MIMO Communications over Relay Channels

Yijia Fan

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
December 2006

Abstract

The use of multiple antennas at both ends of a point-to-point wireless link, called multiple-input multiple-output (MIMO) technology, promises significant improvements in terms of spectral efficiency and link reliability. A large amount of research has been focussed on the point-to-point MIMO systems in the last decade. However, its application in future generation wireless networks has not been thoroughly investigated. It is also widely believed that ad hoc networking or multi-hop cellular networks are important new concepts for future generation wireless networks, where either mobile or fixed nodes (often referred to as relays) are used to help forward the information to the desired user. In this scenario, the information is transmitted in a relay channel rather than a point-to-point channel. The potential benefit of incorporating MIMO technology into relay channels is of significant importance and an interesting topic.

In this thesis we investigate MIMO techniques for relay channels. We first concentrate on the single antenna relay networks, where each node is equipped with a single antenna. In this scenario, a scheme called *cooperative diversity* has been widely discussed, where multiple relays are united in the network as a "virtual antenna array", to mimic a MIMO system. We propose a novel cooperative diversity scheme that can improve the spectral efficiency of the network, especially for high signal to noise ratios (SNR). We analyze the capacity bounds for such schemes and also describe a signalling method to approach this capacity bound.

We then move to the multi-antenna node scenario, where each node is equipped with multiple antennas. We propose different signalling methods and routing protocols for MIMO relay channels and use capacity as a performance metric to evaluate and compare them. The proposed signalling methods can be applied together with the proposed routing schemes. Incorporating them can facilitate the cross-layer design.

Finally, we discuss a network scenario where some nodes are equipped with multiple antennas, others are equipped with single antennas. We constrain ourselves to the case where the source and destination are equipped with a single antenna. We characterize the capacity performance and the diversity-multiplexing tradeoff of such a network. We show that relaying can offer a significant performance advantage over non-relay transmission in certain scenarios, by applying signal combining techniques for the point-to-point MIMO link into relay channels.

Acknowledgements

First of all, I would like to express my appreciation to my supervisor Dr. John S Thompson, without whom I could never complete this work. It is him who offers me the opportunity to do a PhD in the signal and system group, leads me to the world of signal processing and communication research at the beginning of my postgraduate years, teaches me to think in a ture scientific way, and most importantly, offers me significant insights during my second and third years of PhD program. His sense of responsibility and consideration set the model for my future career and life. Working together with him is the most meaningful experience and precious memory in my life.

Secondly, I would like to thank EPSRC and Nortel for offering me the full funding during my three year PhD program. Furthermore, I would like to thank Dr. Mark Naden for his expert help and advice.

Also I would like to thank Professor Halim Yanikomeroglu and Dr. Abdulkareem Adinoyi in the systems and computer engineering department, Carleton university, for their meaningful discussion which gives rise to a part of the contribution in my thesis.

Additionally, I would like to thank my friends Yifan Li and Yuanyuan Fei for their continual help and care. It has been a great time sharing the life with them in Edinburgh.

Last but not least, I would like to thank my family for their emotional support during my PhD. It is their love and encouragement that helps me go through the hard time in the past three years abroad.

Contents

List of figures

List of tables

Acronyms and abbreviations

3G	Third generation
AMPS	Advanced mobile phone service
ARQ	Automatic repeat-request
BLAST	Bell Labs Layered Space-Time Architecture
BS	Base station
CSI	Channel state information
D-BLAST	Diagonal Bell Labs Layered Space-Time Architecture
DBR	Digital beamforming relaying
DDF	Dynamic decode and forward
DR	Digital relaying (decode-and-forward relaying)
FDD	Frequency division duplex
GSM	Globe system for mobile
HBR	Hybrid beamforming relaying
HR	Hybrid relaying (filter and forward relaying)
I.i.d.	identically independent distributed
LAN	Local area networks
LOS	Line of sight
MAR	Modified analogue relaying
MDR	Modified digital relaying
MFR	Matched filter based relaying
MIMO	Multiple-input multiple-output
MISO	Multiple-input single-output
ML	Maximum likehood
MMFR	Modified matched filter relaying
MMSE	Minimum mean squared error
MR	Multicast routing
MRC	Maximal ratio combining
MS	Mobile station
OFDM	Orthogonal frequency division multiplexing

OHR	Optimal hybrid relaying
OSR	Optimal selective routing
PSSR	Pathloss and shadowing based selective routing
QoS	Quality of service
SIMO	Single-input multiple-output
SINR	Signal to noise plus interference ratio
SISO	Single-input single-output
SNR	Signal to noise ratio
SVD	Singular value decomposition
TB	Transmit beamforming
TDD	Time division dual
TDMA	Time division multiple access
TSTNR	Total signal to total noise ratio
V-BLAST	Vertical Bell Labs Layered Space-Time Architecture
ZF	Zero forcing

Nomenclature

[1]

P_{out}	Outage probability
$SINR$	Signal to noise plus interference ratio
SNR	Signal to noise ratio
T_c	Coherence time
T_d	Delay spread
arg	index for the selected value
$\mathbf{G}_k$	Channel transfer matrix from the kth relay to the destination
$\mathbf{H}$	Channel transfer matrix
$\mathbf{H}_k$	Channel transfer matrix from the source to the kth relay
$\mathbf{I}$	Identity matrix
$\mathbf{I}_L$	Identity matrix with dimension L
$\mathbf{h}$	Channel vector
$\mathbf{n}$	Noise vector
$\mathbf{s}$	Transmitted signal vector
δ	standard deviation (dB)
$\det(.)$	determinant of a matrix
η	Transmit power
$\exp\{.\}$	exponential function
γ	pathloss exponent
γ^i	Waterfilling allocations
λ^i	The ith eigenvalue of a matrix
$\lVert\cdot\rVert_F$	Frobenius-norm
ρ_d^1	SNR at the destination when each relay is equipped with one antenna
ρ_d^N	SNR at the destination when the relay is equipped with N antennas
$\rho_d^{m_k}$	SNR at the destination when the kth relay is equipped with m_k antennas
τ	Total signal to total noise ratio

[1]Only partial notations are listed. Some other notations will be given and explained specifically in certain parts of the thesis, while redefinition of the same symbol might occur.

ζ	Normally distributed random variable
$h_{a,b}$	Channel coefficient between node a and b
n	white Gaussian noise
s	Transmitted signal
$\mathbf{R}_{ss}$	Covariance matrix for transmitted signal vector
$\text{diag}\{.\}$	Diagonal matrix
$\text{E}[.]$	Mathematical expectation

Chapter 1
Introduction

1.1 Introduction

Wireless communications began over 100 years ago with the invention of the radiotelegraphy by Guglielmo Marconi, and became a rapidly developing field in industry since the cellular concept was conceived by Ring at Bell Labs in 1947. This idea required dividing the service area into smaller cells. Each cell is covered by a fixed base-station, which offers phone services for the wireless subscribers who are within the cell.

Over the past few decades, wireless cellular systems around the world have gone through several phases, first came the analogue age, such as the advanced mobile phone service (AMPS) proposed in 1970; then came the digital age, such as the global system for mobile (GSM), the time-division-multiple-access (TDMA) standard developed in USA (IS-136), and code-division-multiple-access (CDMA) (IS-95). With increasing use of wireless internet in the late 1990s, the demand for higher spectral efficiency and data rates has led to the development of the so called third generation (3G) wireless technologies. Though cellular systems were originally developed for telephony (voice services), current and future systems (3G and beyond) are designed to handle both data and/or voice [2]. While some of the 3G systems are essentially evolutions of previous cellular systems, others are designed to fulfill different types of data services. An example is the *ad hoc network* [3, 4]. Here, instead of having a base station, all nodes are equivalent. The network organizes itself into links between various pairs of nodes and develops routing tables using these links. This type of system is often used in certain local area networks (LAN) such as battlefield due to its lack of centralized control and limited node capability. However, in future generation wireless systems, it is possible to incorporate the ad hoc network into cellular network structure. The key benefits underlying this combination is improved capacity and coverage enhancement due to *relaying*.

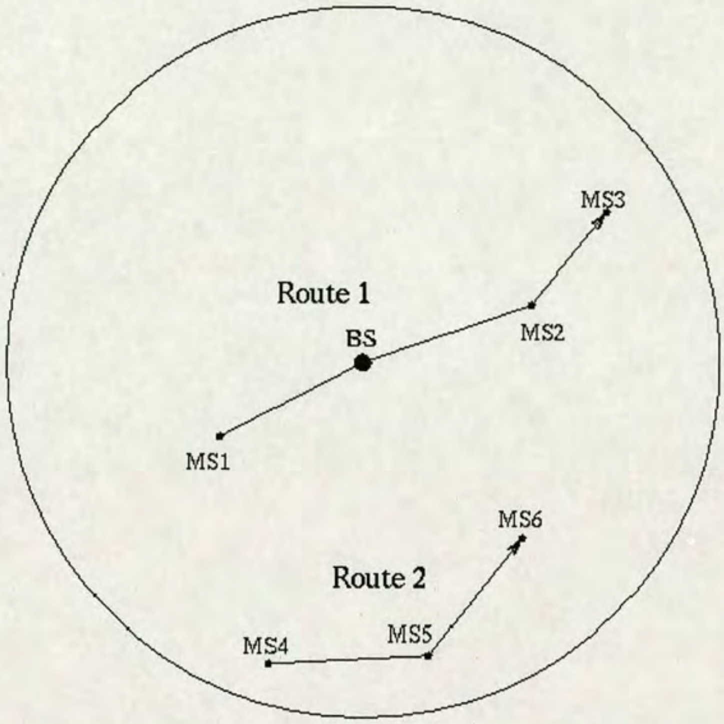


Figure 1.1: *An example of routing in one cell of a multi-hop cellular network, where BS denotes base station, MS denotes mobile station.*

1.2 Relay networks

Relaying is a technique where a node (relay) helps another node (source) to forward the information to the desired node (destination). Relaying can be especially convenient for wireless networks due to the broadcast nature of wireless communications. In a multi-node network, a transmission from any node not only reaches the desired node, but all other nodes. This offers the chance to exploit the space dimension in a wireless network by applying relaying at different nodes to transmit the message cooperatively.

An example of a relay network is the so-called multi-hop cellular or mesh network[5, 6], which combines the benefits of ad hoc networks and cellular networks. Figure 1.1 offers a simplified illustration of a cell. The base station (BS) is in the center, while several mobile stations (MS) are within the cell. When the source or destination are far away from the base station, signals can be forwarded to the desired user via one or more intermediate transceivers (i.e. relays), which are located in a more favorable location relative to the base station (e.g. Route 1). Relays can be either fixed nodes or mobile nodes in this scenario. Furthermore, when the users are located close to each other, an ad hoc network environment can be created so that the nodes can transmit the signals cooperatively, without resorting to the base station. Another example of network structures applied relaying includes the sensor network[7].

Generally speaking, two main benefits might be obtained through applying relaying in a network.

- Capacity and coverage enhancement: One fundamental aspect of wireless communication is the phenomenon of *fading*: The time variation of the channel strengths due to the small-scale effect of multipath fading, as well as larger-scale effects such as path loss via distance attenuation and shadowing by obstacles. The large-scale fading limits the throughput and coverage of the whole network. For example, in a cellular network, users near the cell edge or behind a large obstacle might not be able to communicate with the base station efficiently. In this scenario, using relays which have better links to both the users and the base station will significantly improve the energy efficiency, the capacity between the user and the destination, and the cell coverage.
- Link reliability improvement: Even if the users are located in a good position relative to the base station, the small-scale fading can still impair the link quality whenever the channels are in a *deep fade*. By applying relaying, the destination might receive multiple copies of the same message from different fading channels. Therefore, it is possible to avoid deep fading through exploiting the *diversity* of the relay channel. One way is to apply signal combining at the destination to average out the overall fading effect.

In general, it is widely believed that relaying is going to be a key technology for future generation wireless networks [2, 5].

1.3 Multiple-input multiple-output (MIMO) wireless systems

Another technology exploiting the space dimension of wireless communications is the so-called multiple-input multiple-output (MIMO) technology, where multiple antennas are deployed at both ends of a point-to-point wireless link. MIMO technology originated from the *smart antenna array* techniques dating back several decades[8], and was initially implemented in a military context. The purposes for using antenna arrays were mainly to provide improved space diversity gain and to suppress co-channel interference and noise in adverse fading channels and

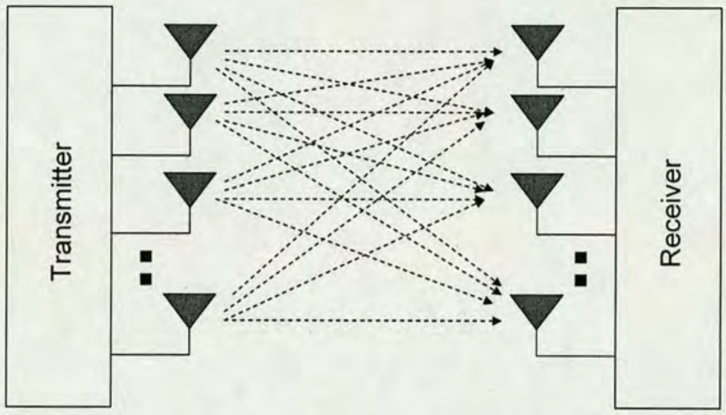


Figure 1.2: *A point-to-point MIMO system.*

hence obtain an acceptable error performance for each user in the system. The concept of antenna arrays was extended as MIMO for use in civilian cellular systems and was first introduced by Jack Winters in 1987[9], for two basic communication systems. These are (1) communication between multiple terminals and base station with multiple antennas and (2) communication between two terminals each with multiple antennas. While research began on the first type in the early 1990s[10–14], since the mid 1990s much more effort has been focused on the second type [15–20], which is shown in Figure 1.2.

The benefits of a MIMO system over single-input single-output (SISO) system are mainly due to two aspects. One is its diversity gain, which is often obtained by transmitting the same message through different sub-channels and then combining them at the receiver. Examples exploiting the diversity of MIMO system include space-time block coding [21–23], transmit beamforming[24] and antenna selection[25]. As discussed in Section 1.2, diversity can significantly enhance the link reliability of the system. The other benefit is the so-called *multiplexing gain*, which can be considered as the number of data streams a MIMO system can simultaneously support. Generally speaking, a MIMO system with M transmit antennas and N receive antennas can support at most $\min(M, N)$ data streams. This also means that for high signal to noise ratio (SNR) the capacity for a MIMO system is $\min(M, N)$ times as that of a SISO system. Multiplexing gain can be maximized by certain spatial multiplexing structures such as BLAST [15, 26].

MIMO technology has been adopted into 3G mobile and fixed wireless standards such as IEEE 802.11, and is promising for future generation wireless systems.

1.4 Contributions

A lot of research has been conducted for both relay networks and MIMO systems. However, there is not much work focusing on the combination of them. The potential benefit of incorporating MIMO technology into relay channels is of great importance, and is the main focus of the thesis. We divide our research into three scenarios, namely single antenna relay channels, multiple antenna relay channels, and single and multiple antenna relay channels.

In single antenna relay channels, each node is equipped with a single antenna. In this scenario, a scheme called *cooperative diversity* has been discussed widely, where multiple relays are united in the network to form a "virtual" MIMO system. We propose a novel transmission scheme that can improve the spectral efficiency of the network, especially for high SNRs. We analyze the capacity bounds for such schemes and also describe a signalling method to approach this capacity bound.

We then move to multiple antenna relay channels, where each node is equipped with multiple antennas. We propose different signalling methods and routing protocols for MIMO relay channels and use capacity as a performance metric to evaluate and compare them. The proposed signalling methods can be applied together with the proposed routing schemes. Incorporating them can facilitate cross-layer design for MIMO relays.

Finally, we discuss a "hybrid" scenario where some nodes are equipped with multiple antennas, and others are equipped with single antennas. We constrain ourselves to the case where the source and destination are equipped with a single antenna. We characterize the capacity performance of such a network. We show that relaying can offer a significant performance advantage over non-relay transmission in certain scenarios by applying signal combining techniques for the point-to-point MIMO link into relay channels.

1.5 Structure of thesis

The rest of the thesis is organized as follows. In Chapter 2, we give an introduction to major topics which will be useful throughout the thesis. This includes a detailed introduction to wireless propagation models, MIMO techniques and relay networks. From Chapter 3 to Chapter 5, we will discuss the three different scenarios mentioned in Section 1.4, with each chapter focusing on one scenario. Finally, we summarize the whole thesis and discuss future work in

Chapter 6.

Chapter 2
Background

In this chapter we offer some background for MIMO systems and relay networks. The content of this chapter will be frequently referred in the rest of the thesis. For the communication systems discussed in this thesis, we always assume a discrete complex baseband model. The reason is that in typical wireless applications most of the processing such as coding/decoding or modulation/demodulation is actually done at the baseband, though communication often occurs in a passband. We will introduce some key concepts for wireless MIMO or relay channels, and also characterize the performances of different signalling and coding schemes for MIMO systems or relay networks.

2.1 Wireless fading channels

We start with an introduction to wireless fading channels. More details for this section can be found in some classic textbooks [27, 28]. Unlike wired connection, in a wireless channel the signal arrives at the destination through different physical propagation paths, which we refer to as multipath. Multipath results from reflection and diffraction caused by objects in the environment or refraction in the medium. We refer to all these distorting mechanisms as "scattering". The signal power variation caused by scattering is often referred to as fading.

There are roughly two main types of fading, depending on their effects on the different space and time scales: a) Large scale fading, which can be further divided to pathloss and shadowing, b) small scale fading, which also refers to multipath fading.

2.1.1 Pathloss

Pathloss presents the mean energy loss of the transmitted signals as function of distance. In cellular environments, the pathloss can be simply approximated by:

$$P_r = P_T x^{-\gamma}, \tag{2.1}$$

where P_T is the transmit power, P_r is the receive power, x denotes the transmission range, and γ is often referred to as the path loss exponent. In practice γ may vary from 2.5 to 6 in different environments [29, 30].

2.1.2 Shadowing

Shadowing effects of large objects such as buildings or hills can also cause signal variation, which we refer to as shadowing. It has been observed that the received signal power under shadowing approaches a log-normal distribution [27], Combined with pathloss, the received power can be expressed by:

$$P_r = P_T x^{-r} 10^{\zeta/10} \tag{2.2}$$

where ζ is a random variable drawn from a normal distribution with a mean of 0 dB and a standard deviation δ (dB).

2.1.3 Multipath fading

Multipath fading refers to the rapid fluctuations of the received signals and is mainly caused by the constructive and destructive interference of the multiple signal paths between the transmitter and receiver. If we assume that there are a large number of independent random scattered paths, then the envelope of the received signal follows a Rayleigh distribution. We refer to this kind of multipath fading as Rayleigh fading. Specifically, in a discrete-time complex baseband model the channel coefficient can be expressed as:

$$h = \alpha + \beta i \tag{2.3}$$

where α and β are independent identically distributed Gaussian random variables with zero mean and variance $\frac{\sigma^2}{2}$. The magnitude $|h|$ has a Rayleigh density function given by:

$$f(x) = \frac{x}{\sigma^2} \exp\left\{\frac{-x^2}{2\sigma^2}\right\}, \;\; x \geq 0 \tag{2.4}$$

and the squared magnitude $|h|^2$ is exponentially distributed with density:

$$f(x) = \frac{1}{\sigma^2} \exp\left\{\frac{-x}{\sigma^2}\right\}, \;\; x \geq 0. \tag{2.5}$$

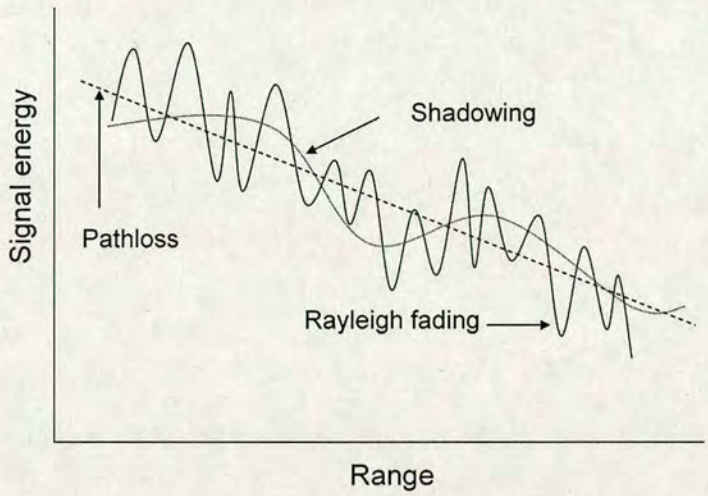


Figure 2.1: *Signal power fluctuation vs range in wireless channels.*

In this thesis the value of σ is always assumed to be 1.

If there is a line of sight (LOS) path present between transmitter and receiver, the Ricean distribution [12] is often used as an alternative model to Rayleigh fading model. In Ricean model, a factor K is used as a ratio of the received signal power for LOS path to that for Non-LOS paths, i.e., scattered paths. However, in this thesis we always use Rayleigh fading model for analysis, which is assumed in most existing literatures. Figure 2.1 shows the combined effects of pathloss and fading on received power in a wireless channel.

One important characteristic of multipath fading is the delay spread T_d. Roughly speaking, it is the difference in propagation time between the longest and shortest path, counting the paths with significant energy [31]. When T_d is considerably longer than one transmission symbol time, the channel acts like a tapped delay line filter and is said to be *frequency-selective*. Intersymbol interference might occur in this scenario as different symbols might arrive at the same time on different propagation paths. When T_d is much less than one transmission symbol time, one single channel filter tap is sufficient to represent the channel and the channel is in a condition called *flat fading*. In a discrete-time complex baseband model, the input-output relation between the transmitter and receiver can be expressed by:

$$y = \sqrt{\eta} h s + n. \tag{2.6}$$

where η is transmit power, s is the input signal, y is the output signal and n is the additive

Gaussian noise generated at the receiver. Both s and n are assumed to have zero mean and unit covariance. In this thesis, we always assume that the channel is flat fading. However, it has been shown that the frequency selective fading channel can be converted to an equivalent frequency flat fading channel by signal processing methods such as Orthogonal Frequency Division Multiplexing(OFDM)[32, 33].

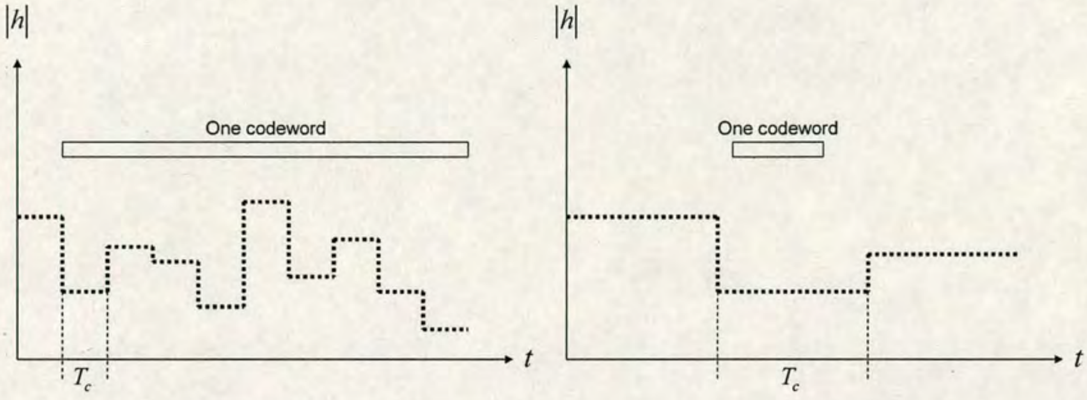
Another important characteristic for multipath fading is called the channel coherence time T_c. This denotes the time duration in which the multipath fading coefficients remain roughly the same. This parameter reflects the time-scale of the variation of the channel and is mainly dictated by the Doppler spread due to movement of the transmitter or receiver.

To build a simple model for a fading channel, the coefficient h is assumed to remain constant over each coherence time T_c of l symbols and is i.i.d. across different coherence periods. This is the so-called *block fading* model, which will be used throughout the thesis. Each value of h is called a channel realization.

Based on different T_c and transmission rates, we can further categorize the channel into *fast fading* and *slow fading*. In wireless communications, information is often coded into codewords, each of which consists of different symbols. In fast fading, one codeword length is often assume to cover a large number of channel blocks. In slow fading, however, one codeword length is often assumed to be less than one channel block length. See Figure 2.2 for illustration of these two concepts. In practice these two scenarios not only depend on the channel itself but might also correspond to applications based on different delay requirements. Fast fading can be assumed for applications that can deal with large delays (e.g. file downloads), while slow fading is often assumed for a low delay requirement, such as voice applications. In this thesis, we mainly concentrate ourselves on the slow fading channel.

2.1.4 Channel state information

We note that in all analysis in this chapter, the receiver is always assumed to have full channel state information (CSI). This means that the receiver can obtain the *exact* channel coefficients for each fading realization. In practice this can be done by using training sequences and channel estimation techniques [34]. The transmitter, might either have full CSI or no CSI, depending on different scenarios. For example, in a time division duplex (TDD) system, the receiver might be able to feed the full CSI back to the transmitter [35]. However, this might require additional



(a) Fast fading

(b) Slow fading

Figure 2.2: *Block fading models. (a) Fast fading. (b) Slow fading.*

resource and signaling overheads.

2.2 Time invariant channel

In this section, we begin our introduction to MIMO systems with the time invariant channel, then we move further into slow fading channel in the next section.

2.2.1 Single-input single-output channel

Before we investigate the performance of MIMO systems, we firstly look at the single-input single-output (SISO) system. We use Shannon capacity to measure the system performance.

Shannon capacity is defined as the maximal transmission rate at which reliable communication can be established [36]. For a bandlimited system discussed in the thesis, the Shannon capacity given a value of *receive* signal to noise ratio (SNR) can be written as:

$$C = \log_2 (1 + SNR), \tag{2.7}$$

where SNR is the value of SNR. Information theory shows that as long as the transmission rate is below C, the information can be recovered with an arbitrary small error rate if the coding block length is long enough. The proof of the Shannon capacity theorem can be conducted in different ways and can be found in several text books (e.g.[36, 37]).

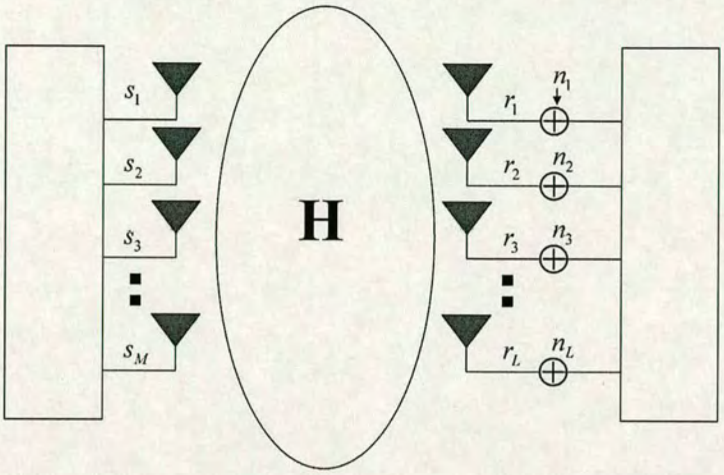


Figure 2.3: *System model for a point-to-point MIMO link with* M *transmit antennas and* L *receive antennas.*

Assuming an input-output relation expressed by (2.6) with fixed value of h, the capacity of SISO system can be expressed by

$$C = \log_2\left(1 + \eta\,|h|^2\right) \tag{2.8}$$

in bits/s/Hz. Here we denote the value of η as SNR [1]. At high SNR, the capacity can be approximated by

$$C \approx \log_2 SNR + \log_2 |h|^2 . \tag{2.9}$$

Later we will use this expression to compare with capacity behavior of the MIMO channel.

2.2.2 MIMO channel

With M transmit antennas and L receive antennas, the input output relation for a MIMO system in a flat fading environment can often be expressed as:

$$\mathbf{r} = \sqrt{\eta}\mathbf{H}\mathbf{s} + \mathbf{n}, \tag{2.10}$$

[1]Note that the SNR can also be defined as $\eta\,|h|^2$, which is often called receive SNR as shown in (2.7). In the Rayleigh fading scenario, it is the instantaneous SNR and its average is equal to the value of η when $\sigma = 1$. Here we use the value of η for simplicity of further analysis.

where $\mathbf{r}$ is the $L \times 1$ received signal vector. η now becomes the power *per transmit antenna* at the source. The vector $\mathbf{s}$ is the $M \times 1$ transmit signal vector and $\mathbf{n}$ is the $L \times 1$ complex circular additive white Gaussian noise vector at relay k which has zero mean and identity covariance matrix. $\mathbf{H}$ is the $L \times M$ channel transfer matrix. Figure 2.3 illustrates the system model.

With different CSI at the transmitter, the capacity for MIMO systems can be different. In the following we will discuss the capacity performance when CSI is either fully available or not available at the transmitter.

2.2.2.1 Full CSI

With full CSI at the transmitter, the capacity for a MIMO channel can be expressed by[16]:

$$C = \max_{\mathrm{Tr}(\mathbf{R}_{ss})=M} \log_2 \det\left(\mathbf{I}_L + \eta \mathbf{H}\mathbf{R}_{ss}\mathbf{H}^H\right), \tag{2.11}$$

where $\mathbf{R}_{ss}$ is the covariance matrix for signal vector $\mathbf{s}$. This equation can be maximized by choosing $\mathbf{R}_{ss}$ optimally. By singular value decomposition (SVD)[38], the matrix $\mathbf{H}$ can be expressed in the form:

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^H \tag{2.12}$$

where matrix $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, and $\mathbf{D}$ is the diagonal matrix which contains the singular values of the matrix $\mathbf{H}$. If we assume $\mathbf{H}$ has full rank, the optimal $\mathbf{R}_{ss}$ can be expressed by:

$$\mathbf{R}_{ss} = \mathbf{V}\Sigma\mathbf{V}^H \tag{2.13}$$

where Σ is an diagonal matrix given by

$$\Sigma = \mathrm{diag}\left\{\gamma^1, \cdots, \gamma^{\min(L,M)}\right\}. \tag{2.14}$$

The elements γ^i in (2.14) are the waterfilling allocations[39]:

$$\gamma^i = \left(\mu - \frac{1}{\eta\lambda^i}\right)^+, \sum_{i=1}^{M} \gamma^i = M.$$

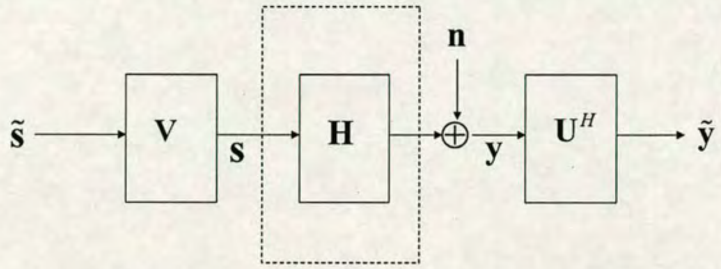


Figure 2.4: *Converting the MIMO channel into parallel channel through SVD.*

The channel capacity in this scenario can be written as:

$$C = \sum_{i=1}^{\min(M,L)} \log_2\left(1 + \eta\lambda^i\gamma^i\right), \tag{2.15}$$

where each λ^i is the square of ith singular value of $\mathbf{D}$. Equation (2.15) implies that the optimal capacity performance can be achieved by a decomposition of the MIMO channel into several parallel channels, while optimally allocating powers to each of them. The realization of this process is shown in Figure 2.4, where the signal vector $\tilde{\mathbf{s}}$, which has the covariance matrix Σ, is first multiplied by $\mathbf{V}$ and transmitted through the channel. The receiver multiplies the receive signal vector by $\mathbf{U}^H$ before performing detection. The details of this process have also been introduced in some textbooks [31, 40].

2.2.2.2 No CSI

When the transmitter has no CSI, it has no preferred channel direction. The vector s in this scenario may be chosen to be statistically non-preferential, i.e., $\mathbf{R_{ss}} = \mathbf{I_M}$. In this scenario, the capacity can be expressed as:

$$C = \log_2 \det\left(\mathbf{I}_L + \eta\mathbf{H}\mathbf{H}^H\right). \tag{2.16}$$

Note that this expression can also be expressed in a sum of parallel channel capacities through singular value decomposition of $\mathbf{H}$:

$$C = \sum_{i=1}^{\min(M,L)} \log_2\left(1 + \eta\lambda^i\right). \tag{2.17}$$

For high SNR (i.e., η) values, (2.17) can be approximated as:

$$C = \min(L, M)\log_2 SNR + \sum_{i=1}^{\min(M,L)} \log_2(\lambda^i). \tag{2.18}$$

Compared with the capacity of SISO channel (2.9), we can see that the MIMO channel offers a gain of $\min(L, M)$ for high SNR. Note that this is also true when the transmitter knows the CSI. The only difference is an enhancement in the second term due to its additional power gain offered by the waterfilling coefficients γ^i (see equation 2.15). One might think of replacing η with η/M for a more fair comparison in terms of total transmit power. However, this still will not change the gain $\min(L, M)$ in the capacity expression. This gain is referred to as the maximum *multiplexing gain* of the MIMO system. The multiplexing gain reflects the increased spatial degrees of freedom of the MIMO channel compared with the SISO channel. A simple way to interpret this is that a SISO channel can effectively support one data stream, while a MIMO channel can effectively support at most $\min(L, M)$ data streams given the same SNR. The capacity for single-input multiple-output (SIMO) or multiple-input single-output (MISO) systems, where multiple antennas are only used at the receiver or transmitter, can be analyzed in the same way but replacing the matrix $\mathbf{H}$ by a channel vector h with dimension $L \times 1$ or $M \times 1$. These will be discussed in more detail in Section 2.3.

2.2.2.3 V-BLAST structure

A simple question may now be posed: Is it possible to achieve the capacity (2.16) through practical signal processing schemes? A structure where the message is de-multiplexed into M multiple streams can realize this capacity performance. By coding and interleaving across different streams and applying a maximum likehood (ML) detector at the receiver [20], this capacity can be achieved if the codeword length is long enough. However, it is well known that the ML detection complexity increases exponentially with the number of transmit antennas. Therefore it might not be a realistic method.

A simpler method to achieve the capacity is to de-multiplex the message to M *independent* streams, each of which is coded independently and transmitted through one antenna. The receiver applies a V-BLAST-MMSE detector [20] to decode each signal stream successively. A detection process consists of M iterations. In each iteration, the detector detects one stream through an MMSE filter, then subtracts it and moves to the next iteration. This process is shown

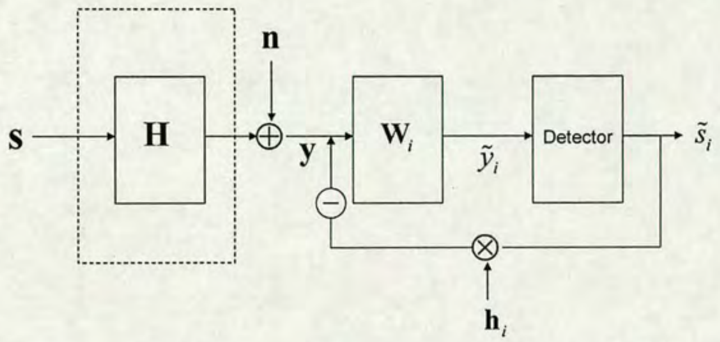


Figure 2.5: *The structure of V-BLAST-MMSE detector.*

in Figure 2.5. The reason for using the MMSE filter is to suppress the inter-stream interference at the receiver. Another popular filter for interference cancellation is the zero forcing (ZF) filter [41], where the MIMO channel is converted to orthogonal channels at the cost of amplifying the noise. The MMSE filter makes a tradeoff between the noise amplification and interference suppression. In fact, it has been shown that the MMSE filter is information lossless [31].

