R_3 \leq E[|c_{311}|^2] \lim_{B \to \infty} \text{var}[X_{11}(B)] - \text{var}[X_{11}(B)|U_3(B)] + E[|c_{312}|^2] \lim_{B \to \infty} \text{var}[X_{11}(B)] - \text{var}[X_{11}(B)|U_3(B)]
\]

(138)

\[
R \leq E[|c_{111}|^2] \lim_{B \to \infty} \text{var}[X_{11}(B)] - \text{var}[X_{11}(B)|U_3(B)] + E[|c_{212}|^2] \lim_{B \to \infty} \text{var}[X_{11}(B)] - \text{var}[X_{11}(B)|U_3(B)]
\]

(139)

\[
R \leq E[|c_{111}|^2] \lim_{B \to \infty} \text{var}[X_{11}(B)] - \text{var}[X_{11}(B)|U_2(B)] + E[|c_{311}|^2] \lim_{B \to \infty} \text{var}[X_{11}(B)] - \text{var}[X_{11}(B)|U_2(B)]
\]

(140)

Define

\[
P_{12} = \lim_{B \to \infty} \text{var}[X_{11}(B)|U_3(B)]
\]

(141)

\[
P_{22} = \lim_{B \to \infty} \text{var}[X_{12}(B)|U_3(B)]
\]

(142)

\[
P_{13} = \lim_{B \to \infty} \text{var}[X_{11}(B)|U_2(B)]
\]

(143)

\[
P_{23} = \lim_{B \to \infty} \text{var}[X_{12}(B)|U_2(B)]
\]

(144)

\[
P_{1c} = \lim_{B \to \infty} \text{var}[X_{11}(B)] - \text{var}[X_{11}(B)|U_3(B)] - \lim_{B \to \infty} \text{var}[X_{11}(B)|U_2(B)]
\]

(145)

\[
P_{2c} = \lim_{B \to \infty} \text{var}[X_{12}(B)] - \text{var}[X_{12}(B)|U_3(B)] - \lim_{B \to \infty} \text{var}[X_{12}(B)|U_2(B)].
\]

(146)

Clearly $P_{ij} \geq 0$, so that we can think of them as powers. Notice that we cannot assume $P_{ic} \geq 0$. However, we
have the constraints

\[
P_{1c} + P_{12} + P_{13} \leq P_1 \tag{147}
\]
\[
P_{2c} + P_{22} + P_{23} \leq P_2 \tag{148}
\]
\[
P_{1c} + P_{12} \geq 0 \tag{149}
\]
\[
P_{1c} + P_{13} \geq 0 \tag{150}
\]
\[
P_{2c} + P_{22} \geq 0 \tag{151}
\]
\[
P_{2c} + P_{23} \geq 0. \tag{152}
\]

With this we can write

\[
R_2 \leq E[|c_{211}|^2]P_{1c} + E[|c_{212}|^2]P_{2c}
+ E[|c_{211}|^2]P_{12} + E[|c_{212}|^2]P_{22} \tag{153}
\]
\[
R_3 \leq E[|c_{311}|^2]P_{1c} + E[|c_{312}|^2]P_{2c}
+ E[|c_{311}|^2]P_{13} + E[|c_{312}|^2]P_{23} \tag{154}
\]
\[
R \leq E[|c_{211}|^2]P_{1c} + E[|c_{212}|^2]P_{2c} + E[|c_{211}|^2]P_{12}
+ E[|c_{212}|^2]P_{22} + E[|c_{311}|^2]P_{13} + E[|c_{312}|^2]P_{23} \tag{155}
\]
\[
R \leq E[|c_{311}|^2]P_{1c} + E[|c_{312}|^2]P_{2c} + E[|c_{211}|^2]P_{12}
+ E[|c_{212}|^2]P_{22} + E[|c_{311}|^2]P_{13} + E[|c_{312}|^2]P_{23}. \tag{156}
\]

We will show that the upper bound can always be achieved by a common/private message solution. First consider the case \(\{E[|c_{211}|^2], E[|c_{212}|^2]\} \leq \{E[|c_{311}|^2], E[|c_{312}|^2]\}\). The optimum solution has \(P_{12} = P_{22} = 0\). Namely, putting \(P_{1c} \rightarrow P_{1c} + P_{12}\) and \(P_{2c} \rightarrow P_{2c} + P_{22}\) will not decrease any rate bounds, while the power bounds are still satisfied. Notice that we can now assume \(P_{1c} \geq 0, P_{2c} \geq 0\). So, we end up with

\[
R_2 \leq E[|c_{211}|^2]P_{1c} + E[|c_{212}|^2]P_{2c} \tag{157}
\]
\[
R_3 \leq E[|c_{311}|^2]P_{1c} + E[|c_{312}|^2]P_{2c}
+ E[|c_{311}|^2]P_{13} + E[|c_{312}|^2]P_{23} \tag{158}
\]
\[
R \leq E[|c_{211}|^2]P_{1c} + E[|c_{212}|^2]P_{2c}
+ E[|c_{311}|^2]P_{13} + E[|c_{312}|^2]P_{23} \tag{159}
\]
\[
R \leq E[|c_{211}|^2]P_{1c} + E[|c_{212}|^2]P_{2c}
+ E[|c_{311}|^2]P_{13} + E[|c_{312}|^2]P_{23}. \tag{160}
\]
or

\[ R_2 \leq E[c_{211}^2]P_{1c} + E[c_{212}^2]P_{2c} \]  \hspace{1cm} (161) \\
\[ R_3 \leq E[c_{211}^2]P_{1c} + E[c_{212}^2]P_{2c} + E[c_{311}^2]P_{13} + E[c_{312}^2]P_{23} \]  \hspace{1cm} (162) \\
\[ R \leq E[c_{211}^2]P_{1c} + E[c_{212}^2]P_{2c} + E[c_{311}^2]P_{13} + E[c_{312}^2]P_{23}. \]  \hspace{1cm} (163)

This can be achieved by transmitting a common message understood by both nodes, and a private message to node 3.

