Recovering Multiplexing Loss Through Successive Relaying Using Repetition Coding

Yijia Fan, Chao Wang, John Thompson, H. Vincent Poor

Abstract

In this paper, a transmission protocol is studied for a two relay wireless network in which simple repetition coding is applied at the relays. Information-theoretic achievable rates for this transmission scheme are given, and a space-time V-BLAST signalling and detection method that can approach them is developed. It is shown through the diversity multiplexing tradeoff analysis that this transmission scheme can recover the multiplexing loss of the half-duplex relay network, while retaining some diversity gain. This scheme is also compared with conventional transmission protocols that exploit only the diversity of the network at the cost of a multiplexing loss. It is shown that the new transmission protocol offers significant performance advantages over conventional protocols, especially when the interference between the two relays is sufficiently strong.

I. INTRODUCTION

A. Background

In the past few years, cooperative diversity protocols [1]–[4], [7]–[12] have been studied intensively to improve the diversity of relay networks. In most of the prior work, a time-division-multiple-access (TDMA) half-duplex transmission is assumed and the most popular transmission
protocol (e.g. [2]) can be described in two steps: In the first step, the source broadcasts the information to all the relays. The relays process the information and forward it to the destination (in either the same or a different time slot) in the second step, while the source remains silent. The destination performs decoding based on the message it received in both steps. We refer to this protocol as the classic protocol throughout the paper.

For digital relaying, where the relay decodes, re-encodes and forwards the message, the simplest coding method is repetition coding [2], [3], where the source and all the relays use the same codebook. This scheme can achieve full diversity and is also practically implementable. Any capacity achieving AWGN channel codes can be used to approach the performance limit of such schemes. The disadvantage of this scheme is that it requires the relays to transmit in orthogonal time slots in the second step in order for the destination to combine effectively the relays’ signals. This will result in a significant multiplexing loss compared with direct transmission. Space-time codes, which were originally applied in multiple-input multiple-output (MIMO) systems, have been suggested for use in relay networks (e.g. [3], [24]). Here all the relays can transmit the signals simultaneously to the destination in the second step and the multiplexing factor is recovered to 1/2. However, this still causes spectral inefficiency for the high signal to noise ratio (SNR) region. In fact, the network capacity in this scenario will become only half of the non-relay network capacity for high SNR, even assuming that the message is always correctly decoded at the relays.

To fully recover the multiplexing loss, much more complicated protocols and coding strategies have been proposed in [9], where new independent random codebooks are used at the relays to transmit the same information as they received from the source, while the relays can adjust their listening times dynamically in the first step. Those approaches, which are based on Shannon’s random coding theory, are currently theoretical and extremely difficult to realize in reality. Practical coding design for relay networks often follows a quite different approach from these theoretical investigations (see [25]–[28] for example).

Instead of using complicated coding schemes, a protocol using repetition coded relaying was proposed in [4](see also [29]) to avoid multiplexing loss for single relay channels. In this protocol,
denoted as protocol I in [4], the source transmits a different message in the second time slot, so that the destination sees a collision of messages from both the relay and the source in the second time slot. Although multiplexing loss is recovered due to the continuous transmission of the source, diversity gain is lost due to the fact that the source transmission in time slot two is not relayed to the destination.

B. Contribution of the Paper

In this paper, we study a transmission protocol based on protocol I in [4] for digital relaying. By adding an additional relay in the network and making the two relays transmit in turn, we show that multiplexing loss can be effectively recovered while diversity/combining gain can still be obtained. Specifically, $L$ codewords can be transmitted in $(L+1)$ time slots with joint decoding at the destination. Our analysis is based on two different scenarios: (a) The instantaneous channel state information (CSI) is known to the receiver and can be fed back to the transmitter; (b) The instantaneous CSI is known to the receiver but is not available to the transmitter. We make the following observations in this paper for the proposed protocol:

- For scenario (a), we derive the achievable rates for this protocol when repetition coding is assumed to be used at the relays. We show that in certain scenarios, the capacity for the network becomes that for a MIMO system with $L$ inputs and $L+1$ outputs and has a multiplexing gain of $L/(L+1)$. Assuming that the relays correctly decode the signal, we show that the proposed protocol offers significant capacity performance advantages over the classic protocol due to its improved multiplexing gain.

- We also discuss the source-relay channel conditions and the interference that arises between the relays. We derive the required channel constraints for the optimal performance of such a protocol as a function of SNR, as well as the achievable rates for different channel conditions. We believe these analyses offer strong insights for adaptive protocol design, where relaying and direct transmission can be combined. Based on our network models we show that the proposed protocol, combined with the direct transmission protocol, can also give a significant capacity performance advantage over the classic protocol, especially when
the two relays are located close to each other.

- We propose a practical low-rate feedback V-BLAST decoding algorithm which approaches the theoretical achievable rates for a slow fading environment.
- For scenario (b), we analyze the diversity multiplexing tradeoff for such a network when $L$ is large, conditioned on the signals being correctly decoded at the relays. We show that in this scenario the network mimics a multiple-input single-output (MISO) system with two transmit and one receive antennas. This means it can offer a maximal diversity gain of two with almost no multiplexing loss.

C. Relations to Previous and Concurrent Work

The idea for successive relaying first appeared in [30]. This study was focused on amplify-and-forward relaying and did not offer insight into the achievable rates and diversity multiplexing tradeoff for such relaying methods. The scheme has been further analyzed for amplify-and-forward relaying in [6], [31] and [32]. In [6] capacity analysis was performed assuming that the direct link is ignored. Hence no diversity can be obtained at the destination. The analysis in very recent papers [31] and [32] make the assumption that the relays are isolated, when there are more than one relay. Also the relay-to-relay link in [31] and [32] acts in a different way from that in our work due to a different relaying mode.