Without loss of generality, we assume that the stream from the first transmit antenna is detected and decoded first. The MMSE vector for ith iteration can be expressed as [42]:

$$\mathbf{W}_i = \left(\mathbf{H}\left(i+1\right)\mathbf{H}\left(i+1\right)^H + \mathbf{I}_L\right)^{-1}\mathbf{h}_i, \tag{2.19}$$

where $\mathbf{h}_i$ is the ith column of $\mathbf{H}$ and $\mathbf{H}\left(i+1\right)$ $(i = 1, 2, \cdots, M)$ is the matrix $[\mathbf{h}_i\mathbf{h}_{i+1}\cdots\mathbf{h}_M]$. The receive signal to interference plus noise ratio (SINR) for the ith iteration can be expressed by:

$$SINR_i = \eta\mathbf{h}_i^H\left(\mathbf{H}\left(i+1\right)\mathbf{H}\left(i+1\right)^H + \mathbf{I}_L\right)^{-1}\mathbf{h}_i. \tag{2.20}$$

It has been proved that if each signal stream is decoded correctly, the capacity of this V-BLAST system approaches the MIMO channel capacity [43]:

$$\sum_{i=1}^{M}\log_2\left(1 + SINR_i\right) = \log_2\det\left(\mathbf{I} + \eta\mathbf{H}\mathbf{H}^H\right). \tag{2.21}$$

Note that in practice in order for the V-BLAST structure to achieve the channel capacity, certain feedback regarding the receive SINR information should always be allowed from the receiver to allow the transmit antennas to match the measured SINR values. This will be further discussed later in this chapter.

2.3 Slow fading channel

When the channel is in a slow fading environment, however, the channel capacity is completely different. For a deep fade channel realization, where the value of $|h|^2$ is extremely small, the capacity can be near zero. We recall that Shannon capacity for a time-variant channel is defined as the maximal rate for communicating reliably over *all* channel realizations. Therefore, the Shannon capacity for the slow fading channel is in fact *zero*.

For any transmission rate R, if the capacity for a channel realization is smaller than R, whatever the code used by the transmitter, the decoding error probability cannot be made arbitrarily small. The system in this condition is said to be in *outage*. Outage probability is often used as a good measure of the system performance in the slow fading scenario:

$$P_{out} \triangleq P\left[C < R\right], \tag{2.22}$$

where C is the Shannon capacity for any specific channel realization.

2.3.1 SISO system

For a SISO system with an input-output relation expressed by (2.6), assuming the channel is Rayleigh fading the outage probability can be directly calculated as:

$$P_{out} = 1 - \exp\left(\frac{-\left(2^R - 1\right)}{\eta}\right). \tag{2.23}$$

We can further obtain the high SNR approximation:

$$P_{out} \approx \frac{2^R - 1}{SNR}, \tag{2.24}$$

where η is replaced by SNR. We can see that the outage performance decays as $1/SNR$.

2.3.2 MIMO systems

As we already mentioned, for a SISO system the outage mainly occurs when the channel is in a deep fade. However, in a MIMO system where multiple channels exist, even if one of the channels is in a deep fade, the others might not. This offers additional *diversity* compared with

SISO channel. We might think to mitigate the deep fading effect by exploiting the benefits of all the channels. We refer to these methods as space diversity schemes. A detailed introduction to space diversity schemes can be found in [31, 40]. Here we begin our discussion with single-input multiple-output (SIMO) and multiple-input single-output (MISO) systems where multiple antennas are only deployed at one side.

2.3.2.1 SIMO system

The input-output relation for SIMO system with L receive antennas can be written as:

$$\mathbf{y} = \sqrt{\eta}\mathbf{h}s + \mathbf{n}, \tag{2.25}$$

where $\mathbf{h}$ is an $L \times 1$ channel vector. An effective way to exploit the diversity of the channel is to combine coherently the signal energy from all L channel dimensions. In order to do this, we multiply the receive signal vector by $\mathbf{h}^H$, so that the receive SNR in this scenario becomes $\eta \left\|\mathbf{h}\right\|^2$. The capacity of the channel is therefore:

$$C = \log_2 \left(1 + \eta \left\|\mathbf{h}\right\|^2\right). \tag{2.26}$$

It is not difficult to see that (2.26) is also the capacity for SIMO channel, when compared with (2.16). The receive SNR in this scenario is maximized by linearly combining L signal branches. This method is often referred to as maximal ratio combining (MRC). The outage probability can be written as:

$$P_{out} = P\left\{\left\|\mathbf{h}\right\|^2 < \frac{2^R - 1}{\eta}\right\}. \tag{2.27}$$

In an i.i.d. Rayleigh fading scenario, $\left\|\mathbf{h}\right\|^2$ is distributed as a Chi-squared random variable with $2L$ degrees of freedom, with density [44]:

$$f(x) = \frac{1}{(L-1)!} x^{L-1} e^{-x}. \tag{2.28}$$

For small x $e^{-x} \approx 1$, it is easy to see that:

$$P\left\{\left\|\mathbf{h}\right\|^2 < \varepsilon\right\} \approx \frac{1}{L!}\varepsilon^L \tag{2.29}$$

for small ε. Therefore, at high SNR (η) the outage probability is given by:

$$P_{out} \approx \frac{\left(2^R - 1\right)^L}{L!SNR^L}. \tag{2.30}$$

Comparing with SISO systems, the outage probability decays like $1/SNR^L$. We refer to L as the maximal *diversity gain* of the system.

2.3.2.2 MISO system

When CSI is not known to the transmitter, a famous diversity scheme is the Alamouti code for a 2×1 channel [21]. It is the simplest orthogonal space-time block code [17, 23]. The Alamouti code transmission consists of two symbol periods, where antenna 1 transmits $[s_1, -s_2^*]$ and antenna 2 transmits $[s_2, s_1^*]$ over the two symbol durations. Assuming the channel stays constant for the two symbol transmission periods, the input-output relation, after some modification, can be expressed as:

$$\begin{bmatrix} y_1 \\ y_2^* \end{bmatrix} = \sqrt{\frac{\eta}{2}} \underbrace{\begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}}_{\hat{\mathbf{H}}} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix} \tag{2.31}$$

where η is defined as the total transmit power at two transmit antennas, h_i is the channel coefficient from antenna i to the receiver, y_i and n_i are the received signal and noise at time slot i (either 1 or 2). We observe that the columns of $\hat{\mathbf{H}}$ are orthogonal. By multiplying the receive vector by $\hat{\mathbf{H}}^H$, the channel can be decomposed into two parallel channels each with a receive SNR $\frac{\eta}{2} \|\mathbf{h}\|^2$, where $\mathbf{h}$ is the 2×1 channel vector $\begin{bmatrix} h_1 & h_2 \end{bmatrix}$. From the previous analysis it is not difficult to see that this scheme can achieve the same maximal diversity gain as that of MRC for $L = 2$.

When the transmitter knows the CSI, it can exploit the channel diversity in a way similar to the MRC scheme for any number of transmit antennas M. Assuming one symbol is transmitted at one time, the signal energy is weighted differently at M transmit antennas, based on their own channel quality. The input-output relation can be expressed as:

$$y = \sqrt{\eta}\mathbf{h}\mathbf{w}s + n, \tag{2.32}$$

where η is now defined as the total transmit power at all transmit antennas. Note that $\mathbf{h}$ becomes

$1 \times M$ instead of $L \times 1$. The weight vector w is:

$$\mathbf{w} = \frac{\mathbf{h}^H}{||\mathbf{h}||}. \tag{2.33}$$

It is not difficult to see that the capacity for this system is the same as (2.26) if $L = M$. Therefore it offers the same diversity gain. This scheme is often referred to as transmit beamforming (TB)[24], as it exploits the channel diversity by beaming the signal energy to the strongest channel direction through a weight vector $\mathbf{w}$.

Compared with Alamouti scheme, we will see a *power gain* for TB. Specifically, there is a factor of two receiver SNR loss for Alamouti scheme compared with TB. This reflects the benefit of CSI at the transmitter, as in Alamouti scheme the signal energy is equally distributed in all channel directions (i.e., equal power allocation) due to the absence of CSI at the transmitter.

2.3.2.3 MIMO system

When CSI is not known at the transmitter, the Alamouti scheme can be easily extended to any $L \times 2$ channel, i.e., a channel with two transmit antennas and L receive antennas. The matrix $\hat{\mathbf{H}}$ now has dimension $2L \times 2$ with two orthogonal columns (see equation (5.37) in [40] for example). Assuming the total transmit power is η, the capacity of the system can be expressed as:

$$C = \log_2\left(1 + \frac{\eta}{2}\,||\mathbf{H}||^2\right), \tag{2.34}$$

where $\mathbf{H}$ is $L \times 2$ channel matrix in (2.10). $||\mathbf{H}||^2$ is Chi-square distributed with $4L$ degrees of freedom. Therefore, similar to the analysis in section 2.3.2.1, we can see that Alamouti scheme offers a maximal diversity gain of $2L$.

When CSI is known to the transmitter, TB can be extended to any $L \times M$ channel. The input-output relation in this scenario can be expressed by:

$$y = \sqrt{\eta}\mathbf{g}^H\mathbf{H}\mathbf{w}s + \mathbf{g}^H\mathbf{n}, \tag{2.35}$$

where $\mathbf{w}$ and $\mathbf{g}$ are the $M \times 1$ and $L \times 1$ weight vectors at the transmitter and the receiver. Recall the singular value decomposition of the channel $\mathbf{H}$ (2.12). It can be proven that the receive SNR is maximized if $\mathbf{w}$ and $\mathbf{g}$ are chosen as the right and left singular vectors (columns of $\mathbf{V}$ and $\mathbf{U}$), corresponding to the largest singular value $\sqrt{\lambda_{\max}}$ of $\mathbf{H}$. The effective input-output relation for

Scheme	MRC	TB	Alamouti coding	V-BLAST-MMSE
Maximal diversity gain	L	$M \times L$	$L \times 2$	$L - M + i$

Table 2.1: *The maximal diversity gain for different schemes for MIMO channel.*

the channel thus reduces to:

$$y = s\sqrt{\eta\lambda_{\max}} + n, \tag{2.36}$$

where n is Gaussian noise with zero mean and unit variance. This equation implies that TB is exploiting the channel diversity by allocating all the transmit power to the strongest channel direction (eigenmode). Since for λ_i we have the following identity:

$$\sum_{i=1}^{\min(L,M)} \lambda_i = \|\mathbf{H}\|^2, \tag{2.37}$$

$\lambda_{\max}$ may be upper and lower bounded by:

$$\frac{\|\mathbf{H}\|^2}{\min(L,M)} \leq \lambda_{\max} \leq \|\mathbf{H}\|^2. \tag{2.38}$$

Now considering the capacity expression and outage probability of this system, it is not difficult to see from (2.38) that TB can offer a maximal diversity gain of $M \times L$, compared with SISO system. Since the MIMO channel itself contains $M \times L$ i.i.d. sub channels, we expect the maximal diversity can be exploited from the channel is $M \times L$. In this sense, TB become the diversity-optimal scheme for a MIMO channel.

V-BLAST-MMSE can also be applied to a slow fading environment. However, it's diversity performance is poor compared with Alamouti coding or TB, mainly due to the co-channel interference at the receiver. It has been shown that for a V-BLAST-ZF detector or V-BLAST-MMSE detector, the diversity gain of the kth detected stream is $L - (M - k)$ [31], where $(M - k)$ is the number of uncanceled interfering streams at the Kth stage and the cause of the loss of diversity. The first stream has the worst diversity gain $L - (M - 1)$, which is the bottleneck in the system.

2.4 Diversity-multiplexing tradeoff

From the discussion in Section 2.2, we see that the MIMO channel can offer a maximal spatial multiplexing gain of $\min(L, M)$ for high SNR. However, we cannot see this benefit from the schemes discussed in the slow fading channel in Section 2.3. The reason is because in a slow fading channel, we mainly focus on extracting the maximal diversity of the systems to average out the fading effect by transmitting at a fixed rate R, which becomes vanishingly small compared with time invariant channel capacity at high SNR (which grows like $\min(L, M)\log_2 SNR$). Therefore all the spatial multiplexing gain is actually sacrificed to maximize the diversity gain. In order to retain some of the diversity and multiplexing gain, one would instead want to communicate at a rate $R \approx r\log_2 SNR$, where r ($0 \le r \le \min(L, M)$) is a fraction of the maximal multiplexing gain, while retaining a certain fraction d of the maximal diversity gain ($0 \le d \le L \times M$). In doing this, we formulate the following diversity-multiplexing tradeoff for a slow fading channel [18]:

Definition 2.1 (Diversity-Multiplexing Tradeoff). *Consider a family of codes C_η operating at SNR η and having rates R bits/s/Hz with large block length. The multiplexing gain and diversity gain are defined as*[2]

$$r \triangleq \lim_{\eta\to\infty} \frac{R}{\log_2 \eta}, \quad d \triangleq -\lim_{\eta\to\infty} \frac{\log_2 P_{out}(R)}{\log_2 \eta}. \tag{2.39}$$

The curve $d(r)$ is the diversity-multiplexing tradeoff of the system.

Once the outage probability is obtained, the derivation of diversity-multiplexing tradeoff is straightforward. For a SISO channel, the diversity-multiplexing tradeoff can be expressed as

$$d = 1 - r, \tag{2.40}$$

For a MIMO channel, It has been proven that the optimal diversity-multiplexing tradeoff of the $L \times M$ i.i.d. Rayleigh fading channel is a piece-wise linear curve joining the (x, y) co-ordinates:

$$(k, (M-k)(L-k)), k = 0, \cdots, \min(L, M). \tag{2.41}$$

Figure 2.6 gives an example for the optimal diversity-multiplexing tradeoff for SISO, MISO (SIMO) and MIMO channels. From the discussion in the previous section, it is not difficult to show that MRC, Alamouti coding or TB can offer the optimal diversity-multiplexing tradeoffs

[2] *Generally speaking, for a specific scheme the diversity d is decided by the error probability $P_e = P_{out} + P'_e$, where P'_e is the detection error probability given the channel is not in outage. It has been shown [18] that for high SNRs this term can be ignored if the block length of the code is large enough.*

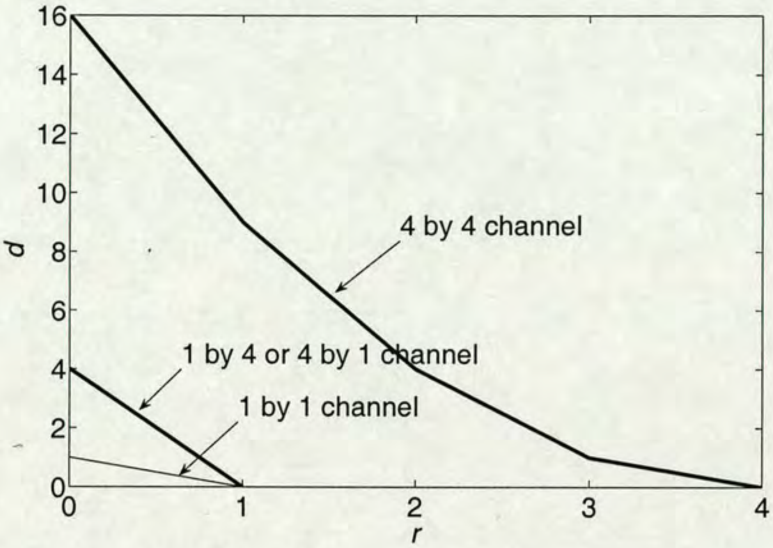
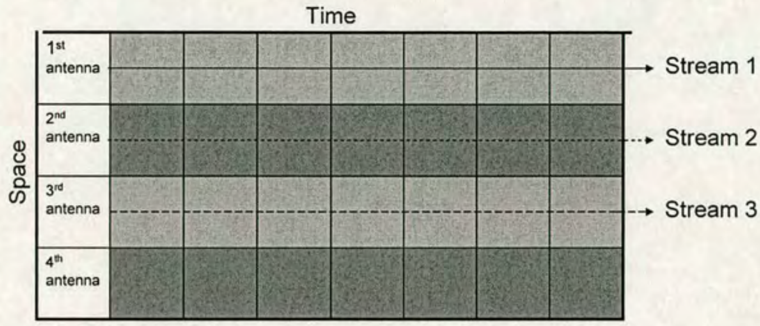


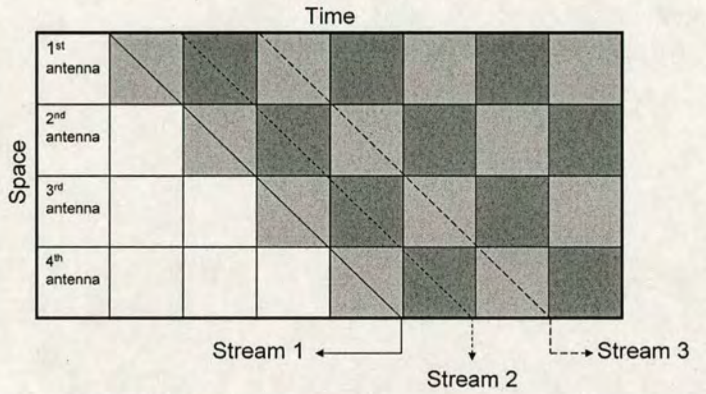
Figure 2.6: *An example of the optimal diversity-multiplexing tradeoffs for point-to-point systems, where L by M channel denotes the channel with M transmit antennas and L receive antennas.*

for SIMO or MISO systems, which are $d = L(1-r)$ for SIMO and $d = M(1-r)$ for MISO. However, none of these schemes can achieve the optimal diversity-multiplexing tradeoff for a MIMO channel. The design of tradeoff-optimal schemes is an active area of research, while most of the approaches concentrate on space-time codes [45–48]. For spatial multiplexing structures, a well known near tradeoff-optimal scheme is the so called Diagonal(D)-BLAST-MMSE detector [15]. The major difference between D-BLAST and V-BLAST lies in the layering of the information streams. In V-BLAST information streams are encoded and transmitted independently within each antenna, while in D-BLAST information streams are encoded and transmitted across different antennas diagonally in the space-time domain using parallel channel codes [31]. When the receiver applies successive interference cancelation schemes, it decodes each information stream diagonally in the space-time domain, then subtracts it from the received signal and then moves to the next diagonal layer, until all the streams are decoded. Figure 2.7 illustrates the structure for D-BLAST-MMSE detector and its difference from V-BLAST structure.

Compared with V-BLAST, D-BLAST can achieve a better diversity gain and has a performance advantage in terms of outage probability. To see this clearly, we formulate the outage probability for both systems. Assuming that the data rate for transmit antenna k is R_k and the total



(a) V-BLAST



(b) D-BLAST

Figure 2.7: *The structure of V-BLAST and D-BLAST detectors for four transmit antennas. (a) V-BLAST. (b) D-BLAST.*

transmission rate is R, the outage probability for V-BLAST can be written as:

$$P_{out} = \Pr\left\{\log_2\left(1 + SINR_k\right) < R_k,\right\} \tag{2.42}$$

where $SINR_k$ is defined as (2.20). From (2.42) we can see that the performance of V-BLAST is actually constrained by the quality of each sub-channel. The system will be in outage whenever at least one of the sub-channels is in outage. For the D-BLAST-MMSE system, the outage probability can be expressed as:

$$P_{out} = \Pr\left\{\sum_{k=1}^{M}\log_2\left(1 + SINR_k\right) < \sum_{k=1}^{M} R_k\right\}, \tag{2.43}$$

which is exactly the outage probability for the MIMO channel:

$$P_{out} = \Pr\left\{\log_2 \det\left(\mathbf{I}_L + \eta \mathbf{H}\mathbf{H}^H\right) < R\right\}. \tag{2.44}$$

Comparing (2.42) with (2.43), we can clearly see a performance advantage of D-BLAST: unlike V-BLAST, even when some of the sub-channels are in poor conditions, the system can still recover itself as long as the other channels are good enough. This benefit is mainly due to the parallel channel coding applied for each information stream. Compared with V-BLAST, D-BLAST requires additional coding complexity and also a small data loss at the beginning and end of the transmission due to its diagonal structure.

2.5 Low-rate feedback system

The discussion (except for TB) in the above section is based on an assumption that the transmitter only knows the second order statistics of the channel. The outage probability or diversity-multiplexing tradeoff can be obtained at the transmitter side provided that the statistics of the channel is available and can help transmitter set up the transmission rate for different quality of service (QoS) and data rate requirements. However, due to the lack of instantaneous channel information, the error probability can never be guaranteed to be arbitrarily small.

In a slow fading environment, it is possible to set up a low rate feedback channel between the transmitter and the receiver to help the transmitter obtain better channel knowledge. This can be especially true for a Frequency-Division-Duplex (FDD) system. In this scenario, the transmitter can perform adaptive modulation and coding to adaptive the rate for each slow fading channel realization. If coding is applied separately for each channel block, outage events can be *completely* avoided if for each channel block the transmission rate is adapted to be below the Shannon capacity for that channel realization. In this scenario, the performance is really determined by the average fading effect of the channel. We use the term *average capacity*, which is the mean value of the channel capacities for each channel realization, to evaluate the system performance. Note that although feedback requires some channel resource and hence decrease the system throughput, in a slow fading context it can be ignored due to the long fading block length.

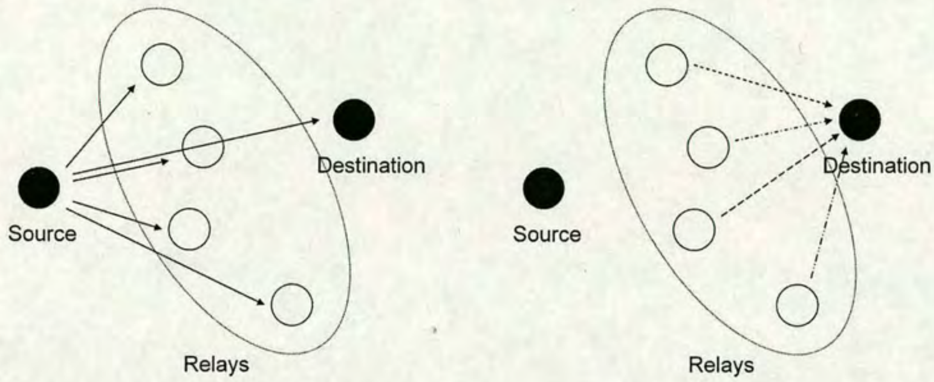
To perform dynamic rate allocation for SISO channels, only one variable containing the value of h needs to be fed back to the transmitter. However, for an $L \times M$ MIMO system, the whole channel matrix might have to be fed back in order to approach the capacity upper bound for each channel realization (2.15). This will clearly cost much more channel resource. Instead, if we can apply V-BLAST structure, then only the M output $SINR$ values need to be fed back to the transmitter. Though there is a performance loss for V-BLAST compared with (2.15), at high SNR level it is easy to show that the loss can be ignored. Certain power allocation schemes for V-BLAST can also help approach (2.15) for arbitrary SNR [42].

Without feedback, the V-BLAST system discussed in the previous section has a low diversity performance compared with D-BLAST or other diversity schemes. However, with the feedback, the outage event can be eliminated by adapting the rate below the instantaneous channel capacity. Therefore, V-BLAST can be the preferred choice both over diversity schemes for its multiplexing gain and D-BLAST for its lower complexity.

2.6 Relay networks: Virtual MIMO systems

So far, we have discussed the point-to-point (one-hop) link, where only two nodes are involved in the communication. In practice, a network often contains multiple nodes and they are often not isolated. As we mentioned in Chapter 1, in an ad-hoc or cellular environment, capacity and coverage might both be improved by making the nodes act as relays to help each other transmit the information. In this section we extend the discussion to two-hop links, where single or multiple relays might exist to help the source to forward the information. We assume a Time-Division-Multiple-Access (TDMA) scenario where the information is transmitted from the source to the destination in two steps. We introduce a simple transmission protocol here. 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 the second step, while the source remains silent. The destination performs decoding based on the message it received in both steps. This transmission protocol is commonly-used in many papers (see [49, 50] for example) and we refer to it as *classic protocol* throughout the thesis. Furthermore, we note that *half-duplex* relaying is always assumed throughout the thesis. This means that the relays *can not* receive and transmit simultaneously.

An example is shown in Figure 2.8. We can see that in both transmission steps, the network forms a point-to-point MIMO system only with a larger dimension. Specially, when each node is equipped with a single antenna, the network becomes a SIMO link for the first time slot and



(a) first step

(b) second step

Figure 2.8: *Classic transmission protocol for relay networks. (a) first step. (b) Second step.*

a MISO link for the second time slot. This implies that even with single antenna at each node, the network might obtain the benefits of a SIMO (MISO) system when relaying is used.

2.6.1 Relaying methods

Before looking further into the characteristics of relay networks, we first introduce the relaying methods, i.e. through which means the relays process the message before forwarding it to the destination.

Generally, there are three kinds of relaying schemes. The first is called amplify-and-forward or analogue relaying. In this relaying mode, after the radio signal is converted to complex baseband, it is sent through an amplifier which amplifies the received signal (together with additive white Gaussian noise), then forwarded to the destination. By this method, the signal energy is amplified by the relay. However, the deficiency is that the receive noise at the relay is also amplified and included at the destination input, which will impair the system performance.

The second relaying mode is called decode-and-forward (digital) relaying. The relay decodes, then re-encodes the information and then forwards it to the destination. The advantage of this mode over amplify-and-forward is that it can completely eliminate the noise generated at the relay. However, if a decoding error happens at the relay, the system will suffer from error propagation, and the source to relay link qualities become very important.

The third relaying mode is called compress-and-forward relaying, in which the relay encodes a quantized version of its received signal, using ideas from source coding with side information [51, 52]. Generally, it is believed that the compress-and-forward relaying performs well if the relay-destination channel quality is high.

In this thesis, we mainly concentrate on the first two relaying modes, i.e., amplify-and-forward and decode-and-forward. For decode-and-forward, how to apply the coding method at each node is an active area of research (see [49, 50, 53–58] for example). Generally, the simplest coding method for a two-hop multiple relay network is *repetition coding*, where the source and all the relays are using the same modulation and coding method (i.e. the same codebook). The destination combines the message received from the source and all the relays before performing decoding. If the relay does not have forward channel information, it requires the relays to transmit in orthogonal time slots in order for the destination to combine effectively the signals for the second hop. Therefore, if there are N relays in total, it will take $N+1$ time slots to complete the transmission. Let us take single antenna network as an example, where each node is equipped with single antenna. Assuming the message is correctly decoded at the relays and the channel coefficients between node a and b is $h_{a,b}$, the capacity of the network can be expressed as

$$C = \frac{1}{N+1}\log_2\left(1+\eta\left|h_{s,d}\right|^2+\eta\sum_{i=1}^{N}\left|h_{r_i,d}\right|^2\right), \tag{2.45}$$

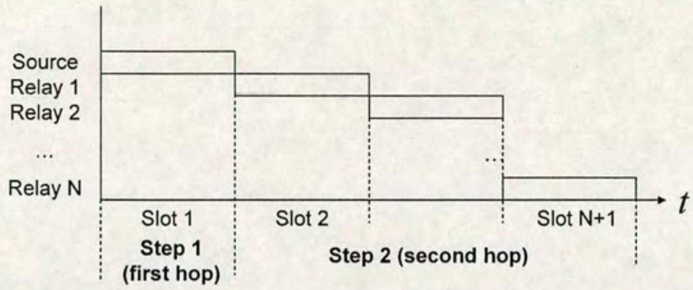
where s, d and r_i denotes the source, destination and relay i. Note that here we assume the transmit power at each relay is the same as at the source. It can be seen from (2.45) that repetition coding is clearly not spectrally efficient due to its use of $N+1$ time slots.

Note that $N+1$ transmission time slots can in fact offer $N+1$ parallel (orthogonal) channels. Parallel channel coding, which has been mentioned for the D-BLAST system in the previous section, can offer a much better spectral efficiency in this scenario if they are applied at the relays. The capacity in this scenario can be expressed as:

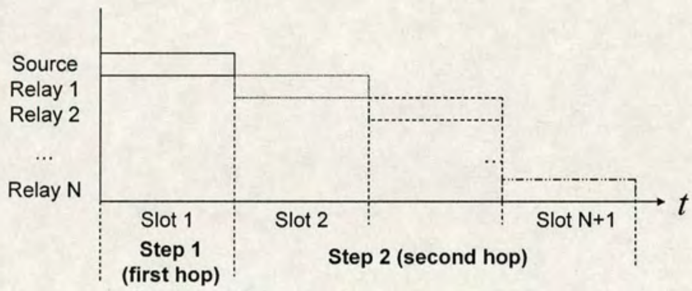
$$C = \frac{1}{N+1}\left(\log_2\left(1+\eta\left|h_{s,d}\right|^2\right)+\sum_{i=1}^{N}\log_2\left(1+\eta\left|h_{r_i,d}\right|^2\right)\right). \tag{2.46}$$

However, the complexity for parallel channel coding is much higher.

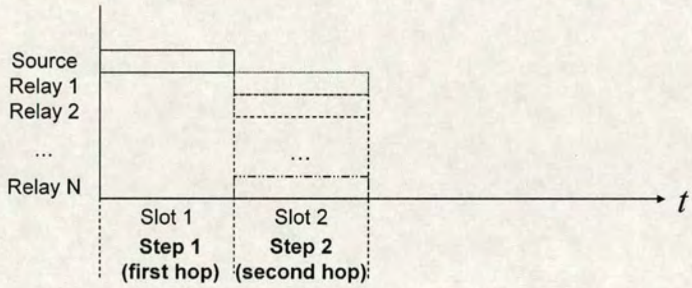
Both repetition coding and parallel coding require $N+1$ time slots to complete transmission. To decrease the number of transmission time slots, a better way is to apply distributed space-time codes at the relays. In theory, if all the nodes in the network are using i.i.d. random codebooks,



(a) Repetition coding: the same codebook are used for all $N + 1$ transmission time slots



(b) Parallel channel coding: Parallel channel codes are used for all $N+$ transmission time slots. An example is that i.i.d. random codebooks are applied at each transmitter.



(c) Space-time distributed coding: i.i.d. random code books are applied at each transmitter, while all the relays are transmitting simultaneously for the second time slot. The network now consists of two parallel channels. One is the source to relay channel, the other is the relay(s) to destination channel, where diversity is achieved through distributed random coding.

Figure 2.9: *Different coding schemes and protocols for relay networks. (a) Repetition coding. (b) Parallel channel coding. (c) Space-time distributed coding.*

the transmission can be completed in two time slots, which is the classic protocol. The capacity in this scenario can be expressed as:

$$C = \frac{1}{2}\left(\log_2\left(1+\eta\left|h_{s,d}\right|^2\right) + \log_2\left(1+\eta\sum_{i=1}^{N}\left|h_{r_i,d}\right|^2\right)\right). \tag{2.47}$$

Comparing this equation with (2.46), we can see that space-time distributed coding might offer even higher spectral efficiency than parallel coding. Figure 2.9 illustrates the three coding schemes discussed above for relay networks.

2.6.2 Diversity and power gain

If we compare the capacity equations (2.45)-(2.47) with those for the point-to-point link in the previous sections, we can see that through relaying similar diversity benefits can be achieved as in the MIMO link, even if each node is equipped with single antenna. Specifically, through repetition coding the network mimics a MISO link (2.26); parallel coding can improve the outage behavior in a similar way to D-BLAST (2.43); the space time distributed coding seems to offer both benefits of repetition coding and parallel coding. All of these diversity benefits result from the fact that the source and relay nodes can form a virtual antenna array during the transmission. The diversity achieved by cooperating the source and relay nodes is often called *user cooperation diversity* or *cooperative diversity* [59, 60].

However, there is a major difference between the relay network and the point-to-point MIMO link: the antenna elements in relay networks belong to different nodes. This difference results in a performance constraint for the relay networks, especially at the first hop, as the nodes are distributed and cannot perform diversity combining schemes such as MRC. Therefore, the capacity performance will be mainly constrained by the first hop. Note the capacity performance equations (2.45)-(2.47) and diversity gains discussed above are conditioned on the fact that the relays correctly decode the message. In fact, it can be seen that no diversity gain can be obtained if the relays are chosen totally at random [49] for each channel realization, or if relaying is always used regardless of the source to relay channel quality [50]. Certain adaptive protocols [50] [53] or relay selection methods [49] need to be developed in order to obtain full diversity gain. An extreme example is that only the best single relay is chosen to forward the message for each channel realization, based on both the source to relay and the relay to destination channel qualities [61]. This method can also achieve the full diversity gain of the network. The diversity achieved through this way is often referred to as *selection diversity*.

We also expect relaying to offer a higher power gain compared with direct transmission. This has been stated in the first chapter. In practice, a relay is often chosen for its better position (e.g. smaller pathloss or shadowing) compared with the destination (e.g. [62, 63]. Therefore, the value of $|h_{r_i,d}|^2$ or $|h_{s,r_i}|^2$ is often larger than $|h_{s,d}|^2$. This can clearly improve the capacity performance.

2.6.3 Multiplexing loss

Although relaying can offer diversity and power gain, it might suffer a multiplexing loss compared with point-to-point link. This is mainly due to the half-duplex nature of relaying. It can be seen from the examples in Section 2.6.1 that transmission of one data should take at least two time slots for the classic protocol. Compared with direct link which only requires one time slot, the data rate is only at most half. To contrast this with a point-to-point MIMO link, which often offers multiplexing gain, we call this deficiency caused by relaying as *multiplexing loss*.

2.7 Conclusions

In this chapter we provided some background on the point-to-point MIMO system and relay channels. It can be seen that in different ways, both MIMO systems and relay networks can offer diversity gain over that for the point-to-point SISO link. However, compared with a point-to-point MIMO system, which can offer a multiplexing gain over a SISO channel, the relay network might suffer a multiplexing loss. How to exploit the tradeoff between diversity gain and multiplexing loss due to relaying is an interesting topic. Moreover, the discussion for relay networks in this chapter are mainly focussed on single antenna nodes, how to effectively relay the signals when some or all of the nodes are equipped with multiple antennas is an active ongoing area of research. These topics will be discussed further in the next three chapters.

Chapter 3
Single antenna relay channels

In this chapter we continue our discussion on the single antenna relay networks, where each node is equipped with a single antenna. As mentioned in Section 2.6, in this scenario the networks can act as a virtual MIMO system. Our analysis is focused on the capacity behavior of the networks, while diversity and multiplexing factors are also discussed to assess impact on the system capacity. More specifically, as mentioned in Chapter 2, in practice feedback of the CSI is always assumed to allow the wireless link to obtain the *average capacity* of the channel in a slow fading environment, which is assumed throughout the thesis. Again, we note that relaying is always assumed to be performed in half-duplex mode throughout the thesis.

3.1 Introduction

As we mentioned in Chapter 2, when compared with direct transmission, the classic protocol might suffer from multiplexing loss, which will result in a loss of spectral efficiency for the high signal to noise ratio (SNR) region. This is especially true for the repetition coding scheme. To see this more clearly, we can write the capacity for direct transmission as $\log_2(1+SNR)$ (bits per channel use), where SNR denotes the receive SNR at the destination for direct transmission. Then, the capacity for repetition coded relaying based on *classic protocols* can be expressed as $0.5 \times \log_2(1+\kappa SNR)$, where κ reflects the power gain (or some diversity gain in the sense of outage probability) achieved due to relaying and 0.5 denotes the multiplexing loss due to the use of two time slots in the classic protocol. The capacity ratio of relaying to direct transmission can be expressed as:

$$G \triangleq \frac{0.5\log_2(1+\kappa SNR)}{\log_2(1+SNR)} \tag{3.1}$$

It is obvious to see that $G \approx \kappa/2$ when $SNR \to 0$; and $G \approx 0.5$ when $SNR \to +\infty$. This means that the classic protocol can improve link capacity only for the low SNR region (with a gain of $\kappa/2$). When the receive SNR for direct link transmission is high, the benefit for

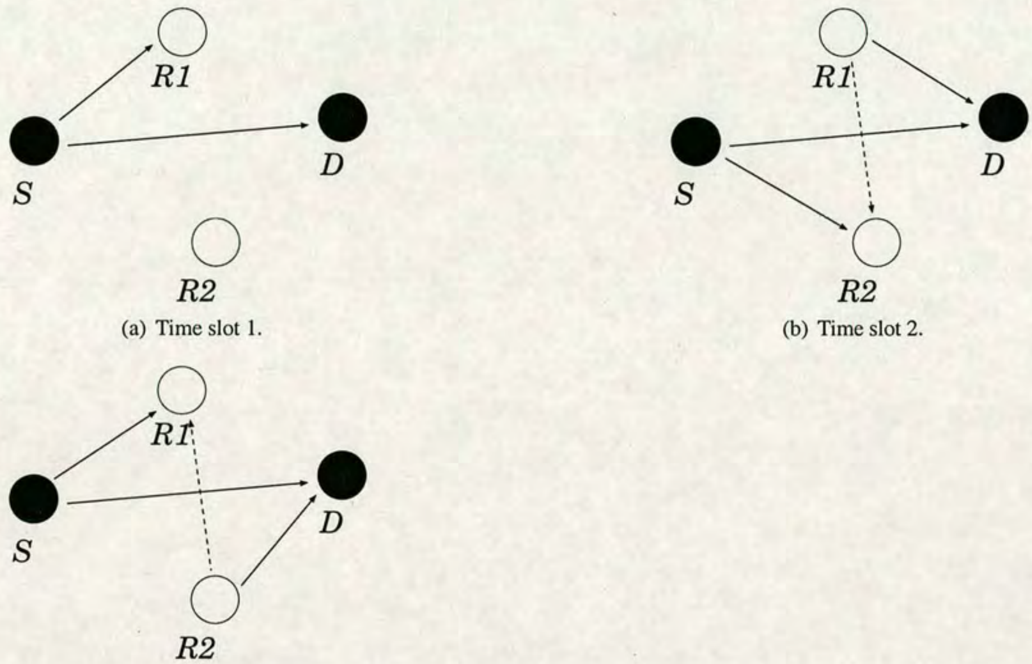
Node/Time Slot	1	2
S	s_1	s_2
R	Receive s_1	s_1
D	Receive s_1	Receive s_1, s_2

Table 3.1: *Transmission schedule for protocol I in [1]*

increasing the link reliability by relaying will not compensate for its multiplexing loss of 0.5. As we mentioned in the last chapter, most of the work concentrated on improving the value of κ by applying adaptive protocols to improve the quality of source to relay link (e.g., [49, 50, 58, 64, 65]). However, few papers concentrate on recovering the multiplexing loss 0.5 due to relaying. Some work applies *independent* random codebooks at the relays to recover the multiplexing loss while retaining some diversity benefits (e.g. the parallel coding and space-time distributed coding discussed in [49] and also introduced in the last chapter, or the dynamic decode and forward (DDF) protocol discussed in [53]). However, those approaches are currently theoretical and extremely difficult to realize in reality [1]. Practical coding design (such as Turbo codes and convolutional codes) for relay networks often follows a quite different approach from these theoretical investigations (see [55–57] for example). One protocol that can avoid multiplexing loss for repetition coded relaying was the one first proposed in [1] (also analyzed in [66]) for single relay channels. In this protocol, denoted as protocol I in [1], 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. This schedule is shown in Table 3.1. 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.