Next consider the case \( E[c_{211}^2] \leq E[c_{311}^2], E[c_{312}^2] \leq |c_{212}| \), with strict inequality in at least one of the inequalities. Then a solution with \( P_{12} = P_{23} = 0 \) is optimum, which can be seen by putting \( P_{1c} \rightarrow P_{1c} + P_{12} \) and \( P_{2c} \rightarrow P_{2c} + P_{23} \). Again we can then assume \( P_{1c} \geq 0, P_{2c} \geq 0 \). Then

\[ R_2 \leq E[c_{211}^2]P_{1c} + E[c_{212}^2]P_{2c} + E[c_{212}^2]P_{22} \]  \hspace{1cm} (164) \\
\[ R_3 \leq E[c_{311}^2]P_{1c} + E[c_{312}^2]P_{2c} + E[c_{312}^2]P_{22} \]  \hspace{1cm} (165) \\
\[ R \leq E[c_{211}^2]P_{1c} + E[c_{212}^2]P_{2c} + E[c_{311}^2]P_{13} + E[c_{312}^2]P_{22} \]  \hspace{1cm} (166) \\
\[ R \leq E[c_{211}^2]P_{1c} + E[c_{212}^2]P_{2c} + E[c_{311}^2]P_{13} + E[c_{312}^2]P_{22}. \]  \hspace{1cm} (167)

We will argue that we can always obtain an optimum solution with the right hand sides of (166) and (167) equal. Assume the right hand side of (166) is smaller than that of (167). We can decrease (167) by putting \( P_{13} \rightarrow P_{13} - \delta, P_{1c} \rightarrow P_{1c} + \delta \). Either the bounds become equal, or we end up with \( P_{13} = 0 \), so

\[ R_2 \leq E[c_{211}^2]P_{1c} + E[c_{212}^2]P_{2c} + E[c_{212}^2]P_{22} \]  \hspace{1cm} (168) \\
\[ R_3 \leq E[c_{311}^2]P_{1c} + E[c_{312}^2]P_{2c} \]  \hspace{1cm} (169) \\
\[ R \leq E[c_{311}^2]P_{1c} + E[c_{312}^2]P_{2c} + c_{212}^2P_{22} \]  \hspace{1cm} (170) \\
\[ R \leq E[c_{211}^2]P_{1c} + E[c_{212}^2]P_{2c} + E[c_{212}^2]P_{22}. \]  \hspace{1cm} (171)

But this can be written as

\[ R_2 \leq E[c_{311}^2]P_{1c} + E[c_{312}^2]P_{2c} + E[c_{212}^2]P_{22} \]  \hspace{1cm} (172) \\
\[ R_3 \leq E[c_{311}^2]P_{1c} + E[c_{312}^2]P_{2c} \]  \hspace{1cm} (173) \\
\[ R \leq E[c_{311}^2]P_{1c} + E[c_{312}^2]P_{2c} + E[c_{212}^2]P_{22}, \]  \hspace{1cm} (174)

which can be achieved by a common message and a private message to node 3.

On the other hand, suppose the right hand side of (166) is larger than that of (167). Then we can decrease (166)
by putting $P_{13} \rightarrow P_{22} - \delta$, $P_{2c} \rightarrow P_{2c} + \delta$. If the two bounds don’t become equal we end up with

$$R_3 \leq E[|c_{211}|^2]P_{1c} + E[|c_{212}|^2]P_{2c}$$  
$$R_4 \leq E[|c_{311}|^2]P_{1c} + E[|c_{312}|^2]P_{2c} + E[|c_{311}|^2]P_{13}$$
$$R \leq E[|c_{211}|^2]P_{1c} + E[|c_{212}|^2]P_{2c} + E[|c_{311}|^2]P_{13}$$

which, as above, can be achieved by common/private messaging.

Theorem 13. Consider the relay channel in Fig. 5 for static channel case. Assume that

$$\min\{|c_{211}|^2, |c_{311}|^2\} \leq |c_{211}^Hc_{311}|.$$  

The deterministic capacity is then given by transmitting a common message to the relays and a private message to one of the relays.

**Proof:** We will prove that the upper bound for the broadcast cut set can be achieved by a common/private message solution. Without loss of generality we can consider the case $|c_{211}| \geq |c_{311}|$. According to Lemma 2 and 3 we have

$$R_2 \leq \lim_{B \to \infty} B \text{var}[c_{21}^H X(B)]$$
$$- \lim_{B \to \infty} B \text{var}[c_{21}^H X(B)|U_2(B)]$$

and

$$R \leq \lim_{B \to \infty} \text{var}[c_{31}^H X(B)|U_3(B)]$$
$$+ \lim_{B \to \infty} \text{var}[c_{21}^H X(B)]$$
$$- \lim_{B \to \infty} \text{var}[c_{31}^H X(B)|U_3(B)]$$

Define

$$X = \lim_{B \to \infty} \text{cov}[X(B)]$$
$$A = \lim_{B \to \infty} \text{cov}[X(B)|U_2(B)]$$
$$B = \lim_{B \to \infty} \text{cov}[X(B)|U_3(B)].$$

The upper bounds can now be stated as
\[ R_2 \leq c_{21}^H (X - A) c_{21} \]
\[ R_3 \leq c_{31}^H (X - B) c_{31} \]
\[ R \leq c_{31}^H (X - B) c_{31} + c_{21}^H B c_{21} \]
\[ R \leq c_{21}^H (X - A) c_{21} + c_{31}^H A c_{31} \] (185)

subject to
\[ \text{tr} X \leq P \] (186)
\[ X \succ A, B \] (187)
\[ A, B \succ 0. \] (188)

Notice that the condition \( \| c_{21} \|^2 \geq |c_{21}^H c_{31}| \geq \| c_{31} \|^2 \) ensures that \( |c_{21}^H v| \geq |c_{31}^H v| \) for any vector \( v \). Consequently \( c_{21}^H A c_{21} \geq c_{31}^H A c_{31} \), and we can put \( A = 0 \) without decreasing any bounds. The bounds then are
\[ R_2 \leq c_{21}^H X c_{21} \] (189)
\[ R_3 \leq c_{31}^H (X - B) c_{31} \] (190)
\[ R \leq c_{31}^H (X - B) c_{31} + c_{21}^H B c_{21} \] (191)
\[ R \leq c_{21}^H X c_{21}. \] (192)

Here the right hand side of (191) is clearly larger than the right hand side of (192), so we can rewrite as
\[ R_2 \leq c_{31}^H C c_{31} + c_{21}^H B c_{21} \] (193)
\[ R_3 \leq c_{31}^H C c_{31} \] (194)
\[ X = C + B. \] (195)

It is clear that this is optimized for \( B = \alpha_2 c_{21} c_{21}^H \) and \( C = \alpha_3 c_{31} c_{31}^H \), and that this can be achieve by transmitting a private message in the direction of \( c_{21} \) and a common message in the direction of \( c_{31} \).