A further work [5] also analyzes the capacity for such schemes when digital relaying is used. One major difference between [5] and our work is that the direct link is ignored in [5] while it is considered in this paper. Therefore the analysis becomes different for the following reasons. First, the scheme in [5] does not offer any cooperative diversity gain, which is a very important benefit that the relay can offer. We will show in this paper that a diversity gain of 2 can be obtained by considering the direct link. Secondly, this also results in very different characteristics in terms of achievable rates and signalling methods due to the additional interference and diversity that the direct link introduces. In this paper we specifically analyze the network capacity under different channel and interference constraints, which were not given in [5]. Also the use of the V-BLAST decoder is unique to our paper. Finally, we note that the capacity analysis discussed in our paper
in fact contains the scenario in [5] as a special case, i.e. the same capacity values as in [5] is obtained on assuming that the channel coefficient for the direct link is zero in our model. Therefore our analysis is more general, and the adaptive protocols introduced here fit better in the context of previous work on this topic [2].

II. Protocol Design

We assume a four-node network model, where one source, one destination and two relays exist in the network. For simplicity, we denote the source as $S$, the destination as $D$, and the two relays as $R1$ and $R2$. We split the source transmission into different frames, each containing $L$ codewords denoted as $s_i$. These $L$ codewords are transmitted continuously by the source, and are decoded and forwarded by two relays successively in turn. Before decoding $L$ codewords, the destination waits for $L+1$ transmission time slots until all $L$ codewords are received, from both direct link and the relay links. It then performs joint decoding of all $L$ codewords. The specific steps for each transmission (reception) time slot for every frame are described as follows:

Time slot 1: $S$ transmits $s_1$. $R1$ listens to $s_1$ from $S$. $R2$ remains silent. $D$ receives $s_1$.

Time slot 2: $S$ transmits $s_2$. $R1$ decodes, re-encodes and forwards $s_1$. $R2$ listens to $s_2$ from $S$ while being interfered with by $s_1$ from $R1$. $D$ receives $s_1$ from $R1$ and $s_2$ from $S$.

Time slot 3: $S$ transmits $s_3$. $R2$ decodes, re-encodes and forwards $s_2$. $R1$ listens to $s_3$ from $S$ while being interfered with by $s_2$ from $R2$. $D$ receives $s_2$ from $R2$ and $s_3$ from $S$. The progress repeats until Time slot $L$.

Time slot $L+1$: $R1$ (or $R2$) decodes, re-encodes and forwards $s_L$. $D$ performs a joint decoding algorithm to decode all $L$ codewords received from the $L+1$ transmission time slots.

The transmission schedule for the first three time slots for each frame is shown in Fig. [1]. Compared with direct transmission, the multiplexing ratio for this protocol is clearly $L/(L+1)$, which approaches 1 for large frame lengths $L$. Unlike protocol III in [4], the destination always receives two copies of each codeword, from both the direct and relay link (a delayed version). This implies that diversity gain can still be realized by this protocol.

The major issue for this protocol to be effectively implemented is to tackle the co-channel
interference at the relays and the destination. As described above, except for the first and last time slot, the relays and the destination always observe collisions from different transmitters (i.e. the source or the relays). Suppression of the interference thus becomes a major problem. We will discuss this problem further in the next two sections.

III. Achievable Rates

When the channel information can be fed back to the transmitter, one can implement adaptive coding and modulation to achieve the system performance limit. Thus capacity is a key measurement in this scenario. We assume a slow, flat, block fading environment, where the channel remains static for each message frame transmission (i.e. $L + 1$ time slots). Note that while this assumption is made for presentation simplicity, the capacity analysis can also be applied to a more relaxed flat block fading scenario, e.g. fast fading where each channel coefficient changes for each time slot. We also assume that each transmitter transmits with equal power (i.e. no power allocation or saving among the source and relays). We denote the channel coefficient between node $a$ and $b$ by $h_{a,b}$, which may contain path-loss, Rayleigh fading, and lognormal shadowing. For simplicity, we denote the capacity function $\log_2 (1 + x)$ by $C(x)$, in which the parameter $SNR$ denotes the ratio of signal power to the noise variance at the receiver.

A. Source-Relay Link

In order for the relays to decode the signals correctly, the source transmission rate should be below the Shannon capacity of the source-relay channels. We express this constraint as

$$R_i \leq C( |h_{S,r_i}|^2 SNR), 1 \leq i \leq L$$

(1)

where $r_i$ is the $i$th element in the $L$ dimensional relay index vector

$$r = \begin{bmatrix} R1 & R2 & R1 & R2 & R1 \cdots \end{bmatrix},$$

(2)

and $R_i$ denotes the achievable rate for $s_i$. 
B. Interference Cancellation Between Relays

One major defect of the protocol is the interference generated among the relays when one relay is listening to the message from the source, while the other relay is transmitting the message to the destination. This situation mimics a two user Gaussian interference channel [13], where two transmitters (the source and one of the relays) are transmitting messages each intended for one of the two receivers (the other relay and the destination). The optimal solution for this problem is still open. We concern ourselves only with suppressing the interference at the relays at this stage (interference suppression at the destination will be left until all \(L\) signals are transmitted). We give a very simple decoding criterion for the relays: if the interference between relays is stronger than the desired signal, we decode the interference and subtract it from the received signals before decoding the desired signal. Otherwise, we decode the signal directly while treating the interference as Gaussian noise.

The achievable rate is therefore based on different channel conditions between the source to relay and the relay to destination links. For example, when \(R_1\) transmits \(s_1\) while \(R_2\) is receiving \(s_2\), if \(|h_{R_1,R_2}| \succ |h_{S,R_2}|\), \(R_2\) first decodes \(s_1\), subtracts it (as the interference), then decodes \(s_2\) (as the desired signal). Therefore, besides the rate constraint proposed in the previous subsection, there will be an additional rate constraint for \(s_1\) to be correctly decoded at \(R_2\), which can be expressed as follows:

\[
R_1 \leq C \left( \frac{|h_{R_1,R_2}|^2 SNR}{1 + |h_{S,R_2}|^2 SNR} \right). \tag{3}
\]

Otherwise if \(s_2\) is decoded directly, treating \(s_1\) as noise, the achievable rate for \(s_2\) is further constrained and can be expressed as

\[
R_2 \leq C \left( \frac{|h_{S,R_2}|^2 SNR}{1 + |h_{R_1,R_2}|^2 SNR} \right). \tag{4}
\]

Note that this decoding criterion applies from the second time slot to the \(L\)th time slot when transmitting each frame. In slot \(i\), inequality (3) can be adapted to a constraint on \(R_{i-1}\) and inequality (4) can be adapted to a constraint on \(R_i\).
C. Space-Time Processing at the Destination