In this chapter, we propose a novel transmission protocol based on protocol I in [1] for decode-and-forward 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/power gain can still be obtained. Specifically, for L symbols transmitted in $(L+1)$ time slots with joint decoding at the destination, the system can be modeled as a Multiple-Input Multiple-Output (MIMO) system with L inputs and $L+1$ outputs. It can offer multiplexing gain of at most $L/(L+1)$ and a diversity/power gain at most 2 compared with direct transmission. Thus it can make relaying more beneficial for the high SNR region while retaining

[1] we will introduce a practical space-time distributed coding method later in the chapter



(a) Time slot 1.

(b) Time slot 2.

(c) Time slot 3.

Figure 3.1: *Transmission schedule for the proposed protocol.*

diversity gain. We give a achievable rate for this protocol, and also propose a practical low-rate feedback V-BLAST detection algorithm which approaches this achievable rate. We also compare this protocol with the classic protocol designed for repetition coded two relay networks under different network geometries. Based on our practical network models we show that the proposed protocol can give a significant performance advantage over the classic protocol, especially when the two relays are located close to each other.

3.2 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 (packages), each containing L symbols denoted as s_l. These L symbols are transmitted continuously by the source, decoded and forwarded by two relays successively in turn. Before decoding L symbols, the destination waits for $L+1$ transmission time slots until all L symbols are received, from both direct link and the relay links. It then performs joint decoding of all L symbols. The

Node/Time Slot	1	2	3	4	5
S	s_1	s_2	s_3	s_4	
$R1$	Receive s_1	s_1	Receive s_3 (s_2)	s_3	
$R2$		Receive s_2 (s_1)	s_2	Receive s_4 (s_3)	s_4
D	Receive s_1	Receive s_1, s_2	Receive s_2, s_3	Receive s_3,s_4	Receive s_4

Table 3.2: *Transmission schedule for the proposed protocol when* $L = 4$

specific steps divided by each transmission (reception) time slot for every frame are described as follows:

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

Time slot 2: S transmits symbol 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 symbol 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 till *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 symbols received from the $L + 1$ transmission time slots.

The transmission schedule for the first three time slots for each frame is shown in Figure 3.1. and an example for $L = 4$ is shown in Table 3.2. 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 I in [1], the destination always receives two copies of each symbol, from both the direct and relay link (a delayed version). This implies that diversity gain can be realized by this protocol, for all transmitted data waveforms.

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). How to suppress the interference thus becomes a major problem. We will discuss this problem further in the next two sections when studying the network capacity and signalling methods for this protocol.

3.3 Achievable rate

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. a faster fading where each channel coefficient can change 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 $h_{a,b}$ as the channel coefficient between node a and b, which may contains path-loss, Rayleigh fading, and lognormal shadowing. For simplicity, we denote $C(x)$ the capacity function $\log_2(1+x)$, and SNR the ratio of transmit power to the noise variance at the receiver. We divide our capacity analysis into three parts, concerning the source to relay links, interference cancellation between relays and joint detection at the destination. We then discuss an interesting scenario where the impact of the interference between relays can be completely ignored. We show that the proposed protocol can give the best performance under this scenario if the quality of the source to relay links are guaranteed for relaying.

3.3.1 Source-relay link

It is clear that 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

$$R_i \le C\left(|h_{S,r_i}|^2 SNR\right), 1 \le i \le L \tag{3.2}$$

where r_i is the ith element in the L dimensional relay index vector

$$\mathbf{r} = \begin{bmatrix} R1 & R2 & R1 & R2 & R1\cdots \end{bmatrix}, \tag{3.3}$$

and R_i denotes the achievable rate for s_i.

3.3.2 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 [67], 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 only concern ourselves 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 upper bound of the reliable transmission rate is therefore based on different channel conditions between the source to relay and the relay to destination links. For example, when $R1$ transmits s_1 while $R2$ is receiving s_2, if $|h_{R1,R2}| \succ |h_{S,R2}|$, $R2$ firstly 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 additional rate constraint for s_1 to be correctly decoded at $R2$, which can be expressed as follows:

$$R_1 \leq C\left(\frac{|h_{R1,R2}|^2\, SNR}{1 + |h_{S,R2}|^2\, SNR}\right). \tag{3.4}$$

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

$$R_2 \leq C\left(\frac{|h_{S,R2}|^2\, SNR}{1 + |h_{R1,R2}|^2\, SNR}\right). \tag{3.5}$$

Note that this decoding criterion applies from the second time slot to the Lth time slot when transmitting each frame. In slot i, equation (3.2) can be adapted to a constraint on R_{i-1}, equation (3.5) can be adapted to a constraint on R_i.

We note that the interference between the relays might be eliminated by using advanced coding schemes at the source, such as dirty paper coding [68] across the signals transmitted from time slot 2 to L. However, this will also result in a much higher complexity. Here we emphasize that the advantage of our scheme is its *simplicity*, i.e., no complicated coding across different transmitted signals is needed.

3.3.3 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:

$$\mathbf{y} = P \underbrace{\begin{bmatrix} h_{S,D} & 0 & 0 & 0 & 0 \\ h_{r_1,D} & h_{S,D} & 0 & 0 & 0 \\ 0 & h_{r_2,D} & h_{S,D} & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & h_{r_{L-1},D} & h_{S,D} \\ 0 & 0 & 0 & 0 & h_{r_L,D} \end{bmatrix}}_{\mathbf{H}} \mathbf{s} + \mathbf{n}, \tag{3.6}$$

where $\mathbf{y}$ is the $(L+1)\times 1$ receive signal vector, $\mathbf{s}$ is the $L \times 1$ transmit signal vector and $\mathbf{n}$ is the $(L+1)\times 1$ complex circular additive white Gaussian noise vector at the destination, P is the transmit power. Unlike conventional multiple access MIMO channels, the dimension of $\mathbf{y}$, $\mathbf{s}$ and $\mathbf{n}$ is expanded in the time domain rather than the space domain. However, the capacity region should be the same, which can be expressed as follows [69]:

$$\begin{aligned} R_k &\le \log_2\left(\det\left(\mathbf{I} + \mathbf{h}_k \mathbf{h}_k^H SNR\right)\right), & (3.7)\\ R_{k_1} + R_{k_2} &\le \log_2\left(\det\left(\mathbf{I} + SNR\left(\mathbf{h}_{k_1}\mathbf{h}_{k_1}^H + \mathbf{h}_{k_2}\mathbf{h}_{k_2}^H\right)\right)\right), & (3.8)\\ &\cdots \\ \sum_{k=1}^{L} R_k &\le \log_2\left(\det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^H SNR\right)\right), & (3.9) \end{aligned}$$

where $\mathbf{h}_k$ denotes the kth column of $\mathbf{H}$. As it is extremely complicated to give an exact description for the rate region of each signal when $L > 2$, we will only concentrate on inequalities (3.7) and (3.9) 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 achievable when a space-time V-BLAST algorithm is applied at the destination to decode the signals.

3.3.4 Network achievable rate

Combining the transmission rate constraints proposed by the previous three subsections, we provide a way of calculating the network achievable rate for the proposed protocol. The specific steps are described as follows:

First, we impose a rate constraint R_i for each transmitted symbol s_i.

In the first time slot (initialization), we write:

$$R_{S,r_1} \leq C\left(|h_{S,r_1}|^2 SNR\right) \tag{3.10}$$

For the $(i+1)$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` $h_{R1,R2} \succ h_{S,r_{i+1}}$,

$$R_i \leq \min\left(C\left(\frac{|h_{R1,R2}|^2 SNR}{1+\left|h_{S,r_{i+1}}\right|^2 SNR}\right), R_{S,r_i}, C\left(|h_{S,D}|^2 SNR + |h_{r_i,D}|^2 SNR\right)\right),$$
$$R_{S,r_{i+1}} \leq C\left(\left|h_{S,r_{i+1}}\right|^2 SNR\right); \tag{3.11}$$

`else`

$$R_i \leq \min\left(R_{S,r_i}, C\left(|h_{S,D}|^2 SNR + |h_{r_i,D}|^2 SNR\right)\right),$$
$$R_{S,r_{i+1}} \leq C\left(\frac{\left|h_{S,r_{i+1}}\right|^2 SNR}{1+|h_{R1,R2}|^2 SNR}\right); \tag{3.12}$$

`end`.

Note that the term $C\left(|h_{S,D}|^2 SNR + |h_{r_i,D}|^2 SNR\right)$ represents the constraint expressed by (3.7). The purpose of the `if` statement is to select the decoding order at the relay and to decide whether equation (3.4) or (3.5) is the correct constraint to apply.

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

$$R_L \leq \min\left(R_{S,r_L}, C\left(|h_{S,D}|^2 SNR + |h_{r_L,D}|^2 SNR\right)\right). \tag{3.13}$$

Combining these constraints with the sum capacity constraint expressed by (3.9), a network achievable rate for $L+1$ time slots can then be written as

$$C_{upper} = \min\left(\max_{R_1 \cdots R_L}\left\{\sum_{i=1}^{L} R_i\right\}, \log_2\left(\det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^H SNR\right)\right)\right). \tag{3.14}$$

The first term in the min function comes from the calculation described above, the second one comes from equation (3.9). If the signal is correctly decoded and transmitted by the relays, the first term in (3.14) is omitted and the system mimics a MIMO system with L transmit antennas and $L+1$ receive antennas, for which the maximum multiplexing order is L. Compared with direct transmission over $L+1$ time slots which has multiplexing order of $L+1$, the multiplexing gain of relaying over direct transmission is $L/(L+1)$, which approaches 1 for large L. We also expect the diversity/array gain achieved by this protocol to be 2, since each signal transmission involves two independent fading channels (the direct link and a relay link).

3.3.5 Interference free transmission

From the above discussion on 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 two user Gaussian interference channel, if the interference is sufficiently strong, the network can perform the same as an interference free network [37, 67]. We mentioned in the previous subsection that the interference channel between the relays mimics a Gaussian interference channel, so the same conclusion can be made if we use the interference cancellation criterion developed in the previous subsection.

Specifically, if the interference between relays (i.e. the value of $|h_{R1,R2}|$) is so large that the following inequality holds

$$\frac{|h_{R1,R2}|^2 SNR}{1 + |h_{S,r_{i+1}}|^2 SNR} \geq \min\left(|h_{S,r_i}|^2 SNR, \left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right) SNR\right), i = 1 \cdots L \tag{3.15}$$

the relay can always correctly decode the interference and subtract it before decoding the desired message, without affecting the whole network capacity. In this situation, the capacity analysis for ith ($1 \leq i \leq L$) transmitted signal as expressed by (3.11)-(3.13) can be simplified to

$$R_i \leq \min\left(C\left(|h_{S,r_i}|^2 SNR\right), C\left(\left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right) SNR\right)\right). \tag{3.16}$$

It is obvious that the rate provided by (3.16) is significantly larger than that provided by (3.11)-(3.13).

It has been discussed before that the quality of the source to relay link (i.e. h_{S,r_i}) is an important factor that may constrain the network capacity. Similar to previous work (e.g. [49, 50]), 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:

$$|h_{S,r_i}|^2 \geq |h_{S,D}|^2 + |h_{r_i,D}|^2, 1 \leq i \leq L. \tag{3.17}$$

The rate expressed by (3.14) can be simplified to

$$C_{upper} \leq \min\left(\sum_{i=1}^{L} C\left(\left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right) SNR\right), \log_2\left(\det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^H SNR\right)\right)\right). \tag{3.18}$$

By Jensen's inequality [37] it is clear that

$$\sum_{i=1}^{L} C\left(\left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right) SNR\right) \geq \log_2\left(\det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^H SNR\right)\right). \tag{3.19}$$

Therefore the rate is equal to the MIMO channel capacity equation:

$$C_{upper} \leq \log_2\left(\det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^H SNR\right)\right). \tag{3.20}$$

This result shows that the proposed protocol can offer the best capacity performance conditioned on (3.15) and (3.17), which guarantees that the relays will correctly decode the message without affecting the network capacity.

It should be noted that this large interference scenario (i.e. condition (3.15)) is *not uncommon* in reality. A practical example is when the two relays (e.g. mobiles) are located close to each other. If routing techniques are developed to choose these relays, the capacity performance can be significantly improved by applying the proposed protocol. To satisfy condition (3.17), an adaptive protocol can be developed from the proposed protocol to guarantee that the relays are only used when (3.17) holds, otherwise direct transmission is assumed. However, for a large dense network of relays, it is not difficult to find two relays satisfying both (3.15) and (3.17). A simple example is a fixed relay network scenario [5], 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 (3.15) and (3.17) can be met by choosing two nearby fixed relays.

3.4 The V-BLAST algorithm

For the achievable rate proposed in Section 3.3, is it possible to approach it by using advanced signal processing algorithms? The answer is positive. In this section we apply the low-rate feedback V-BLAST MMSE algorithm for detecting the signals at the destination. We show by analysis and simulations that this detector can approach the capacity upper bound provided in Section 3.3.

As we introduced in Chapter 2, the V-BLAST algorithm was initially designed for spatial multiplexing MIMO systems. 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 time dimension rather than the space dimension. When considering the rate R_i, the same analysis should be made as in Section 3.3. We summarize the capacity calculation process as follows:

The initialization step is the same as (3.10).

For the $(i+1)$th time slot (for $1 \leq i \leq L-1$), based on the same interference cancellation criterion as in Section 3.3.2, the rate calculation can be performed as:

`if` $h_{R1,R2} \succ h_{S,r_{i+1}}$

$$\begin{aligned} R_i \leq \min\left(C\left(\frac{|h_{R1,R2}|^2 SNR}{1+|h_{S,r_{i+1}}|^2 SNR} \right), R_{S,r_i}, \log_2\left(1+SINR_{r_i}\right) \right), \\ R_{S,r_{i+1}} \leq C\left(|h_{S,r_{i+1}}|^2 SNR \right); \end{aligned} \tag{3.21}$$

```
else
```

$$R_i \leq \min\left(R_{S,r_i}, \log_2\left(1 + SINR_{r_i}\right)\right), R_{S,r_{i+1}} \leq C\left(\frac{\left|h_{S,r_{i+1}}\right|^2 SNR}{1 + \left|h_{R1,R2}\right|^2 SNR}\right); \qquad (3.22)$$

```
end.
```

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

$$R_L \leq \min\left(R_{S,r_L}, \log_2\left(1 + SINR_{r_L}\right)\right). \qquad (3.23)$$

The $SINR_{r_i}$ denotes the SINR for s_i, which is decoded, encoded and forwarded by relay r_i. The network capacity is therefore

$$C_{upper} = \max_{R_1 \cdots R_L} \left\{\sum_{i=1}^{L} R_i\right\}. \qquad (3.24)$$

The condition for interference free transmission discussed in Section 3.3.5 can be expressed as:

$$C\left(\frac{\left|h_{R1,R2}\right|^2 SNR}{1 + \left|h_{S,r_{i+1}}\right|^2 SNR}\right) \geq \min\left(C\left(\left|h_{S,r_i}\right|^2 SNR\right), \log_2\left(1 + SINR_{r_i}\right)\right). \qquad (3.25)$$

The rate for the ith ($1 \leq i \leq L$) signal under this condition can be expressed as:

$$R_i \leq \min\left(C\left(\left|h_{S,r_i}\right|^2 SNR\right), \log_2\left(1 + SINR_{r_i}\right)\right). \qquad (3.26)$$

Similar to the discussion in section 3.3.5, we can further apply adaptive protocols or make relay selections in the network to enhance the source to relay links:

$$C\left(\left|h_{S,r_i}\right|^2 SNR\right) \geq \log_2\left(1 + SINR_{r_i}\right), \qquad (3.27)$$

it is clear that (3.24) equals (3.20) under conditions (3.27) and (3.25). This implies that the V-BLAST algorithm can achieve the network capacity upper bound (3.20) 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 (3.25) and (3.27) have a higher probability of being fulfilled than those in (3.15) and (3.17) due to the following observation:

$$SINR_{r_i} \leq \left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right) SNR. \tag{3.28}$$

This further implies that the conditions in (3.25) and (3.27) are better suited to assist the VBLAST algorithm to achieve the capacity upper bound in (3.20), than those in (3.15) and (3.17). However, these conditions also imply an increased signalling overhead between the source, relays and destination in order to obtain the required SINR information.

3.5 Comparison with classic protocols

We now introduce two classic protocols using repetition coding for the two relay network in this section and compare their performance with the proposed protocol in certain scenarios.

3.5.1 Classic protocol I

The first classic protocol was presented by Laneman and Wornell [49], which has been discussed in Section 2.6.1. In a two relay scenario, 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:

$$\begin{aligned} C = \frac{1}{3} \times \min\Big(\quad & C\left(|h_{S,R1}|^2 SNR\right), C\left(|h_{S,R2}|^2 SNR\right), \\ & C\left(\left(|h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2\right) SNR\right)\Big), \end{aligned} \tag{3.29}$$

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

3.5.2 Classic protocol II

A simple improvement of Classic protocol I is to apply distributed Alamouti codes at the relays [70, 71]. 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 $R1$

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 3.3: *Comparison of the different transmission schemes for the two relay case*

transmits $[s_1, -s_2^*]$ and $R2$ transmits $[s_2, s_1^*]$. The destination uses maximum ratio combining to combine the signals received from all the four time slots in order to detect and decode them. The capacity achieved by this protocol can be written as:

$$\begin{aligned} C = \frac{1}{2} \times \min\Big(\quad & C\left(|h_{S,R1}|^2 SNR\right), C\left(|h_{S,R2}|^2 SNR\right), \\ & C\left(\left(|h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2\right) SNR\right)\Big), \end{aligned} \tag{3.30}$$

it is clear that (3.30) outperforms (3.29) as it has the same diversity gain but less multiplexing loss compared with direct transmission.

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

$$\min\left(C\left(|h_{S,R1}|^2 SNR\right), C\left(|h_{S,R2}|^2 SNR\right)\right) \geq C\left(\left(|h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2\right) SNR\right) \tag{3.31}$$

when relaying is used. The network under this condition can achieve the best capacity performance (i.e. the third term in (3.29) and (3.30)). This result clearly mimics the performance of a 3×1 single-input multiple-output (SIMO) or multiple-input single-output (MISO) system, which obtains a diversity/array gain of 3 compared with direct transmission. The maximum diversity and multiplexing gains can be achieved by the two classic relaying schemes are compared with these of the proposed transmission scheme and direct transmission in Table 3.3 for the two relay case.

3.5.3 Performance comparison

It can be seen that if the two relays are close to each other so that (3.15) holds, condition (3.17) is more likely to hold than (3.31). This implies that the best capacity (3.20) for the proposed protocol can be achieved with a higher probability than that for the classic protocols. We will

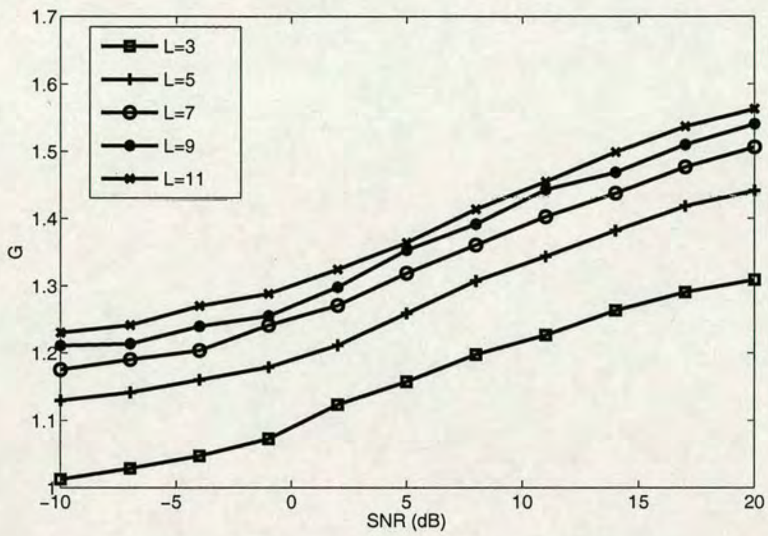


Figure 3.2: *Capacity gain of the proposed protocol over classic protocol II.*

now demonstrate an immediate and clear performance advantage of the proposed protocol over classic protocol II in this sense. We assume the adaptive protocol is applied so that relaying is only used when (3.31) holds, otherwise direct transmission will always be assumed. We calculate the following capacity gain

$$G \triangleq \frac{\frac{1}{L+1}\mathrm{E}\left[\log_2\left(\det\left(\mathbf{I}+\mathbf{H}\mathbf{H}^H SNR\right)\right)\right]}{0.5\times\mathrm{E}\left[C\left(\left(|h_{S,D}|^2+|h_{R1,D}|^2+|h_{R2,D}|^2\right)SNR\right)\right]}, \tag{3.32}$$

where $\mathrm{E}[\bullet]$ denotes the expectation and we assume each $h_{a,b}$ is an i.i.d. complex, zero mean Gaussian random variable with unit variance. G is plotted as a function of SNR in Fig. 3.2 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 less multiplexing loss and offer higher capacity gains. For smaller values of L and lower SNR, the advantage of the proposed protocol becomes less significant due to its reduced multiplexing gain and diversity loss compared with classic protocol.

It is not difficult to see that the capacity gain of proposed protocol is even higher when the relaying condition is changed to (3.17), as there will be more chance for the proposed protocol to achieve its best capacity performance. More detailed simulation comparisons for scenarios based on different network geometries will be found in the next section.

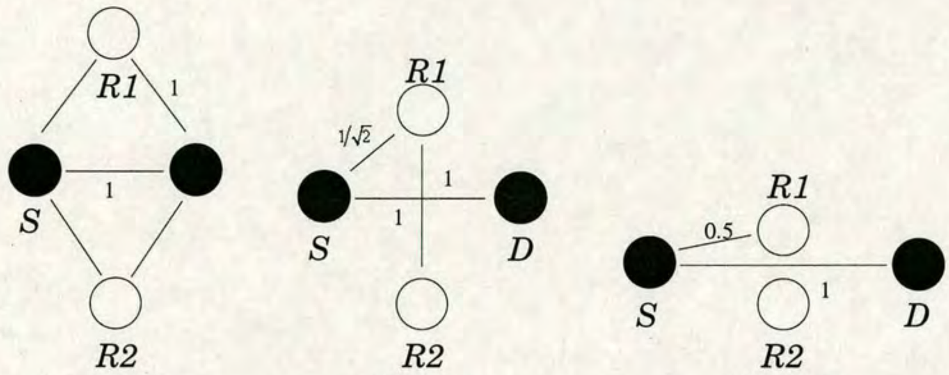


Figure 3.3: *Network models for different geometries.*

3.6 Simulation results

In this section we make further comparison of the protocols for different network geometries. Since classic protocol II is clearly better than classic protocol I, we only compare classic protocol II with the proposed protocol. As mentioned previously, to achieve a 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 develop adaptive protocols, we provide three examples here, which might be the easiest to occur to in terms of link capacity: (a) $\min\left(|h_{S,R1}|, |h_{S,R2}|\right) \geq |h_{S,D}|$, i.e. the source to relay link is better than the direct link; (b) condition (3.17) holds or (c) condition (3.31) holds. Although (b) and (c) fit better to the analysis in this chapter, condition (a) appears the *simplest* since it does not require knowledge of the relays to destination links. In the following we will only adopt (a) in the simulations. However, *similar* curve behaviors can be found if condition (b) or (c) is adopted. For a fair comparison, we combine the same adaptive transmission protocol with the proposed relaying protocol. Note that the proposed protocol can perform even better if it is modified to an adaptive protocol that considers the interference between the relays when selecting a transmission format.

Our simulations are based on three network geometries case I, II and III, which are shown in Figure 3.3. We assume that each $h_{a,b}$ contains Rayleigh fading, pathloss and independent lognormal shadowing terms. It can be written as

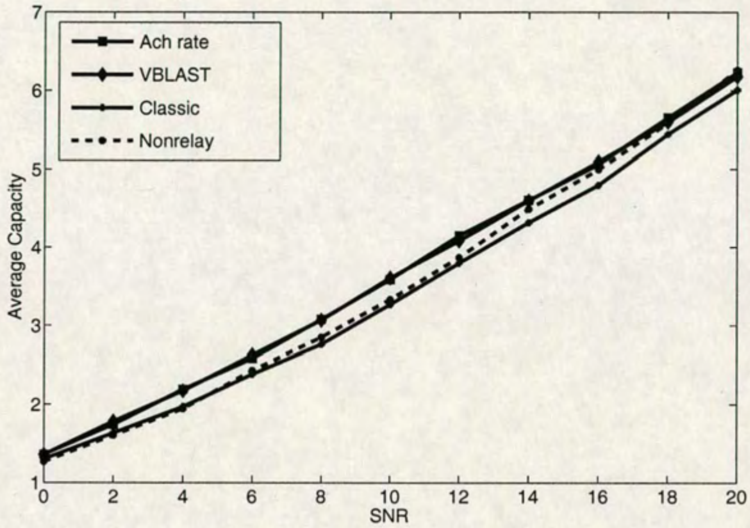
$$h_{a,b} = \sqrt{v {x_{a,b}}^{-\gamma} 10^{\varsigma/10}}, \tag{3.33}$$

where v is an i.i.d complex Gaussian random variable with unit variance, $x_{a,b}$ is the distance between the nodes a and b. The scalar γ denotes the path loss exponent (in this paper it is always set to 4). The lognormal shadowing term, ζ, is a random variable drawn from a normal distribution with a mean of 0 dB and a standard deviation δ (dB). In our simulations we use $\delta = 8$ dB, which is a value typical of shadowing deviations in urban cellular environments. We assume that the distance between the source and destination is normalized. In case I, the distance between the source to relays and relays to destination are all normalized, the distance between the two relays is therefore $\sqrt{3}$. In case II, the distance between relays are 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 0.5 while the distance between the relays can be negligible compared with the source to relays links (i.e. near 0). 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.

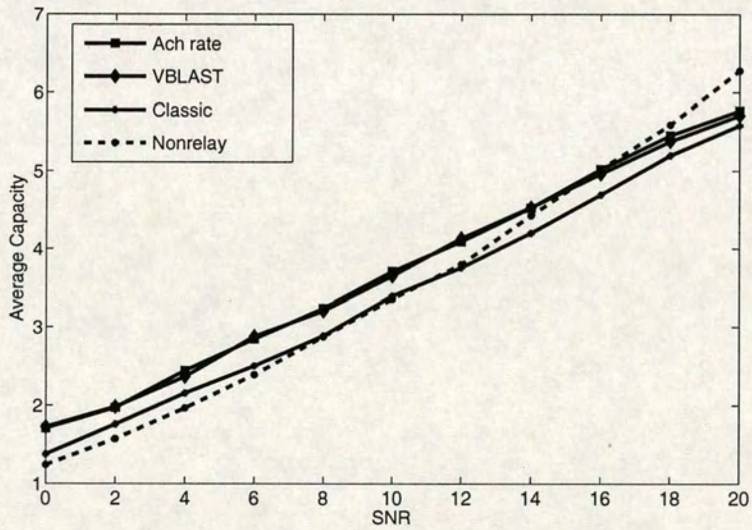
We assume $L = 7$ in the simulation, and the performance for the proposed protocol will certainly increase as L increases. Figure 3.4 and Figure 3.5 show the achievable rate for the proposed protocols (Ach rate), the capacity achieved by V-BLAST MMSE detection (VBLAST), the classic protocols (classic) and direct transmission (nonrelay), all averaged over 1000 channel realizations. It can be clearly seen from all three figures that the V-BLAST algorithm approaches the achievable rate introduced for the proposed protocol in the paper.

Figure 3.4(a) implies 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 the direct transmission for both high and low SNR region in this case.

Compared with case I, in case II (Figure 3.4(b)) the source to relay links become slightly better, therefore for the low SNR region we can see a slight capacity improvement for the proposed protocol. However, since the interference between the relays grows larger (but not large enough to meet the interference free condition), at high SNR level the increased interference between the relays dominates thermal noise and will degrade the performance of the proposed protocol. However, it still gives better performance than the classic protocol at both low and high SNR regions in this case.



(a) Case I



(b) Case II

Figure 3.4: *Average capacity of the network for different network geometries. (a) Case I. (b) Case II.*

Compared with case I and II, in case III the source to relay links are much stronger, and the relays become very close to each other so that the interference is sufficiently strong to allow interference free transmission, which has been discussed previously. It can be clearly seen in Figure 3.5 that the proposed protocol gives a significant performance advantage over direct transmission for both low and high SNR regions due to its array gain and negligible multiplexing loss. The classic protocol still performs worse than direct transmission at high SNRs due to its significant multiplexing loss compared with that case, although its performance gain over direct transmission for the low SNR region is improved.

3.7 Conclusions

In this chapter we have developed a transmission protocol for a two relay network, where the two relays forward the message successively in turn. Our capacity analysis showed that this protocol can maintain power/diversity gain while recovering the multiplexing loss associated with the classic protocol. We used a low complexity V-BLAST detection algorithm to help implement this protocol effectively. From the simulation study based on different geometries in the chapter, we can draw two main conclusions: (a) For both the proposed and classic protocols, selecting the relays in the middle region between the source and destination seems a good choice to improve the average network capacity; (b) in this scenario, while the classic protocol still loses its performance advantage for high SNR region, the proposed protocol scheme can give significant performance advantages for both the low and high SNR region.

We argue that to exploit the benefits of relaying in a practical wireless network, relay selection or adaptive transmission protocols should be jointly applied with either classic protocols or the proposed protocol. In this sense, our proposed protocols can be implemented in different network environments to achieve the best performance. This can be achieved, for example, in a fixed relay network, by deploying two fixed relays close to each other, or in a mobile relay network, by choosing two mobile relays closely located in the middle region between the source and destination.

Some recent work presented by Rankov and Wittneben [72, 73] presents a *similar* relaying strategy to the one proposed in this chapter. However, there are some major differences between their work and the work presented in this chapter: a) They ignore the direct (source to destination) link. Therefore the diversity order of the network for decode and forward relaying

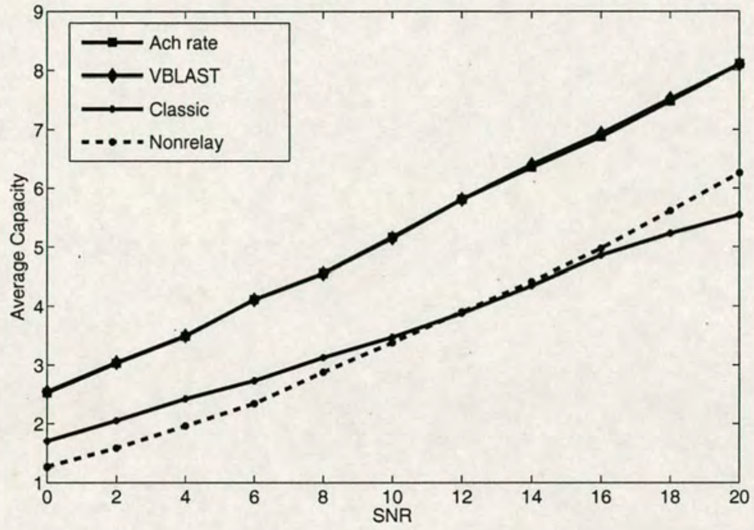


Figure 3.5: *Average capacity of the network for case III.*

is 1, as each signal is transmitted through only one (relay) route. Here the impact of both the relay and direct links are considered. Combining both routes at the destination provides the network with diversity order 2, which can provide a significant performance advantage over their proposed scheme. b) As a consequence of point (a), in the model discussed in this chapter, signal collision happens not only at the relay, but also at the destination. The proposed space-time VBLAST detector is specifically designed to tackle this issue. c) In this chapter we studied strong interference relay links and gave specific conditions for an interference-free transmission scenario. d) By considering the effect of direct link, our protocol can be modified to become an adaptive protocol which can select between direct transmission and relay transmission strategies, depending on the current channel conditions. This means our analysis fits in well with previous work (e.g. [50]).

Chapter 4
Multiple antenna relay channels

4.1 Introduction

In the last chapter, we were concentrating on single antenna systems, where each node in the network is equipped with a single antenna only. In this chapter, we extend the study to the multi-antenna systems, i.e. each node in the network is equipped with multiple antennas.

The extension of point-to-point MIMO systems to multi-node MIMO systems is largely concentrated on a multi-user scenarios. They are either MIMO broadcast channels [74–78], where one node broadcasts the information to different users, or multiple access channels [79, 80], where multiple users wish to communicate with one node simultaneously. The MIMO relay channel appears to combine some characteristics of both broadcast channels and multiple access channels and is currently an important research topic. Specially in a multi-relay scenario, the source to relay (destination) link is in the same form as a MIMO broadcast channel, while the relay to destination link is in the same form as a MIMO multiple access channel. However, the difference is that instead of receiving different information independently, the relays may receive and transmit the same information cooperatively.

A few papers have considered MIMO relay channels in the last couple of years. The approaches are basically similar to that of point-to-point systems. They are either using space-time coding [81] or a spatial multiplexing configuration. In this chapter, our study concentrates on spatial multiplexing. For spatial multiplexing structures, capacity bounds for single relay MIMO channels are presented in [82, 83], although a naive assumption of full duplex transmission is held. For multiple relay channels, capacity bounds are shown in [84], where a cooperative relaying scheme is developed to approach this capacity bound when the number of relays goes to infinity. The analysis is extended to multiple source-destination scenarios in [85, 86], where the energy efficiency of the MIMO multiple relay network considering multiple source-destination pairs is further investigated.

Most of these papers are aimed at obtaining information-theoretic limits for different relay cooperation protocols in the network. However, in wireless networks, a desirable goal is to develop suboptimal, but more practical, approaches for routing and signal processing. In this chapter we discuss practical signalling and routing schemes for MIMO relay channels in terms of network capacity. Our discussions on signalling methods are based on three relaying modes. For decode and forward relaying, we apply a V-BLAST/beamforming signalling and detection structure in the relay channel and show through simulation that its performance approaches the information-theoretic upper bound of relay channels, especially when the CSI of the relay to destination channel is available at the relay. We propose a novel *hybrid relaying* concept for MIMO relay channels, which combines the benefit of decode and forward relaying and amplify and forward relaying. We develop optimal and suboptimal hybrid relaying schemes and compare them with amplify and forward relaying and decode and forward relaying schemes. We show that hybrid relaying outperforms amplify and forward relaying, and is a good suboptimal choice compared with decode and forward relaying, especially when the number of antennas at the relay is larger than at the source and destination.