On the other hand, it is shown in the Appendix that Theorem 13 is not necessarily true if the condition (179) is not satisfied.

V. DETERMINISTIC CAPACITY OF MULTIFLOW NETWORKS IN THE LOW POWER REGIME

In this section we will extend the investigation of deterministic capacity to channels with multiple information flows. The setup is the same as in [11] Section 15.10: Node \( i \) has a message \( W_{ji} \) intended for node \( j \); to allow for multicast we may have \( W_{ji} = W_{kj} \). The encoder at node \( i \) is a function \( X_i[m](W_{i1}, \ldots, W_{iN}, Y_i[1] \ldots Y_i[m-1]) \), \( m \in \{1, \ldots, n\} \), a function of the messages known to node \( i \) and past received signals. As for single flow networks we introduce a corresponding deterministic code:
\[ \bar{X}_i[n](W_{11}, W_{12}, \ldots, W_{NN}). \] (196)

That is, the the deterministic code is a function solely of all messages in the total network. We now make the following straightforward generalization of Definition [1]
Definition 14. A set of rates \{R_{ij}\} is said to be achievable by deterministic codes if there exists a sequences of codes \{\hat{X}_i[n], i = 1 \ldots N\} and sequences of corresponding deterministic codes\{\hat{X}_i[n], i = 1 \ldots N\}, so that

\[
\forall i \in \{1, \ldots, N\} \lim_{n \to \infty} P(\hat{X}_i[n] \neq X_i[n]) = 0 \tag{197}
\]

\[
\lim_{n \to \infty} P(W_{ji}(\hat{X}_j[n]) \neq W_{ji}) = 0. \tag{198}
\]

The deterministic capacity region is the closure of the set of all rate sets \{R_{ij}\} achievable by deterministic codes. We say that a sequence of codes \hat{X}_i[n] is reliable if there exist a corresponding sequence of deterministic codes \hat{X}_i[n] so that (197) is satisfied.

In the following we only consider networks with at most two messages, which we will denote \(W_1 \in \{1, \ldots, 2^{nR_1}\}\) and \(W_2 \in \{1, \ldots, 2^{nR_2}\}\) instead of the more general notation above. When a node needs to transmit two messages, we consider the following three different strategies for achievable rate

- **Superposition**: The node transmits \(X = X_1(W_1) + X_2(W_2)\) (where the addition is over the real numbers). A receiving node can decode either \(W_1\) or \(W_2\), or both.

- **Network coding\[13\]**: The node calculates a non-invertible function \(f(W_1, W_2)\) and transmits \(X(f(W_1, W_2))\). A receiving node decodes \(f(W_1, W_2)\). An example is that nodes calculate \(W_1 \oplus W_2\), where the messages are considered a stream of bits and addition is over \(GF(2)\).

- **Multiplexed coding\[15\], \[7\], \[25\]**: The node encodes \(X(W_1, W_2)\). A receiving node can decode \(W_1\) and \(W_2\) if the link has a capacity \(C > R_1 + R_2\), but if the node already knows \(W_1\), it can decode \(W_2\) if just \(C > R_2\).

We emphasize the difference between network coding and multiplexed coding. Network coding is done purely on the messages, at the networking layer so-to-say. At the decoder, the channel code is first decoded to get the message \(f(W_1, W_2)\), and then possible side information is applied. Differently, multiplexed coding is done at the channel coding layer. Possible side information is applied when the channel code is decoded, not afterwards. The consequence is that different information can be decoded depending on the channel state. This exemplifies the difference between network coding and reliable coding: reliable coding can be thought of as network coding at the channel coding layer. Or, if we think of network coding as source coding, reliable coding can be thought of as joint source-channel coding.

Of course, reliable coding allows much more general operations than multiplexed coding, just as network coding allows more general operations than \(W_1 \oplus W_2\). However, in the examples presented here, combinations of the above three simple strategies will be shown to be optimum in terms of deterministic capacity. But none of the strategies are sufficient alone.

For the multiflow channels we consider the fading model only, as even this is complex.

A. Cooperative Interference Channels

In this section we consider cooperative interference channels in the low power regime. It might seem more natural to consider cooperative MAC and BC channels as the primary example of multiflow channels, but it can be verified that for the small networks where we can find deterministic capacity, MAC and BC channels do not really provide any new perspectives compared to the relay channels considered previously (e.g., a MAC with two sources turns into a superposition of two relay channels). On the other hand, interference channels behave quite differently than relay channels.

We consider at first the interference channel in \[6\] where the transmitters can overhear each other’s transmission. This channel was considered in \[15\]. For this channel we have the following partial result.
Theorem 15. In the interference channel with transmitter cooperation under the scaled fading model, Fig. 6 if
\[
E[|c_{31}|^2] < E[|c_{21}|^2] < E[|c_{41}|^2]
\]
\[
E[|c_{42}|^2] < E[|c_{12}|^2] < E[|c_{32}|^2]
\]
(199)
multiplexed coding achieves the deterministic capacity in the low power regime. Neither superposition nor network-
ing coding achieves deterministic capacity.