If the transmission rate is below the Shannon capacity proposed by the previous two subsections, the relays can successfully decode and retransmit the signals for all the \( L + 1 \) time slots. The input output channel relation for the relay network is equivalent to a multiple access MIMO channel, which can be expressed as

\[
y = \sqrt{SNR} \begin{bmatrix}
    h_{s,D} & 0 & 0 & 0 & 0 \\
    h_{r_1,D} & h_{s,D} & 0 & 0 & 0 \\
    0 & h_{r_2,D} & h_{s,D} & 0 & 0 \\
    0 & 0 & \ddots & \ddots & \ddots \\
    0 & 0 & 0 & h_{r_{L-1},D} & h_{s,D} \\
    0 & 0 & 0 & 0 & h_{r_{L},D}
\end{bmatrix} s + n, \tag{5}
\]

where \( y \) is the \((L+1) \times 1\) received signal vector, \( s \) is the \( L \times 1 \) transmitted signal vector and \( n \) is an \((L+1) \times 1\) complex circular additive white Gaussian noise vector at the destination. Unlike conventional multiple access MIMO channels, the dimensions of \( y, s \) and \( n \) are expanded in the time domain rather than the space domain. However, the capacity region should be the same, which can be expressed as follows [14]:

\[
R_k \leq \log_2 \left( \det \left( I + h_k h_k^H SNR \right) \right), \tag{6}
\]

\[
R_{k_1} + R_{k_2} \leq \log_2 \left( \det \left( I + SNR \left( h_{k_1} h_{k_1}^H + h_{k_2} h_{k_2}^H \right) \right) \right), \tag{7}
\]

\[
\ldots
\]

\[
\sum_{k=1}^{L} R_k \leq \log_2 \left( \det \left( I + HH^H SNR \right) \right), \tag{8}
\]

where \( h_k \) denotes the \( k \)th column of \( H \). As it is extremely complicated to give an exact description for the rate region of each signal when \( L > 2 \), we will concentrate only on inequalities (6) and (8) to give a sum capacity upper bound for the network in the next subsection. However, as will be shown later in the paper, this bound is extremely tight and is achievable when a space-time
V-BLAST algorithm is applied at the destination to decode the signals in a slow fading scenario.

D. Network Achievable Rates

Combining the transmission rate constraints proposed by the previous three subsections, we provide a way of calculating the network capacity upper bound for the proposed protocol. First, we impose a rate constraint $R_i$ for each transmitted codeword $s_i$. In the first time slot (initialization), we write

$$R_{S,r_1} \leq C \left( |h_{S,r_1}|^2 \text{SNR} \right).$$

(9)

For $(i + 1)$th time slot (for $1 \leq i \leq L - 1$), we calculate the rate constraints based on the decoding criterion at the relays. The calculation can be written as a logical if statement as follows:

if $h_{R_1,R_2} \succ h_{S,r_{i+1}}$,

$$R_i \leq \min \left( C \left( \frac{|h_{R_1,R_2}|^2 \text{SNR}}{1 + |h_{S,r_{i+1}}|^2 \text{SNR}} \right), R_{S,r_i}, C \left( |h_{S,D}|^2 \text{SNR} + |h_{r_i,D}|^2 \text{SNR} \right) \right),$$

$$R_{S,r_{i+1}} \leq C \left( |h_{S,r_{i+1}}|^2 \text{SNR} \right);$$

(10)

else

$$R_i \leq \min \left( R_{S,r_i}, C \left( |h_{S,D}|^2 \text{SNR} + |h_{r_i,D}|^2 \text{SNR} \right) \right),$$

$$R_{S,r_{i+1}} \leq C \left( \frac{|h_{S,r_{i+1}}|^2 \text{SNR}}{1 + |h_{R_1,R_2}|^2 \text{SNR}} \right);$$

(11)

end.

Note that the term $C \left( |h_{S,D}|^2 \text{SNR} + |h_{r_i,D}|^2 \text{SNR} \right)$ represents the constraint expressed by (6). The purpose of the if statement is to select the decoding order at the relay and to decide whether equation (3) or (4) is the correct constraint to apply.

In the $(L + 1)$th time slot, we have

$$R_L \leq \min \left( R_{S,r_L}, C \left( |h_{S,D}|^2 \text{SNR} + |h_{r_L,D}|^2 \text{SNR} \right) \right).$$

(12)
Combining these constraints with the sum capacity constraint expressed by (8), an achievable rate per time slot can then be written as

\[ C_{ach} = \frac{1}{L+1} \min \left( \max_{R_1 \cdots R_L} \left\{ \sum_{i=1}^{L} R_i \right\}, \log_2 \left( \det \left( I + HH^H SNR \right) \right) \right). \] (13)

The first term in the \( \min \) function comes from the calculation described above, the second one comes from equation (8).

E. Interference Free Transmission

From the above discussion of the proposed protocol, it is clear that the interference between relays is one major and obvious factor that can significantly degrade the network capacity performance. However, it has been shown that for a Gaussian interference network, if the interference is sufficiently strong, the network can perform the same as an interference free network [15]. Specifically, for the scenario discussed in our model, if the interference between relays (i.e. the value of \(|h_{R1,R2}|\)) is so large that the following inequality holds

\[ \frac{|h_{R1,R2}|^2 SNR}{1 + |h_{S,r_i}|^2 SNR} \geq \min \left( |h_{S,r_i}|^2 SNR, \left( |h_{S,D}|^2 + |h_{r_i,D}|^2 \right) SNR \right), i = 1 \cdots L, \] (14)

then the relay can always correctly decode the interference and subtract it before decoding the desired message, without affecting the overall network capacity. In this situation, the capacity analysis for the \( i \)th (\( 1 \leq i \leq L \)) transmitted signal as expressed by (10)-(12) can be simplified to

\[ R_i \leq \min \left( C \left( |h_{S,r_i}|^2 SNR \right), C \left( \left( |h_{S,D}|^2 + |h_{r_i,D}|^2 \right) SNR \right) \right). \] (15)

It is obvious that the rate bounds provided by (15) are significantly larger than those provided by (10)-(12).