For the routing schemes designed for multiple relay channels, unlike most of the papers cited above, we assume that each relay processes and forwards the signals independently; no advanced protocol considering relay cooperation was used. We investigate multiple relay channels in a different way from previous work where the signals are multi-casted by multiple relays. In this chapter the spatial diversity of the network is exploited through adaptive routing techniques, i.e. by selecting the most preferred single relay from all candidate relays to forward the signals. We exploit the *selection diversity* for the relay channel, instead of *cooperative diversity*, through relay selection. We propose and compare the optimal routing scheme with a suboptimal routing scheme designed for MIMO relay channels. It is shown that the proposed suboptimal routing scheme approaches the performance upper bound of the optimal routing scheme with large array sizes, and can be the preferred choice for its tradeoff between performance and complexity. We also show that with a total power constraint at the relays, the multi-cast routing scheme, which uses all the relays to forward the message, is not preferable to a relay selection routing scheme when the relays are not allowed to cooperate.

Note that our proposed signalling methods for single relay channels can also be applied with our proposed selective routing schemes in a multi-relay scenario. In the last part of the chapter, we compare the spatial multiplexing array processing technique with the diversity scheme called transmit beamforming, which has been introduced in Chapter 2, in the relay scenario. We show that transmit beamforming might give higher capacity performance at low receive

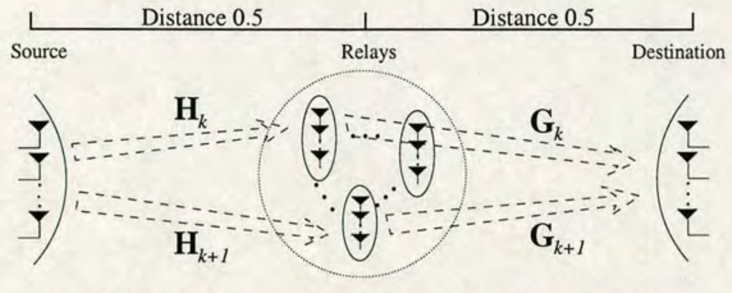


Figure 4.1: *Basic system model of a MIMO two hop network.*

signal to noise ratio (SNR) values. We indicate that it might be beneficial to combine transmit beamforming with relaying, as relaying is often applied for low SNR (for direct link). However, if the number of the relays in the network increases, this choice is likely to change.

4.2 System model

The basic system model for a single user two hop relay network is shown in Fig. 4.1. We consider a two hop network model with one source, one destination and K relays located randomly and uniformly within the middle region between the source and the destination. For easier analysis, we ignore the direct link between the source and the destination due to the larger distance and additional pathloss compared to the relay links. We also assume that the total transmit power for the source and the relays is the same; it is equally distributed among the relays. Each relay processes the received signals independently. For notational simplicity, we assume in the chapter that the source and destination have the same number of transmit and receive antennas M, while each relay has $N \geq M$ antennas. The results can be extended to a more general case where different numbers of antennas are available among each transmitter or receiver.

We restrict our discussion to the case of a slow, flat block fading model. The data transmission is over two time slots using two hops. In the first transmission time slot, the source broadcasts the signal to all the relay terminals. The input/output relation for the source to the kth relay is given by

$$\mathbf{r}_k = \sqrt{\eta}\mathbf{H}_k\mathbf{s} + \mathbf{n}_k, \tag{4.1}$$

where $\mathbf{r}_k$ is the $N \times 1$ received signal vector and η denotes the power per transmit antenna at the source. The vector $\mathbf{s}$ is the $M \times 1$ transmit signal vector with covariance matrix $\mathbf{I}_M$ and $\mathbf{n}_k$ is the $N \times 1$ complex circular additive white Gaussian noise vector at relay k which has zero mean and identity covariance matrix $\mathbf{I}_N$. $\mathbf{H}_k$ is the $N \times M$ channel transfer matrix from the source to the kth relay. $\mathbf{H}_k$ can be further expressed as $\mathbf{H}_k = \sqrt{\alpha_k}\tilde{\mathbf{H}}_k$, where the entries of $\tilde{\mathbf{H}}_k$ are i.i.d. complex Gaussian random variables with unit variance. Each factor α_k contains the pathloss and independent lognormal shadowing terms. As introduced in Chapter 2, it can be written as $\alpha_k = x^{-\gamma}10^{\zeta_k/10}$, where x is the distance between the source and relay k. The pathloss exponent γ is always set to 4 in the chapter. We also assume the lognormal shadowing term ζ_k has a mean of 0 dB and a standard deviation $\delta = 8$ dB. We normalized the range between the source and destination so that x is 0.5.

Each relay processes their received signals and transmits them to the destination. The signal received at the destination can be written as:

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{G}_k \mathbf{d}_k + \mathbf{n}_d, \tag{4.2}$$

where the matrix $\mathbf{G}_k$ is the channel matrix from kth relay to the destination, which might also be written as $\mathbf{G}_k = \sqrt{\beta_k}\tilde{\mathbf{G}}_k$, where each entry of $\tilde{\mathbf{G}}_k$ is an i.i.d. complex Gaussian random variable with unit variance. β_k contains the same pathloss modeled as α_k and independent lognormal shadowing terms with the same mean and deviation as in α_k. The vector $\mathbf{n}_d$ is the $M \times 1$ complex circular additive white Gaussian noise at the destination with identity covariance matrix. The vector $\mathbf{d}_k$ is the transmit signal vector at relay k, which should meet the total transmit power constraint:

$$\mathrm{E}\left[\|\mathbf{d}_k\|_F^2\right] \le \frac{\eta M}{K}, \tag{4.3}$$

where $\|\bullet\|_F$ denotes the Frobenius-norm. Note that this power constraint make the total transmit power at all the relays the same as that at the source. Unless specifically stated, we assume that the source does not know the channel information; the destination knows all the channel information. Based on the different CSI available at relays, we place MIMO relay channels into the following two categories[84]:

- **Non-coherent Relay Channels**, where the kth relay only has full knowledge of the channel matrix $\mathbf{H}_k$.
- **Coherent Relay Channels**, where the kth relay can obtain full knowledge of both $\mathbf{H}_k$ and $\mathbf{G}_k$.

Note that as mentioned in Chapter 2, in practice for the signalling or routing schemes to be effectively implemented for rate adaptive transmission, feedback should be allowed from the receiver to the transmitter. We will specifically state the feedback required by the particular techniques when we introduce them later in the chapter.

4.3 Relaying schemes for non-coherent single relay channels

In this section we discuss the relaying configurations for non-coherent single relay channels. For simplicity, we replace the notation $\mathbf{H}_k$ and $\mathbf{G}_k$ by $\mathbf{H}$ and $\mathbf{G}$.

4.3.1 Decode-and-forward relaying (DR)

In this scheme, we apply the V-BLAST structure introduced in Chapter 2 at the relay and destination. The message at the source is multiplexed into M different signal streams, each independently encoded and transmitted to the relay. We apply a low-rate feedback V-BLAST MMSE detector at the receiver to decode the signal streams, where the signal to interference plus noise ratios (SINRs) at the receiver are fed back to the transmitter. Specifically, the relay uses N antennas to detect each signal stream through successive interference cancellation, which consists of M iterations, each aimed at decoding one signal stream. It then decodes and re-encodes each signal stream to forward the M signal streams to the destination using *arbitrary* M antennas. The destination can then apply the V-BLAST MMSE detector to detect and decode each signal stream in the same way. The Shannon capacity from the source to the relay can be achieved by the V-BLAST MMSE detector if we assume that each signal is correctly decoded (see also (2.21)):

$$C = \log_2 \det \left(\mathbf{I} + \eta \mathbf{H}\mathbf{H}^H\right) = \sum_{i=1}^{M} \log_2 \left(1 + \eta \psi_{\mathbf{H}}^i\right) \tag{4.4}$$

in bits per channel use, where $\psi_{\mathbf{H}}^i$ is the output SINR for signal i in the V-BLAST detector at the relay. Note that this capacity (4.4) can be achieved regardless of the decoding order [43]. In this chapter, we choose the decoding order as follows: for each iteration, the receiver detects and decodes the *strongest* signal (with the highest SINR in that iteration). The main reason we recommend this decoding order is to reduce possible error propagation (i.e. improve bit error ratio) in the V-BLAST-MMSE detector in a real system. However, with rate allocation and long code block lengths, error propagation can be ignored. Maximum SINR ordering is also useful for coherent relay channels, which will be studied in the later section.

The detection at the destination is performed in the same way, and the capacity from relay to destination can also be expressed as $C = \sum_{i=1}^{M} \log_2 \left(1 + \eta\psi_{\mathbf{G}}^i\right)$, where $\psi_{\mathbf{G}}^i$ is the output SINR for signal i at destination. Note that the channel matrix $\mathbf{G}$ is of size $M \times M$ instead of $M \times N$. The network capacity for this relay configuration can be easily observed as:

$$C = 0.5 \times \sum_{i=1}^{M} \log_2 \left(1 + \eta \times \min\left(\psi_{\mathbf{H}}^i, \psi_{\mathbf{G}}^i\right)\right), \tag{4.5}$$

where the factor 0.5 denotes the half multiplexing loss due to relaying compared with a relay-free scenario. We note that in order to achieved the capacity of the system, the SINR information ($\psi_{\mathbf{H}}^i$ or $\psi_{\mathbf{G}}^i$) needs to be fed back to the transmitter for rate adaptive transmission. If the relay knows the SINR values $\psi_{\mathbf{G}}^i$ at the destination, e.g., through overhearing the channel feedback from destination to source, equation (4.5) can be maximized by ranking the $\psi_{\mathbf{H}}^i, \psi_{\mathbf{G}}^i$ in the same order (e.g. both are monotonically decreasing sequences).

With the power constraint proposed by (4.3), the mutual information upper bound for the non-coherent relay channel is

$$C = 0.5 \min\left(\log\left(\mathbf{I} + \eta\mathbf{H}\mathbf{H}^H\right), \log\left(\mathbf{I} + \eta\frac{M}{N}\mathbf{G}\mathbf{G}^H\right)\right) \tag{4.6}$$

where the channel matrix $\mathbf{G}$ has dimension $M \times N$ and the factor M/N represents the power constraint at the relay. From a practical point of view, this upper bound could be achieved by joint coding/interleaving across antennas and maximum likelihood (ML) detection at the relay and destination. However, as mentioned in Chapter 2, it is well known that the ML detection complexity increases exponentially with the number of antennas.

The capacity result in equation (4.6) could also be achieved by the V-BLAST MMSE structure, when the relay has knowledge of the channel SINR values $\psi_{\mathbf{G}}^i$. However, unlike the ML detector, this may require breaking the M received data packets and reconstructing N different data packets for retransmission to the destination, in order to exploit optimally the relay-destination SINRs $\psi_{\mathbf{G}}^i$. This results in a higher complexity and might raise implementation issues in practice. Here we take the Automatic Repeat-reQuest (ARQ) protocol as an example, which is commonly used in the communications systems.

ARQ [63, 87, 88] is an error control method for data transmission in which the receiver detects transmission errors in a message and automatically requests a retransmission from the transmitter. Usually, when the transmitter receives the ARQ, the transmitter retransmits the message until it is either correctly received or the error persists beyond a predetermined number of retransmissions. If an end-to-end (source to destination) ARQ protocol is employed, it implies that one ARQ process must be used for all M source packets, rather than M or N separate ARQ processes. An error in one of the M or N packets means that all M packets must be retransmitted from source, reducing throughput compared to the case where only the error packet is retransmitted [88]. Alternatively, separate ARQ processes must be set up for the source-relay and relay-destination links. Omitting this breaking/reconstructing process simplifies the task of the relay and the required ARQ process and is therefore assumed in our scheme, which achieves the capacity in (4.5).

As will be observed in later simulations, our configuration performs almost the same as the upper bound in (4.6) when $N = M$. This is mainly because with uncorrelated lognormal shadowing considered on both links in the network, one link will usually experience much higher signal power than the other link. This means the minimum SINRs of the two channel matrices in the V-BLAST MMSE detectors are often all for the same link, which means that (4.5) equals (4.6) for that channel realization. In the unlikely event that the shadowing coefficients are similar in magnitude, the minimum SINRs for different streams may be chosen from both links. However, ψ_{G}^i and ψ_{H}^i are likely to be similar in value as i increases. This is mainly due to the increasing diversity order, which can help average out the fading effect, in each iteration of the V-BLAST detector. As mentioned in Chapter 2, the diversity gain obtained by the V-BLAST-MMSE detector without ordering is shown to be $(N - M + i)$, which equals to i when $N = M$. It has also been speculated that changing the detection order will not affect the diversity order [89, 90], if not increasing the diversity order.

The situation is slightly different when $N > M$ due to increased degrees of freedom for retransmission at the relays. However, with the power constraint at the relay, the performance advantage of $M \times N$ channels over that for $M \times M$ channels is limited, as it still only has M eigenvalues and can only effectively support M data streams.

4.3.2 Amplify-and-forward relaying (AR)

In this scheme, the relay amplifies its received signal vector by:

$$\sqrt{\rho} = \sqrt{\frac{M \times \eta}{\eta \, \|\mathbf{H}\|_F^2 + N}} \tag{4.7}$$

to meet the power constraint described by (5.4) and forwards it to the destination.[1] Compared with *DR*, one obvious defect for *AR* is that while the relays amplify the signals, they also amplify the receiver noise. The average total signal to total noise ratio (TSTNR) τ at the relay can be defined as:

$$\tau \triangleq \frac{\eta \mathrm{E}\left[\|\mathbf{H}\|_F^2\right]_{\tilde{\mathbf{H}}}}{N} = \eta \alpha_k M. \tag{4.8}$$

However, this scheme requires no decoding at the relay. This means there is no decoding delay at the relay and it requires less processing power at the relay compared with *DR*, which employs full decoding and re-encoding at the relay. As will be described below, *AR* can be regarded as the simplest *hybrid relaying* scheme.

4.3.3 Hybrid relaying (HR)

In the *HR* schemes, the relays only decode the training sequence from the source to obtain full CSI, then filter the received signals based on the knowledge of CSI without decoding them. After multiplying the signal vector by the filtering weight matrix $\mathbf{W}$, the relay then amplifies and forwards the filtered signals to the destination. The amplifying factor now can be written as:

$$\sqrt{\rho} = \sqrt{\frac{M \times \eta}{\eta \, \|\mathbf{W}\mathbf{H}\|_F^2 + \|\mathbf{W}\|_F^2}}. \tag{4.9}$$

Note that for *AR*, $\mathbf{W} = \mathbf{I}$. The input/output relation from the source to destination for *HR* can be expressed as:

$$\mathbf{y} = \sqrt{\eta}\left(\sqrt{\rho}\mathbf{G}\mathbf{W}\mathbf{H}\right)\mathbf{s} + \left(\mathbf{G}\sqrt{\rho}\mathbf{W}\mathbf{n}_1 + \mathbf{n}_d\right), \tag{4.10}$$

[1]Note that another suboptimal but more convenient way of scaling power is to take expectation of $\|\mathbf{H}\|_F^2$ regarding $\tilde{\mathbf{H}}$ so that $\sqrt{\rho} = \sqrt{\frac{M \times \eta}{\eta \alpha_k M N + N}}$

where $\mathbf{n}_1$ is the white Gaussian noise vector at the relay. Thus we can treat the whole system as a point-to-point MIMO link. The network capacity can be written as follows[12]:

$$C = 0.5\log_2 \det\left(\eta\rho\left(\mathbf{GWH}\right)^H\left(\mathbf{I}+\rho\mathbf{GW}\left(\mathbf{GW}\right)^H\right)^{-1}\mathbf{GWH}+\mathbf{I}\right). \tag{4.11}$$

The capacity can be achieved by perfect decoding using the V-BLAST detector at the destination. One way to choose $\mathbf{W}$ is to enhance the average SNR at the relay. This can be done by applying a matched filter at relay. This relaying scheme is referred to as *Matched Filter based Relaying (MFR)*. By setting $\mathbf{W}=\mathbf{H}^H$, τ can be written as

$$\tau \triangleq \frac{\eta \mathrm{E}\left[\left\|\mathbf{H}^H\mathbf{H}\right\|_F^2\right]_{\tilde{\mathbf{H}}}}{\mathrm{E}\left[\left\|\mathbf{H}^H\right\|_F^2\right]_{\tilde{\mathbf{H}}}} = \eta\alpha_k\left(N+M\right), \tag{4.12}$$

which is obtained from the identity [91], $\mathrm{E}\left[\left\|\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right\|_F^2\right]_{\tilde{\mathbf{H}}} = NM\left(N+M\right)$. It can be seen that the τ for *MFR* is $(N+M)/M$ times that for *AR*, which implies that *MFR* has a larger value of τ than *AR*. However, it should also be noted that MFR has the defect of correlating the signals at each antenna, which makes it more difficult for the destination to separate the signals compared with *AR*.

One may consequently consider applying an MMSE filter at the relay. However, we have found that unlike a point-to-point MIMO link, applying MMSE filtering at the relay performs worse than matched filtering. More details can be found in [92].

It is worthwhile to mention that the term "hybrid relaying" might also refer to a class of relaying schemes which apply soft decoding or estimation of the message received at the relays [93, 94]. We note that our hybrid relaying schemes do not perform any kind of soft or hard decoding of the desired message, except for the training sequences.

4.4 Relaying schemes for coherent single relay channels

Compared with the non-coherent scenario, the relay has the freedom to explore and coordinate both source to relay and relay to destination channels in a coherent scenario. In the following we will investigate both digital and hybrid relaying.

4.4.1 Modified decode-and-forward relaying (MDR)

In this scheme the relay uses all the N antennas to retransmit the M streams to exploit beamforming gain on the relay to destination channels. We use the singular value decomposition of the $M \times N$ channel matrix $\mathbf{G} = \mathbf{U}_\mathbf{G}\mathbf{D}_\mathbf{G}\mathbf{V}_\mathbf{G}^H$, where $\mathbf{D}_\mathbf{G}$ is a diagonal matrix containing the singular values of $\mathbf{G}$. After decoding the signals in the same way as in non-coherent channels, the relay multiplies the decoded signal vector $\mathbf{d}$ by an $N \times M$ unitary matrix $\mathbf{V}_{\mathbf{G},M}$, which contains the M columns of $\mathbf{V}_\mathbf{G}$ corresponding to the M nonzero singular values of $\mathbf{G}$. The destination then multiplies the received signal vector by $\mathbf{U}_\mathbf{G}$. The source to destination MIMO channels become M parallel channels and each antenna branch can perform detection independently. To achieve the optimal capacity from the relay to destination link, a waterfilling algorithm is applied at the relay transmitter. The capacity of this scheme can therefore be expressed as

$$C = 0.5 \times \sum_{i=1}^{M} \log_2 \left(1 + \eta \times \min\left(\psi_\mathbf{H}^i, \frac{M}{N}\gamma^i \lambda_\mathbf{G}^i\right)\right), \tag{4.13}$$

where $\lambda_\mathbf{G}^i$ is the ith eigenvalue[2] of matrix $\mathbf{G}\mathbf{G}^H$ and γ^i represents the power allocation at the relay transmit antennas for stream i and can be expressed as:

$$\gamma^i = \left(\mu - \frac{N}{M\eta\lambda^i}\right)^+, \sum_{i=1}^{M}\gamma^i = N.$$

Equation (4.13) can be maximized by ranking the columns of $\mathbf{U}_\mathbf{G}$ and $\mathbf{V}_{\mathbf{G},M}$ so that $\psi_\mathbf{H}^i$ and $\gamma^i\lambda_\mathbf{G}^i$ are ranked in the same order (e.g. both are monotonically decreasing sequences). It is not difficult to see that the maximum SINR decoding order at the relay suggested in Section 4.3.1 might help improve the capacity (4.13) in this scenario, as it can maximize the smallest $\psi_\mathbf{H}^i$ by decoding the weakest signal last.

The mutual information upper bound for coherent relay channels is:

$$C = 0.5 \min\left(\log\left(\mathbf{I} + \eta\mathbf{H}\mathbf{H}^H\right), \log\left(\mathbf{I} + \eta\frac{M}{L}\mathbf{G}\Sigma\mathbf{G}^H\right)\right) \tag{4.14}$$

where $\Sigma = \operatorname{diag}\left\{\gamma^1, \cdots, \gamma^M\right\}$ denotes the digonal matrix generated from the iterative waterfilling algorithm conducted at the relay before retransmission. Note that this upper bound again requires de-multiplexing and re-multiplexing of the data streams at the relay. It can also

[2] In this chapter we assume the eigenvalues of matrices are ordered arbitrarily unless specifically stated, as we will discuss the eigenvalue ordering problem in the following section

be observed that (4.13) approaches (4.14) for the same reason as stated in section 4.3.1 for digital relaying for non-coherent relay channels, though this time it is true not only for $N = M$ but also for $N > M$. Note that in a point-to-point MIMO link for $N = M$ the benefit of waterfilling is quite small and even then is only useful for low receive SNR values [40]. In a relay scenario where lognormal shadowing terms are considered, even this benefit is negligible due to the minimum function in capacity calculation (i.e. the average value of (4.6) is about the same as that of (4.14)). However, the benefit of waterfilling grows when N becomes larger than M, as energy is only allocated to the non-zero eigenmodes of the channel matrix. As can be seen from (4.13) and (4.14), when $N > M$ the capacity performance is mainly constrained by the relay to destination link due to the scaling factor M/N, thus waterfilling can significantly improve the network capacity by increasing the gain of γ^i as N increases.

4.4.2 Optimal hybrid relaying (OHR)

We now give an information-theoretic study on the optimal configuration for hybrid relaying based on knowledge of both $\mathbf{G}$ and $\mathbf{H}$ at the relay. We first replace $\mathbf{GW}$ with $\mathbf{M}$ in (4.11). We write the singular value decomposition $\mathbf{M} = \mathbf{UDV}^H$ and $\mathbf{H} = \mathbf{U_H D_H V_H^H}$. Recall the identity

$$\det(\mathbf{I} + \mathbf{AB}) = \det(\mathbf{I} + \mathbf{BA}). \tag{4.15}$$

The capacity can then be rewritten as follows

$$\begin{aligned}
C=&0.5\log_2\det\left(\eta\rho\mathbf{V_H D_H^H U_H^H V D^H U^H}\left(\mathbf{I}+\rho\mathbf{U}\Lambda\mathbf{U}^H\right)^{-1}\mathbf{UDV}^H\mathbf{U_H D_H V_H^H}+\mathbf{I}\right)\\
=&0.5\log_2\det\left(\eta\mathbf{V_H D_H^H U_H^H V}\left(\frac{\rho\Lambda}{1+\rho\Lambda}\right)\mathbf{V}^H\mathbf{U_H D_H V_H^H}+\mathbf{I}\right)\\
=&0.5\log_2\det\left(\eta\Lambda_{\mathbf{H}}^{1/2}\mathbf{U_H^H V}\left(\frac{\rho\Lambda}{1+\rho\Lambda}\right)\mathbf{V}^H\mathbf{U_H}\Lambda_{\mathbf{H}}^{1/2}+\mathbf{I}\right)
\end{aligned} \tag{4.16}$$

where $\Lambda = \mathbf{DD}^H = \operatorname{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_N\}$, contains the eigenvalues of $\mathbf{MM}^H$ and $\Lambda_{\mathbf{H}} = \operatorname{diag}\left\{\lambda_{\mathbf{H}}^1, \cdots, \lambda_{\mathbf{H}}^N\right\}$. Using Hadamard's inequality [37], the capacity is maximized when

$$\mathbf{U_H} = \mathbf{V}. \tag{4.17}$$

Thus $\mathbf{M} = \mathbf{UDU_H^H}$ and (4.16) can be written as:

$$C=0.5\sum_{i=1}^{M}\log_2\left(1+\eta\lambda_{\mathbf{H}}^{i}\frac{\rho\lambda_i}{\rho\lambda_i+1}\right) \tag{4.18}$$

$$=0.5\sum_{i=1}^{N}\log\left(1+\eta\lambda_{\mathbf{H}}^{i}\left(\frac{M\eta\lambda_i}{J+M\eta\lambda_i}\right)\right), \tag{4.19}$$

where $J=\eta\left\|\mathbf{WH}\right\|_F^2+\left\|\mathbf{W}\right\|_F^2$ is the signal energy received at the relay. For every fixed λ_i, maximizing C is equivalent to minimizing J. By replacing $\mathbf{W}$ with $\mathbf{W}=\mathbf{G}^{\dagger}\mathbf{M}$, it follows that

$$J=\mathrm{tr}\left(\eta\mathbf{WHH}^H\mathbf{W}^H+\mathbf{WW}^H\right)$$

$$=\mathrm{tr}\left(\mathbf{G}^{\dagger}\mathbf{UD}\left(\eta\Lambda_{\mathbf{H}}+\mathbf{I}\right)\mathbf{D}^H\mathbf{U}^H\left(\mathbf{G}^{\dagger}\right)^H\right) \tag{4.20}$$

$$=\mathrm{tr}\left(\mathbf{U}^H\left(\mathbf{U_G}\Lambda_{\mathbf{G}}^{\dagger}\mathbf{U}_{\mathbf{G}}^H\right)\mathbf{U}\left(\mathbf{D}\left(\eta\Lambda_{\mathbf{H}}+\mathbf{I}\right)\mathbf{D}^H\right)\right), \tag{4.21}$$

where the identity $\mathrm{tr}\left(\mathbf{AB}\right)=\mathrm{tr}\left(\mathbf{BA}\right)$ is used to go from (4.20) to (4.21). To minimize (4.21), note that for a unitary matrix $\mathbf{U}$, $N\times N$ Hermitian matrix $\mathbf{S}$ and diagonal matrix $\mathbf{\Sigma}$, we have the following identity [95]:

$$\mathrm{tr}\left(\mathbf{U}^H\mathbf{SU\Sigma}\right)\geq\sum_{i=1}^{N}a_ib_{N-i+1}, \tag{4.22}$$

where $a_1\leq\cdots\leq a_N$ are eigenvalues of $\mathbf{S}$ and $b_1\leq\cdots\leq b_N$ are the diagonal entries of Σ. So we see that the minimum J should be in the form:

$$J=\sum_{i=1}^{M}\left(\eta\lambda_{\mathbf{H}}^{i}+1\right)\lambda_{\mathbf{G}^{\dagger}}^{i}\lambda_i, \tag{4.23}$$

which can be obtained by choosing

$$\mathbf{U}=\mathbf{U_G}, \tag{4.24}$$

where $\lambda_{\mathbf{H}}^{i}$ and $\lambda_{\mathbf{G}^{\dagger}}^{i}$ are the M non-zero eigenvalues of $\mathbf{HH}^H$ and $\left(\mathbf{GG}^H\right)^{\dagger}$, respectively. Note that an ordering of $\lambda_{\mathbf{G}^{\dagger}}^{i}$ corresponds to the reverse ordering of $\lambda_{\mathbf{G}}^{i}$.

To obtain the maximum C in (4.19), we should follow three steps:

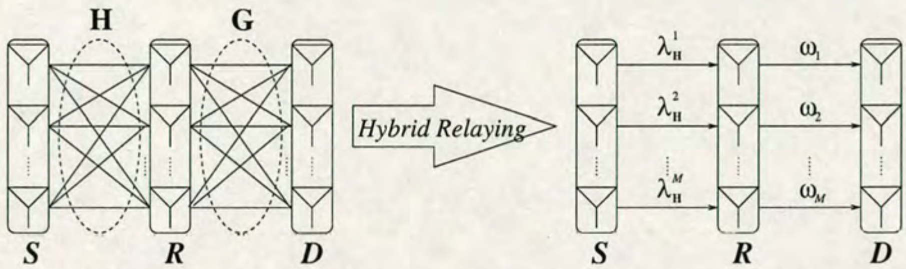


Figure 4.2: *The process of hybrid relaying for coherent relay channels. The MIMO relay channels are decomposed into several parallel channels each with gain $\lambda_{\mathbf{H}}^i\omega_i$, where $\omega_i = \frac{\rho\lambda_i}{1+\rho\lambda_i}$. For modified amplify-and-forward relaying, λ_i is $\lambda_{\mathbf{G}}^i$. For modified matched filter relaying, λ_i is $\lambda_{\mathbf{H}}^i\lambda_{\mathbf{G}}^i$.*

Step 1: We calculate J in equation (4.23) as a function of λ_i for every ordering [3]of $\lambda_{\mathbf{H}}^i$ and $\lambda_{\mathbf{G}^\dagger}^i$. Thus we obtain $M!$ expressions for J.

Step 2: For every expression for J, we evaluate (4.19). Then the capacity becomes a function of λ_i. We then calculate the maximum value of this function. There are $M!$ maximum values of C and we denote each one of them as $C_{\max}^i$.

Step 3: The final optimal C_{opt} is obtained as $\max\left(C_{\max}^1, \cdots, C_{\max}^{M!}\right)$.

Though complicated in the calculation process, the underlying ideas behind these steps are very simple: (a) to optimally match the source to relay and relay to destination eigenmodes; (b) to find the optimal power allocation at the relay transmitters based on these matched eigenmodes. A closed form solution for each value of $C_{\max}^i$ might be extremely complicated, as we shall first obtain the optimum relation between each λ_i by solving the M differential equations $\partial C/\partial\lambda_i = 0$. In practice we can calculate $C_{\max}^i$ numerically (i.e. by *fminbnd* function in Matlab). The calculation complexity for J is $M!$, which is also extremely high for $M > 2$. However, this approach gives us the theoretical capacity upper bound for MIMO coherent relay channels when using hybrid relaying schemes.

We know that in order to maximize C the matrix $\mathbf{GW}$ should be in the form of $\mathbf{U_G D U_H^{\mathit{H}}}$. The MIMO relay channel can thus be decomposed into several parallel channels each with gain $\lambda_{\mathbf{H}}^i\omega_i$, where $\lambda_{\mathbf{H}}^i$ is the source to relay channel gain and $\omega_i = \rho\lambda_i/(1+\rho\lambda_i)$ can be regarded as the relay to destination channel gain. The scalar ω_i is optimized by the weight matrix $\mathbf{W}$ under

[3]We conjecture from our simulation results that arranging $\lambda_{\mathbf{H}}^i$ and $\lambda_{\mathbf{G}}^i$ both in decreasing order results in the optimal C

the power constraint at the relay. A visual description for this process is shown in Fig. 4.2. Based on this discovery, we now propose two practical suboptimal hybrid relaying schemes.

4.4.3 Suboptimal hybrid relaying schemes

4.4.3.1 Modified amplify-and-forward relaying (MAR)

One simple way to make $\mathbf{GW}$ have the form of $\mathbf{U_G D U_H^{\mathit{H}}}$ is to make

$$\mathbf{W} = \mathbf{V_G U_H^{\mathit{H}}}. \tag{4.25}$$

Then $\mathbf{GW} = \mathbf{U_G D_G U_H^{\mathit{H}}}$ and $\mathbf{GWH} = \mathbf{U_G D_G D_H V_H^{\mathit{H}}}$. The capacity can be expressed as:

$$C{=}0.5\sum_{i=1}^{M}\log_2\left(1+\eta\lambda_{\mathbf{H}}^{i}\frac{\rho\lambda_{\mathbf{G}}^{i}}{\rho\lambda_{\mathbf{G}}^{i}+1}\right). \tag{4.26}$$

$$={0.5}\sum_{i=1}^{M}\log_2\left(1+\eta\lambda_{\mathbf{H}}^{i}\frac{\eta M\lambda_{\mathbf{G}}^{i}}{\eta M\lambda_{\mathbf{G}}^{i}+\eta\sum\limits_{m=1}^{M}\lambda_{\mathbf{H}}^{m}+N}\right). \tag{4.27}$$

Compared with (4.18), *MAR* simply replaces λ_i with $\lambda_{\mathbf{G}}^{i}$. To maximize C, the columns of $\mathbf{U_H}$ can be ordered to make $\lambda_{\mathbf{H}}^{1} \leq \cdots \leq \lambda_{\mathbf{H}}^{i} \leq \cdots \leq \lambda_{\mathbf{H}}^{M}$; the columns of $\mathbf{V_G}$ can be ordered in a similar fashion. The amplifying factor ρ can be written as:

$$\rho = \frac{M\eta}{\eta\left\|\mathbf{V_G U_H^{\mathit{H}} H}\right\|_F^2 + \left\|\mathbf{V_G U_H^{\mathit{H}}}\right\|_F^2} = \frac{M\eta}{\eta\left\|\mathbf{H}\right\|_F^2 + N} \tag{4.28}$$

where we use the identity

$$\|\mathbf{UA}\|_F = \|\mathbf{A}\|_F \tag{4.29}$$

for any unitary matrix $\mathbf{U}$. This is the same value for ρ as in *AR*. We thus denote this scheme as modified analogue relaying. We also note that J and τ for *MAR* have the same form as in *AR* by employing (4.29). Compared with *AR*, it can be seen from (4.26) that the relay is able to decompose the channels and coordinate the backward channels with the forward channels to optimize the sum capacity for M parallel data streams.

4.4.3.2 Modified Matched Filter Relaying (MMFR)

As in *MAR*, we design a new $\mathbf{W}$ based on the matched filter weight matrix $\mathbf{H}^H$. If we make $\mathbf{W} = \tilde{\mathbf{W}}\mathbf{H}^H$ and write the following:

$$\mathbf{G}\mathbf{W} = \mathbf{U}_\mathbf{G}\mathbf{D}_\mathbf{G}\mathbf{V}_\mathbf{G}^H\tilde{\mathbf{W}}\mathbf{V}_\mathbf{H}\mathbf{D}_\mathbf{H}^H\mathbf{U}_\mathbf{H}^H. \tag{4.30}$$

we can see that to make $\mathbf{M}$ have the form $\mathbf{U}_\mathbf{G}\mathbf{D}\mathbf{U}_\mathbf{H}$, we can make

$$\tilde{\mathbf{W}} = \mathbf{V}_{\mathbf{G},M}\mathbf{V}_\mathbf{H}^H, \tag{4.31}$$

and $\mathbf{D}$ becomes $\mathrm{diag}\big(d_\mathbf{G}^1\lambda_\mathbf{H}^1, \cdots, d_\mathbf{G}^i\lambda_\mathbf{H}^i, \cdots, d_\mathbf{G}^M\lambda_\mathbf{H}^M\big)$, where $d_\mathbf{G}^i$ are the singular values of $\mathbf{G}$ and $\mathbf{V}_{\mathbf{G},M}$ has been introduced in section 4.4.1. The capacity can be written as

$$C = 0.5\sum_{i=1}^{M}\log_2\left(1 + \eta\lambda_\mathbf{H}^i\frac{\rho\lambda_\mathbf{H}^i\lambda_\mathbf{G}^i}{\rho\lambda_\mathbf{H}^i\lambda_\mathbf{G}^i + 1}\right). \tag{4.32}$$

By comparing (4.32) with (4.18), we see that *MMFR* simply replaces λ_i with $\lambda_\mathbf{H}^i\lambda_\mathbf{G}^i$. To maximize C, the columns of $\mathbf{V}_\mathbf{H}$ can be ordered to make $\lambda_\mathbf{H}^1 \leq \cdots \leq \lambda_\mathbf{H}^i \leq \cdots \leq \lambda_\mathbf{H}^M$; the columns of $\mathbf{V}_{\mathbf{G},M}$ can be ordered in a similar fashion. We also note that J, ρ and τ for *MMFR* have the same form as in *MFR* by employing (4.29). Compared with *MAR*, this scheme has the advantage of enhancing the SNR at the relay.

It should be noted that when M reduces to 1, equation (4.32) can be rewritten as:

$$C = 0.5 \times \log_2\left(1 + \eta\lambda_\mathbf{H}^1\frac{\eta\lambda_\mathbf{G}^1}{\eta\lambda_\mathbf{G}^1 + \eta\lambda_\mathbf{H}^1 + 1}\right). \tag{4.33}$$

It is not hard to see that (4.33) equals (4.19) if we replace J in (4.19) with the expression in (4.23) for $M = 1$. This means that *MMFR* becomes the optimal hybrid relaying scheme for $M = 1$. However, for $M \geq 2$, the signals become more correlated due to the matched filter factor $\mathbf{H}^H$ in the weight matrix $\mathbf{W}$, which impairs the sum capacity.