Proof: First, following the argument in the proof of Theorem 7, we can argue that node 2 is able to decode message \( W_1 \) since \( E[|c_{31}|^2] \leq E[|c_{21}|^2] \). Similarly, node 1 can decode message \( W_2 \). We will now argue that additionally node 3 can decode message \( W_2 \). Also this argument is similar to the argument in the proof of Theorem 7 but it is subtly different, so we will provide it here. Define
\[
k_1 = \frac{E[|c_{12}|^2]}{E[|c_{32}|^2]} < 1
\]
\[
k_2 = \frac{E[|c_{21}|^2]}{E[|c_{41}|^2]} < 1.
\]
(200)

Now consider the received signals at nodes 1 and 3,
\[
Y_3[1] = c_{31}[1]X_1[1](W_1) + c_{32}[1]X_2[1](W_2) + Z_3[1]
\]
\[
Y_1[1] = c_{12}[1]X_2[1](W_2) + Z_1[1]
\]
\[
Y_3[2] = c_{31}[2]X_1[2](W_1, Y_1[1]) + c_{32}[2]X_2[2](W_2, Y_2[1]) + Z_3[2]
\]
\[
Y_1[2] = c_{12}[2]X_2[2](W_2, Y_2[1]) + Z_1[2]
\]
\[
Y_3[3] = c_{31}[3]X_1[3](W_1, Y_1[1], Y_1[2]) + c_{32}[3]X_2[3](W_2, Y_2[1], Y_2[2]) + Z_3[3]
\]
\[
Y_1[3] = c_{12}[3]X_2[3](W_2, Y_2[1], Y_2[2]) + Z_1[3]
\]
\[
\vdots
\]
(201)
We will argue that node 3 can estimate $Y_1[n]$. By assumption, node 3 can decode $W_1$. It can therefore calculate

$$Y_3'[1] = Y_3[1] - c_{31}[1]X_1[1](W_1) = c_{32}[2]X_2[1](W_2) + Z_3[1]$$

$$W_1' = k_1Y_3'[1] + Z_1'[1] = c_{12}[1]X_2[1] + k_1Z_3[1] + Z_1'[1]$$

$$Y_3'[2] = Y_3[2] - c_{31}[2]X_1[2](W_1, Y_1'[1]) = c_{32}[2]X_2[1](W_2) + c_{31}[2](X_1[2](W_1, Y_1[1]) - X_1[2](W_1, Y_1'[1])) + Z_3[1]$$

$$W_2' = k_1Y_3'[2] + Z_2'[2] = c_{12}[2]X_2[1] + k_1Z_3[2] + Z_2'[2] + k_1c_{31}[2](X_1[2](W_1, Y_1[1]) - X_1[2](W_1, Y_1'[1]))$$

where $Z_t[n] \sim \mathcal{N}(0, (1 - k_1) N_0 B)$ is i.i.d. Under the deterministic condition, with high probability $X_1[n](W_1, Y_1'[n]) = X_1[n](W_1, W_2)$. Under this condition, $Y_1'[n]$ and $Y_1[n]$ have the same distribution, and therefore with high probability $X_1[n](W_1, Y_1'[n]) = X_1[n](W_1, W_2)$. Since $Y_1'[n]$ and $Y_1[n]$ have the same distribution, and node 1 can decode $W_2$, so can node 3 with high probability. Similarly it can be argued that node 4 can decode $W_1$. We therefore get the bounds

$$R_1 \leq I(X_1; Y_2) \quad (203)$$

$$R_2 \leq I(X_2; Y_1) \quad (204)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y_3) \quad (205)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y_4), \quad (206)$$

which by Lemma 2 result in the low power bounds

$$R_1 \leq |c_{31}|^2 P_1 \quad (207)$$

$$R_2 \leq |c_{32}|^2 P_2 \quad (208)$$

$$R_1 + R_2 \leq |c_{31}|^2 P_1 + |c_{32}|^2 P_2 \quad (209)$$

$$R_1 + R_2 \leq |c_{41}|^2 P_1 + |c_{42}|^2 P_2. \quad (210)$$

This rate is achievable with Block-Markov multiplexed coding according to row 1 of Table III in [15].

Multiplexed coding is not always optimum for the above channel, as the following simple example shows.

**Example 16.** Consider the cooperative interference channel without direct links, Fig. 7. Each transmitter node sees as degraded broadcast channel, and it is therefore clear that each transmitter node should transmit a superposition of $W_1$ and $W_2$. It would seems obvious that if the direct links are weak, this would still be the case, but outer bounds are not strong enough to prove this.

We now consider the interference channel with a relay that can help both transmissions, Fig. 8. The new ingredient in this channel is that node 1 knows only $W_1$ and node 2 only $W_2$. In the previous channel, Fig. 6, both nodes 1

---

7A more formal argument follows the argument in the proof of Theorem 7.
and 2 quickly learn each other’s messages. Because nodes 1 and 2 in Fig. 8 do not have full message knowledge, this is not a straightforward broadcast channel. Still, we have the following partial result

**Theorem 17.** Consider the interference channel with relay, Fig. 8 under a scaled fading model. If

\[
E[c_{31}^2] \leq E[c_{51}^2] \leq E[c_{32}^2] \\
E[c_{42}^2] \leq E[c_{52}^2] \leq E[c_{41}^2]
\]  
(211)

network coding (or multiplexed coding) achieves the low power deterministic capacity. On the other hand, if

\[
E[c_{31}^2] < E[c_{51}^2] \\
E[c_{42}^2] < E[c_{52}^2]
\]  
(212)

and

\[
\frac{E[c_{31}^2]}{E[c_{35}^2]} - \frac{E[c_{41}^2]}{E[c_{45}^2]} \geq 0 \\
\frac{E[c_{42}^2]}{E[c_{45}^2]} - \frac{E[c_{32}^2]}{E[c_{35}^2]} \geq 0
\]  
(213)

superposition coding achieves the low power deterministic capacity under a total power constraint.