From the above capacity analysis, it can also be seen that the quality of the source to relay link (i.e. \( h_{S,r_i} \)) is also an important factor that may constrain the network capacity. This has also been justified and discussed in many papers (e.g. [2], [3], [7], [9], [12]). Similar to this previous work, we suggest that \( h_{S,r_i} \) should be compared with \( h_{S,D} \) or \( h_{r_i,D} \) before deciding to relay or
not. For the interference free scenario discussed here, the constraint becomes

$$|h_{S,r_i}|^2 \geq |h_{S,D}|^2 + |h_{r_i,D}|^2, \quad 1 \leq i \leq L. \quad (16)$$

The capacity expressed by (13) can be simplified to

$$C_{ach} = \frac{1}{L + 1} \min \left( \sum_{i=1}^{L} C \left( \left(|h_{S,D}|^2 + |h_{r_i,D}|^2 \right) \text{SNR} \right), \log_2 \left( \det \left( I + HH^H \text{SNR} \right) \right) \right). \quad (17)$$

By Jensen’s inequality [15] it is clear that

$$\sum_{i=1}^{L} C \left( \left(|h_{S,D}|^2 + |h_{r_i,D}|^2 \right) \text{SNR} \right) \geq \log_2 \left( \det \left( I + HH^H \text{SNR} \right) \right). \quad (18)$$

Therefore the rate is equal to the MIMO channel capacity equation with a multiplexing scaling factor:

$$C_{ach} = \frac{1}{L + 1} \log_2 \left( \det \left( I + HH^H \text{SNR} \right) \right). \quad (19)$$

This result shows that the proposed protocol can offer the best capacity performance conditioned on (14) and (16), which guarantees that the relays will correctly decode the message without affecting the network capacity. To summarize, we have the following theorem.

**Theorem 1:** Conditioned on (14) and (16), the capacity for the successive relaying scheme can be expressed as

$$C_{ach} = \frac{1}{L + 1} \log_2 \left( \det \left( I + HH^H \text{SNR} \right) \right)$$

where $H$ denotes the channel matrix in (5).

It should be noted that this high interference scenario (i.e. condition (14)) is *not uncommon* in reality. A practical example is when the two relays (e.g. mobiles) are located close to each other. If the routing techniques are developed to choose these relays, the capacity performance can be significantly improved by applying the proposed protocol. To satisfy condition (16), an adaptive protocol can be developed from the proposed protocol to guarantee that the relays are used only when (16) holds, otherwise direct transmission is assumed. However, for a large dense network of relays, it is even not difficult to find two relays satisfying both (14) and (16). A simple
example is a fixed relay network scenario [16], where the source to relay links are often assumed to be significantly better than the corresponding relay to destination links and the direct link. Therefore both (14) and (16) can be met by choosing the two nearby fixed relays. Furthermore, studies have shown that for a large relay network where many relays exist, choosing the best one or few relays will be preferable to using all the relays in many situations (e.g. [10], [17], [18], [20]). Therefore it is possible that the proposed relay protocol can be combined with relay selection techniques to achieve an even higher capacity gain over the classic multi-cast relay protocol, especially for high SNR conditions.

F. The V-BLAST Algorithm

In this section we apply the low-rate feedback V-BLAST minimum mean squared error (MMSE) algorithm for detecting the signals at the destination. The V-BLAST algorithm was initially designed for spatial multiplexing MIMO systems [21]. For a system with $M$ transmit and $N$ receive antennas, the message at the transmitter is multiplexed into $M$ different signal streams, each independently encoded and transmitted to the receiver. The receiver uses $N$ antennas to detect and decode each signal stream by a V-BLAST MMSE detector [22]. The V-BLAST MMSE detection consists of $M$ iterations, each aimed at decoding one signal stream. For each iteration, the receiver applies the MMSE algorithm to detect and decode the strongest signal while treating the other signals as interference, then subtracts it from the received signal vector. The detection continues until all $M$ signal streams are decoded. The Shannon capacity of this system can be achieved if we assume that each signal is correctly decoded [23]:

$$C = \log_2 \det (I + HH^H \text{SNR}) = \sum_{i=1}^{M} \log_2 (1 + \text{SINR}_i),$$

(20)

where $\text{SINR}_i$ is the output signal to interference plus noise ratio (SINR) for signal $s_i$ in the V-BLAST detector. In order for each signal to be correctly decoded, a low-rate feedback channel can be used to feed the value of $\text{SINR}_i$ back to the transmitter. Adaptive modulation and coding should be applied to make the transmission rate for $s_i$ lower than $\log_2 (1 + \text{SINR}_i)$.

Unlike traditional MIMO systems, when we apply this V-BLAST MMSE detector at the
destination for the proposed protocol, each signal stream is independently encoded along the \textit{time dimension} rather than the space dimension. When considering the rate bound \( R_i \), the same analysis should be made as in Section III. The initialization step is the same as (9). For the \((i + 1)\)th time slot (for \(1 \leq i \leq L - 1\)), based on the same interference cancellation criterion as in Section III, the rate calculation can be performed as follows:

\[
\text{if } h_{R1,R2} \succ h_{S,r_{i+1}}
\]

\[
R_i \leq \min \left( C \left( \frac{\left| h_{R1,R2} \right|^2 \text{SNR}}{1 + \left| h_{S,r_{i+1}} \right|^2 \text{SNR}} \right), R_{S,r_i}, \log_2 (1 + \text{SINR}_{r_i}) \right), \quad R_{S,r_{i+1}} \leq C \left( \left| h_{S,r_{i+1}} \right|^2 \text{SNR} \right); \quad (21)
\]

\text{else}

\[
R_i \leq \min \left( R_{S,r_i}, \log_2 (1 + \text{SINR}_{r_i}) \right), R_{S,r_{i+1}} \leq C \left( \frac{\left| h_{S,r_{i+1}} \right|^2 \text{SNR}}{1 + \left| h_{R1,R2} \right|^2 \text{SNR}} \right); \quad (22)
\]

end.

In the \((L + 1)\)th time slot, we have

\[
R_L \leq \min \left( R_{S,r_L}, \log_2 (1 + \text{SINR}_{r_L}) \right). \quad (23)
\]

The \( \text{SINR}_{r_i} \) denotes the SINR for \( s_i \), which is decoded, encoded and forwarded by relay \( r_i \).