As we mentioned before, compared with analogue relaying, hybrid relaying needs a filter at the relay to refine the messages. However, it can be seen in this section that by filtering the signals at the relays, the MIMO relay channel can be decomposed to several independent parallel channels. This will significantly reduce the detection complexity at the destination compared with analogue relaying, as each stream can be detected separately in parallel and no non-linear

	Filtering at R	Non-linear detection at R	Non-linear detection at D
AR	No	No	Yes
HR	Yes	No	No
DR	Yes	Yes	Yes

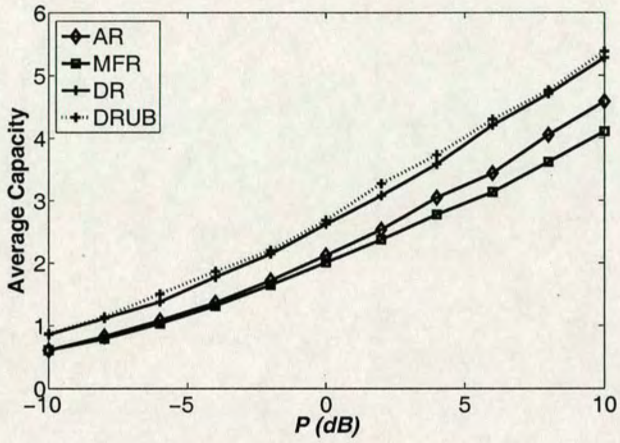
Table 4.1: *Processes for different relaying methods, where R denotes the relay and D denotes the destination*

detector such as V-BLAST is required. Table 4.1 briefly compares the processes for different relaying methods.

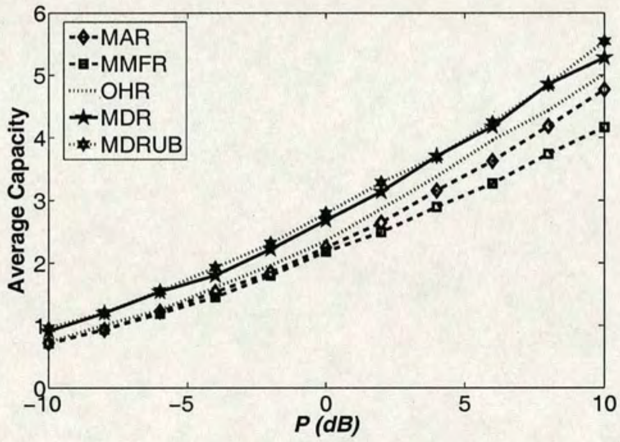
4.4.4 Comparison of Relaying Schemes for Single Relay Channels

We calculate the average Shannon capacity (in bits per channel use) for 1000 channel realizations and we define $P = M\eta$ as the total transmit power at the source. For simplicity, we neglect the training interval in the capacity calculation, assuming that the maximum channel Doppler frequency is much less than the signalling frequency. Some examples for the impact of training on MIMO capacity performance can be found in [96] and [97]. Though its specific impact on MIMO relay channels is beyond the scope of this chapter. Figure 4.3 shows the performance results for different relaying schemes for either coherent or non-coherent relay channels when $N = M$. It can be seen that digital schemes offer the best capacity performance. For either non-coherent channels or coherent channels, *AR (MAR)* outperforms *MFR (MMFR)*, especially for higher SNR values. This implies that weakening the noise at the relays cannot compensate for the disadvantage of signal correlation in *MMFR* for $L = M$. It can also be seen that the relaying schemes designed for coherent relay channels give only a small performance advantage over those for non-coherent relay channels. In particular, the performance of digital relaying schemes (i.e. *DR*, capacity upper bound of *DR* (*DRUB*), capacity upper bound for *MDR* (*MDRUB*)) are all about the same. The reason has been stated in Section 4.3.1 and 4.4.1.

However, the situation is different for $N > M$ and Fig. 4.4 shows the simulation results for this case. It can be seen that unlike the $N = M$ case, relaying schemes for coherent channels offer a significant advantage over schemes for non-coherent channels by exploiting λ_{G}^{i} on the relay to destination channels. For non-coherent relay channels, *MFR* outperforms *AR* and approaches the performance of *DR*. Since signal correlation becomes less important as the ratio N/M increases, the value of τ for *MFR* in (4.12) increases compared to that for *AR* in (4.8). The upper bound for digital relaying (*DRUB*) is still close to the performance of *DR*, the reason for

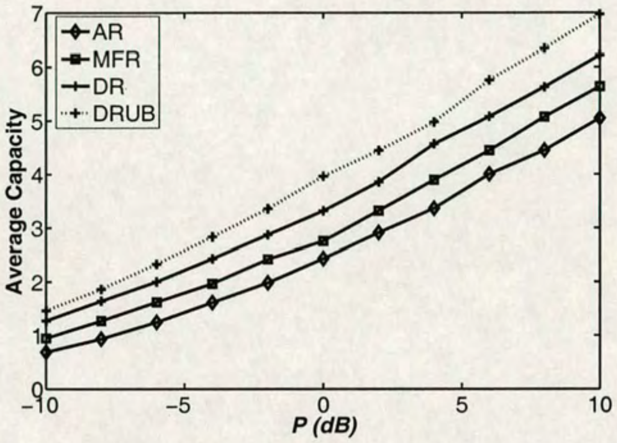


(a) Non-coherent relay channel

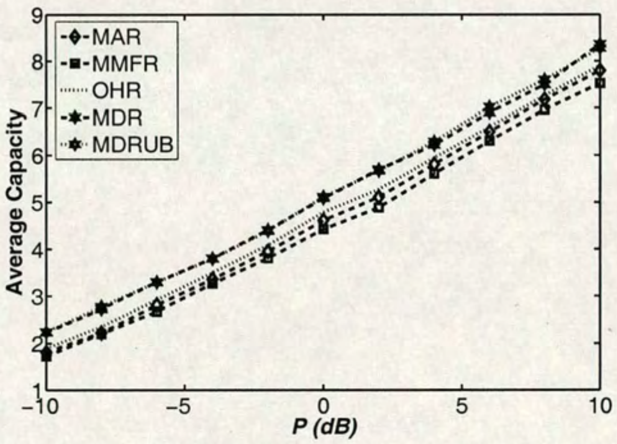


(b) Coherent relay channel

Figure 4.3: *Average capacity of single MIMO relay channels when $N = M = 2$. (a) Non-coherent relay channel. (b) Coherent relay channel.*



(a) Non-coherent relay channel



(b) Coherent relay channel

Figure 4.4: *Average capacity of single MIMO relay channels when $M = 2$, $N = 8$. (a) Non-coherent relay channel . (b) Coherent relay channel.*

this has been given in section 4.3.1. For coherent relay channels, digital relaying schemes still perform best; however, their advantage over modified hybrid relaying schemes is smaller than for $N = M$. As previously discussed, *MMFR* is the optimal hybrid scheme for $M = 1$. For $M > 1$, it can be seen that both *MAR* and *MMFR* approaches the optimal hybrid scheme as the ratio N/M increases.

From the above discussion we conclude that by increasing the number of antennas at the relays and obtaining the forward CSI of the relay channels, we can significantly increase the network capacity, especially for the digital and hybrid schemes. The hybrid schemes give a closer performance to the digital schemes when $N > M$ and are attractive if we consider the tradeoff between performance and complexity.

4.5 Routing for multiple-relay channels

In this section we extend the discussion to multiple relay channels. We assume that the relays do not communicate with each other, which is the most practical case. We will discuss and compare the capacity performance of two routing schemes based on different relay selection criteria. In contrast to *cooperative diversity* as defined in the literature, we have defined the diversity achieved by these schemes as *selection diversity*. We will compare the performance of selective routing with that of multi-cast routing where all the relays are used to forward the message, and the classic protocol II exploiting *cooperative diversity* introduced in Chapter 3. We will show the advantage of selective routing over the other two schemes in certain scenarios.

4.5.1 Optimal selective routing (OSR)

We choose the single best relay through which the highest network capacity can be obtained. We denote the capacity for the kth single relay channel as C_k, where only relay k is used. The network capacity can be expressed as:

$$C = \max\left(C_1, \cdots, C_k, \cdots, C_K\right). \tag{4.34}$$

We can apply selective routing for analogue relaying, hybrid relaying and digital relaying, where for each relaying scheme C_k is calculated according to the capacity formula provided in the previous two sections. In practice, the channel capacity of each relay link must be fed back to the source. The source then decides which relay to choose. This scheme should be optimal in the sense that it selects the single best relay channel that maximizes the network capacity. However, this scheme might require extremely high signaling overhead since for every channel realization, all K relay channel capacities have to be tested and compared before the best relay is chosen. This might only be practical for a very slow fading environment.

4.5.2 Pathloss and shadowing based selective routing (PSSR)

Instead of calculating C for each single relay route, in this scheme the relay is chosen with respect to only pathloss and shadowing coefficients of the channels, i.e. we neglect $\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{G}}_k$ which contain the Rayleigh fading coefficients. In the following we will give detailed explanation of this routing scheme and show that it approaches the performance upper bound of *OSR* for large values of N or M.

By the law of large numbers [44], we have the following identity for the $N \times M$ matrix $\tilde{\mathbf{H}}_k$:

$$\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H \overset{M\to\infty}{\to} M\mathbf{I}_N, \tilde{\mathbf{H}}^H\tilde{\mathbf{H}} \overset{N\to\infty}{\to} N\mathbf{I}_M.$$

Therefore, it is not difficult to see that for large values of both M and N the capacity upper bound of each relay channel for both *DR* and *MDR* becomes[4]

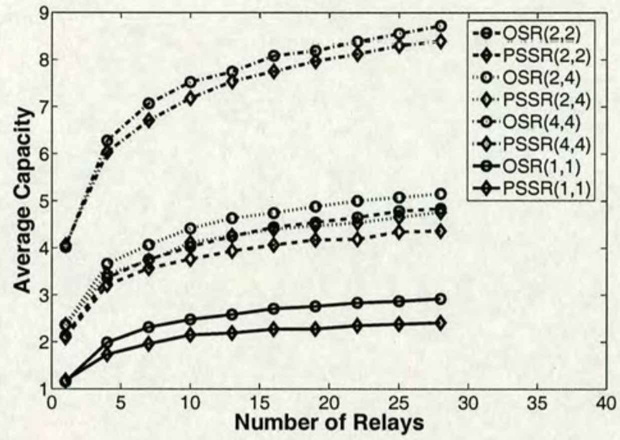
$$C = 0.5 \min\left(M \log\left(1 + \eta M \alpha_k\right), M \log\left(1 + \eta M \beta_k\right)\right)$$

Therefore we set the selection criterion as follows:

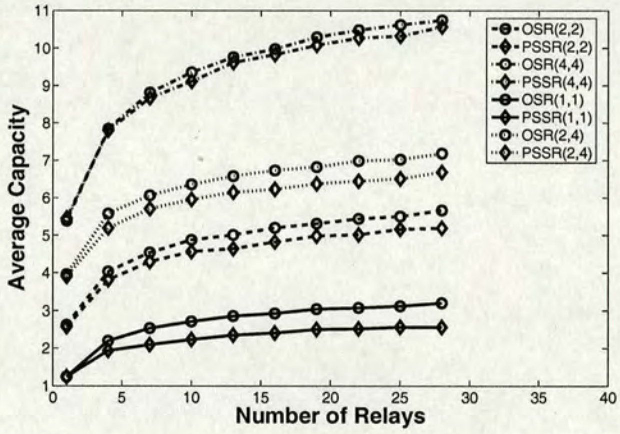
$$\begin{aligned}\Re =& \arg \max_{\Re_i \in \{1,2,\cdots,K\}} \left(M \log_2\left(1 + \eta \min\left(M\alpha_{\Re_i}, M\beta_{\Re_i}\right)\right)\right) \\ =& \arg \max_{\Re_i \in \{1,2,\cdots,K\}} \left(\min\left(\alpha_{\Re_i}, \beta_{\Re_i}\right)\right).\end{aligned} \tag{4.35}$$

For hybrid relaying (analogue relaying) schemes, note that for large values of N, $\lambda^i_{\mathbf{H}}$ and $\lambda^i_{\mathbf{G}}$ tend towards the value N. From the analysis of section 4.4.2. it is not difficult to see that $\mathbf{W} = \mathbf{I}$ leads to the optimal hybrid relaying solution, which is equal to *MAR* in this scenario.

[4]For large N only, the capacity for *DR* is $C = 0.5 \min\left(M \log\left(1 + \eta N \alpha_k\right), M \log\left(1 + \eta M \beta_k\right)\right)$ and for *MDR* is $C = 0.5 \min\left(M \log\left(1 + \eta N \alpha_k\right), M \log\left(1 + \eta N \beta_k\right)\right)$.



(a) Analogue relaying for non-coherent relay channels.



(b) Modified digital relaying for coherent relay channels.

Figure 4.5: *Comparison of average capacity of multiple-relay channels for different selective routing schemes for different antenna number allocations (M,N). The circled marked curves denote the optimal selective routing, and the diamond marked curves denote the pathloss and shadowing based selective routing.* $P = 0dB$. *(a) non-coherent relay channels. (b) coherent relay channels. Note that similar curves are obtained for digital and hybrid relaying in both cases.*

Therefore equation (4.11) can be modified as follows for large N and M:

$$C = 0.5M \log_2 \left(1 + \eta M \times \frac{M\eta\alpha\beta}{\eta M(\alpha+\beta)+1}\right). \tag{4.36}$$

The selection criterion is thus:

$$\Re = \arg \max_{\Re_i \in \{1,2,\cdots,K\}} \left(\frac{\alpha_{\Re_i}\beta_{\Re_i}}{\eta M(\alpha_{\Re_i}+\beta_{\Re_i})+1}\right). \tag{4.37}$$

Compared with the *OSR* scheme, *PSSR* does not require knowledge of $\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{G}}_k$. Since the calculation is based on shadowing coefficients which will change much more slowly than the fading channel realizations, fewer routing updates are required. This will significantly reduce signalling and computation overhead compared to *OSR*.

Figure 4.5 gives some examples of the capacity performance for *OSR* and *PSSR* schemes as the number of relays increases, while P is set to 0dB. We can see that PSSR gives very good performance even for small values of M and N. As the values of M and N both increase, the absolute capacity gap between the *OSR* and *PSSR* schemes for the same number of relays becomes smaller, and the percentage gap shrinks. Also, for a fixed value of M, increasing N can also decrease the capacity gap (e.g. $M = 2, N = 4$). These observations also indicate that MIMO configurations can significantly reduce the variability of the instantaneous channel capacity caused by Rayleigh fading.

4.5.3 Multi-cast routing (MR)

We also give a short discussion on possible multi-cast routing schemes, where all the relays are used to forward the message. For digital relaying, every relay decodes the signals it received, and forwards them to the destination. The destination receives the signals from all the relays and decodes them, so the effective MIMO channel from relays to destination can be written as $\mathbf{G}_{sum} = \sum_{i=1}^{K} \mathbf{G}_k$. For the destination to decode the signals correctly, each relay has to decode the signals correctly. For each signal s_m, we denote the Shannon capacity from source to relay k by $C_{m,\mathbf{H}_k}$, and from the relays to the destination by $C_{m,\mathbf{G}_{sum}}$. The Shannon capacity of the network for s_m can be expressed by:

$$C_m = \min\left(\min\left(C_{m,\mathbf{H}_1}, \cdots, C_{m,\mathbf{H}_K}\right), C_{m,\mathbf{G}_{sum}}\right). \tag{4.38}$$

The network capacity for all the M signals are $C = \sum_{m=1}^{M} C_m$. It can be seen from (4.38) that the capacity for each signal is constrained by the worst channel among the K source to relay channels and the relays to destination channel. As the number of antennas increases, the capacity for each signal will reduce. Therefore, the *MR* scheme turns out to be the poorest routing choice in this scenario. This observation also explains again why the source to relay link is often tested before relaying the message in most of the existing literature discussing digital relaying, as mentioned in the previous chapters.

For hybrid or analogue relaying schemes, the destination receives the filtered or amplified signals from all the relays available in the network. The source to destination channel can be written as:

$$\mathbf{y} = \sqrt{\eta}\left(\sum_{k=1}^{K} \sqrt{\rho_k}\mathbf{G}_k\mathbf{W}_k\mathbf{H}_k\right)\mathbf{s} + \left(\sum_{k=1}^{K} \sqrt{\rho_k}\mathbf{W}_k\mathbf{G}_k\mathbf{n}_k + \mathbf{n}_d\right) \tag{4.39}$$

where ρ_k is the power amplifying factor for relay k, which is the $1/K$ times the value for single relay channels expressed by (4.9).

Note that the relay channels can not be made orthogonal to each other due to the different singular vectors for each relay channel when $M > 1$ and $N > 1$, unless each relay obtains the knowledge of the other $2 \times (K-1)$ backward and forward channel matrices (i.e. the relays cooperate). However, this would involve extremely large signalling overheads. How to choose the proper weight at each relay to suppress the co-channel interference thus remains an open topic. In [84] a suboptimal scheme which requires no joint detection at the destination is presented. However, when $M = 1$, it is possible to combine the signals effectively at the destination. This will be further discussed in the next chapter.

If we simply choose $\mathbf{W}_k$ designed for single relay channels as those for multiple relay channels, each product $\mathbf{G}_k\mathbf{W}_k\mathbf{H}_k$ and $\mathbf{G}_k\mathbf{W}_k$ are i.i.d. random matrices. Therefore, for large K the contribution of the equivalent multi-cast MIMO channel to the capacity becomes the average over all single K relay channels. So the capacity of multi-cast routing schemes will be upper bounded by that of the selective routing schemes, for which only the best (near best for *PSSR*) relay channel is chosen. This trend is shown in Figure 4.6 where *PSSR* is compared to *MR*. For *MR* schemes, the signals become less correlated at the destination for larger number of relays. So matched filter based relaying outperforms analogue relaying due to the reduction in the amplified noise generated at the relays.

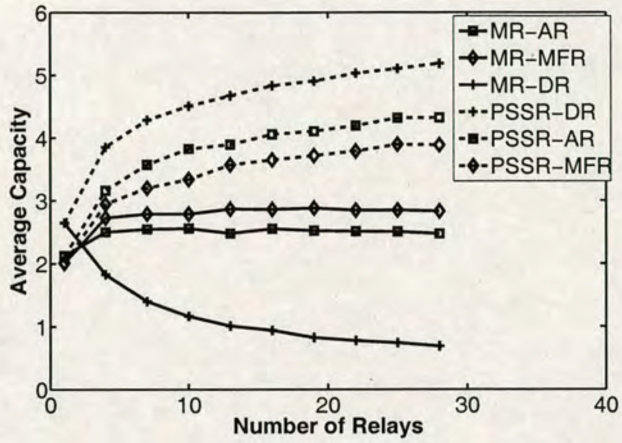


Figure 4.6: *Average capacity of non coherent multiple-relay channels for multi-cast routing and shadowing based selective routing shemes, $P = 0dB$, $N = M = 2$. Similar curve behavior can be found for modified relaying schemes for coherent multiple relay channels.*

4.5.4 Selection Diversity vs. Cooperative Diversity

We now compare selective routing with the classic protocol II introduced in Chapter 3, where Alamouti codes are applied at the relays to exploit the cooperative diversity of the network. This comparison can be regarded as a specific example of a comparison between the benefits of *cooperative diversity* and *selection diversity*.

We study the classic protocol II in Chapter 3 in a multi-antenna scenario. The source uses one antenna to transmit signals s_1 and s_2 for the first two symbol slots. The two relays use multiple antennas to decode the two signals (i.e. maximum ratio combining) and resend them using Alamouti encoding. The destination then decodes the signals by multiplying the received signal vector by the corresponding weight matrix. It will be more difficult to apply the Alamouti encoding to noisy analogue waveforms in analogue or hybrid relaying, so they are not discussed here.

After decoding and re-encoding, relay 1 uses one antenna to transmit $[s_1, -s_2^*]$ and relay 2 transmits $[s_2, s_1^*]$ over two symbol durations. Assuming the channel stays constant for the two symbol transmission periods, the capacity of the relay to destination channels for each signal can be written as

$$C_{r,d}^{STBC} = 0.5 \log_2 \left(1 + 0.5\eta \sum_{i=1}^{2} \sum_{m=1}^{M} |g_{m,i}|^2 \right), \tag{4.40}$$

where $g_{m,i}$ are the channel coefficients from relay i to antenna m at the destination. The factor 0.5 right before η denotes the half power scaling factor for each relay. For optimal selective routing, the destination uses maximum ratio combining to enhance the received signal power. The capacity can then be written as:

$$C^{DSR} = \max\left(\min\left(C_{s,r_1}, C_{r_1,d}\right), \min\left(C_{s,r_2}, C_{r_2,d}\right)\right). \tag{4.41}$$

The capacity C_{s,r_i} is the capacity of the channel between the source and relay i. The capacity $C_{r_i,d}$ is the capacity of the channel between relay i and the destination and can be written as:

$$C_{r_i,d} = \log_2\left(1 + \eta Q_i\right), \tag{4.42}$$

where $Q_i = \sum\limits_{m=1}^{M} |g_{m,i}|^2$. Note that equations (4.41) and (4.42) also hold for $M = 1$, which is the single antenna scenario. Without loss of generality, if we assume that $Q_1 \geq Q_2$, it can be seen that

$$C_{r_1,d} \geq C_{r,d}^{STBC}. \tag{4.43}$$

It follows that

$$C^{DSR} \geq \min\left(C_{s,r_1}, C_{r_1,d}\right) \geq \min\left(C_{s,r_1}, C_{s,r_2}, C_{r,d}^{STBC}\right). \tag{4.44}$$

The right hand side of this inequality is the capacity of each signal for the classic protocol II. Therefore optimal selective relaying actually outperforms classic protocol II in this example. Specifically, inequality (4.43) suggests that selection diversity can offer a better *power gain* over cooperative diversity if there is a power constraint at the relays. A higher capacity can be achieved if we give all the transmission power to the single best relay instead of splitting it equally among different relays, even if full cooperative diversity can be achieved at the relay to destination link. Inequality (4.44) implies that selection diversity might offer more diversity gain if the relays are *randomly* chosen. This is because there is only one source to relay link being considered in the minimum function for selective routing; but for space-time coded (Alamouti coded) relaying, all the source to relay links have to be considered in the capacity function due to its multi-casting nature. We note that these two inequalities (4.43) and (4.44) might be generalized to the scenarios where direct link is included and the number of relays is more than 2 (e.g. the cooperative diversity schemes discussed in [49]).

4.6 Spatial multiplexing vs. transmit beamforming

So far we have mainly concentrated on spatial multiplexing systems where the spatial multiplexing gain M is fully achieved for MIMO communications. In this section we discuss another MIMO configuration which exploits the spatial diversity of MIMO channels. This has been referred to as *transmit beamforming* in Chapter 2.3.2.3. The transmitter uses all of its antennas to transmit one signal instead of multiplexing different signals simultaneously. The weights for transmitter and receiver are chosen as the columns of right and left singular vectors of channel corresponding to the largest singular value of the MIMO channel matrix. The channel gain at the receiver is the square of the largest singular value of channel matrix which can achieve full diversity of $N \times M$. It should be noted that this scheme requires CSI at the transmitter; thus, it can only be applied to coherent channels. Furthermore, the source to relay CSI is required at the source when $M > 1$.

If we apply transmit beamforming to MIMO relay channels, we can also define decode-and-forward beamforming relaying (*DBR*) and hybrid beamforming relaying (*HBR*). For *DBR*, the relay decodes the signal, re-encodes it, weights it and forwards it to the destination. The capacity of a single relay channel can be written as:

$$C = 0.5 \times \log_2 \left(1 + M\eta \times \min\left(\lambda_{\mathbf{H}}^{\max}, \lambda_{\mathbf{G}}^{\max}\right)\right) \tag{4.45}$$

where $\lambda_{\mathbf{H}}^{\max}$ and $\lambda_{\mathbf{G}}^{\max}$ are the largest eigenvalues of $\mathbf{H}\mathbf{H}^H$ and $\mathbf{G}\mathbf{G}^H$. For *HBR*, after multiplying the received signal vector by the corresponding left singular vector of the channel $\mathbf{H}$, the relay amplifies, weights and forwards the signal to the destination. The capacity for the *HBR* scheme can be expressed as:

$$C{=}0.5 \log_2 \left(1 + M\eta\lambda_{\mathbf{H}}^{\max} \frac{\rho\lambda_{\mathbf{G}}^{\max}}{1 + \rho\lambda_{\mathbf{G}}^{\max}}\right) \tag{4.46}$$

$$={0.5} \log_2 \left(1 + M\eta\lambda_{\mathbf{H}}^{\max} \frac{\eta M\lambda_{\mathbf{G}}^{\max}}{1 + \eta M\lambda_{\mathbf{H}}^{\max} + \eta M\lambda_{\mathbf{G}}^{\max}}\right) \tag{4.47}$$

where ρ is the amplifying factor at the relay:

$$\rho = \frac{M\eta}{1 + M\eta\lambda_{\mathbf{H}}^{\max}}. \tag{4.48}$$

It should be noted that *HBR* reduces to *MMFR* when $M = 1$.

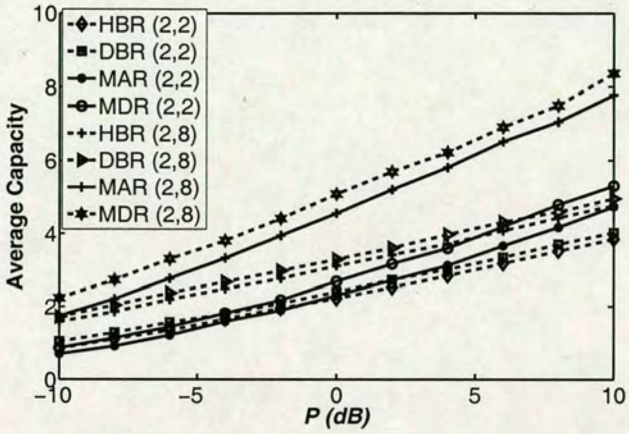


Figure 4.7: *Average capacity of MIMO single relay channels for single signal beamforming schemes and spatial multiplexing schemes when $M = N = 2$ (2,2) and $M = 2$, $N = 8$ (2,8).*

We first compare *DBR* with *MDR* for spatial multiplexing systems with source to relay CSI available at the source. For *MDR*, the source can apply the same beamforming operations as at the relay. Thus $\psi_{\mathbf{H}}^i$ is replaced by $\lambda_{\mathbf{H}}^i$ in equation (4.13). We find that *DBR* outperforms *MDR* for low output SNR values at the destination receiver. This is because $\log_2(1 + SNR_{receiver}) \approx SNR_{receiver}$ for small values of $SNR_{receiver}$, which is the receive SNR for the direct link. The capacity of *MDR* then becomes $0.5\eta \times \sum_{i=1}^{M} \min\left(\lambda_{\mathbf{H}}^i, \lambda_{\mathbf{G}}^i\right)$ for low output SNR, and this is smaller than $0.5M\eta \times \min\left(\lambda_{\mathbf{H}}^{\max}, \lambda_{\mathbf{G}}^{\max}\right)$, which approximates the capacity of *DBR* expressed in (4.45) for low output SNR.

For hybrid relaying schemes, we can also find that *HBR* outperforms *MAR* in a similar way at low transmit power levels. It can be seen that (4.27) is smaller than (4.47) for low $SNR_{receiver}$ with the inequalities:

$$\sum_{i=1}^{M} \eta\lambda_{\mathbf{H}}^i \frac{\eta M \lambda_{\mathbf{G}}^i}{N + \eta \sum_{m=1}^{M} \lambda_{\mathbf{H}}^i + \eta M \lambda_{\mathbf{G}}^i} \leq \sum_{i=1}^{M} \eta\lambda_{\mathbf{H}}^i \frac{\eta M \lambda_{\mathbf{G}}^{\max}}{1 + \sum_{m=1}^{M} \eta\lambda_{\mathbf{H}}^i + \eta M \lambda_{\mathbf{G}}^{\max}} \tag{4.49}$$

$$\leq M\eta\lambda_{\mathbf{H}}^{\max} \frac{M \times \eta M \lambda_{\mathbf{G}}^{\max}}{1 + \eta M \lambda_{\mathbf{H}}^{\max} + \eta M \lambda_{\mathbf{G}}^{\max}}, \tag{4.50}$$

which can be proved by showing that $x/(1+x)$ is monotonically increasing with increasing x.

Similar to the discussion for single antenna networks in Section 3.1, the performance gain of relaying over non-relaying for either spatial multiplexing or single signal beamforming configurations can generally be expressed as follows:

$$G \approx \frac{0.5 \log_2 (1 + \kappa SNR_{receiver})}{\log_2 (1 + SNR_{receiver})} \tag{4.51}$$

where κ denotes the link gain due to relaying (restricted to the worse link between the source-relay and the relay-destination links). $G \approx \kappa/2$ when $SNR_{receiver} \rightarrow 0$; and $G \approx 0.5$ when $SNR_{receiver} \rightarrow +\infty$. This means that relaying should be used for low SNR scenarios. When the receive SNR for direct link transmission is high, the benefit for increasing the link reliability by relaying will not compensate for its bandwidth loss of 0.5 due to the half duplex nature of the relay. Therefore, considering the performance advantage of transmit beamforming over spatial multiplexing for low receive SNR, it is better to combine transmit beamforming with relaying when multiple antennas are applied at each node in this sense. However, if we increase the number of antennas at the relays, or apply relay selection in the networks, the receive SNR in every point-to-point link in the relay network (i.e. source to relay link or relay to destination link) will be improved. In this scenario, spatial multiplexing will outperform transmit beamforming and can still be the preferred choice to be combined with relaying.

The simulation results confirms this discussion. Figure 4.7 gives the simulation results for single relay channels for different values of (M, N). It can be seen that at low values of η transmit beamforming schemes outperform spatial multiplexing schemes. However, this advantage is weakened when a larger number of antennas N are deployed at the relay. This is because larger values of N can enhance the output SNR at the destination receiver by increasing $\lambda_{\mathbf{H}}^{i}$ and $\lambda_{\mathbf{G}}^{i}$. Increasing the number of relays can also result in a larger performance improvement for spatial multiplexing compared with transmit beamforming. In the multiple relay scenario there is more freedom to obtain large values of $\lambda_{\mathbf{H}}^{i}$ and $\lambda_{\mathbf{G}}^{i}$ by selecting the proper relay. This trend is clearly shown in Fig. 4.8 where *PSSR* schemes are applied.

4.7 Conclusions

In this chapter we study MIMO spatial multiplexing configurations for relay channels. For single relay channels, we show that increasing the number of antennas at the relay can be beneficial. First it will make the proposed hybrid relaying schemes good suboptimal choices for either coherent or non-coherent relay channels, when compared with digital relaying schemes.

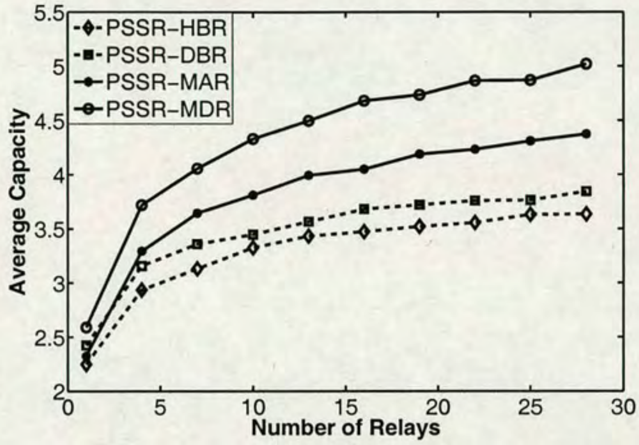


Figure 4.8: *Average capacity of MIMO multiple relay channels for single signal beamforming and spatial multiplexing systems, where shadowing based selective routing schemes are used.* $M = N = 2$*,* $P = 0dB$*.*

Second, it is shown that in such scenarios knowledge of forward CSI at the relay can help improve the network capacity significantly, if we exploit and coordinate the backward and forward channel eigenmodes at the relay. For multiple relay channels, selective routing schemes will give better performance than multi-cast or cooperative diversity schemes when the total transmit power at the relays is fixed. Then multiple relay channels can be simplified to single relay channels where only one relay is used. In this case, the conclusions for single relay channels in the paper can also be applied to multiple relay channels. We also compare the spatial multiplexing solution with transmit beamforming solution for exploiting the diversity of a MIMO link. For low receive SNR values, the transmit beamforming scheme performs better than spatial multiplexing scheme and can be the preferred choice to be combined with relaying.

Chapter 5
Single and multiple antenna relay channels

5.1 Introduction

In this chapter we continue our discussion from Section 4.5.3, where multi-cast routing is assumed. As pointed out in Section 4.5.3, when multiple antennas are deployed at the source and the destination, it is difficult for the relay to combine the signals effectively, due to the co-channel interference. In this chapter we discuss a scenario where only single antenna is applied at the source and the destination. We will show that the signals can be combined efficiently at the destination in this scenario.

Specifically, we apply two kinds of antenna combining techniques at the relay, i.e. MRC (see Section 2.3.2.1) for reception and TB (see Section 2.3.2.2) for transmission. As was been introduced in Chapter 2, those techniques were often used in point-to-point single-input multiple-output (SIMO) or multiple-input single-output (MISO) wireless links and have been shown to achieve the information theoretic upper bound and optimal diversity multiplexing tradeoff. In a relay context, we move the multiple antennas to the relays, while the source and the destination are only equipped with a single antenna. Similar to the previous chapters, our investigation is based on repetition-coded decode-and-forward transmission, where each relay fully decodes the source message, re-encodes it with the same codebook as the source and forwards it to the destination. As mentioned before, this method avoids any form of space-time coding or network coding and is easy to implement in practice. We analyze the performance of this system based on a slow fading scenario, where the transmitter only knows the second order statistics of the channel. More specifically, we examine the outage probability and the diversity multiplexing tradeoff of the network.

5.2 System model

We consider a two hop network model with one source, one destination and K relays. For simplicity we ignore the direct link between the source and the destination. The extension of all the results to include the direct link is straightforward, but is not discussed further here. We assume that the source and destination are deployed with single antennas, while relay k is deployed with m_k antennas; the total number of antennas at all relays is fixed to N[1]. This can be expressed as

$$\sum_{k=1}^{K} m_k = N. \tag{5.1}$$

We restrict our discussion to the case where the channels are slow, frequency-flat fading. The data transmission is over two time slots using two hops. In the first transmission time slot, the source broadcasts the signal to all the relay terminals. The input/output relation for the source to the kth relay is given by

$$\mathbf{r}_k = \sqrt{\eta}\mathbf{h}_k s + \mathbf{n}_k, \tag{5.2}$$

where $\mathbf{r}_k$ is the $m_k \times 1$ receive signal vector, and η denotes the transmit power at the source. The scalar s is the unit mean power transmit signal and $\mathbf{n}_k$ is the $m_k \times 1$ complex circular additive white Gaussian noise vector at relay k with identity covariance matrix $\mathbf{I}_{m_k}$. The vector $\mathbf{h}_k$ is the $m_k \times 1$ channel transfer vector from source to the kth relay. The entries of $\mathbf{h}_k$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. In the second hop, each relay processes its received signals and re-transmits them to the destination. The signal received at the destination can be written as:

$$y = \sum_{i=1}^{K} \mathbf{g}_k \mathbf{d}_k + n_d, \tag{5.3}$$

where the vector $\mathbf{g}_k$ is the channel vector from kth relay to the destination, of which each entry is an i.i.d. complex Gaussian random variable with unit variance. The scalar n_d is the complex additive white Gaussian noise at the destination with unit variance. The vector $\mathbf{d}_k$ is the transmit signal vector at relay k, which should meet the total transmit power constraint:

$$\mathrm{E}\left[\|\mathbf{d}_k\|_F^2\right] \leq \frac{\eta m_k}{N}. \tag{5.4}$$

[1]Note that in Chapter 4, N denoted the number of antennas at one relay in a multiple antenna relay scenario.