**Proof:** If the condition (211) is satisfied, we can argue as in the proof of Theorem [7] that the relay must be
able to decode both $W_1$ and $W_2$. We therefore have the outer bounds

\[
R_1 \leq I(X_1; Y_5) \tag{214} \\
R_2 \leq I(X_2; Y_5) \tag{215} \\
R_1 \leq I(X_1, X_5; Y_3) \tag{216} \\
R_2 \leq I(X_2, X_5; Y_4), \tag{217}
\]

which give the low power bounds

\[
R_1 \leq E[|c_{51}|^2]P_1 \tag{218} \\
R_2 \leq E[|c_{52}|^2]P_2 \tag{219} \\
R_1 \leq E[|c_{31}|^2]P_1 + E[|c_{35}|^2]P_5 \tag{220} \\
R_2 \leq E[|c_{42}|^2]P_2 + E[|c_{45}|^2]P_5. \tag{221}
\]

This can be achieved as follows. Node 1 transmits $X_1(W_1)$ and node 2 transmits $X_2(W_2)$ using random Gaussian codebooks, so that $R_1 \leq E[|c_{51}|^2]P_1$ and $R_2 \leq E[|c_{52}|^2]P_2$. Since $E[|c_{51}|^2] \leq E[|c_{35}|^2]$ and $E[|c_{52}|^2] \leq E[|c_{41}|^2]$, node 3 can decode $W_2$ and node 4 $W_1$. The relay calculates $W_1 \oplus W_2$, and transmits this. Since node 3 knows $W_2$ it can now decode $W_1$ when (220) is satisfied. Similarly for node 4.

Consider now the second case. Again, we can argue that the relay can decode both $W_1$ and $W_2$. Seen from the two sources and the relay, this is then a two receiver non-degraded broadcast channel, and we can use Proposition 5 to obtain the following bounds,

\[
R_1 \leq I(X_1; Y_5) \\
R_1 \leq I(U_1; Y_3) \\
R_1 \leq I(X_1, X_5; Y_3) \\
R_2 \leq I(X_2; Y_5) \\
R_2 \leq I(U_2; Y_4) \\
R_2 \leq I(X_2, X_5; Y_4) \\
R_1 + R_2 \leq I(U_1; Y_3) + I(U_2; Y_4) \tag{222}
\]

\[
R_1 + R_2 \leq I(U_2; Y_4) + I(U_1; Y_5)
\]

\footnote{In the low power regime, the sum rate constraint on the MAC into node 5 disappears.}
for some distribution \(p(u_1, u_2)p(x_1, x_2, x_5|u_1, u_2)\). Using lemmas 2-3, this gives the low power bounds

\[
\begin{align*}
R_1 &\leq E[|c_{51}|^2] \lim_{B \to \infty} \var[X_1] \\
R_3 &\leq E[|c_{31}|^2] \lim_{B \to \infty} \var[X_1] + E[|c_{35}|^2] \lim_{B \to \infty} (\var[X_2] - \var[X_2|U_1]) + E[|c_{55}|^2] \lim_{B \to \infty} (\var[X_5] - \var[X_5|U_1]) \\
R_1 &\leq E[|c_{31}|^2] \lim_{B \to \infty} \var[X_1] + E[|c_{35}|^2] \lim_{B \to \infty} \var[X_5] \\
R_2 &\leq E[|c_{52}|^2] \lim_{B \to \infty} \var[X_2] \\
R_2 &\leq E[|c_{42}|^2] \lim_{B \to \infty} \var[X_2] + E[|c_{41}|^2] \lim_{B \to \infty} (\var[X_1] - \var[X_1|U_2]) + E[|c_{45}|^2] \lim_{B \to \infty} (\var[X_5] - \var[X_5|U_2]) \\
R_2 &\leq E[|c_{42}|^2] \lim_{B \to \infty} \var[X_2] + E[|c_{45}|^2] \lim_{B \to \infty} \var[X_5] \\
R_1 + R_2 &\leq E[|c_{31}|^2] \lim_{B \to \infty} \var[X_1|U_2] + E[|c_{35}|^2] \lim_{B \to \infty} \var[X_5|U_2] + E[|c_{42}|^2] \lim_{B \to \infty} \var[X_2|U_1] + E[|c_{41}|^2] \lim_{B \to \infty} (\var[X_1] - \var[X_1|U_2]) + E[|c_{45}|^2] \lim_{B \to \infty} (\var[X_5] - \var[X_5|U_2]) + E[|c_{55}|^2] \lim_{B \to \infty} (\var[X_5|U_1]).
\end{align*}
\] (223)

Define

\[
\begin{align*}
P_1 &= \lim_{B \to \infty} \var[X_1] \\
P_{12} &= \lim_{B \to \infty} \var[X_1] - \var[X_1|U_2] \\
P_2 &= \lim_{B \to \infty} \var[X_2] \\
P_{21} &= \lim_{B \to \infty} \var[X_2] - \var[X_2|U_1] \\
P_3 &= \lim_{B \to \infty} \var[X_5] \\
P_{51} &= \lim_{B \to \infty} \var[X_5] - \var[X_5|U_1] \\
P_{52} &= \lim_{B \to \infty} \var[X_5] - \var[X_5|U_2].
\end{align*}
\] (224)
Then
\[ R_1 \leq E|c_{31}|^2 P_1 \]
\[ R_1 \leq E|c_{32}|^2 P_1 + E|c_{35}|^2 P_{31} + E|c_{35}|^2 P_{51} \]
\[ R_1 \leq E|c_{31}|^2 P_1 + E|c_{35}|^2 P_5 \]
\[ R_2 \leq E|c_{32}|^2 P_2 \]
\[ R_2 \leq E|c_{42}|^2 P_2 + E|c_{41}|^2 P_{12} + E|c_{45}|^2 P_{52} \]
\[ R_2 \leq E|c_{42}|^2 P_2 + E|c_{45}|^2 P_3 \]
\[ R_1 + R_2 \leq E|c_{31}|^2 (P_1 - P_{12}) + E|c_{35}|^2 (P_5 - P_{52}) + E|c_{42}|^2 P_2 + E|c_{41}|^2 P_{12} + E|c_{45}|^2 P_{52} \]
\[ R_1 + R_2 \leq E|c_{42}|^2 (P_2 - P_{21}) + E|c_{45}|^2 (P_5 - P_{51}) + E|c_{31}|^2 P_1 + E|c_{32}|^2 P_{21} + E|c_{35}|^2 P_{51}. \]