The network capacity is therefore

\[
C_{ach,\text{BLAST}} = \frac{1}{L + 1} \max_{R_1, \ldots, R_L} \left\{ \sum_{i=1}^{L} R_i \right\}. \quad (24)
\]

The condition for interference free transmission discussed in Section \[\text{III-E}\] can be expressed as

\[
C \left( \frac{\left| h_{R1,R2} \right|^2 \text{SNR}}{1 + \left| h_{S,r_{i+1}} \right|^2 \text{SNR}} \right) \geq \min \left( C \left( \left| h_{S,r_i} \right|^2 \text{SNR}, \log_2 (1 + \text{SINR}_{r_i}) \right) \right). \quad (25)
\]
The rate for the $i$th ($1 \leq i \leq L$) signal under this condition can be expressed as

$$R_i \leq \min \left( C \left( |h_{S,r_i}|^2 SNR \right), \log_2 \left( 1 + SINR_{r_i} \right) \right).$$

(26)

Similar to the discussion in Section III-E, we can further apply adaptive protocols or make relay selections in the network to enhance the source to relay links; i.e.,

$$C \left( |h_{S,r_i}|^2 SNR \right) \geq \log_2 \left( 1 + SINR_{r_i} \right),$$

(27)

and it is clear from (20) that (24) equals (19) under conditions (27) and (25). This implies that the V-BLAST algorithm can achieve rate (19) for the protocol if the interference channel between relays and source to relay channels are sufficiently strong.

It can be seen that the conditions in (25) and (27) have a higher probability of being fulfilled than those in (14) and (16) due to the following observation:

$$SINR_{r_i} \leq \left( |h_{S,D}|^2 + |h_{r_i,D}|^2 \right) SNR.$$

(28)

This further implies that the conditions in (25) and (27) are better suited to assist the V-BLAST algorithm to achieve the rate in (19), than those in (14) and (16). We note that in practice these conditions also imply a signalling overhead among the source, relays and destination in order to obtain the required SINR information. Furthermore, we note that V-BLAST might be applied only to a slow fading scenario in which the channel remains unchanged at least in every $L+1$ transmission time slots. This is due to the fact that SINR has to be fed back to the transmitters before the source starts transmitting at the beginning of the $L+1$ time slots.

G. Comparison with Classic Protocols

1) Classic Protocol I: The first classic protocol was presented by Laneman and Wornell [3], where each message transmission is divided into three time slots. In the first time slot, the source broadcasts the message to the two relays and the destination. In the next two time slots, each relay retransmits the message to the destination in turn after decoding and re-encoding it by repetition coding. The destination combines the signals it receives in the three time slots. The
network capacity for this protocol can be written as:

\[ C = \frac{1}{3} \times \min \left( C \left( |h_{S,R1}|^2 SNR \right), C \left( |h_{S,R2}|^2 SNR \right), C \left( \left( |h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2 \right) SNR \right) \right), \]  

(29)

where the term \( \frac{1}{3} \) denotes the multiplexing loss compared with direct transmission.

2) **Classic Protocol II:** A simple improvement of Classic Protocol I is to apply distributed Alamouti codes at the relays [8]. The system uses four time slots to transmit two signals. In the first two time slots the source broadcasts \( s_1 \) and \( s_2 \) to both the relays and the destination. In the next two time slots \( R1 \) transmits \([s_1, -s_2] \) and \( R2 \) transmits \([s_2, s_1] \). The destination uses maximal ratio combining to combine the signals received from all four time slots in order to detect and decode them. The capacity achieved by this protocol can be written as

\[ C = \frac{1}{2} \times \min \left( C \left( |h_{S,R1}|^2 SNR \right), C \left( |h_{S,R2}|^2 SNR \right), C \left( \left( |h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2 \right) SNR \right) \right), \]  

(30)

It is clear that (30) outperforms (29) as it has the same diversity gain but reduced multiplexing loss compared with direct transmission.

In practice, both protocols can be combined with relay selection or adaptive relaying protocols to make sure that

\[ \min \left( C \left( |h_{S,R1}|^2 SNR \right), C \left( |h_{S,R2}|^2 SNR \right) \right) \geq C \left( \left( |h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2 \right) SNR \right) \]  

(31)

when relaying is used. The network under this condition can achieve the best capacity performance (i.e. the third term in (29) and (30)). This result clearly mimics the performance of a 3 × 1 single-input multiple-output (SIMO) or multiple-input single-output (MISO) system.

3) **Performance Comparison:** It can be seen that if the two relays are close to each other so that (14) holds, then condition (16) is more likely to hold than (31). This implies that the best capacity (19) for the proposed protocol can be achieved with a higher probability than that for the classic protocols. We now simply compare the best capacities can be achieved by both
proposed protocol and Classic Protocol II:

\[
G \triangleq \frac{1}{L+1} \mathbb{E} \left[ \log_2 \left( \det \left( I + HH^H SNR \right) \right) \right] \times 0.5 \times \mathbb{E} \left[ C \left( \left( |h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2 \right) SNR \right) \right].
\]

(32)

where \( \mathbb{E} [\bullet] \) denotes the expectation and we assume \( \{ h_{a,b} \} \) is a set of identically, independent distributed (i.i.d), complex, zero mean Gaussian random variables with unit variances. \( G \) is plotted as a function of \( SNR \) in Fig. 2 for different values of \( L \). It is clear that the capacity gain increases as the value of SNR increases. Larger values of \( L \) lead to reduced multiplexing loss and offer higher capacity gains.

IV. DIVERSITY MULTIPLEXING TRADEOFF

When the instantaneous CSI is not known to the transmitter, outages will occur. In this scenario diversity-multiplexing tradeoff is a powerful tool to measure the balance between the rates and error probability. In this section we study further the diversity multiplexing tradeoff [33] for such a protocol. For simplicity our analysis is based on the assumption that the signals are correctly decoded at the relays. We note that this analysis can provide insights on the best possible performance this scheme can offer. We summarize our results in the following.