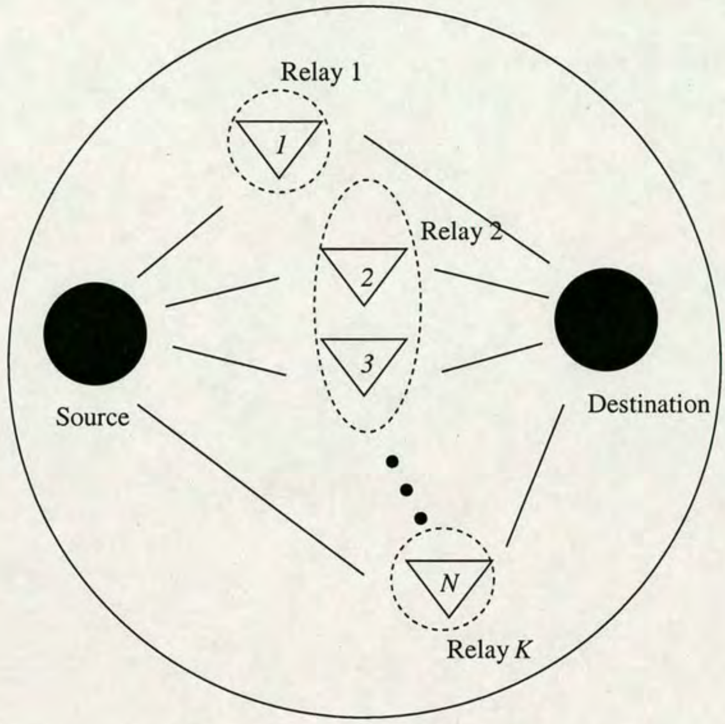


Figure 5.1: *System model for a two hop network: Source and destination are each deployed with 1 antenna. Totally N antennas are deployed at K relays. For each channel realization, either backward or forward channel coefficients for all N antennas remains the same regardless of the number of relays K.*

The same as constraint (4.3), this power constraint means that the total transmit power at the relays are the same as that at the source. Furthermore, it means that the power is allocated at each relay in proportion to its number of antennas. Note that this power allocation can be meaningful in practice, as users with larger number of antennas usually have a higher transmit power. For presentation simplicity we assume here that the total power at all relays is fixed to be η, i.e. the same as at the source. However, all the conclusions in this chapter also hold when the total power at all relays is fixed to an arbitrary constant. We assume a coherent relay channel configuration context where the kth relay can obtain full knowledge of both the backward channel vector $\mathbf{h}_k$ and the forward channel vector $\mathbf{g}_k$. Note that the forward channel knowledge can be obtained if the relay-destination link operates in a Time-Division-Duplex (TDD) mode. For a fair comparison, we also assume that for each channel realization, all the backward and forward channel coefficients for all N antennas remain the same regardless of the number of relays K. Figure 5.1 shows the system model.

5.3 Antenna diversity techniques in relay channels

In this section we apply MRC and TB techniques to the system model described in Section 5.2. We assume that each relay performs MRC of the received signals, by multiplying the received signal vector by the vector $\mathbf{h}_k^H/\|\mathbf{h}_k\|_F$. The signal at the output of the relay receiver is given by

$$\tilde{r}_k = \sqrt{\eta \sum_{i=1}^{m_k} |h_{i,k}|^2} s + \frac{\sum_{i=1}^{m_k} h_{i,k}^* n_{i,k}}{\sqrt{\sum_{i=1}^{m_k} |h_{i,k}|^2}} \tag{5.5}$$

where $h_{i,k}$ denotes the ith antenna at relay k, and $n_{i,k}$ denotes the noise factor for ith receiver input branch. The SNR at the output of the receiver can be written as:

$$\rho_k^{m_k} = \eta \sum_{i=1}^{m_k} |h_{i,k}|^2. \tag{5.6}$$

After the relays decode the signals, each relay then performs TB of the decoded waveform. If we denote the transmitted signals as t_k with unit variance, the transmitted signal vector $\mathbf{d}_k$ for relay k can be written as

$$\mathbf{d}_k = \sqrt{\frac{\eta m_k}{N}} \frac{\mathbf{g}_k^H}{\|\mathbf{g}_k\|_F} t_k. \tag{5.7}$$

The destination receiver simply detects the combined signals from all K relays. If the signals are correctly decoded at all the relays (i.e. $t_k = s$), the output signal at the destination can be written as:

$$y = s \sum_{k=1}^{K} \sqrt{\frac{\eta m_k}{N} \sum_{i=1}^{m_k} |g_{i,k}|^2} + n_d = s \sum_{k=1}^{K} \tilde{g}_k + n_d \tag{5.8}$$

It can be seen from (5.8) that by applying antenna diversity schemes at the relays, the networks can be decomposed to K diversity channels, each with channel gain $\tilde{g}_k$. The output SNR at the destination receiver can therefore be written as:

$$\rho_d^{m_k} = \left(\sum_{k=1}^{K} \sqrt{\frac{\eta m_k}{N} \sum_{i=1}^{m_k} |g_{i,k}|^2} \right)^2. \tag{5.9}$$

When all the relays are deployed with a single antenna, there is no traditional maximum ratio combining gain at the relays and the destination. However, the destination still observes a set of equal gain combined [98] amplitude signals from all relays.[2] Since we assume that the backward and forward channel coefficients for each antenna are kept the same for different values of K and m_i, the output SNR at the destination can be rewritten as

$$\rho_d^1 = \frac{\eta}{N}\left(\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|\right)^2 ; \tag{5.10}$$

when all the antennas are deployed in one relay (i.e. $K = 1$ and $m_1 = N$), full diversity gain is achieved among all the N antennas at the relay and also at the destination. The SNR for this case can be rewritten as

$$\rho_d^N = \eta \sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2. \tag{5.11}$$

5.4 Outage analysis

As we mentioned in Chapter 2, when the channel fading is slow, i.e. codewords span less than one channel block, the Shannon capacity for the Rayleigh fading channels is zero. Therefore a certain outage probability must be allowed for communicating at any finite data rate. The outage probability can be defined as (2.22). In this section we will analyze the outage probability for the relay networks, and then further characterize the diversity-multiplexing tradeoffs for different relaying schemes. For the sake of simple presentation, we put all the proofs in Appendix A.

We first study a simple protocol, in which all the relays participate in the decoding and forwarding process. We refer to this protocol as multi-cast decoding. An outage occurs whenever any relay or the destination fails to decode the signals. Before starting the outage analysis, we firstly introduce a lemma on the bounds of the value of $\rho_d^{m_k}$, i.e. the output SNR at the destination given that the signal is correctly decoded at all relays:

Lemma 1. *For any m_k, $\rho_d^1 \leq \rho_d^{m_k} \leq \rho_d^N$.*

Proof. See Appendix A.1. □

This lemma is important throughout the analysis in the chapter, as it implies that the increased

[2]Unlike [98], the equal gain combining for the relay channels is applied at the transmitter instead of the receiver.

"equal gain combining" gain at the destination can not compensate for the loss of maximum ratio combining gain at the relay and the destination when the number of relays K is increased and the numbers of antennas at each relay are reduced. Based on **Lemma 1**, we now begin our outage analysis with the following lemma:

Lemma 2. *Conditioned on all the relays correctly decoding the messages, the outage probability for the relay channels is bounded by:*

$$\frac{1}{N!}\left(\frac{N\left(2^{2R}-1\right)}{\eta}\right)^{N} \geq P_{out}^{m_k} \geq \frac{1}{N!}\left(\frac{2^{2R}-1}{\eta}\right)^{N} \tag{5.12}$$

Proof. See Appendix A.2. □

Lemma 2 indicates that the full diversity of N can be achieved regardless of the number of relays K, provided that the signals are correctly decoded at the relays. However, the diversity of the network might decrease if certain detection error happens at the relays. This is especially true for the multi-cast decoding protocol, for which we have the following theorem:

Theorem 1. *For large η, the outage probability for the multi-cast decoding is bounded by:*

$$N\left(\frac{2^{2R}-1}{\eta}\right) \geq P_{out} \geq \frac{2}{N!}\left(\frac{2^{2R}-1}{\eta}\right)^{N} \tag{5.13}$$

with equality to the right-hand side if $K=1$, to the left-hand side if $K=N$.

Proof. See Appendix A.3. □

This theorem implies that for multi-cast decoding, having more relays and less antennas per relay actually loses diversity. This is because requiring all the relays to fully decode the source information limits the performance of decode and forward to that of the poorest source to relay link. This has also been observed in Section 4.5.3, where multiple antennas are deployed at the source and the destination. Specifically, it can be seen that for $K=N$ no diversity gain is offered by relaying i.e. the SNR exponent is -1, as no diversity gain can be obtained for the source to relay links in this case. However, for $K=1$ the full diversity of N can be achieved, as the diversity gain for the source to relay link is also N. In terms of the diversity multiplexing tradeoff, we have the following theorem.

Theorem 2. *The diversity-multiplexing tradeoff curve for the multi-cast decoding scheme is bounded by*

$$1 - 2r \leq d \leq N\left(1 - 2r\right), 0 \leq r \leq 0.5 \tag{5.14}$$

with equality to right-hand side if $K = 1$, to left-hand side if $K = N$.

Proof. For large η, replace R with $r \log_2 \eta$ in (5.13), the proof is straightforward. □

It can be seen from Theorem 3 that when $K = N$, the diversity-multiplexing tradeoff for multi-cast decoding is strictly worse than that for direct transmission, which is $d = 1 - r$ as been shown in Chapter 2. When $K = 1$, however, the diversity-multiplexing tradeoff is the same as the space-time distributed coding schemes proposed in [49], where a single antenna network with N relays is considered. It is shown in [49] that full diversity N can be achieved if the message is only forwarded by the relays that can correctly decode it. As mentioned in Chapter 2, this kind of relay selection offers a *selection diversity* in the network. In fact, we can combine the antenna diversity schemes with a protocol *similar to* the one proposed by [49], which exploit further the diversity of source to relay channels by selecting the qualified relays that meet the transmission rate R, to improve the network performance when $K > 1$. Specifically, the protocol for the antenna diversity schemes is proposed as follows:

Protocol 1. ***(Selection Decoding)*** *Select $\tilde{K}$ relays with a total number of antennas $\tilde{N}$, denoted as $\Re\left(\tilde{N}, \tilde{K}\right)$, that could successfully decode the source message at a transmission rate R, to decode and forward the messages.*

We can obtain the outage probability for selection decoding by the following theorem:

Theorem 3. *For large η, the outage probability for the selection decoding scheme is bounded by:*

$$\left(\frac{2^{2R}-1}{\eta}\right)^N \times \sum_{\tilde{N}=1}^{N} \binom{N}{\tilde{N}} \frac{1}{\tilde{N}!}\tilde{N}^{\tilde{N}} \geq P_{out}^{m_k} \geq \left(\frac{2^{2R}-1}{\eta}\right)^N \sum_{\tilde{N}=1}^{N} \binom{N}{\tilde{N}} \frac{1}{\tilde{N}!\left(N-\tilde{N}\right)!}, \tag{5.15}$$

while the upper bound is met when the selected $\tilde{N}$ antennas are all within one relay.

Proof. See Appendix A.4. □

It can be seen from Theorem 3 that for selection decoding full diversity can always be achieved regardless of the number of relays K. This is clearly an advantage over the multi-cast decoding scheme. Replacing R with $r \log_2 \eta$ in (5.15), we can directly obtain the diversity-multiplexing tradeoff for selection decoding:

Theorem 4. *The diversity-multiplexing tradeoff curve for selection decoding is*

$$d = N\left(1 - 2r\right), 0 \leq r \leq 0.5, \tag{5.16}$$

which is the same as that for the space-time distributed coding protocol proposed in [49].

We claim that compared with distributed space-time coding, the messages for antenna combining techniques are simply repetition coded. Therefore it is much easier to implement than space-time coding in practice, provided that each relay antenna can obtain its forward (relay to destination) CSI. Fig. 5.2 shows the diversity multiplexing tradeoff for different protocols discussed in the paper, when $N = 5$.

Note that one different assumption between our approach and the space-time coded approach in [49] is that we allow multiple antennas to be deployed at the relays. However, it has been shown that the diversity gain that can be achieved by our approach is the same as that of the space-time coded approach in [49], as long as the total number of antennas at all relays is fixed, regardless of the number of antennas at each relay and the number of relays. Thus, the same diversity gain can be achieved even when each relay is deployed with a single antenna. Therefore the application of our scheme is quite general.

5.5 Discussion for amplify and forward relaying

This chapter has mainly concentrated on decode and forward relaying protocols. Some recent work has shown that for amplify and forward relaying, full diversity N can also be achieved for a N relay single antenna relay network. This can be done by either using N relays to retransmit the signal in N orthogonal time slots [99] or selecting the best relay to retransmit [61]. In a multi-antenna scenario where multiple antennas can be deployed at the relays, the analysis for the decode-and-forward mode can not be directly extended to the amplify-and-forward mode, as they have different relaying mechanisms. In fact, the analysis for amplify and forward relaying is more difficult due to the following two reasons: (a) the source-relay link and the relay-

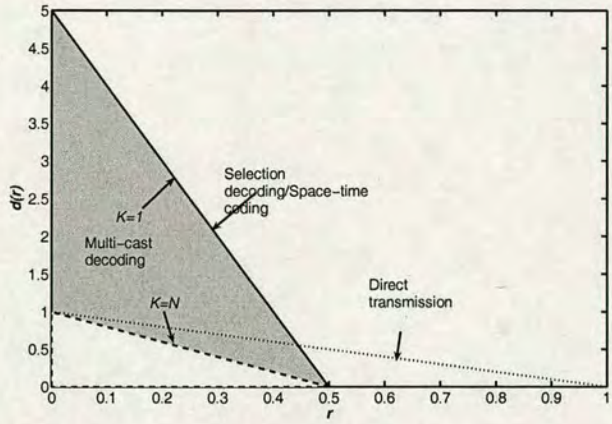


Figure 5.2: *The diversity multiplexing tradeoff for different protocols, when* $N = 5$.

destination link in amplify and forward mode can not be considered separately, as decoding is not performed at the relays; (b) the impact of the noise component generated at the relays is complicated and play a vital role in the capacity performance.

However, we can conjecture the same conclusion as of the decode-and-forward mode by imagining an extreme case, i.e. a very high transmit power level. The performance of amplify and forward mode in this scenario mimics that of decode and forward mode, as the noise component generated at the relays becomes negligible. In fact, similar conclusions for the output SNR at the destination for amplify-and-forward mode can be made as in **Lemma 1** for the decode-and-forward mode. Some convenient capacity lower bounds can also be obtained for amplify-and-forward relaying. *We put the analysis regarding these capacity bounds for amplify-and-forward relaying in Appendix B.* We also conjecture that the diversity N can be obtained regardless the number of relays K, if MRC and TB are applied at the relays.

5.6 Conclusions

From the above analysis, we can draw several conclusions regarding the antenna combining techniques introduced in this chapter: (a) provided the messages are successfully decoded at the relays, having less relays will offer better performance due to increased combining (power) gain at the destination, though the full diversity N of the network can be achieved regardless the number of antennas; (b) if all the relays participate in the decoding and forwarding process, the network performance will degrade as the number of relays increases, as the performance is always restricted to the worst source to relay link. In this sense, deploying all the antennas at a single relay is the optimal choice; (c) however, full diversity can be achieved if we apply the relay selection schemes to choose the potential relays. More specifically, the diversity-multiplexing tradeoff achieved by the antenna combining techniques is the same as that achieved by more complicated space-time distributed coding schemes. In this scenario, deploying more antennas at fewer relays is still a better choice due to improved combining (power) gain. We further note that the recently proposed fixed relay concept [5] in mesh networks allows the possibility to deploy a large number of antennas at the relay. This provides a good application for the antenna combining techniques discussed in the chapter.

Chapter 6
Conclusions

In this chapter, we firstly summarize the work and also highlight the contributions in each of the previous chapters. Then we discuss the possible options for future research.

6.1 Summary

Chapter 1 offered some basic introduction to the history of wireless communications, as well as the key technologies that will be used in current and future generation wireless networks. Specially, we introduced relaying and MIMO technologies, both of which exploit the space dimension of wireless communications. We discussed some of their important advantages over conventional systems that can improve the capacity and link reliability. We indicated that the combination of the two techniques could bring significant changes for future wireless systems.

Chapter 2 offered a detailed background for MIMO systems and relay networks. It began with an introduction to the wireless fading channel models, where different fading characteristics were modeled and discussed. We then provided a capacity analysis for a fixed channel realization. We showed that the MIMO channel can obtain a multiplexing gain and show its capacity advantage over the SISO channel for high SNR. For a fading channel scenario, we showed that MIMO channel can offer a diversity gain that can significantly improve the outage performance. We then showed that multiplexing gain and diversity gain can be obtained simultaneously by characterizing the diversity-multiplexing tradeoff for the MIMO channel. For relay networks, we indicated that a virtual MIMO system might be formed by allowing the different nodes in the network to cooperate. We introduced several transmit protocols and coding schemes that can achieve diversity gain over non-relay networks. However, we also indicated that unlike a point-to-point system, relaying might result in a multiplexing loss. This is mainly due to the half-duplex nature of relaying, i.e. it takes two time slots to transfer every signal to the destination.

Chapter 3 concentrated on a single antenna network where each node is equipped with a single antenna. We designed a new transmission protocol that can recover the multiplexing loss due to relaying, while retaining some diversity gain. We applied simple repetition coding at two relays and show that the network forms a MIMO system provided that the signals can be correctly decoded at the relays. One major issue that constrains the performance of the new protocol is interference between the relays, while one of the relays is transmitting the message and the other is listening to the source message. We provided a method to suppress the co-channel interference and showed that the network can be considered as interference free if the interference is sufficiently strong. We showed that unlike conventional transmission protocols which offer capacity advantage only for low SNR, our new transmission scheme can offer high capacity performance for both low and high SNR, especially when the relays are located closed to each other.

Chapter 4 moved the discussion to a multiple antenna network where each node is equipped with multiple antennas. We firstly discussed signalling methods for the single relay channel. For digital relaying, we applied a V-BLAST/beamforming signalling and detection structure in the relay channel and show through simulation that its performance approaches the information-theoretic upper bound of relay channels, provided that the CSI of the relay to destination channel is available at the relay. We then proposed a novel hybrid relaying concept for MIMO relay channels, which combines the benefit of digital relaying and analogue relaying. We developed optimal and suboptimal hybrid relaying schemes and compare them with analogue relaying and digital relaying schemes. We showed that hybrid relaying outperforms analogue relaying, and is a good suboptimal choice compared with digital relaying, especially when the number of antennas at the relay are larger than at the source and destination. We then moved to a multiple relay scenario, where we exploited the diversity of the relay channel by selecting the most preferred single relay from all candidate relays to forward the signals. We proposed both optimal and suboptimal relay selection schemes and show that their performance converges when a large number of antennas is deployed at each node in the network. We also showed that with a total power constraint at the relays, the multi-cast routing scheme, which uses all the relays to forward the message, is not preferable to the relay selection routing scheme when the relays are not allowed to cooperate. Finally, we compared the spatial multiplexing system with transmit beamforming in the relay scenario. We showed that transmit beamforming might give a higher capacity performance at low SNR. We indicated that it might be beneficial to combine transmit beamforming with relaying, as relaying is often applied for low SNR.

Chapter 5 studied a special case where the source and destination are equipped with one antenna, while relays might be equipped with multiple antennas. We investigated the simple repetition-coded decode-and-forward protocol and applied two antenna combining techniques at the relays: maximal ratio combining on receive and transmit beamforming. We assumed that the total number of antennas at all relays is fixed. With a total power constraint at the relays, we show that the antenna combining techniques can exploit the full spatial diversity of the relay channels and can achieve the same diversity multiplexing tradeoff as achieved by more complex space-time distributed coding techniques, such as those proposed by Laneman and Wornell [49]. We also showed that the outage probability is minimized if all the antennas are deployed at a single relay, and is maximized if each relay is deployed with a single antenna. We also indicated that similar results can be found for the amplify-and-forward relaying protocol.

6.2 Future work

For the new protocol proposed in chapter 3, one very important factor that impairs the capacity performance is the 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. More advanced coding schemes such as dirty paper coding might be used *at the source* to help eliminate the interference and thus improve the network capacity. For the adaptive protocol discussed in the paper, it is also worthwhile to develop new forms of the protocol that explicitly account for the impact of interference between relays on the network capacity. These two topics will remain as interesting future work. Furthermore, we constrain ourselves on simple repetition coding *at the relays*, it would be interesting to see how the performance would be further improved if more complicated coding schemes such as parallel channel coding are applied. In terms of the performance, we have conjectured the maximal diversity and multiplexing gain offered by the proposed protocol(see Table 3.3). It is possible to derive the diversity-multiplexing tradeoff for such scheme in future.

In Chapter 4 where multi-antenna multi-relay channel is considered, we mainly concentrated on the signaling algorithm or routing protocol design. The overall diversity and multiplexing (tradeoff) behaviors of such system is not discussed due to its complexity in analysis. This is because unlike a point-to-point MIMO channel where the antennas at the transmitter or receiver are centralized, here the antennas are only partially centralized (distributed). However, the diversity-multiplex tradeoff for this system is clearly a very important topic for future research, as it can offer significant insights on the fundamental performance limits of such system. Some other future topics for Chapter 4 include. (a) effective signalling for digital and hybrid relaying when only partial forward CSI is available at the relay for single relay channels; (b) effective signal combining at the destination when the effect of the direct link is taken into account; (c) effective signalling for digital and hybrid relaying when multi-cast routing is used.

For the system model discussed in chapter 5, the diversity-multiplexing tradeoff for amplify-and-forward relaying protocol is not derived. Compared with decode-and-forward relaying protocol, the amplify-and-forward relaying protocol does not suffer from error propagation. The capacity shall always increase as the number of the relays increase, provided that the signal waveforms can be effectively combined at the destination. We conjecture that by MRC-TB, full diversity gain could be achieved in two time slots, without any form of relay selection. The validation of this conjecture remains as future work.

Appendix A
Proof of lemmas and theorems

A.1 Proof of Lemma 1

We firstly prove that $\rho_d^1 \leq \rho_d^{m_k}$. We write the following

$$\sqrt{\rho_d^{m_k}} - \sqrt{\rho_d^1} = \sum_{k=1}^{K} \left(\underbrace{\sqrt{\frac{\eta m_k}{N} \sum_{i=1}^{m_k} |g_{i,k}|^2}}_{A_k} - \underbrace{\sqrt{\frac{\eta}{N}} \sum_{i=1}^{m_k} |g_{i,k}|}_{B_k} \right). \tag{A.1}$$

To compare A_k with B_k, we write

$$A_k^2 - B_k^2 = \frac{\eta}{N} \left(m_k \sum_{i=1}^{m_k} |g_{i,k}|^2 - \left(\sum_{i=1}^{m_k} |g_{i,k}| \right)^2 \right) \tag{A.2}$$

$$= \frac{\eta}{N} \left((m_k - 1) \sum_{i=1}^{m_k} |g_{i,k}|^2 \right) - \frac{\eta}{N} \left(\sum_{i,j=1;i\neq j}^{m_k} |g_{i,k}| \, |g_{j,k}| \right). \tag{A.3}$$

Note that

$$(m_k - 1) \sum_{i=1}^{m_k} |g_{i,k}|^2 = \sum_{i=1}^{m_k} \sum_{j=1,j\neq i}^{m_k} |g_{j,k}|^2 \tag{A.4}$$

$$= 0.5 \sum_{i,j=1;i\neq j}^{m_k} \left(|g_{i,k}|^2 + |g_{j,k}|^2 \right). \tag{A.5}$$

So (A.3) can be further written as:

$$A_k^2 - B_k^2 = \frac{\eta}{2N} \sum_{i,j=1;i\neq j}^{m_k} \left(|g_{i,k}|^2 - 2 \, |g_{i,k}| \, |g_{j,k}| + |g_{j,k}|^2 \right)$$

$$= \frac{\eta}{2N} \sum_{i,j=1;i\neq j}^{m_k} \left(|g_{i,k}| - |g_{j,k}| \right)^2 \geq 0.$$

So $A_k \geq B_k$ and therefore $\rho_d^1 \leq \rho_d^{m_k}$.

Next we prove that $\rho_d^N \geq \rho_d^{m_i}$. For simplicity, we denote

$$a_k = \sum_{i=1}^{m_k} |g_{i,k}|^2 \tag{A.6}$$

in equation (5.9) and (5.11). Then $\rho_d^N - \rho_d^{m_i}$ can be written as

$$\rho_d^N - \rho_d^{m_i} = \frac{\eta}{N}\left(\sum_{k=1}^{K}(N-m_k)\,a_k - \sum_{i,j=1;i\neq j}^{K} \sqrt{m_i m_j a_i a_j}\right). \tag{A.7}$$

Note the constraint by (5.1) in section II, we have the following:

$$(N - m_k) = \sum_{i=1,i\neq k}^{K} m_i. \tag{A.8}$$

Putting (A.8) into (A.7), we have the following:

$$\rho_d^N - \rho_d^{m_i} = \frac{\eta}{N}\left(\sum_{k=1}^{K}\sum_{i=1,i\neq k}^{K} m_i a_k - \sum_{i,j=1;i\neq j}^{K} \sqrt{m_i a_i m_j a_j}\right). \tag{A.9}$$

Note the following:

$$\begin{aligned}\sum_{k=1}^{K}\sum_{i=1,i\neq k}^{K} m_i a_k =& \sum_{i,j=1;i\neq j}^{K} (m_i a_j)\\ =&0.5\left(\sum_{i,j=1;i\neq j}^{K} (m_i a_j) + \sum_{i,j=1;i\neq j}^{K} (m_j a_i)\right).\end{aligned}$$

We can further write (A.9) as follows:

$$\begin{aligned}\rho_d^N - \rho_d^{m_i} =& \frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{K}(m_i a_j)\\ &-\frac{\eta}{N}\sum_{i,j=1;i\neq j}^{K}\sqrt{m_i a_j m_j a_i}\\ &+\frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{K}(m_j a_i)\\ =&\frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{K}\left(\sqrt{m_i a_j}-\sqrt{m_j a_i}\right)^2.\end{aligned}$$

Therefore $\rho_d^N \geq \rho_d^{m_i}$ and $\rho_d^1 \leq \rho_d^{m_k} \leq \rho_d^N$.

A.2 Proof of Lemma 2

Based on Lemma 1, it is clear that

$$P_{out}^1 \geq P_{out}^{m_k} \geq P_{out}^N, \tag{A.10}$$

where P_{out}^1 denotes the outage probability for N relay case and P_{out}^N for 1 relay case, given that the signals are correctly decoded at all the relays. Note that

$$\rho_d^1 \geq \frac{\eta}{N}\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2 = \frac{\rho_d^N}{N}, \tag{A.11}$$

inequality (A.10) can be extended as:

$$P\left[\frac{1}{2}\log_2\left(1+\frac{\rho_d^N}{N}\right) < R\right] \geq P_{out}^{m_k} \geq P\left[\frac{1}{2}\log_2\left(1+\rho_d^N\right) < R\right]. \tag{A.12}$$

After some modification, (A.12) can be rewritten as:

$$P\left[\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2 \leq \frac{N\left(2^{2R}-1\right)}{\eta}\right] \geq P_{out}^{m_k} \geq P\left[\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2 \leq \frac{2^{2R}-1}{\eta}\right]. \tag{A.13}$$

Since $\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2$ is chi-square distributed with dimension $2N$, for small ε it is easy to show that [31]

$$P\left(\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2 \leq \varepsilon\right) \approx \frac{1}{N!}\varepsilon^N. \tag{A.14}$$

Put (A.14) to (A.13) finishes the proof of *Lemma* 2.

A.3 Proof of Theorem 1

If we denote $C_r^{k,m_k} = 0.5\log_2\left(1+\rho_k^{m_k}\right)$ as the Shannon capacity from source to relay k channel for each channel realization. The outage probability is given by:

$$\begin{aligned} P_{out} &= P\left[\min\left(C_r^{k,m_k}\right) < R\right] + P\left[\min\left(C_r^{k,m_k}\right) > R\right] P_{out}^{m_k} \qquad \text{(A.15)} \\ &= 1 - \prod_{k=1}^{K}\left(1 - P\left[\sum_{i=1}^{m_i}|h_{i,k}|^2 < \frac{2^{2R}-1}{\eta}\right]\right) + P_{out}^{m_k}\prod_{k=1}^{K}\left(1 - P\left[\sum_{i=1}^{m_i}|h_{i,k}|^2 < \frac{2^{2R}-1}{\eta}\right]\right) \\ &\overset{\eta\to\infty}{\approx} 1 - \prod_{k=1}^{K}\left(1 - \frac{1}{m_k!}\left(\frac{2^{2R}-1}{\eta}\right)^{m_k}\right) + P_{out}^{m_k}, \qquad \text{(A.16)} \end{aligned}$$

where $P_{out}^{m_k}$ is bounded by (5.12). For large η, retaining only the term containing the lowest exponent of $1/\eta$ in the first term, (A.16) can be further modified as

$$P_{out} \approx \sum_{k=1}^{K}\frac{1}{m_k!}\left(\frac{2^{2R}-1}{\eta}\right)^{m_k} + P_{out}^{m_k}. \tag{A.17}$$

Observing that $\frac{1}{a!\eta^a} < \frac{1}{b!\eta^b}$ when $a > b$, P_{out} is maximized when $m_k = 1, K = N$. Therefore for large η

$$P_{out} \leq N\left(\frac{2^{2R}-1}{\eta}\right), \tag{A.18}$$

where $P_{out}^{m_k}$ is omitted due to its higher exponent. P_{out} is minimized when $m_k = N, K = 1$ and $P_{out}^{m_k} = P_{out}^{N}$. We obtain the lower bound

$$P_{out} \geq \frac{2}{N!}\left(\frac{2^{2R}-1}{\eta}\right)^N \tag{A.19}$$

and thus complete the proof.

A.4 Proof of Theorem 3

Since $\Re\left(\tilde{N},\tilde{K}\right)$ is a random set, we utilize the total probability law and write

$$P_{out} = \sum_{\Re\left(\tilde{N},\tilde{K}\right)} P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] P_{out}^{m_k|\Re\left(\tilde{N},\tilde{K}\right)}, \tag{A.20}$$

where $P_{out}^{m_k|\Re\left(\tilde{N},\tilde{K}\right)}$ denotes the outage probability conditioned on $\Re\left(\tilde{N},\tilde{K}\right)$ is chosen, and can be bounded by (5.12) by replacing N with $\tilde{N}$. The probability for any relay to be chosen can be expressed as:

$$P\left[r \in \Re\left(\tilde{N},\tilde{K}\right)\right] = P\left[\sum_{i=1}^{m_k} |h_{i,k}|^2 \geq \frac{2^{2R}-1}{\eta}\right] = 1 - P\left[\sum_{i=1}^{m_k} |h_{i,k}|^2 \leq \frac{2^{2R}-1}{\eta}\right]. \tag{A.21}$$

Therefore any $\Re\left(\tilde{N},\tilde{K}\right)$ exists with a probability that can be written as:

$$\begin{aligned} P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] = & \prod_{r\in\Re\left(\tilde{N},\tilde{K}\right)} \left(1 - P\left[\sum_{i=1}^{m_k} |h_{i,k}|^2 \leq \frac{2^{2R}-1}{\eta}\right]\right) \\ & \times \prod_{r\in\Re\left(N-\tilde{N},K-\tilde{K}\right)} P\left[\sum_{i=1}^{m_k} |h_{i,k}|^2 \leq \frac{2^{2R}-1}{\eta}\right]. \end{aligned} \tag{A.22}$$

Based on (A.14), at high SNR, $P\left[\Re\left(\tilde{N},\tilde{k}\right)\right]$ can be approximated as

$$P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] \approx \left(\frac{2^{2R}-1}{\eta}\right)^{N-\tilde{N}} \prod_{r\in\Re\left(N-\tilde{N},K-\tilde{K}\right)} \frac{1}{m_k!}, \tag{A.23}$$

Which can be bounded by:

$$\frac{1}{\left(N-\tilde{N}\right)!}\left(\frac{2^{2R}-1}{\eta}\right)^{N-\tilde{N}} \leq P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] \leq \left(\frac{2^{2R}-1}{\eta}\right)^{N-\tilde{N}} \tag{A.24}$$

Note that the bounds are independent of K. Putting (A.24) and (5.12) into (A.20), we obtain the bounds (5.15) and thus complete the proof.

Appendix B
Capacity bounds for amplify-and-forward relaying

In this appendix we discuss the capacity bounds for amplify-and-forward relaying for the system model discussed in Chapter 5, where we apply MRC and TB at the relays.

B.1 Amplify-and-forward relaying

For amplify-and-forward relaying, after each relay receiver performing MRC of the signal vector, it amplifies the signal (5.5) by a factor that can meet the power constraint (5.4). The amplifying factor can be computed as:

$$\gamma_k = \frac{\eta \frac{m_k}{N}}{\eta \sum_{i=1}^{m_k} |h_{i,k}|^2 + 1}. \tag{B.1}$$

The transmitted signal t_k at each relay can be expressed by $t_k = \gamma_k \tilde{r}_k$. Note that unlike decode and forward mode, it now becomes a combination of the source signal and the noise at the relay. The relay then applies transmit beamforming to t_k to form the transmit signal vector $\mathbf{d}_k$, which now can be expressed as:

$$\mathbf{d}_k = \frac{\mathbf{g}_k^H}{\|\mathbf{g}_k\|_F} \gamma_k \tilde{r}_k \tag{B.2}$$

The destination receiver receives the sum of the signals from all K relays and performs data detection. The output signal at the destination (5.3), after some modification, can be written as:

$$y = s\sum_{k=1}^{K}\sqrt{\eta\sum_{i=1}^{m_k}|g_{i,k}|^2\sum_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}} \\ \underbrace{+\sum_{k=1}^{K}\frac{\sum\limits_{i=1}^{m_k}h_{i,k}^{*}n_{i,k}}{\sqrt{\sum\limits_{i=1}^{m_k}|h_{i,k}|^2}}\sqrt{\sum_{i=1}^{m_k}|g_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}}}_{n_r}+n_d, \tag{B.3}$$

where we denote n_r as the equivalent noise generated from the relays. It can be seen from (B.3) that compared with decode and forward mode, the signals can also be coherently combined at the destination, with a channel gain which takes into account the source to relay channel gains at the cost of additional noise n_r. Furthermore, we can observe from n_r that the noise generated at different relays *is not* coherently combined at the destination, though the signal can. Beamforming at the relays works only for the signal but not for the noise. Therefore while the signal is enhanced by the beamforming, the noise generated at the relays is not. This implies that besides the equal gain combining gain, beamforming the signal from the different relays can offer an *additional coherent combining gain* for reducing the impact of the noise generated at different relays specially for the amplify-and-forward mode. Specifically, the SNR at the destination can be written as:

$$\rho_a^{m_k} = \frac{\left(\sum\limits_{k=1}^{K}\sqrt{\eta\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\sum\limits_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}}\right)^2}{\sum\limits_{k=1}^{K}\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}+1}, \tag{B.4}$$

where we can clearly see an additional coherent combining gain of the signal power over the noise power generated at the relays.

Similar to the analysis for decode and forward mode, when all the relays are deployed with single antenna, the SNR at the destination can be rewritten as:

$$\rho_a^1 = \frac{\eta \left(\sum_{k=1}^{K} \sum_{i=1}^{m_k} |g_{i,k}| \, |h_{i,k}| \sqrt{\frac{\frac{\eta}{N}}{\eta |h_{i,k}|^2 + 1}} \right)^2}{\sum_{k=1}^{K} \sum_{i=1}^{m_k} |g_{i,k}|^2 \frac{\frac{\eta}{N}}{\eta |h_{i,k}|^2 + 1} + 1}. \tag{B.5}$$

It can be seen that no maximal ratio combining gain can be obtained in this case. However, the additional coherent combining gain and equal gain combining gain are maximized as the number of relays is maximized. When all the antennas are deployed in one relay, the SNR can be rewritten as

$$\rho_a^N = \frac{\eta \sum_{k=1}^{K} \sum_{i=1}^{m_k} |g_{i,k}|^2 \sum_{k=1}^{K} \sum_{i=1}^{m_k} |h_{i,k}|^2 \frac{\eta}{\eta \sum_{k=1}^{K} \sum_{i=1}^{m_k} |h_{i,k}|^2 + 1}}{\frac{\eta}{\eta \sum_{k=1}^{K} \sum_{i=1}^{m_k} |h_{i,k}|^2 + 1} \sum_{k=1}^{K} \sum_{i=1}^{m_k} |g_{i,k}|^2 + 1}. \tag{B.6}$$

In this case there is no equal gain combining gain or additional coherent combining gain, as all the antennas belongs to one relay. However, the maximal ratio combining gain can be obtained due to the full cooperation of the antennas at the relay.