The odd quantities in this bound are \( P_{21} \) and \( P_{12} \). We can interpret \( P_{21} \) as the amount of power node 1 uses on transmitting \( W_2 \). Of course, node 1 doesn’t have any knowledge of \( W_2 \), so logically we should have \( P_{21} = 0 \). Yet, because of the way the single letter bound in Proposition 5 is developed, this cannot be concluded right away. We will show that \( P_{12} = P_{21} = 0 \) is optimum when the condition (213) is satisfied and total power is considered. First, if \( E|c_{31}|^2 > E|c_{35}|^2 \) it is more efficient to transmit \( W_1 \) directly, and similarly if \( E|c_{42}|^2 > E|c_{45}|^2 \). We will therefore consider the case \( E|c_{31}|^2 \leq E|c_{35}|^2 \) and \( E|c_{31}|^2 \leq E|c_{35}|^2 \). In that case, \( P_1 \) and \( P_2 \) should be as small as possible, just large enough for the relay to be able to decode,

\[ P_1 = \frac{R_1}{E|c_{31}|^2} \]
\[ P_2 = \frac{R_2}{E|c_{32}|^2} \] (226)

The total power consumption is now determined by \( P_5 \). We can solve the inequalities (225) to get \( P_5 \),

\[ P_5 \geq \frac{R_2}{E|c_{42}|^2} + \frac{R_1}{E|c_{45}|^2} - \frac{E|c_{31}|^2 P_1}{E|c_{35}|^2} - \frac{E|c_{42}|^2 P_2}{E|c_{45}|^2} + \left( \frac{E|c_{31}|^2}{E|c_{35}|^2} - \frac{E|c_{41}|^2}{E|c_{45}|^2} \right) P_{12} \]
\[ P_5 \geq \frac{R_2}{E|c_{42}|^2} + \frac{R_1}{E|c_{45}|^2} - \frac{E|c_{31}|^2 P_1}{E|c_{35}|^2} - \frac{E|c_{42}|^2 P_2}{E|c_{45}|^2} + \left( \frac{E|c_{31}|^2}{E|c_{35}|^2} - \frac{E|c_{32}|^2}{E|c_{35}|^2} \right) P_{21} \]
\[ P_5 \geq P_{51} \geq \frac{R_1}{E|c_{35}|^2} - \frac{E|c_{31}|^2 P_1}{E|c_{35}|^2} - \frac{E|c_{32}|^2}{E|c_{35}|^2} P_{21} \]
\[ P_5 \geq P_{52} \geq \frac{R_2}{E|c_{45}|^2} - \frac{E|c_{42}|^2 P_2}{E|c_{45}|^2} - \frac{E|c_{41}|^2}{E|c_{45}|^2} P_{12}. \] (227)

If the condition (213) is satisfied, then it is clear that \( P_5 \) is minimized for \( P_{12} = P_{21} = 0 \), and this in turn implies that \( P_5 = P_{51} + P_{52} \). This is exactly the condition for superposition to be optimum.

What is interesting in this last result is that if the cross links are so strong that opposite nodes can decode, this helps transmission, even when the nodes are not interested in the information they decode. On the other hand, if the cross links are so weak that opposite nodes cannot decode, the information they receive on the cross links does not help at all. Information in the low power regime is only useful if it can be decoded.

**B. Butterfly network**

We finally consider the famous butterfly network from [12], Fig. 9 under a scaled fading model. Node 1 is the source and nodes 7 and 8 are destinations that both need to decode the message \( W \). In the setup in [12] it was
shown that network coding at node 4 was optimum and better than routing. The setup in [12] was a pure networking layer setting. We investigate this network from a physical layer perspective, and in particular from a deterministic capacity point of view. Network coding is a particular case of deterministic capacity. But as we have seen for the network in Fig. 6 networking coding is not sufficient to achieve deterministic capacity. Deterministic capacity allows much more general coding functions. In spite of this, Theorem 18 below will show that routing is optimum from a total power perspective.

Without loss of generality we assume that $E[|c_{21}|^2] \geq E[|c_{31}|^2]$. We further assume that

$$E[|c_{42}|^2] \geq E[|c_{72}|^2]$$
$$E[|c_{43}|^2] \geq E[|c_{83}|^2].$$

As we have seen in several previous example, these conditions ensure that the relay node 4 is able to decode the message $W$, which makes the example non-trivial.

**Theorem 18.** Under a total power constraint, routing achieves the deterministic capacity for the Butterfly network in Fig. 9 when (228) is satisfied.

**Proof:** We have the following outer bounds

$$R_3 \leq I(U_3; Y_3)$$
$$R_{23} \leq I(X_1; Y_2|U_3)$$
$$R \leq I(X_2, X_3; Y_4)$$
$$R_{23} \leq I(X_2; Y_4|X_3)$$
$$R_5 \leq I(X_4; Y_5)$$
$$R \leq I(X_2, X_5; Y_7)$$
$$R_{25} \leq I(X_2; Y_7|X_5)$$
$$R \leq I(X_3, X_5; Y_8)$$
$$R_{35} \leq I(X_3; Y_8|X_5)$$
$$R_{53} \leq I(X_3; Y_8|X_3).$$

These bounds are a straightforward combination of degraded BC bounds and MAC bounds from [11]; see also the

---

**Figure 9.** Butterfly network.
proof of Proposition [10] We also have the relationships

\[ R \leq R_3 + R_5 \]
\[ R_{35} + R_5 = R \]
\[ R_{53} + R_3 = R \]
\[ R_{23} + R_3 = R \]
\[ R_{25} + R_5 = R. \]  (230)

These come from corresponding entropy relationships, e.g. \( H(W) = H(W_3, W_5) = H(W_3|W_5) + H(W_5) \) as node 8 must be able to decode \( W \) from data received from nodes 3 and 5. Using Lemma 23 these result in the low power bounds

\[ R_3 \leq E[|c_{31}|^2]P_{13} \]
\[ R - R_3 \leq E[|c_{21}|^2]P_{12} \]
\[ R \leq E[|c_{21}|^2]P_{12} + E[|c_{43}|^2]P_3 \]
\[ R - R_3 \leq E[|c_{43}|^2]P_2 \]
\[ R_5 \leq E[|c_{54}|^2]P_4 \]
\[ R \leq E[|c_{75}|^2]P_5 + E[|c_{72}|^2]P_2 \]
\[ R - R_5 \leq E[|c_{72}|^2]P_2 \]
\[ R \leq E[|c_{83}|^2]P_3 \]
\[ R - R_5 \leq E[|c_{83}|^2]P_3 \]
\[ R - R_3 \leq E[|c_{83}|^2]P_5 \]
\[ P_1 = P_{12} + P_{13} \]
\[ R \leq R_3 + R_5. \]  (231)

By routing we mean that either a node decodes the full message and forwards it, or does not participate in transmission at all. In terms of the outer bounds, this means that the partial rates \( R_3 \) and \( R_5 \) are either 0 or \( R \). We will prove that that is indeed the case. Fix \( R \) and consider minimization of total power consumption. Notice that the problem is a linear programming problem with (231) the constraints (together with positivity constraints on powers and rates). The solution is is on a vertex; we will just have to argue that this vertex has \( R_3 \) and \( R_5 \) either 0 or \( R \).