Theorem 2: Define the diversity gain \( d \) and multiplexing gain \( r \) as those in [33]. Conditioned on the relays correctly decoding the signals (i.e, (14) and (16)), the diversity multiplexing tradeoff for the successive relaying scheme in a slow fading scenario, where the channel coefficients remain the same for \( L + 1 \) time slots, can be expressed as:

\[
d(r) = 2 \left( 1 - \frac{L + 1}{L - r} \right)^+.
\]

(33)

Proof: See Appendix.

As predicted in the previous section, we can see from this theorem that a maximal diversity gain of 2 can be obtained, while the multiplexing gain can be recovered to nearly 1 for large \( L \). This will offer a significant advantage in terms of spectral efficiency, which will be shown through

\footnote{Note that the probability that (14) holds decreases as the SNR increases. Therefore, the theorem offers an upper bound on the performance of such a system at high SNR.}
simulations in the next section. Table I compares the maximal diversity and multiplexing gains between the successive relaying protocol and the classic protocols in a slow fading scenario. Note that for a faster fading scenario where the channel coefficient changes in every transmission time slot, the same theorem still holds if the signal transmitted in each time slot is independently encoded. Furthermore, we note that it is possible to obtain the same diversity-multiplexing tradeoff performance if proper adaptive protocols similar to those in [2] are used to consider the conditions of the source to relay links [36].

V. SIMULATION RESULTS

In this section we make further comparison of the above protocols for different network geometries in terms of achievable rates. We compare only Classic Protocol II with the proposed protocol. As mentioned previously, to achieve better capacity performance in practice, the classic protocols should be combined with adaptive protocols so that relaying is applied only if the source to relay channels are good. There are a number of ways to enable adaptive protocols. Three examples are to base adaptation on one of the following conditions: (a) \( \min (|h_{S,R1}|, |h_{S,R2}|) \geq |h_{S,D}| \), i.e. the source to relay link is better than the direct link; (b) condition (16) holds; or (c) condition (31) holds. Although (b) and (c) fits better with the analysis in this paper, condition (a) is the simplest since it does not require knowledge of the relay to destination links. In the following we will only adopt (a) in the simulations. I.e, if condition (a) is not met, the system will use direct transmission. Similar results would be obtained if condition (b) or (c) were to be adopted instead.

Our simulations are based on three network geometries: cases I, II and III, which are shown in Fig.3. We assume that each \( h_{a,b} \) contains Rayleigh fading, pathloss and independent lognormal shadowing terms. These terms can be written as \( h_{a,b} = v_{a,b} \sqrt{x_{a,b}}^{-\gamma} 10^{\zeta_k/10} \), where \( \{v_{a,b}\} \) is a set of i.i.d. complex Gaussian random variables with unit variances, and \( x_{a,b} \) is the distance between the nodes \( a \) and \( b \). The scalar \( \gamma \) denotes the path loss exponent (in this paper it is always set to 4). The lognormal shadowing term \( \zeta_k \) is a random variable drawn from a normal distribution with a mean of 0 dB and a standard deviation \( \delta = 8 \) (dB). We assume that the distance between
the source and destination is normalized to unit distance. In case I, the distances between the source to relays and relays to destination are all normalized, so the distance between the two relays is therefore \(\sqrt{3}\). In case II, the distance between relays is normalized, while the distance between the source to relays and relays to destinations is \(1/\sqrt{2}\). In case III, the relays are located in the middle region between the source and destination, so that the distance between the source and relays is 1/2 while the distance between the relays is negligible compared with the source to relays links. For the proposed protocol, these three cases represent a meaningful tradeoff between the strength of source to relay channels and the interference channel between the two relays.

We assume \(L = 7\) in the simulation, and the performance for the proposed protocol will certainly increase as \(L\) increases. Fig.4 shows the achievable rates for the proposed protocols (ach rate), the capacity achieved by V-BLAST MMSE detection (VBLAST), the classic protocols (classic) and direct transmission (direct), all averaged over 10,000 channel realizations. It can be clearly seen from all three figures that the V-BLAST algorithm approaches the capacity bounds for the protocol proposed in this paper.

Both Fig.4(a) and Fig. 4(b) imply that it is generally not helpful to implement relaying protocols when the source to relay link is about the same quality as the source to destination (direct) link, as the link gain due to relaying is small in this case. However, the proposed protocol still offers a performance gain over direct transmission for both the high and low SNR regions in these cases. Compared with case I and II, in case III the source to relay links are much stronger, and the relays become close to each other so that the interference is sufficiently strong to allow interference free transmission, as discussed in Sections III and IV. It can be clearly seen in Fig. 4(c) that the proposed protocol gives a significant performance advantage over direct transmission for both low and high SNR regions due to its combining gain and negligible multiplexing loss. The classic protocol still performs worse than direct transmission due to its significant multiplexing loss compared with direct transmission, although its performance gain over direct transmission for the low SNR region is improved.
VI. CONCLUSIONS

In this paper we have analyzed successive relaying protocol. Our analysis shows that this protocol can maintain combining/diversity gain while recovering the multiplexing loss associated with the classic protocol. We have proposed the use of a low complexity V-BLAST detection algorithm to help implement this protocol effectively. From the simulation study based on different geometries, we can draw two main conclusions: (a) For both the proposed and classic protocols, the network capacity increases when the source-relay link becomes stronger; (b) in this scenario, while the classic protocol still loses its performance advantage for the high SNR region, the proposed protocol can give significant performance advantages for both the low and high SNR regions.

Note that one very important factor that impairs the capacity performance of the proposed protocol is interference between the two relays. Our capacity analysis does not offer the optimal capacity results for this protocol because the optimal method of suppressing the interference between the relays is not known in general. For the adaptive protocol discussed in the paper, it is also worthwhile to develop alternative forms of the protocol that explicitly account for the impact of interference between relays on the network capacity. Also it should be interesting to extend the analysis into a more than two relay scenario. These are interesting topics for future work.