B.2 Capacity bounds

The network capacity for amplify-and-forward relaying for each channel realization can be written as:

$$C_A^{m_k} = 0.5 \log_2 \left(1 + \rho_a^{m_k}\right) \tag{B.7}$$

The same upper bound for the output SNR at the destination for amplify and forward mode can be made as in **Lemma 1** for the decode-and-forward mode. This is stated in the following lemma:

Lemma 3. *For any m_k, $\rho_a^{m_k} \leq \rho_a^N$.*

Proof. To efficiently prove this lemma, we firstly introduce the following new lemma.

Lemma 4. *For any positive real numbers x_1, x_2, y_1, y_2, a, if*

$$x_1 \geq x_2 \text{ and } \frac{x_1}{y_1} \geq \frac{x_2}{y_2}, \tag{B.8}$$

then

$$\frac{x_1}{y_1+a} \geq \frac{x_2}{y_2+a}. \tag{B.9}$$

Proof. The proof is straightforward by showing $\frac{x_1}{y_1+a} - \frac{x_2}{y_2+a} \geq 0$ □

Compare numerators of (B.4) and (B.6), we have the following:

$$\begin{aligned}
\left(\sum_{k=1}^{K}\sqrt{\eta\sum_{i=1}^{m_k}|g_{i,k}|^2\sum_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}}\right)^2 &\overset{(a)}{\leq}\sum_{k=1}^{K}\sum_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta}{\eta\sum\limits_{k=1}^{K}\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}\\
&\quad\times\left(\sum_{k=1}^{K}\sqrt{\eta\sum_{i=1}^{m_k}|g_{i,k}|^2\frac{m_k}{N}}\right)^2\\
&\overset{(b)}{\leq}\sum_{k=1}^{K}\sum_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta}{\eta\sum\limits_{k=1}^{K}\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}\\
&\quad\times\eta\sum_{k=1}^{K}\sum_{i=1}^{m_k}|g_{i,k}|^2,
\end{aligned}$$

where inequality (a) holds due to the fact that $\frac{x}{1+x}$ is monotonically increasing with x, inequality (b) can be found in the proof of **Lemma 1** in Appendix A.1. From **Lemma 4**, we can now remove factor 1 in the denominators of (B.4) and (B.6). After some modifications, it can be seen that to compare ρ_a^N and $\rho_a^{m_k}$ is to compare

$$\sum_{k=1}^{K}\sum_{i=1}^{m_k}|h_{i,k}|^2$$

with

$$\frac{\left(\sum\limits_{k=1}^{K}\sqrt{\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\sum\limits_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}}\right)^2}{\sum\limits_{k=1}^{K}\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}},$$

which is equivalent to comparing $\sum\limits_{k=1}^{K} c_k \sum\limits_{k=1}^{K} d_k$ with $\left(\sum\limits_{k=1}^{K}\sqrt{c_k d_k}\right)^2$, where

$$c_k = \sum_{i=1}^{m_k}|h_{i,k}|^2,\ \ d_k = \sum_{i=1}^{m_k}|g_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}.$$

By the Cauchy-Schwarz inequality, it is clear that

$$\sum_{k=1}^{K} c_k \sum_{k=1}^{K} d_k \geq \left(\sum_{k=1}^{K}\sqrt{c_k d_k}\right)^2, \tag{B.10}$$

the proof is thus completed. □

Based on this lemma, we directly obtain the following theorem for the amplify-and-forward relaying mode:

Theorem 5. *If we denote the network capacity for $K = 1$ as C_A^N, for any m_k, $C_A^{m_k} \leq C_A^N$.*

The capacity lower bound, however, is difficult to obtain. The reason is that unlike decode-and-forward relaying, it is much more complicated to compare $\rho_a^{m_k}$ with ρ_a^1 due to the additional coherent combining gain among different relays for amplify-and-forward relaying. However, we can still give a comparison between the highest achievable capacity for both cases. In the following we derive two tight (achievable) upper bounds for both $\rho_a^{m_k}$ and ρ_a^1, and show that the upper bound for $\rho_a^{m_k}$ is strictly larger than that for ρ_a^1.

Lemma 5. *We have the following upper bounds for $\rho_a^{m_k}$ and ρ_a^1.*

$$\rho_a^{m_k} \leq \rho_{upper}^{m_k} \triangleq \sum_{k=1}^{K}\frac{\eta\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\sum\limits_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}}{\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}+\frac{1}{K}}; \tag{B.11}$$

$$\rho_a^1 \leq \rho_{upper}^1 \triangleq \sum_{k=1}^{K} \frac{\left(\sum\limits_{i=1}^{m_k} |g_{i,k}|\,|h_{i,k}|\sqrt{\frac{\frac{\eta}{N}}{\eta|h_{i,k}|^2+1}}\right)^2}{\sum\limits_{i=1}^{m_k} |g_{i,k}|^2 \frac{\frac{\eta}{N}}{\eta|h_{i,k}|^2+1} + \frac{1}{K}}. \tag{B.12}$$

Proof. The proof is straightforward if the following identity is noticed:

Lemma 6. *For positive real sequences $\{a_n\}$, $\{b_n\}$, the following inequality holds:*

$$\sum \frac{a_n}{b_n} \geq \frac{\left(\sum \sqrt{a_n}\right)^2}{\sum b_n}, \tag{B.13}$$

with equality if $\{a_n\}$, $\{b_n\}$ are constant values that do not change with n.

Proof. The proof can be directly obtained by showing $\sum \frac{a_n}{b_n} - \frac{(\sum \sqrt{a_n})^2}{\sum b_n} \geq 0$. □

□

Based on these bounds, we have the following identity:

Lemma 7. *For any m_k, $\rho_{upper}^{m_k} \geq \rho_{upper}^1$.*

Proof. The proof can be started by comparing each element in the summation of $\rho_{upper}^{m_k}$ and ρ_{upper}^1. i.e. by comparing

$$SNR_{m_k}^{m_k} \triangleq \frac{\eta \sum\limits_{i=1}^{m_k} |g_{i,k}|^2 \sum\limits_{i=1}^{m_k} |h_{i,k}|^2 \frac{\eta \frac{m_k}{N}}{\eta \sum\limits_{i=1}^{m_k} |h_{i,k}|^2+1}}{\sum\limits_{i=1}^{m_k} |g_{i,k}|^2 \frac{\eta \frac{m_k}{N}}{\eta \sum\limits_{i=1}^{m_k} |h_{i,k}|^2+1} + \frac{1}{K}}$$

with

$$SNR_{m_k}^1 \triangleq \frac{\left(\sum\limits_{i=1}^{m_k} |g_{i,k}|\,|h_{i,k}|\sqrt{\frac{\frac{\eta}{N}}{\eta|h_{i,k}|^2+1}}\right)^2}{\sum\limits_{i=1}^{m_k} |g_{i,k}|^2 \frac{\frac{\eta}{N}}{\eta|h_{i,k}|^2+1} + \frac{1}{K}} \tag{B.14}$$

It can be seen from the proof of **Lemma 3** in Appendix B.1 that $SNR_{m_k}^{m_k} \geq SNR_{m_k}^1$ by setting $K = m_k$ and $m_k = 1$. The proof is therefore completed. □

We consequently has the following theorm regarding the capacity of the network:

Theorem 6. *If we denote the* maximal achievable *network capacity for* $K = N$ *as* C_A^1*, for any* m_k*,* $C_A^1 \leq C_A^{m_k}$.

We note that for both amplify-and-forward relaying and decode-and-forward relaying discussed in Chapter 5, the three capacity theorems can hold for *any* choice of fading distribution, as long as each $h_{i,k}$ ($g_{i,k}$) is i.i.d. distributed.

Appendix C
Publications

All publications can be found behind the references at the end of the thesis.

Journals

1. Y. Fan, J. S. Thompson, A. B. Adinoyi and H. Yanikomeroglu, "Antenna combining for multi-antenna multi-relay channels", to appear in European Transactions on Telecommunications, no. 6, vol. 18, 2007, (invited paper).

2. Y. Fan and J. S. Thompson, "MIMO configurations for relay channels: theory and practice", IEEE Transactions on Wireless Communications, vol. 6, no. 3, March 2007.

3. J. S. Thompson, P. M. Grant and Y. Fan, "An introduction to multi-hop multi-antenna communications", Frequenz (Journal of RF-Engineering and Telecommunications), Vol. 60, Issue 5-6, pp. 103-106, 2006, (invited paper).

4. Y. Fan, J. S. Thompson, A. B. Adinoyi and H. Yanikomeroglu, "On the diversity-multiplexing tradeoff for multi-antenna multi-relay channels", submitted to IEEE Transactions on Wireless Communications, Dec. 2006.

5. Y. Fan, J. S. Thompson, "Recovering multiplexing loss through successive relaying", submitted to IEEE Transactions on Wireless Communications, June. 2006, under revision.

Conference Proceedings

1. F.A. Onat, A. Adinoyi, Y. Fan, H. Yanikomeroglu, J. S. Thompson, "On the optimal threshold of digital cooperative relaying schemes", WCNC07, Mar. 2007, HongKong, China.

2. Y. Fan, J. S. Thompson, "Recovering multiplexing loss in relay networks", WWRF17, 15-17 Nov. 2006, Heidelberg, Germany.

3. Y. Fan, J. S. Thompson, A. B. Adinoyi and H. Yanikomeroglu, "On the diversity-multiplexing tradeoff for multi-antenna multi-relay channels", IEEE ICC 2007, Glasgow, UK.

4. Y. Fan, J. S. Thompson, "Recovering multiplexing loss through successive relaying", submitted to IEEE Globecomm 2007, Washington D.C., USA.

5. Y. Fan and J. S. Thompson, "On the performance of MIMO spatial multiplexing relay channels", IEEE ICC 2007, Glasgow, UK.

6. Y. Fan, J. S. Thompson, A. B. Adinoyi and H. Yanikomeroglu, "Space diversity for multi-antenna multi-relay channels", European Wireless Conference 2006, April 2-5 2006, Athens, Greece.

7. Y. Fan, J. S. Thompson and M. Naden, "Relay configurations for spatial multiplexing MIMO channels", WWRF15, 08-09 December 2005, Paris, France.

8. Y. Fan and J. S. Thompson, "On the outage capacity of MIMO multihop networks", IEEE Globecom 2005, 28 November-2 December 2005, St. Louis, MO, USA.

References

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

[2] M. Frodigh *et al.*, "Future-generation wireless networks," *IEEE Personal Comm.*, vol. 8, no. 5, pp. 10–17, Oct. 2001.

[3] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks wireless communications," *IEEE Personal Comm.*, vol. 9, no. 4, pp. 8–27, Aug. 2002.

[4] M. Frodigh *et al.*, "Self-organising packet radio ad hoc networks with overlay," *IEEE Comm. Mag.*, vol. 40, no. 6, pp. 149–157, June 2002.

[5] R. Pabst *et al.*, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Comm. Mag.*, vol. 42, no. 9, pp. 80–89, Sept. 2004.

[6] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Comm. Mag.*, vol. 43, no. 9, pp. S23–S30, Sept. 2005.

[7] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Comm. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[8] W. F. Gabriel, "Adaptive arrays - an introduction," *Proc. IEEE*, vol. 64, pp. 239–272, 1976.

[9] J. H. Winters, "Optimum combining for indoor radio systems with multiple users," *IEEE Trans. Comm.*, vol. 35, no. 11, pp. 1222–1230, Nov. 1987.

[10] S. C. Swales *et al.*, "The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems," *IEEE Trans. V. Tech.*, vol. 39, no. 1, pp. 56–67, Feb. 1990.

[11] J. S. Thompson, P. Grant, and B. Mulgrew, "Smart antenna arrays for cdma systems," *IEEE Pers. Comm. Mag.*, vol. 3, no. 5, pp. 16–25, Oct. 1996.

[12] F. R. Farrokhi, G. J. Foschini, A. Lozano, and R. A. Valenzuela, "Link-optimal space-time processing with multiple transmit and receive antennas," *IEEE Comm. Lett.*, vol. 5, no. 3, Mar. 2001.

[13] J. H. Winters, "Smart antennas for wireless systems," *IEEE Pers. Comm.*, vol. 5, no. 1, pp. 23–27, Feb. 1998.

[14] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Sig. Proc. Mag.*, vol. 14, no. 6, pp. 49–83, Nov. 1997.

[15] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Laboratories Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[16] I. E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.

[17] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[18] 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.

[19] D. Gesbert *et al.*, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Comm.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

[20] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of mimo communications: A key to gigabit wireless," *Proc. of the IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[21] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Comm.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[22] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Sel. Areas Comm.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.

[23] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[24] J. B. Andersen, "Antenna arrays in mobile communications: gain, diversity, and channel capacity," *IEEE Antennas Propagat. Mag.*, vol. 42, no. 2, pp. 12–16, Apr. 2000.

[25] A. Gorokhov, D. A. Gore, and A. J. Paulraj, "Receive antenna selection for MIMO spatial multiplexing: theory and algorithms," *IEEE Trans. Sig. Proc.*, vol. 51, no. 11, pp. 2796–2807, Nov. 2003.

[26] G. Foschini, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Comm.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[27] W. Jakes, *Microwave Mobile Communications*. New York, NY: Wiley, 1974.

[28] T. Rappaport, *Wireless Communications: Principles and practice*. Upper Saddle River, NJ: Prentice Hall, 1996.

[29] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukua, "Field strength and its variability in UHF and VHF land-mobile radio service," *Rev. Elec. Commun. Lab.*, vol. 16, no. 9, pp. 825–873, 1968.

[30] V. Erceg *et al.*, "An empirically based path loss model for wireless channels in suburban environments," *IEEE J. Sel. Areas Comm.*, vol. 17, pp. 1205–1211, July 1999.

[31] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. UK: Cambridge university press, 2005.

[32] R. W. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *Bell System Technical Journal*, vol. 45, pp. 1775–1796, 1966.

[33] B. R. Saltzberg, "Performance of an efficient parallel data transmission system," *IEEE Trans. Comm.*, vol. 15, pp. 805–811, 1967.

[34] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing*. New York, NY: John Wiley and Sons, Inc., 1997.

[35] B. Wang, J. Zhang, and L. Zheng, "Achievable rates and scaling laws of wideband sensory relay networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4084–4104, Sept.

[36] W. W. C. E. Shannon, *Mathematical Theory of Communication*. Champaign, IL: University of Illinois Press, 1963.

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

[38] R. A. Horn and C. R. Johnson, *Matrix Analysis*. UK: Cambridge university press, 1985.

[39] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Comm.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.

[40] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. UK: Cambridge university press, 2003.

[41] P. W. Wolniansky *et al.*, "V-blast: an architecture for realizing very high data rates overthe rich-scattering wireless channel," in *ISSSE 98*, Pisa, Italy, 29 Sep. - 2 Oct. 1998.

[42] T. Chung, A. Lozano, H. C. Huang, A. Sutivong, and J. M. Cioffi, "Approaching the mimo capacity with a low-rate feedback channel in v-blast," *EURASIP J. App. Sig. Proc.*, vol. 5, pp. 762–771, 2004.

[43] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the gaussian multiple-access channel," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Monterey, Calif, USA, Nov. 1997, pp. 1405–1409.

[44] A. Papoulis, *Probability, Random Variables, and Stochastic Process*. New York, NY: McGraw-Hill, 1984.

[45] H. Yao and G. Wornell, "Achieving the full mimo diversity-multiplexing frontier with rotation-based space-time codes," in *Proc. Allerton Conf. Comm., Contr. and Comp.*, Monticello, IL, Oct. 2003.

[46] S. Tavildar and P. Viswanath, "Approximately universal codes over slow fading channels," *IEEE Trans. Inf. Theory*, 2006, to be published.

[47] H. E. Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity-vs-multiplexing tradeoff of mimo channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 968–985, June 2004.

[48] J. C. Belfiore, "The golden code: a 2 × 2 full-rate space-time code with nonvanishing determinants," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1432–1436, Apr. 2005.

[49] 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.

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

[51] V. Stankovic, A. Host-Madsen, and Z. Xiong, "Cooperative diversity for wireless ad hoc networks: capacity bounds and code designs," *IEEE Sig. Proc. Mag.*, vol. 22, pp. 37–49, Sept.

[52] Z. Xiong, A. D. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Sig. Proc. Mag.*, vol. 21, no. 5, pp. 80–94, Sept.

[53] K. Azarian, H. E. Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.

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

[55] M. Janani *et al.*, "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE. Trans. Signal. Process.*, vol. 52, no. 2, pp. 362–371, Feb. 2004.

[56] A. Stefanov and E. Erkip, "Cooperative coding for wireless networks," *IEEE Trans. Comm.*, vol. 52, no. 9, pp. 1470–1476, Sept. 2005.

[57] T. E. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Trans. wireless Comm.*, vol. 5, no. 2, pp. 283– 289, Feb. 2006.

[58] Z. Lin, E. Erkip, and A. Stefanov, "Cooperative regions and partner choice in coded cooperative systems," *IEEE Trans. Comm.*, vol. 54, no. 4, p. 760, Apr. 2006.

[59] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperative diversity-part i: system description," *IEEE Trans. Commun.*, vol. 51, pp. 1927–1938, Nov. 2003.

[60] ——, "User cooperative diversity-part ii: implementation aspects and performance analysis," *IEEE Trans. Commun.*, vol. 51, pp. 1939–1948, Nov. 2003.

[61] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Comm.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

[62] M. Zorzi and R. R. Rao, "Geographic random forwarding (geraf) for ad hoc and sensor networks: Multihop performance," *IEEE Trans. Mobile Comput.*, vol. 2, no. 4, p. 337C348, 2003.

[63] B. Zhao and M. C. Valenti, "Practical relay networks: A generalization of hybrid-arq," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, p. 7C18, Jan. 2005.

[64] E. Zimmermann, P. Herhold, and G. Fettweis, "On the performance of cooperative diversity protocols in practical wireless systems," in *IEEE VTC 2003-Fall*, Oct. 2003, pp. 2212– 2216.

[65] J. Cho and Z. J. Haas, "On the throughput enhancement of the downstream channel in cellular radio networks through multihop relaying," *IEEE J. Sel. Comm.*, vol. 22, no. 7, pp. 1206–1219, Sept. 2004.

[66] N. Prasad and M. K. Varanasi, "Diversity and multiplexing tradeoff bounds for cooperative diversity protocols," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, June 2004, p. 268.

[67] M. Costa, "On the gaussian interference channel," *IEEE. Trans. Inf. Theory*, vol. 31, no. 5, pp. 607–615, Sept. 1985.

[68] ——, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441.

[69] B. Suard *et al.*, "Uplink channel capacity of space-division-multiple-access schemes," *IEEE. Trans. Inf. Theory*, vol. 44, no. 4, July 1998.

[70] P. Anghel, G. Leus, and M. Kavehl, "Multi-user space-time coding in cooperative networks," in *ICASSP'03*, vol. 4, Apr. 2003, pp. IV– 73–6.

[71] P. A. Anghel and M. Kaveh, "On the performance of distributed space-time coding systems with one or two non-regenerative relays," *IEEE Trans. Wireless Comm.*, vol. 5, no. 3, pp. 682–692, Mar. 2006.

[72] 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.

[73] ——, "Spectral efficient protocols for non-regenerative half-duplex relaying," in *Proc. Allerton Conf. Comm., Contr. and Comp.*, Monticello, IL, Oct. 2005.

[74] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1691–1706, July 2003.

[75] S. Vishwanath and A. J. G. N. Jindal, "Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[76] N. Jindal *et al.*, "Sum power iterative water-filling for multi-antenna gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[77] ——, "Zero-forcing methods for downlink spatial multiplexing in multiuser mimo channels," *IEEE Trans. Signal Processing*, vol. 52, no. 2, Feb. 2004.

[78] W. Yu and J. M. Cioffi, "Sum capacity of gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, Sept. 2004.

[79] A. Goldsmith *et al.*, "Capacity limits of mimo channels," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 684 – 702, June 2003.

[80] B. Zhao and M. C. Valenti, "On the capacity of multiuser wireless channels with multiple antennas," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, Oct. 2003.

[81] S. Yang and J.-C. Belfiore, "Optimal space-time codes for the mimo amplify-and-forward cooperative channel," *IEEE Trans. Inf. Theory*, 2006, to be published.

[82] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of mimo relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

[83] C. K. Lo, S. Vishwanath, and J. R. W. Heath, "Rate bounds for mimo relay channels using precoding," in *IEEE Globecom 2005*, St. Louis, MO, USA, 28 Nov. - 2 Dec. 2005.

[84] H. Bolcskei *et al.*, "Capacity scaling laws in mimo relay networks," *IEEE Trans. Wireless Comm.*, vol. 5, no. 6, pp. 1433–1444, June 2006.

[85] O. Oyman and A. J. Paulraj, "Energy efficiency in mimo relay networks under processing cost," in *Conf. Inf. Science and Systems*, Johns Hopkins University.

[86] ——, "Power-bandwidth tradeoff in linear multiantenna interference relay networks," *IEEE Trans. Wireless Comm.*, 2007, to be published.

[87] H. E. Gamal, G. Caire, and M. O. Damen, "The mimo arq channel: diversity-multiplexing-delay tradeoff," *IEEE Trans. Wireless Comm.*, vol. 52, no. 8, pp. 3601–3619, Aug. 2006.

[88] H. Zheng, A. Lozano, and M. Haleem, "Multiple arq processes for mimo systems," *EURASIP J. App. Sig. Proc.*, vol. 5, pp. 772–782, 2004.

[89] S. Loyka and F. Gagnon, "Performance analysis of the v-blast algorithm: an analytical approach," *IEEE Trans. Wireless Comm.*, vol. 3, no. 4, pp. 1326–1337, July 2006.

[90] H. Zhang, H. Dai, and B. L. Hughes, "On the diversity-multiplexing tradeoff for ordered sic receivers over mimo channels," in *IEEE ICC'06*, Istanbul (Turkey), June 2006.

[91] A. M. Tulino and S. Verdu, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, 2004.

[92] Y. Fan and J. S. Thompson, "On the outage capacity of mimo multihop networks," in *IEEE Globecom 2005*, St. Louis, MO, USA, 28 Nov. - 2 Dec. 2005.

[93] X. Bao and J. Li, "Decode-amplify-forward (daf): A new class of forwarding strategy for wireless relay channels," in *Proc. IEEE SPAWC'05*, New York, NY, June 2005, pp. 836–840.

[94] K. S. Gomadam and S. A. Jafar, "Optimizing soft information in relay networks," in *Proc. Asilomar Conf. Sig. Sys. Comput.*, Monterey, CA, Oct. 2006.

[95] J. Dattorro, *Convex Optimization and Euclidean Distance Geometry*. USA: Meboo Publishing, 2005.

[96] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[97] T. L. Marzetta, "Blast training: estimating channel characteristics for high capacity space-time wireless," in *Proceedings of 1999 Allerton Conference*, Monticello, IL,, USA, 1999.

[98] Y. Chen and C. Tellambura, "Performance analysis of l-branch equal gain combiners in equally correlated rayleigh fading channels," *IEEE Commun. Lett.*, vol. 8, no. 3, pp. 150–152, Mar. 2004.

[99] Y. Zhao, R. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: Optimal power allocation versus selection," *IEEE Trans. Wireless. Commun.*, 2007, to be published.

EUROPEAN TRANSACTIONS ON TELECOMMUNICATIONS
Euro. Trans. Telecomms. **00**: 1–11 (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/ett.0000

Special Issue on EW2006 Best Papers

Antenna combining for multi-antenna multi-relay channels

Yijia Fan[1]*, Abdulkareem Adinoyi[2], John S Thompson[1], Halim Yanikomeroglu[2]

[1] *Institute for Digital Communications, University of Edinburgh, Edinburgh, EH9 3JL, UK*
[2] *Broadband Communications and Wireless Systems (BCWS) centre, Department of Systems and Computer Engineering, Carleton University, Ottawa, K1S 5B6, Canada*

SUMMARY

In this paper we analyze the performance of multiple relay channels when multiple antennas are deployed only at relays. We apply two antenna diversity techniques at relays, namely maximum ratio combining (MRC) on receive and transmit beamforming (TB). We show that for both decode-and-forward and amplify-and-forward relaying protocols, with K relays the network can be decomposed into K diversity channels each with a different channel gain, and that the signals can be effectively combined at the destination. We assume that the total number of antennas at all relays is fixed at N. With a reasonable power constraint at the relays, the network capacity will be lower bounded by that of N relay channels each with single antenna, and upper bounded by that of single relay channel with N antennas. Copyright © 2007 AEIT

1. INTRODUCTION

It is widely believed that ad hoc networking [1] or multi-hop cellular networks [2] are important new concepts for future generation wireless systems [3], where either mobile or fixed nodes (often referred to as relays) are used to help forward the information to the desired user. One advantage of these structures are that it is possible to unite multiple relays in the network as a "virtual antenna array" to forward the information cooperatively, while appropriate combining at the destination realizes diversity gain. The diversity achieved in this way is often named as *user cooperation diversity* or *cooperative diversity* [4], as it mimics the performance advantages of multiple-input multiple-output (MIMO) systems[5] in exploiting the spatial diversity of the relay channels. The performance limits of space-time codes, which can exploit cooperative diversity, are discussed in [6, 7, 8] for single-antenna relay networks. For multiple-antenna relay channels where every terminal in the network can be deployed with multiple antennas, studies are mainly concentrated on spatial multiplexing systems [9, 10, 11].

In this paper we exploit the spatial diversity of the relay channels in a different way from space-time coding based approach. We apply two kinds of antenna combining techniques at the relay, namely maximum ratio combining (MRC) [12] for reception and transmit beamforming (TB)[13] for transmission. Those techniques were often used in point-to-point single-input multiple-output(SIMO) or multiple-input single-output (MISO) wireless links, where either the transmitter or receiver is equipped with multiple antennas. It has been shown that MRC (TB) is able to achieve information theoretic upper bound of SIMO (MISO) systems [14]. In a relay context, we move the multiple antennas to the relays, while the source and the destination are only equipped with a single antenna. Compared with the single antenna relay network where every node is equipped with one antenna, our system model allows certain wired cooperation between the antennas at the relays, which is clearly a advantage. However, it will be shown that our schemes can also work effectively for single antenna networks. More specifically, we will show

*Correspondence to: Institute for Digital Communications, University of Edinburgh, Edinburgh, EH9 3JL, UK. E-mail: y.fan@ed.ac.uk

Prepared using ettauth.cls *[Version: 2007/01/05 v1.00]*

Accepted 5 February 2007

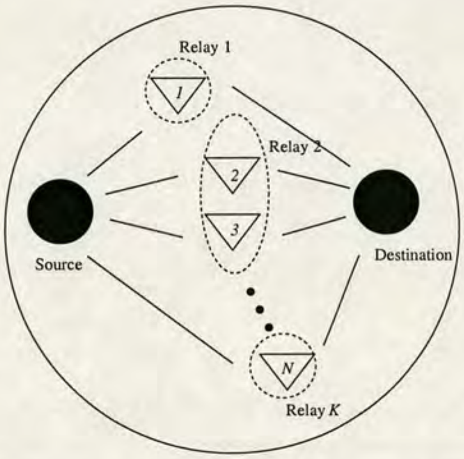


Figure 1. System model for a two hop network: Source and destination are each deployed with 1 antenna. A total of N antennas are deployed at K relays. For each channel realization, both backward and forward channel coefficients for all N antennas remain the same regardless of the number of relays K.

that by applying the antenna combining techniques in the relay network, a network with K relays can be decomposed into K diversity channels each a with different channel gain, and the signals from all K branches can be effectively combined at the destination. We derive the capacity bounds for this signal combining techniques and our analysis results can be applied to both ergodic capacity and outage capacity [15] performance. Our investigation is based on both the decode-and-forward (digital) relaying mode, where the relays decode, re-encode and re-transmit the signals, and the amplify-and-forward (analogue) mode, where the relays amplify and forward the signals without decoding the message.

The rest of this paper is organized as follows. In Section II, the basic system model and assumptions are introduced. Section III introduces the antenna combining techniques. The capacity performance analysis are made in section IV. Section V presents and discusses simulation results and finally, conclusions are drawn in Section VI.

2. SYSTEM MODEL

We consider a two hop network model with one source, one destination and K relays. For simpler presentation we ignore the direct link between the source and destination. However, the extension of all the results to include the direct link is straightforward. We assume that the source and destination are deployed with single antennas, while relay k is deployed with m_k antennas; the total number of antennas at all relays is fixed to N. This can be expressed as

$$\sum_{k=1}^{K} m_k = N. \tag{1}$$

We restrict our discussion to the case where the channels are frequency-flat fading. The data transmission is over two times slots using two hops. In the first transmission time slot, the source broadcasts the signal to all the relay terminals. The input/output relation for the source to the kth relay is given by

$$\mathbf{r}_k = \sqrt{\eta}\mathbf{h}_k s + \mathbf{n}_k, \tag{2}$$

where $\mathbf{r}_k$ is the $m_k \times 1$ receive signal vector, and η denotes the transmit power at the source. The scalar s is the unit mean power transmit signal and $\mathbf{n}_k$ is the $m_k \times 1$ complex circular additive white Gaussian noise vector at relay k with identity covariance matrix $\mathbf{I}_{m_k}$. The vector $\mathbf{h}_k$ is the $m_k \times 1$ channel transfer vector from the source to the kth relay. The entries of $\mathbf{h}_k$ are independent and identically distributed (i.i.d) random variables. In the second hop, each relay processes its received signals and re-transmits them to the destination. The signal received at the destination can be written as:

$$y = \sum_{i=1}^{K} \mathbf{g}_k \mathbf{d}_k + n_d, \tag{3}$$

where the $1 \times m_k$ vector $\mathbf{g}_k$ is the channel vector from the kth relay to the destination, of which each entry is an i.i.d. random variable. The scalar n_d is the complex additive white Gaussian noise at the destination with unit variance. The $m_k \times 1$ vector $\mathbf{d}_k$ is the transmit signal vector at relay k, which should meet the total transmit power constraint:

$$\mathrm{E}\left[\|\mathbf{d}_k\|_F^2\right] \leq \frac{\eta m_k}{N}, \tag{4}$$

where $\|\bullet\|_F$ denotes the Frobenius norm and $\mathrm{E}[\bullet]$ denotes the expectation. This power constraint means that the power is allocated at each relay in proportion to its number of antennas. For presentation simplicity we assume here that the total power at all relays is fixed to be η, i.e. the same as at the source. However, all the conclusions in the paper also hold when the total power at all relays is fixed to an arbitrary constant. We assume a coherent relay channel configuration context where the kth relay can obtain full knowledge of both backward channel vector $\mathbf{h}_k$ and forward channel vector $\mathbf{g}_k$. For fair comparison,

DOI: 10.1002/ett

we also assume that for each channel realization, all the backward and forward channel coefficients for all N antennas remain the same regardless of the number of relays K. Fig. 1 shows the system model.

3. ANTENNA COMBINING TECHNIQUES IN RELAY CHANNELS

In this section we apply MRC and TB techniques to the system model described in section II. We discuss both decode-and-forward and amplify-and-forward relaying modes below.

3.1. Decode-and-forward relaying

We assume that each relay performs MRC of the received signals, by multiplying the received signal vector by the vector $\mathbf{h}_k^H/\|\mathbf{h}_k\|_F$, where $\mathbf{h}_k^H$ denotes the complex conjugate transpose of $\mathbf{h}_k$. The signal at the output of the relay receiver is given by

$$\tilde{r}_k = s\sqrt{\eta\sum_{i=1}^{m_k}|h_{i,k}|^2} + \frac{\sum_{i=1}^{m_k} h_{i,k}^* n_{i,k}}{\sqrt{\sum_{i=1}^{m_k}|h_{i,k}|^2}} \tag{5}$$

where $h_{i,k}$ denotes the ith antenna at relay k, and $n_{i,k}$ denotes the noise factor for ith receiver input branch. We denote $h_{i,k}^*$ the complex-conjugate of $h_{i,k}$. The signal to noise ratio (SNR) at the output of the receiver can be written as:

$$\rho_k^{m_k} = \eta\sum_{i=1}^{m_k}|h_{i,k}|^2. \tag{6}$$

After the relays decode the signals, each relay then performs TB of the decoded waveform. If we denote the transmitted signals as t_k with unit variance, the transmitted signal vector $\mathbf{d}_k$ for relay k can be written as

$$\mathbf{d}_k = \sqrt{\frac{\eta m_k}{N}}\frac{\mathbf{g}_k^H}{\|\mathbf{g}_k\|_F}t_k. \tag{7}$$

The destination receiver simply detects the combined signals from all K relays. If we adjust the transmission data rate so that the signals are correctly decoded at all the relays (i.e. $t_k = s$), the output signal at the destination can be written as:

$$y = s\sum_{k=1}^{K}\sqrt{\frac{\eta m_k}{N}\sum_{i=1}^{m_k}|g_{i,k}|^2} + n_d = s\sum_{k=1}^{K}\tilde{g}_k + n_d \tag{8}$$

It can be seen from (8) that by applying antenna diversity schemes at relays, the networks can be decomposed to K diversity channels each with channel gain $\tilde{g}_k$. The output SNR at the destination receiver can therefore be written as:

$$\rho_d^{m_k} = \left(\sum_{k=1}^{K}\sqrt{\frac{\eta m_k}{N}\sum_{i=1}^{m_k}|g_{i,k}|^2}\right)^2. \tag{9}$$

When all the relays are deployed with a single antenna, there is no traditional maximum ratio combining gain at the relays and the destination. However, the destination still observes a set of equal gain combined [16] amplitude signals from all relays.† Since we assume that the backward and forward channel coefficients for each antenna are kept the same for different values of K and m_i, the output SNR at the destination can be rewritten as

$$\rho_d^1 = \frac{\eta}{N}\left(\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|\right)^2; \tag{10}$$

when all the antennas are deployed in one relay (i.e. $K = 1$ and $m_1 = N$), the network can be separated into a SIMO (source-relay) and a MISO (relay-destination) channel. Using MRC and TB can achieve the information theoretic upper bound of each link. The SNR for this case can be rewritten as

$$\rho_d^N = \eta\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2. \tag{11}$$

3.2. Amplify-and-forward relaying

For amplify-and-forward relaying, after each relay receiver performing MRC of the signal vector, it amplifies the signal (5) by a factor that can meet the power constraint (4). The amplifying factor can be computed as:

$$\gamma_k = \frac{\eta\frac{m_k}{N}}{\eta\sum_{i=1}^{m_k}|h_{i,k}|^2 + 1}. \tag{12}$$

The transmitted signal t_k at each relay can be expressed by $t_k = \gamma_k\tilde{r}_k$. Note that unlike decode and forward mode, it now becomes a combination of the source signal and the noise at the relay. The relay then applies transmit

†Unlike [16], the equal gain combining weights for the relay channels are applied at the transmitter(s) instead of the receiver.

beamforming to t_k to form the transmit signal vector $\mathbf{d}_k$, which now can be expressed as:

$$\mathbf{d}_k = \frac{\mathbf{g}_k^H}{\|\mathbf{g}_k\|_F}\gamma_k \tilde{r}_k \tag{13}$$

The destination receiver receives the sum of the signals from all K relays and performs data detection. The output signal at the destination (3), after some modification, can be written as:

$$y = s\sum_{k=1}^{K}\sqrt{\eta\sum_{i=1}^{m_k}|g_{i,k}|^2\sum_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum_{i=1}^{m_k}|h_{i,k}|^2+1}} + \underbrace{\sum_{k=1}^{K}\frac{\sum_{i=1}^{m_k}h_{i,k}^*n_{i,k}}{\sqrt{\sum_{i=1}^{m_k}|h_{i,k}|^2}}\sqrt{\sum_{i=1}^{m_k}|g_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum_{i=1}^{m_k}|h_{i,k}|^2+1}}}_{n_r} + n_d, \tag{14}$$

where we denote n_r as the equivalent noise generated from the relays. It can be seen from (14) that compared with decode and forward mode, the signals can also be coherently combined at the destination, with a channel gain which takes into account the source to relay channel gains at the cost of additional noise n_r. Furthermore, we can observe from n_r that the noise generated at different relays *is not* coherently combined at the destination, though the signal can. Beamforming at the relays works only for the signal but not for the noise. Therefore while the signal is enhanced by the beamforming, the noise generated at the relays is not. This implies that besides the equal gain combining gain, beamforming the signal from the different relays can offer an *additional coherent combining gain* for reducing the impact of the noise generated at different relays specially for the amplify-and-forward mode. Specifically, the SNR at the destination can be written as:

$$\rho_a^{m_k} = \frac{\left(\sum_{k=1}^{K}\sqrt{\eta\sum_{i=1}^{m_k}|g_{i,k}|^2\sum_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum_{i=1}^{m_k}|h_{i,k}|^2+1}}\right)^2}{\sum_{k=1}^{K}\sum_{i=1}^{m_k}|g_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum_{i=1}^{m_k}|h_{i,k}|^2+1}+1}, \tag{15}$$

where we can clearly see an additional coherent combining gain of the signal power over the noise power generated at the relays.