To do so we solve the optimization problem. We find total power consumption as a function of \((R_3, R_5)\), and then minimize over \((R_3, R_5)\). At first we ignore the three direct bounds on \( R \). Since all the other bounds each contain
only one power $P$, we directly get the minimum total power solution as

$$
\begin{align*}
P_{13} &= \frac{R_3}{E[|c_{31}|^2]} \\
P_{12} &= \frac{R - R_3}{E[|c_{21}|^2]} \\
P_2 &= \max \left\{ \frac{R - R_3}{E[|c_{42}|^2]}, \frac{R - R_5}{E[|c_{72}|^2]} \right\} \\
P_3 &= \frac{R - R_5}{E[|c_{83}|^2]} \\
P_4 &= \frac{R_5}{E[|c_{54}|^2]} \\
P_5 &= \frac{R - R_3}{E[|c_{85}|^2]} \\
R &\leq R_3 + R_5.
\end{align*}
$$

We now have a linear programming problem in $(R_3, R_5)$, which must have a solution at a vertex, that is $(R_3, R_5) = (0, R)$ or $(R_3, R_5) = (R, 0)$. What remains is to prove that the solution satisfies the bounds on $R$.

Consider $(R_3, R_5) = (R, 0)$ giving

$$
\begin{align*}
P_1 &= \frac{R}{E[|c_{31}|^2]} \\
P_2 &= \frac{R}{E[|c_{72}|^2]} \\
P_3 &= \frac{R}{E[|c_{83}|^2]} \\
P_4 &= 0 \\
P_5 &= 0.
\end{align*}
$$

Since $P_4$ and $P_5$ are zero, nodes 4 and 5 can be eliminated together with the corresponding constraint on $R$. The remaining two constraints on $R$ are

$$
\begin{align*}
R &\leq E[|c_{75}|^2]P_5 + E[|c_{72}|^2]P_2 \\
R &\leq E[|c_{85}|^2]P_5 + E[|c_{83}|^2]P_3,
\end{align*}
$$

which are satisfied.

Consider $(R_3, R_5) = (0, R)$ giving

$$
\begin{align*}
P_1 &= \frac{R}{E[|c_{21}|^2]} \\
P_2 &= \frac{R}{E[|c_{42}|^2]} \\
P_3 &= 0 \\
P_4 &= \frac{R}{E[|c_{54}|^2]} \\
P_5 &= \frac{R}{E[|c_{85}|^2]}.
\end{align*}
$$
In this case, all three sum rate constraints are in effect,
\[ R \leq E[|c_{42}|^2]P_2 + E[|c_{43}|^2]P_3 \]
\[ R \leq E[|c_{75}|^2]P_5 + E[|c_{72}|^2]P_2 \]
\[ R \leq E[|c_{85}|^2]P_5 + E[|c_{83}|^2]P_3. \] (236)

The first and last are satisfied. To satisfy the second it may be necessary to increase either \( P_2 \) or \( P_5 \), depending on which of \( E[|c_{75}|^2] \) and \( E[|c_{72}|^2] \) is larger. However, increasing \( P_2 \) or \( P_5 \) does not change that fact that \( (R_3, R_5) = (0, R) \) for the solution.

In conclusion, the optimum solution has \( R_3 \) and \( R_5 \) either 0 or \( R \), and this is sufficient to conclude that routing is optimum.

VI. DISCUSSION OF RESULTS AND CONCLUSION

This paper has defined the concept of deterministic capacity. One can think of deterministic capacity as a generalization, or perhaps more precise, and abstraction of decode-forward and network coding. The paper has also introduced novel techniques for finding (deterministic) capacity in the low-power regime, by directly calculating limits of single letter bounds. The deterministic capacity of a number of example networks were found. While we do not have a general theory for deterministic capacity of networks, we can identify a number of key principles for networks in the low power regime, principles that also give insight into information theory in general:

- **Message splitting**: A key point is whether or not messages should be split into multiple parts, or in a more abstract sense if relays should forward non-invertible functions of messages. The MIMO relay channels show that for larger networks that is necessary. However, it could still be possible that under a total power constraint, message splitting is not needed. The butterfly network is an example of that, but the bounds we have is not strong enough to show this for general relay or multicast networks.

- **Power splitting**: The real issue is if a message is split, or if a node needs to transmit multiple messages, should it also split its power among these messages. This is highly non-trivial. In the low power regime, the main conclusion about the degraded broadcast channel (using either the single letter bound in [11] or the entropy power inequality in [25]) is that it needs to split its power among the messages. It cannot just transmit a single message from which each receiver can decode different kind of information depending on their channel condition. Whether or not the transmitter uses superposition, TDMA or FDMA is irrelevant in the low power regime; but the power splitting is essential. While this may seem obvious, the examples with the interference channel shows that it is not. With multiplexed coding, for example, a single message is used to carry different kind of information to different receivers without any power splitting: each receiver uses the total power of the received signal to decode what they need. In this context, one can interpret the role of the auxiliary random variable in the bounds for the broadcast channel (either the degraded on in [11] or Proposition 5) as inducing a power split; in fact, that is the only contribution from the auxiliary random variable in all the proofs in this paper. The author believes that this dichotomy between power splitting and multiplexing is one of the essential elements of low power networks.