APPENDIX

PROOF OF Theorem 1

As mentioned in Section III.E, conditioned on the event that the relays correctly decode the message, the successive relaying protocol mimics a multiple access MIMO channel (5) with a capacity constraints (6) - (8). For each constraint there is a probability of not meeting it. The probability of outage is the highest among all these probabilities. Therefore there are \((2^L - 1)\) diversity-multiplexing tradeoffs for all those conditions and the lowest curve within the range of multiplexing gain is the optimal tradeoff curve for the system [35]. To characterize the diversity-multiplexing tradeoff achieved by each constraint, we consider an \((m + 1) \times m\) MIMO channel.
matrix $H_m$ in the same form as in (5). Define $v_0$ as the exponential order [9] of $1/|h_{S,D}|^2$ and $v_k$ as the exponential order of $1/|h_{r_k,D}|^2$. Furthermore, Let $M_{m+1} = I + \frac{1}{2} \Sigma_S \Sigma_n^{-1}$, where $\Sigma_S$ and $\Sigma_n$ denote the covariance matrices of the observed signal and noise components at the receiver, respectively. We assume that each source message $s_i$ is chosen from a Gaussian random codebook of codeword length $l$. When $m = 1$, the upper bound on the ML conditional pair-wise error probability (PEP) can be calculated by

$$P_{PE|v_0,v_1} \leq \det \left( I + \frac{1}{2} \Sigma_S \Sigma_n^{-1} \right)^{-l}$$

$$= \left( 1 + \frac{1}{2} \rho |h_{S,D}|^2 + \frac{1}{2} \rho |h_{r_1,D}|^2 \right)^{-l}$$

$$= \rho^{-l \left( \max \{1 - v_0, 1 - v_1 \} \right)^+}$$

(34)

where $\doteq$ denotes the exponential equality [33] and SNR is replaced by $\rho$ for notational simplicity.

We assume each $s_i$ is transmitted with data rate $R$ bits in each transmission time slot. Since the successive relaying protocol uses $(L + 1)$ time slots to transmit $L$ different symbols, the average transmission rate is $\overline{R} = \frac{L+1}{L+1} R$. On assuming that average transmission rate changes as $\overline{R} = r \log \rho$ with respect to $\rho$, then it is easy to see $R = \frac{L+1}{L} r \log \rho$. Therefore, we have a total of $\rho^{\frac{L+1}{L} r l}$ codewords. Thus, the error probability can be bounded by

$$P_{E|v_0,v_1} \leq \rho^{-l \left( \max \{1 - v_0, 1 - v_1 \} \right)^+} \cdot \overline{R}.$$  

(35)

Next, we want to find the set in which the outage event always dominates the error probability performance. The analysis regarding this is similar to that in [9] and is thus omitted here. This set is given by

$$O^+ = \left\{ (v_0, v_1) \in R^2^+ \middle| \left( \max \{1 - v_0, 1 - v_1 \} \right)^+ \leq \frac{L+1}{L} r \right\}.$$  

(36)

Then, for any error event which belongs to the non-outage set, we can choose $l$ to make its probability sufficiently small to ensure that the error performance is dominated by the outage probability, which can be expressed as $\rho^{-d_o(r)}$ for $d_o(r) = \inf_{(v_0,v_1) \in O^+} (v_0 + v_1)$. Now using (36),
$d_o(r)$ can be calculated as

$$d_o(r) = 2 \left( 1 - \frac{L + 1}{L} r \right)^+$$  \hspace{1cm} (37)$$

which represents the diversity-multiplexing tradeoff in the case $m = 1$. When $m \geq 2$, the analysis of the determinant of $M_{m+1}$ can be conducted in a way similar to that in [32]; so we omit the specific calculation due to limited space. Define $D_k := \det(M_{(k)})$, where $M_{(k)}$ denotes a $k \times k$ sub-matrix formed by the first $k$ rows and $k$ columns from the upper left-most corner of $M$. The coefficients of $D_{m+1}$ can be calculated recursively as

$$D_{m+1} \left( \frac{1}{2} \rho |h_{S,D}|^2 \right)^m + \prod_{j=1}^{m} \left( 1 + \frac{1}{2} \rho |h_{r_j,D}|^2 \right) + P \left( \frac{1}{2} \rho |h_{S,D}|^2 \right)$$

where $P(\frac{1}{2} \rho |h_{S,D}|^2)$ is a polynomial in $\frac{1}{2} \rho |h_{S,D}|^2$ and is always nonnegative. Thus, we have

$$D_{m+1} \geq \left( \frac{1}{2} \rho |h_{S,D}|^2 \right)^m + \prod_{k=1}^{m} \left( 1 + \frac{1}{2} \rho |h_{r_k,D}|^2 \right).$$  \hspace{1cm} (38)$$

Since we assume a slow fading environment, $v_1 = v_3 = \ldots$ and $v_2 = v_4 = \ldots$. On setting $v = \max\{v_1, v_2\}$, it can be seen that

$$\det(I + \Sigma_{\text{S}} \Sigma_{\text{n}}^{-1}) \geq \rho^{\max\{m(1-v_0)^+, m(1-v)^+\}}.$$  \hspace{1cm} (39)$$

If we define $\det(I + \Sigma_{\text{S}} \Sigma_{\text{n}}^{-1}) \triangleq \rho^{f(v_0, v_1, v_2)}$ and

$$\rho^{\max\{m(1-v_0)^+, m(1-v)^+\}} \geq \rho^{g(v_0, v_1, v_2)},$$

then we have

$$f(v_0, v_1, v_2) \geq g(v_0, v_1, v_2), \quad \forall (v_0, v_1, v_2) \in \mathbb{R}^3^+.$$  \hspace{1cm} (41)$$

Similarly to the analysis for $m = 1$, $O_f^+$ should be defined as

$$O_f^+ = \left\{ (v_0, v_1, v_2) \in \mathbb{R}^3^+ \mid f(v_0, v_1, v_2) \leq \frac{L + 1}{L} mr \right\}.$$
where $m$ denotes that $m$ symbols are transmitted and the equivalent data rate $R = \frac{L+1}{L}mr \log \rho$.

We define

$$O_g^+ = \left\{ (v_0, v_1, v_2) \in R^{3^+} \mid f(v_0, v_1, v_2) \leq \frac{L+1}{L}mr \right\}. \quad (43)$$

Because of (41), it can be seen that $O_f^+ \subseteq O_g^+$. Therefore

$$\inf_{(v_0, v_1, v_2) \in O_f^+} (v_0 + v_1 + v_2) \geq \inf_{(v_0, v_1, v_2) \in O_g^+} (v_0 + v_1 + v_2),$$

which means that the diversity gain calculated from $O_f^+$ is always larger than that calculated from $O_g^+$.

From (40) and (43), it is not difficult to show that

$$\inf_{(v_0, v_1, v_2) \in O_f^+} (v_0 + v_1 + v_2) \geq 2 \left( 1 - \frac{L+1}{L}r \right)^+.$$

(44)

Comparing (37) and (44), we can see that the diversity gain achieved by a multiple access MIMO channel with channel matrix $H_m$ ($m > 1$) is always larger than that for $H_1$.

Now we consider the product of the determinants of $n$ matrices $\prod_{i=1}^{n} \det(I + \frac{1}{2}\Sigma S_{m_i} \Sigma^{-1})$, which is related to all other rate constraints from (6) - (8). Using (39), it is easy to obtain

$$\prod_{i=1}^{n} \det \left( I + \frac{1}{2}\Sigma S_{m_i} \Sigma^{-1} \right) \geq \rho_{\max} \{(\Sigma_{i=1}^{n} m_i)(1-v_0)^+, (\Sigma_{i=1}^{n} m_i)(1-v_1)^+, (\Sigma_{i=1}^{n} m_i)(1-v_2)^+\}.$$

Define $\rho_f^{n}(v_0, v_1, v_2) \equiv \prod_{i=1}^{n} \det(I + \frac{1}{2}\Sigma S_{m_i} \Sigma^{-1})$ and $\rho_g^{n}(v_0, v_1, v_2) \equiv \rho_{\max} \{(\Sigma_{i=1}^{n} m_i)(1-v_0)^+, (\Sigma_{i=1}^{n} m_i)(1-v_1)^+, (\Sigma_{i=1}^{n} m_i)(1-v_2)^+\}$. It can be seen that

$$f_n(v_0, v_1, v_2) \geq g_n(v_0, v_1, v_2), \quad \forall (v_0, v_1, v_2) \in R^{3^+}. \quad (45)$$

Similarly, applying $O_g^+ = \{(v_0, v_1, v_2) \in R^{3^+} \mid g(v_0, v_1, v_2) \leq \frac{L+1}{L}(\Sigma_{i=1}^{n} m_i)r \}$, we have that

$$\inf_{v_0, v_1, v_2 \in O_g^+} (v_0 + v_1 + v_2) \geq 2 \left( 1 - \frac{L+1}{L}r \right)^+. \quad (45)$$

The determinant of the matrix $(I + \frac{1}{2}\Sigma S \Sigma^{-1})$ can always be decomposed into the product of the determinants of several submatrices $(I + \frac{1}{2}\Sigma S_{m_i} \Sigma^{-1})$. Therefore the error exponent is always
larger than or equal to \(2(1 - \frac{L+1}{L} r)^+\) and the proof is complete.

**REFERENCES**

[1] A. Sendonaris, E. Erkip and B. Aazhang, “User cooperative diversity-Part I and II,” *IEEE Trans. Commun*, vol. 51, no. 11, pp. 1927-1938, Nov. 2003.

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

[3] J. N. Laneman and G. W. Wornell, “Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks,” *IEEE Trans. Inf. Theory*, vol. 49, pp. 2415-2425, Oct. 2003.

[4] R. U. Nabar, et al., “Fading relay channels: Performance limits and space-time signal design,” *IEEE J. Sel. Areas Comm.*, vol. 22, no. 6, pp. 1099-1109, Aug. 2004.

[5] B. Rankov and A. Wittneben, “Spectral efficient signaling for half-duplex relay channels,” in *Proc. Asilomar Conf. Signals, syst., comput.*, Pacific Grove, CA, Nov., 2005.

[6] B. Rankov and A. Wittneben, “Spectral efficient protocols for non-regenerative half-duplex relaying,” in *Proc. 43th Annual Allerton Conf. Comm., Contr. and Comp.*, Monticello, IL, Oct. 2005.

[7] E Zimmermann, P. Herhold and G. Fettweis, “On the performance of cooperative diversity protocols in practical wireless systems,” *Proc. 2003 IEEE Vehicular Tech. Conf. - Fall*, Orlando, Florida, Oct. 2003.

[8] P. Razaghi and W. Yu. “Bilayer low-density parity-check codes for decode-and-forward in relay channels,” submitted to *IEEE Trans. Info. Theory*, February 18, 2013.
[29] A. Wittneben and B. Rankov, “Impact of cooperative relays on the capacity of rank-deficient MIMO channels,” *Proc. 12th IST Summit on Mobile and Wireless Commun.*, Aveiro, Portugal, pp. 421-425, June 2003.

[30] T. Oechtering and A. Sezgin, “A new cooperative transmission scheme using the space-time delay code,” *Proc. ITG Workshop Smart Antenna*, pp. 41-48, Zurich, Mar. 2004.

[31] S. Yang, J.-C. Belfiore, “Towards the optimal amplify and forward cooperative diversity scheme,” presented at the *35th IEEE Comm. Theory. Workshop*, Dorado, Puerto Rico, May 2006.

[32] S. Yang, J.-C. Belfiore, “On slotted amplify-and-forward cooperative diversity schemes,” *Proc. 2006 IEEE Int’l. Sym. Inf. Theory*, Seattle, USA, July 2006.

[33] L. Zheng and D. Tse, “Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels,” *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073-1096, May 2003.

[34] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

[35] D. Tse, P. Viswanath and L. Zheng, “Diversity multiplexing tradeoff in multiple access channels,” *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1859 - 1874, Sep 2004.

[36] C. Wang, Y. Fan, J. S. Thompson and H. V. Poor, “Adaptive protocols for relay channels,” in preparation.
Fig. 1. Transmission schedule for the proposed protocol.

| Schemes/Maximum Gain | Multiplexing | Diversity |
|----------------------|--------------|-----------|
| Direct transmission  | 1            | 1         |
| Classic I            | 1/3          | 3         |
| Classic II           | 1/2          | 3         |
| Proposed scheme      | $L/(L+1)$    | 2         |

**TABLE I**

Comparison of the different transmission schemes for the two relay case.
Fig. 2. Capacity gain of the proposed protocol over classic protocol II.

Fig. 3. Network models for different geometries.
Fig. 4. Average capacity of the network for different network geometries in bits per transmission time slot.