- **Side information**: the examples for the interference channel shows that side-information is useful in increasing rate. If a node is able to decode a message it does not need, it can use that in helping to decode the message it needs. This is the principle behind index coding [27]. The fundamental question here is if the signal it receives is not strong enough to actual decode the undesired message, can the information in the received signal still be useful? In the channel Fig. 8 in the second case, the received undesired signal is too weak to enable decoding,
and the outer bound shows that the side-information does not help rate in any way. Another example is the degraded broadcast channel: the stronger node can decode the message of the weaker node, and use that to help decode its own message, but the weaker node cannot use the information contained about the stronger node’s message in the received signal for anything; in that sense this is related to power splitting in the sense that power splitting is needed exactly because the side-information cannot be utilized. The author believes this could be a general principle in low-power networks: side-information is only useful if it can be decoded. But of course, the small examples we have is to little to prove this in general.

The type of networks where the current methodology is applicable is mainly limited by the difficulty in finding single letter bounds for networks. In particular, the Marton-Körner type bounds for the broadcast channel [23], [24], Proposition 5 do not have any known generalizations to channels with more than two receivers. Such generalizations, except in particular cases [7], seem very hard to develop. Yet, with such a generalization, the class of networks where deterministic capacity can be found could be dramatically expanded. This is a further motivation to develop single-letter bounds. In fact, another way to look at the results in this paper is that they give another, more concrete, use of single letter bounds. Often, even when single letter bounds can be found, they are infeasible to evaluate. The low power regime gives a new way to evaluate bounds.

REFERENCES


9In fact, in the proof of Theorem 6 we do develop a bound for three receivers.
In this appendix we will show that the upper bound of Theorem 5 is not necessarily achievable with common/private message transmission in the synchronous case. Specifically we will prove that the bound (185) is not achievable by common/private message transmission. Strictly speaking it does not prove that Theorem 5 is not achievable, as it is not proven that any set of positive definite matrices \((X, A, B)\) (with \(X \succ A, B\)) is a valid set of covariance matrices. However, it indicates that the proof technique does not work in general for the synchronous case.

We put \(\|c_{21}\|^2 = \|c_{31}\|^2 = 1\) and \(\angle(c_{21}, c_{31}) = \alpha = 0.4\). We consider the real subspace spanned by \(c_{21}, c_{31}\), and in this define an orthonormal basis by \(\{c_{21}, c_{21}^\perp\}\). Then

\[
\begin{align*}
\ c_{31} &= \begin{bmatrix} \cos \alpha \\ \sin \alpha \end{bmatrix} \\
&= \begin{bmatrix} 0.9211 \\ 0.3894 \end{bmatrix}.
\end{align*}
\]

We now define

\[
\begin{align*}
\ v &= \begin{bmatrix} \cos 0.208 \\ \sin 0.208 \end{bmatrix} = \begin{bmatrix} 0.9784 \\ 0.2065 \end{bmatrix} \\
\ X - B &= vv^H \\
\ B &= c_{21}c_{21}^H \\
\ A &= 0.05c_{31}c_{21}^H \\
\ X &= B + (X - B) = vv^H + c_{21}c_{21}^H \\
\ X - A &= vv^H + c_{21}c_{21}^H - 0.05c_{31}c_{31}^H \\
&= \begin{bmatrix} 1.9149 \\ 0.1841 \end{bmatrix}. \\
\end{align*}
\]
The eigenvalues of $X - A$ are $(0.01720, 1.9328)$, so $X - A$ is positive definite. We now have

$$R_2 = c_{21}^H (X - A)c_{21} = \cos^2(0.208) + 1 - 0.05 \cos^2 \alpha = 1.9149$$

$$R_3 = c_{31}^H (X - B)c_{31} = \cos^2(\alpha - 0.208) = 0.9636$$

$$R_a = c_{31}^H (X - B)c_{31} + c_{21}^H Bc_{21} = 0.9636 + 1 = 1.9636$$

$$R_b = c_{21}^H (X - A)c_{21} + c_{31}^H Ac_{31} = 1.9149 + 0.05 = 1.9649.$$  \hfill (246)

Consider the achievable rate by a common/private messaging scheme in the Gaussian channel. We transmit the common message along a unit vector $u$ and beamform the private messages to their respective destinations. This scheme achieves the following rates

$$R_0 \leq \min \{ c_{21}^H uuc_{21}, c_{31}^H uuc_{31} \} P_0 \tag{247}$$

$$R_2 \leq \| c_{21} \|^2 P_2 + R_0 \tag{248}$$

$$R_3 \leq \| c_{31} \|^2 P_3 + R_0 \tag{249}$$

$$R \leq \| c_{21} \|^2 P_2 + \| c_{31} \|^2 P_3 + R_0 \tag{250}$$

subject to the constraint $P_0 + P_2 + P_3 \leq P$. Put

$$c_0^2 = \max_{\| u \|=1} \min \{ c_{21}^H uuc_{21}, c_{31}^H uuc_{31} \}. \tag{251}$$

In this case it’s easy to see that

$$c_0^2 = \cos^2(\alpha/2) = 0.9605. \tag{252}$$

Now consider the problem of given a rate triple $(R_2, R_3, R)$ minimizing the total power $P$. We have to solve

$$\min \left\{ \frac{R_0}{c_0} + \frac{R_2'}{\| c_2 \|^2} + \frac{R_3'}{\| c_3 \|^2} \right\} \tag{253}$$

subject to

$$R_2' + R_0 \geq R_2 \tag{254}$$

$$R_3' + R_0 \geq R_3 \tag{255}$$

$$R_2' + R_3' + R_0 \geq R. \tag{256}$$

It’s easy to see that the optimum solution is

$$P = \frac{R_2 + R_3 - R}{c_0} + \frac{R - R_3}{\| c_2 \|^2} + \frac{R - R_2}{\| c_3 \|^2}. \tag{257}$$
Inserting \((R_2, R_2, R) = (1.9149, 0.9636, 1.9636)\) from (246) we get

\[
P = \frac{R_2 + R_3 - R}{e_0} + R - R_2 + R - R_3
\]

\[
= 2.0011 > \text{tr}X = 2, \tag{258}
\]

which shows that the solution (246) is not achievable by common/private message transmission.