Similar to the analysis for decode and forward mode, when all the relays are deployed with single antenna, the SNR at the destination can be rewritten as:

$$\rho_a^1 = \frac{\eta\left(\sum_{k=1}^{K}\sum_{i=1}^{m_k}|g_{i,k}|\,|h_{i,k}|\sqrt{\frac{\frac{\eta}{N}}{\eta|h_{i,k}|^2+1}}\right)^2}{\sum_{k=1}^{K}\sum_{i=1}^{m_k}|g_{i,k}|^2\frac{\frac{\eta}{N}}{\eta|h_{i,k}|^2+1}+1}. \tag{16}$$

It can be seen that no maximal ratio combining gain can be obtained in this case. However, the additional coherent combining gain and equal gain combining gain are maximized as the number of relays is maximized. When all the antennas are deployed in one relay, the SNR can be rewritten as

$$\rho_a^N = \frac{\eta\sum_{k=1}^{K}\sum_{i=1}^{m_k}|g_{i,k}|^2\sum_{k=1}^{K}\sum_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta}{\eta\sum_{k=1}^{K}\sum_{i=1}^{m_k}|h_{i,k}|^2+1}}{\frac{\eta}{\eta\sum_{k=1}^{K}\sum_{i=1}^{m_k}|h_{i,k}|^2+1}\sum_{k=1}^{K}\sum_{i=1}^{m_k}|g_{i,k}|^2+1}. \tag{17}$$

In this case there is no equal gain combining gain or additional coherent combining gain, as all the antennas belongs to one relay. However, the maximal ratio combining gain can be obtained due to the full cooperation of the antennas at the relay. It has been shown that in this scenario using MRC and TB can achieve the information theoretic upper bound of the single relay channel [17] if the direct link is ignored.

4. CAPACITY BOUNDS

4.1. Decode and forward relaying

The network capacity for decode and forward relaying for each channel realization can be written as

$$C_D^{m_k} = \min\left(C_r^{1,m_1}, C_r^{2,m_2}, \cdots, C_r^{K,m_k}, C_d^{m_k}\right) \tag{18}$$

where $C_r^{k,m_k} = 0.5\log_2\left(1+\rho_k^{m_k}\right)$ denoting the Shannon capacity for the source to the kth relay channel, and $C_d^{m_k} = 0.5\log_2\left(1+\rho_d^{m_k}\right)$ denoting the Shannon capacity for the relay to destination channels.‡ The factor 0.5 denotes the half multiplexing factor compared with non-relay channels.

We firstly analyze channel capacity for the relays to destination link by bounding $\rho_d^{m_k}$,i.e. the output SNR at the destination.

‡Here the SNR for the direct link should be included inside the log function when direct link is accounted for.

DOI: 10.1002/ett

Lemma 1 *For any m_k, $\rho_d^1 \leq \rho_d^{m_k} \leq \rho_d^N$.*

Proof See Appendix A. ∎

From **Lemma 1**, we can see that

$$C_d^1 \leq C_d^{m_k} \leq C_d^N, \tag{19}$$

where C_d^1 denotes the capacity for the relays to destination channel when $K = N$, and C_d^N denotes the capacity for the relays to destination channel when $K = 1$. Now also considering the capacity for the source to relay links and extending the analysis to the whole network scenario, we have the following theorem:

Theorem 1 *If we denote the network capacity for $K = N$ as C_D^1 and for $K = 1$ as C_D^N, for any m_k, $C_D^1 \leq C_D^{m_k} \leq C_D^N$.*

Proof Considering the SNR $\rho_k^{m_k}$ for the source to the kth relay link, if we denote it as ρ_n^1 for $K = N$ and ρ_1^N for $K = 1$, it can be shown that

$$\min\left(\rho_n^1\right) \leq \min\left(\rho_k^{m_k}\right) \leq \rho_1^N. \tag{20}$$

Therefore, we have the following:

$$\begin{aligned}\min\left(C_r^{1,1}, \cdots, C_r^{N,1}\right) &\leq \min\left(C_r^{1,m_1}, \cdots, C_r^{K,m_k}\right) \\ &\leq C_r^{1,N}.\end{aligned} \tag{21}$$

Combining (21) and (19), we thus complete the proof. ∎

From the above analysis we have shown that for the antenna combining techniques discussed in the paper, the network capacity will be lower bounded by that of N relay channels each with a single antenna, and upper bounded by that of a single relay channel with N antennas. This means that even there are more relays, the increased "equal gain combining" gain at the destination can not compensate for the loss of maximum ratio combining gain at the relay and the destination when the numbers of antennas at each relay are reduced.

4.2. Amplify and forward relaying

The network capacity for amplify-and-forward relaying for each channel realization can be written as:

$$C_A^{m_k} = 0.5 \log_2\left(1 + \rho_a^{m_k}\right) \tag{22}$$

The capacity analysis for the decode-and-forward mode can not be directly extended to the amplify-and-forward mode, as they have different relaying mechanisms. In fact, the analysis for amplify and forward relaying is more difficult due to the following two reasons: (a) the source-relay link and the relay-destination link in amplify and forward mode can not be considered separately, as decoding is not performed at the relays; (b) the impact of the noise component generated at the relays is complicated and play a vital role in the capacity performance.

However, we can conjecture the same conclusion as of the decode-and-forward mode by imagining an extreme case, i.e. a very high transmit power level. The performance of amplify and forward mode in this scenario mimics that of decode and forward mode, as the noise component generated at the relays becomes negligible. In fact, the same upper bound for the output SNR at the destination for amplify and forward mode can be made as in *Lemma 1* for the decode-and-forward mode. This is stated in the following lemma:

Lemma 2 *For any m_k, $\rho_a^{m_k} \leq \rho_a^N$.*

Proof See Appendix B. ∎

Based on this lemma, we directly obtain the following theorem for the amplify-and-forward relaying mode:

Theorem 2 *If we denote the network capacity for $K = 1$ as C_A^N, for any m_k, $C_A^{m_k} \leq C_A^N$.*

The capacity lower bound, however, is difficult to obtain. The reason is that unlike decode-and-forward relaying, it is much more complicated to compare $\rho_a^{m_k}$ with ρ_a^1 due to the additional coherent combining gain among different relays for amplify-and-forward relaying as stated in section III. However, we can still give a comparison between the highest achievable capacity for both cases. In the following we derive two tight (achievable) upper bounds for both $\rho_a^{m_k}$ and ρ_a^1, and show that the upper bound for $\rho_a^{m_k}$ is strictly larger than that for ρ_a^1.

Lemma 3 *We have the following upper bounds for $\rho_a^{m_k}$ and ρ_a^1.*

$$\rho_a^{m_k} \leq \rho_{upper}^{m_k} \triangleq \sum_{k=1}^{K} \frac{\eta \sum_{i=1}^{m_k} |g_{i,k}|^2 \sum_{i=1}^{m_k} |h_{i,k}|^2 \frac{\eta \frac{m_k}{N}}{\eta \sum_{i=1}^{m_k} |h_{i,k}|^2 + 1}}{\sum_{i=1}^{m_k} |g_{i,k}|^2 \frac{\eta \frac{m_k}{N}}{\eta \sum_{i=1}^{m_k} |h_{i,k}|^2 + 1} + \frac{1}{K}}; \tag{23}$$

DOI: 10.1002/ett

$$\rho_a^1 \le \rho_{upper}^1 \triangleq \sum_{k=1}^{K} \frac{\left(\sum\limits_{i=1}^{m_k} |g_{i,k}|\,|h_{i,k}|\sqrt{\frac{\frac{\eta}{N}}{\eta|h_{i,k}|^2+1}}\right)^2}{\sum\limits_{i=1}^{m_k} |g_{i,k}|^2 \frac{\frac{\eta}{N}}{\eta|h_{i,k}|^2+1} + \frac{1}{K}}. \tag{24}$$

Proof See Appendix C. ∎

Based on these bounds, we have the following identity:

Lemma 4 *For any* m_k, $\rho_{upper}^{m_k} \ge \rho_{upper}^1$.

Proof See Appendix D.

We consequently has the following theorm regarding the capacity of the network:

Theorem 3 *If we denote the* maximal achievable *network capacity for* $K = N$ *as* C_A^1*, for any* m_k*,* $C_A^1 \le C_A^{m_k}$.

We note that for both analogue and digital relaying, the three capacity theorems can hold for *any* choice of fading distribution, as long as each $h_{i,k}$ ($g_{i,k}$) is i.i.d. distributed. For the simulations in the next section, we choose Rayleigh fading as an example to confirm the analysis.

5. SIMULATION RESULTS

We consider both fast fading and slow fading scenarios. For fast fading, we calculate the ergodic capacity (in bits per channel use), which is the minimum of the average channel capacity for each link in the network. For slow fading, we calculate the 10% outage capacity. We define that an outage occurs whenever the transmission rate is above the channel capacity for the worst link in the network. We consider 1000 channel realizations for each value of η, denoted as *P* in the figures. We assume that the distance between source and destination is normalized. The relays are uniformly and randomly located in the middle region between the source and destination. Taking into account the pathloss and Rayleigh fading, each channel realization can be expressed as:

$$h_{i,k} = \sqrt{0.5^{-4}}\tilde{h}_{i,k}, \quad g_{i,k} = \sqrt{0.5^{-4}}\tilde{g}_{i,k},$$

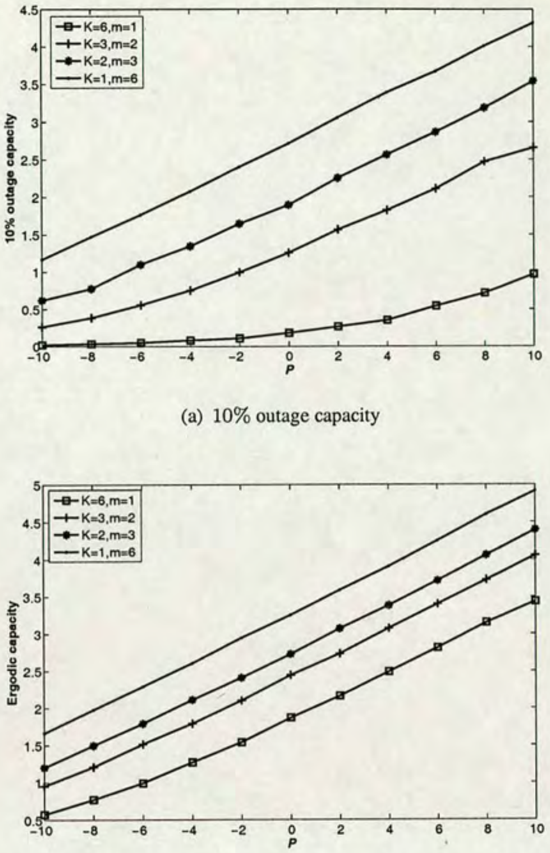
where $\sqrt{0.5^{-4}}$ denotes the pathloss (with exponent 4), the entries of $\tilde{h}_{i,k}$ ($\tilde{g}_{i,k}$) are i.i.d. complex Gaussian variables with zero mean and unit variance. We assume the total number of antennas at relays (N) is 6 and we also assume that all K relays have the same number of antennas m. Fig. 2 and Fig. 3 show the capacity performance for digital relaying and analogue relaying. We can see that for different (K, m), the capacity is always upper bounded by $(1, 6)$ and lower bounded by $(6, 1)$. These results verify the analysis made in this paper. Furthermore, we can see through the simulation that larger m and small K might give larger benefit, since larger m allows more freedom of cooperation among the antennas at each relay. Therefore when m reaches N (K reduces to 1), full cooperation is made among all the antennas to give rise to the best performance.

Comparing the performance of digital relaying with that of analogue relaying, we can see that as K increases, the performance of digital relaying decays faster than that of analogue relaying. This is mainly because with more relays the capacity of digital relaying is constrained by the worst source-relay link in order for the signals to be correctly decoded at all relays. However, the capacity of analogue relaying takes into account the impact of all the K relay channels as decoding is only performed at the destination. Indeed one should note from (18) that the performance advantage for digital relaying is significant only when $\min\left(C_r^{k,m_k}\right)$ is no less than the relay-destination link capacity. In this sense, deploying all the antennas on one relay turns out to be the optimal choice. This observation also confirms the suggestion in many existing papers, which argue that relays should be properly selected before being used for digital relaying (e.g. [6, 7]).

6. CONCLUSIONS

In this paper we analyze the performance of multiple relay channels when multiple antennas are deployed only at relays. We apply antenna diversity techniques at the relays, which are known as maximum ratio combining and transmit beamforming. If we assume that the total number of antennas at all relays is fixed to N and the total transmit power at all relays is fixed to a constant, the network capacity will be lower bounded by that of N relay channels each with single antenna, and upper bounded by that of a single relay channel with N antennas.

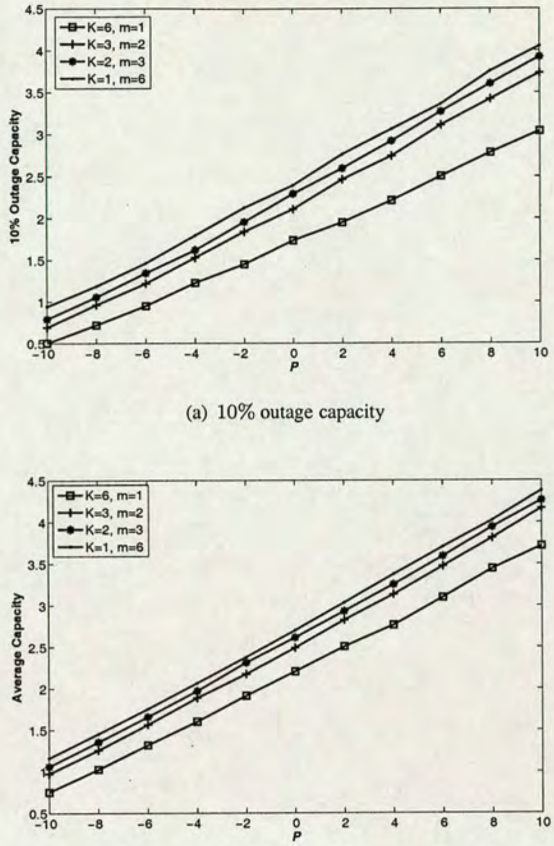
The analysis in the paper also implies that given a certain amount of available antennas in the network, wired cooperation (i.e. all the antennas belong to one terminal) outperforms wireless cooperation (i.e. each antenna belongs to different terminals). We further note that the recently proposed fixed relay concept [2] in mesh networks allows the possibility to deploy a large number of



(a) 10% outage capacity

(b) Ergodic capacity

Figure 2. Capacity of relay channels using decode-and-forward relaying for different number of relays K, while each relay is deployed with m antennas. (a) 10% outage capacity. (b) Ergodic capacity.



(a) 10% outage capacity

(b) Ergodic capacity

Figure 3. Capacity of relay channels using amplify-and-forward relaying for different number of relays K, while each relay is deployed with m antennas. (a) 10% outage capacity. (b) Ergodic capacity.

antennas at the relay. This provide a good application for the antenna combining techniques discussed in the paper.

A. PROOF OF LEMMA 1

We firstly prove that $\rho_d^1 \leq \rho_d^{m_k}$. We write the following

$$\sqrt{\rho_d^{m_k}} - \sqrt{\rho_d^1} = \sum_{k=1}^{K} \left(\underbrace{\sqrt{\frac{\eta m_k}{N} \sum_{i=1}^{m_k} |g_{i,k}|^2}}_{A_k} - \underbrace{\sqrt{\frac{\eta}{N}} \sum_{i=1}^{m_k} |g_{i,k}|}_{B_k} \right).$$

To compare A_k with B_k, we write

$$A_k^2 - B_k^2 = \frac{\eta}{N} \left(m_k \sum_{i=1}^{m_k} |g_{i,k}|^2 - \left(\sum_{i=1}^{m_k} |g_{i,k}| \right)^2 \right) \quad (25)$$

$$= \frac{\eta}{N} \left((m_k - 1) \sum_{i=1}^{m_k} |g_{i,k}|^2 \right) - \frac{\eta}{N} \left(\sum_{i,j=1; i \neq j}^{m_k} |g_{i,k}| \, |g_{j,k}| \right). \quad (26)$$

DOI: 10.1002/ett

Note that

$$(m_k-1)\sum_{i=1}^{m_k}|g_{i,k}|^2=\sum_{i=1}^{m_k}\sum_{j=1,j\neq i}^{m_k}|g_{j,k}|^2 =0.5\sum_{i,j=1;i\neq j}^{m_k}\left(|g_{i,k}|^2+|g_{j,k}|^2\right).$$

So (26) can be further written as:

$$A_k^2-B_k^2=\frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{m_k}\left(|g_{i,k}|^2-2|g_{i,k}||g_{j,k}|+|g_{j,k}|^2\right) =\frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{m_k}\left(|g_{i,k}|-|g_{j,k}|\right)^2\geq 0.$$

So $A_k \geq B_k$ and therefore $\rho_d^1 \leq \rho_d^{m_k}$.

Next we prove that $\rho_d^N \geq \rho_d^{m_i}$. For simplicity, we denote

$$a_k=\sum_{i=1}^{m_k}|g_{i,k}|^2 \tag{27}$$

in equation (9) and (11). Then $\rho_d^N-\rho_d^{m_i}$ can be written as

$$\rho_d^N-\rho_d^{m_i}=\frac{\eta}{N}\left(\sum_{k=1}^{K}(N-m_k)\,a_k -\sum_{i,j=1;i\neq j}^{K}\sqrt{m_i m_j a_i a_j}\right). \tag{28}$$

Noting the constraint (1) in section II, we have the following:

$$(N-m_k)=\sum_{i=1,i\neq k}^{K}m_i. \tag{29}$$

Putting (29) into (28), we then have:

$$\rho_d^N-\rho_d^{m_i}=\frac{\eta}{N}\left(\sum_{k=1}^{K}\sum_{i=1,i\neq k}^{K}m_i a_k -\sum_{i,j=1;i\neq j}^{K}\sqrt{m_i a_i m_j a_j}\right). \tag{30}$$

Note the following:

$$\sum_{k=1}^{K}\sum_{i=1,i\neq k}^{K}m_i a_k=\sum_{i,j=1;i\neq j}^{K}(m_i a_j) =0.5\left(\sum_{i,j=1;i\neq j}^{K}(m_i a_j)+\sum_{i,j=1;i\neq j}^{K}(m_j a_i)\right).$$

We can further write (30) as follows:

$$\rho_d^N-\rho_d^{m_i}=\frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{K}(m_i a_j) -\frac{\eta}{N}\sum_{i,j=1;i\neq j}^{K}\sqrt{m_i a_j m_j a_i} +\frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{K}(m_j a_i) =\frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{K}\left(\sqrt{m_i a_j}-\sqrt{m_j a_i}\right)^2.$$

Therefore $\rho_d^N \geq \rho_d^{m_i}$ and $\rho_d^1 \leq \rho_d^{m_k} \leq \rho_d^N$.

B. PROOF OF LEMMA 2

To efficiently prove **Lemma 2**, we firstly introduce the following new lemma.

Lemma 5 *For any positive real numbers x_1, x_2, y_1, y_2, a, if*

$$x_1 \geq x_2 \text{ and } \frac{x_1}{y_1} \geq \frac{x_2}{y_2}, \tag{31}$$

then

$$\frac{x_1}{y_1+a} \geq \frac{x_2}{y_2+a}. \tag{32}$$

Proof The proof is straightforward by showing $\frac{x_1}{y_1+a}-\frac{x_2}{y_2+a}\geq 0$. ∎

Compare numerators of (15) and (17), we have the following:

DOI: 10.1002/ett

$$\left(\sum_{k=1}^{K}\sqrt{\eta\sum_{i=1}^{m_k}|g_{i,k}|^2\sum_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}}\right)^2$$
$$\overset{(a)}{\leq}\sum_{k=1}^{K}\sum_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta}{\eta\sum\limits_{k=1}^{K}\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}\times\left(\sum_{k=1}^{K}\sqrt{\eta\sum_{i=1}^{m_k}|g_{i,k}|^2\frac{m_k}{N}}\right)^2$$
$$\overset{(b)}{\leq}\sum_{k=1}^{K}\sum_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta}{\eta\sum\limits_{k=1}^{K}\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}\times\eta\sum_{k=1}^{K}\sum_{i=1}^{m_k}|g_{i,k}|^2,$$

where inequality (a) holds due to the fact that $\frac{x}{1+x}$ is monotonically increasing with x, inequality (b) can be found in the proof of **Lemma 1** in Appendix A. From **Lemma 3**, we can now remove factor 1 in the denominators of (15) and (17). After some modifications, it can be seen that to compare ρ_a^N and $\rho_a^{m_k}$ is to compare

$$\sum_{k=1}^{K}\sum_{i=1}^{m_k}|h_{i,k}|^2$$

with

$$\frac{\left(\sum\limits_{k=1}^{K}\sqrt{\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\sum\limits_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}}\right)^2}{\sum\limits_{k=1}^{K}\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}},$$

which is equivalent to comparing $\sum\limits_{k=1}^{K}c_k\sum\limits_{k=1}^{K}d_k$ with $\left(\sum\limits_{k=1}^{K}\sqrt{c_kd_k}\right)^2$, where

$$c_k=\sum_{i=1}^{m_k}|h_{i,k}|^2,\;\; d_k=\sum_{i=1}^{m_k}|g_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}.$$

By the Cauchy-Schwarz inequality, it is clear that

$$\sum_{k=1}^{K}c_k\sum_{k=1}^{K}d_k\geq\left(\sum_{k=1}^{K}\sqrt{c_kd_k}\right)^2, \qquad (33)$$

the proof is thus completed.

C. Proof of *Lemma 3*

The proof is straightforward if the following identity is noticed:

Lemma 6 *For positive real sequences $\{a_n\}$, $\{b_n\}$, the following inequality holds:*

$$\sum\frac{a_n}{b_n}\geq\frac{\left(\sum\sqrt{a_n}\right)^2}{\sum b_n}, \qquad (34)$$

with equality if $\{a_n\}$, $\{b_n\}$ are constant values that do not change with n.

Proof The proof can be directly obtained by showing $\sum\frac{a_n}{b_n}-\frac{(\sum\sqrt{a_n})^2}{\sum b_n}\geq 0$. ▮

D. Proof of *Lemma 4*

The proof can be started by comparing each element in the summation of $\rho_{upper}^{m_k}$ and ρ_{upper}^{1}. i.e. by comparing

$$SNR_{m_k}^{m_k}\triangleq\frac{\eta\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\sum\limits_{i=1}^{m_k}|h_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}}{\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\frac{\eta\frac{m_k}{N}}{\eta\sum\limits_{i=1}^{m_k}|h_{i,k}|^2+1}+\frac{1}{K}}$$

with

$$SNR_{m_k}^{1}\triangleq\frac{\left(\sum\limits_{i=1}^{m_k}|g_{i,k}|\,|h_{i,k}|\sqrt{\frac{\frac{\eta}{N}}{\eta|h_{i,k}|^2+1}}\right)^2}{\sum\limits_{i=1}^{m_k}|g_{i,k}|^2\frac{\frac{\eta}{N}}{\eta|h_{i,k}|^2+1}+\frac{1}{K}} \qquad (35)$$

It can be seen from the proof of *Lemma 2* in Appendix II that $SNR_{m_k}^{m_k}\geq SNR_{m_k}^{1}$ by setting $K=m_k$ and $m_k=1$. The proof is therefore completed.

REFERENCES

1. A. J. Goldsmith, S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks Wireless Communications", *IEEE Pers. Commun.*, Vol. 9, no. 4, pp. 8 - 27, Aug. 2002.
2. R. Pabst et al, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, Sept. 2004.
3. M. Frodigh, S. Parkvall, C. Roobol, P.Johansson, Larsson, "Future-generation wireless networks" *IEEE Pers. Commun.*, vol. 8, no. 5, pp. 10 - 17, Oct. 2001.
4. J. N. Laneman, G. W. Wornell and D. N. C. Tse, "An efficient protocol for realizing cooperative diversity in wireless networks," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001.
5. D. Gesbert, M. Shafi, Shiu Da-shan, P.J.Smith, A.Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 281 - 302, April 2003.
6. J. N. Laneman, D. N. C. Tse and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, Volume 50, Issue 12, Dec. 2004 Page(s):3062 - 3080.
7. 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.
8. 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.
9. B. Wang, J. Zhang, and A. Host-Madsen, "On capacity of MIMO relay channel," *IEEE Trans. Inf. Theory*, Volume 51, Issue 1, Jan. 2005 Page(s):29 - 43.
10. H. Bolcskei et al, "Capacity Scaling Laws in MIMO Relay Networks," *IEEE Trans. Wireless Comm.*, Volume 5, Issue 6, June 2006 Page(s):1433 - 1444.
11. O. Oyman and A. J. Paulraj, "Energy efficiency in MIMO relay networks under processing cost", in *Conf. Inf. Science and Systems*, The Johns Hopkins University, March 16-18, 2005.
12. J. G. Proakis, *Digital Communications: Fourth Edition*, 2001.
13. J. Bach Andersen, "Antenna arrays in mobile communications: gain, diversity, and channel capacity", *IEEE Antennas Propagat. Mag.*, vol. 42, no. 2, pp. 12 - 16, April 2000.
14. D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005.
15. W. Rhee, J. M. Cioffi, "On the Capacity of Multiuser Wireless Channels With Multiple Antennas", *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2580 - 2595, October 2003.
16. Y. Chen, C. Tellambura, "Performance analysis of L-Branch equal gain combiners in Equally correlated Rayleigh fading channels", *IEEE Commun. Letters*, vol. 8, no. 3, pp. 150 - 152, March 2004.
17. Y. Fan, J. S. Thompson, "MIMO configurations for relay channels: theory and practice", accepted to *IEEE Trans. Wireless. Commun.*, 2007.

AUTHORS' BIOGRAPHIES

Yijia Fan received his BEng degree in electrical engineering from Shanghai Jiao Tong University (SJTU), Shanghai, P.R.China, in July 2003. Since October 2003, he has been a PhD candidate at the Institute for Digital Communications, University of Edinburgh. His PhD project is fully funded by Engineering and Physical Sciences Research Council (EPSRC), UK. His current research interests include signal processing and information theory for MIMO wireless systems and their applications in future generation wireless networks.

Abdulkareem Adinoyi received a B.Eng degree from the University of Ilorin, Nigeria, in 1992, M.S degree from the King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 1998 and Ph.D degree from Carleton University, Ottawa, Canada, in 2006, all the degrees are in electrical engineering. He was with Dubi Oil Limited, Port Harcourt, Nigeria as an Instrument/Electrical Engineer from April 1993 to August 1995. He was with KFUPM between September 1995 and October 1998 as a Research Assistant. Between January 1999 and August 2002 he held the position of a lecturer at the department of Electrical Engineering, KFUPM. Between January 2004 and December 2006 he participated in the European Union 6th Framework integrated project - the WINNER under the auspices of the Department of Systems and Computer Engineering at Carleton University. From May to December 2006, he was a senior research associate. The WNNER is dedicated to researching, developing and demonstrating a seamless multi-scenario next generation wireless air interface. Dr. Adinoyi is now an Assistant Professor at Qassim University, Saudi Arabia. His research interest is in wireless communication networks with a special emphasis on infrastructure-based multihop and relay networks, cooperative diversity schemes and protocols.

John S Thompson received his BEng and PhD degrees from the University of Edinburgh in 1992 and 1996, respectively. From July 1995 to August 1999, he worked as a postdoctoral researcher at Edinburgh, funded by the UK Engineering and Physical Sciences Research Council (EPSRC) and Nortel Networks. Since September 1999, he has been a lecturer at the School of Engineering and Electronics at the University of Edinburgh. In October 2005, he was promoted to the position of reader. His research interests currently include signal processing algorithms for wireless systems, antenna array techniques and multihop wireless communications. He has published approximately 100 papers to date including a number of invited papers, book chapters and tutorial talks, as well as co-authoring an undergraduate textbook on digital signal processing. He is currently an editor-in-chief of IEE Proceedings on Vision, Image and Signal Processing and is the technical programme co-chair for the IEEE International Conference on Communications (ICC) 2007, to be held in Glasgow.

Halim Yanikomeroglu received a B.S. degree in electrical and electronics engineering from the Middle East Technical University, Ankara, Turkey, in 1990, and an M.A.S. degree in electrical engineering (now ECE), and a Ph.D. degree in electrical and computer engineering from the University of Toronto, Canada, in 1992 and 1998, respectively. He was with the Research and Development Group of Marconi Kominikasyon A.S., Ankara, Turkey, from January 1993 to July 1994. Since 1998, he has been with the Department of Systems and Computer Engineering at Carleton University, Ottawa, where he is now an Associate Professor and Associate Chair for Graduate Studies. His research interests include almost all aspects of wireless communications with a special emphasis on infrastructure-based multihop/mesh/relay networks. He has been involved in the steering committees and technical program committees of numerous international conferences in communications; he has also given several tutorials in such conferences. He was the Technical Program Cochair of the IEEE Wireless Communications and Networking Conference 2004 (WCNC04). He was an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS during 2002 - 2005, and a Guest Editor for Wiley Journal on Wireless Communications & Mobile Computing; he was an Editor for IEEE COMMUNICATIONS SURVEYS & TUTORIALS for 2002 - 2003. Currently he is serving as the Chair of the IEEE Communications Societys Technical Committee on Personal Communications (TCPC), he is also a Member of IEEE ComSocs Technical Activities Counsel (TAC).

MIMO Configurations for Relay Channels: Theory and Practice

Yijia Fan, *Student Member, IEEE,* and John Thompson, *Member, IEEE*

***Abstract*— In this paper we discuss and compare different signalling and routing methods for multiple-input multiple-output (MIMO) relay networks in terms of the network capacity, where every terminal is equipped with multiple antennas. Our study for signalling includes the two well known *digital (decode and forward) relaying*, *analogue (amplify and forward) relaying*, and a novel *hybrid (filter, amplify and forward) relaying*. We propose both optimal and suboptimal *hybrid relaying* schemes which avoid full decoding of the message at the relay. We show that they outperform *analogue relaying* and give similar performance to *digital relaying*, particularly when the relay has forward channel state information (CSI) or larger number of antennas than the source and destination. For the routing schemes designed for multiple relay channels, we use relay selection schemes to exploit the spatial diversity, which we call *selection diversity* of the networks. We propose both optimal and suboptimal relay selection schemes and show that their performance converges when a large number of antennas is deployed at each node in the network. We also compare relay selection routing with a space-time coded relay cooperation protocol and show the performance advantage of *selection diversity* over *cooperative diversity* in certain scenarios. Finally, we give a brief discussion on the application of another MIMO structure called single signal beamforming in the relay scenario. Its performance will be compared with that of spatial multiplexing.**

***Index Terms*— Multiple-input multiple-output (MIMO), relay, capacity, wireless networks.**

I. Introduction

THE use of multiple antennas at both ends of a wireless link, called multiple-input multiple-output (MIMO) technology, promises significant improvements in terms of spectral efficiency and link reliability. A number of research projects have been conducted in the last decade on point-to-point MIMO channels (see [1] and references therein). Recently, the characteristics of MIMO channels in a multi-user context have also been theoretically studied [2]–[8].

More recently, applying MIMO techniques to relay networks [9], [10] has also come under consideration. For single antenna relay networks, the so-called *cooperative diversity*, which mimics the performance advantages of MIMO systems in exploiting the spatial diversity of relay channels, has been studied through defining effective protocols or using space-time codes for channelisation [11]–[15]. For multiple-antenna or MIMO relay channels where every terminal in the network can be deployed with multiple antennas, studies are mainly concentrated on spatial multiplexing systems. Capacity bounds for single relay MIMO channels are presented in [16]. Quantitative capacity results for a multiple MIMO relay network have been reported in [17], where diversity is achieved again through cooperation among all the relays available in the network. The analysis is extended to multiple source-destination scenarios in [18], where the energy efficiency of the MIMO multiple relay network considering multiple source-destination pairs is further discussed.

So far most research on multiple antenna relay channels is aimed at obtaining information-theoretic limits for different protocols exploiting relay cooperative diversity in the network. In this paper we discuss practical signalling and routing schemes for MIMO relay channels in terms of network capacity. Our discussions on signalling methods are based on three relaying modes. The first two kinds are well known: *analogue (amplify and forward) relaying* [13], where the relays simply amplify the signals, and *digital (decode and forward) relaying* [13] where the relays decode, re-encode, and re-transmit the signals. We also investigate a novel one called *hybrid (filter, amplify and forward) relaying*, where the relays only decode the training sequence from the source or the feedback from the destination to obtain the full CSI for either source to relay (backward) or relay to destination (forward) channels. Then the relay applies a spatial filter to the received signals based on the CSI without decoding them and retransmits the filter outputs. The major contributions of this paper for relay signalling are as follows:

- For digital relaying, we apply a V-BLAST/beamforming signalling and detection structure in the relay channel and show through simulation that its performance approaches the information-theoretic upper bound of relay channels, especially when the CSI of the relay to destination channel is available at the relay.
- We propose a novel hybrid relaying concept for MIMO relay channels, which combines the benefit of digital relaying and analogue relaying. We develop optimal and suboptimal hybrid relaying schemes and compare them with analogue relaying and digital relaying schemes. We show that hybrid relaying outperforms analogue relaying, and is a good suboptimal choice compared with digital relaying, especially when the number of antennas at the relay are larger than at the source and destination.

For the routing schemes designed for multiple relay channels, unlike most of the papers cited above, we assume that each relay processes and forwards the signals independently; no advanced protocol considering relay cooperation was used.

Manuscript received August 30, 2005; revised March 27, 2006 and August 30, 2006; accepted October 3, 2006. The associate editor coordinating the review of this paper and approving it for publication was J. Zhang. This work was supported by the EPSRC Grant GR/S58782/01, UK.

The authors are with the Institute for Digital Communications, University of Edinburgh, Edinburgh, EH9 3JL, UK (e-mail: {y.fan, john.thompson}@ed.ac.uk).

Digital Object Identifier 10.1109/TWC.2007.05618.

We investigate multiple relay channels in a different way from previous work where the signals are multi-casted by multiple relays. In this paper the spatial diversity of the network is exploited through adaptive routing techniques, i.e. by selecting the most preferred single relay from all candidate relays to forward the signals. The contribution of this paper regarding this aspect is summarized as follows:

- We exploit the *selection diversity* for the relay channel, instead of *cooperative diversity*, through relay selection. We propose and compare the optimal routing scheme with a suboptimal routing scheme designed for MIMO relay channels. It is shown that the proposed suboptimal routing scheme approaches the performance upper bound of the optimal routing scheme with large array sizes, and can be the preferred choice for its tradeoff between performance and complexity.
- We give an example comparing relay selection routing with a protocol exploiting *cooperative diversity* through distributed space-time coding in a two relay scenario. We show that *selection diversity* is preferable to *cooperative diversity* when the total transmit power at the relays is fixed.
- We also show that with a total power constraint at the relays, the multi-cast routing scheme, which uses all the relays to forward the message, is not preferable to the relay selection routing scheme when the relays are not allowed to cooperate.

Note that our proposed signalling methods for single relay channels can also be applied with our proposed selective routing schemes in a multi-relay scenario. In the last part of the paper, we compare the spatial multiplexing array processing technique with another popular MIMO structure called *single signal beamforming* in the relay scenario. We show that *single signal beamforming* might give higher capacity performance at low receive signal to noise ratio (SNR) values. We indicate that it might be beneficial to combine *single signal beamforming* with relaying, as relaying is often applied for low SNR. However, if the number of the relays in the network increases, this choice is likely to change.

The rest of this paper is organized as follows. In Section II, the basic system model and assumptions for MIMO relay channels is introduced. Section III discusses some basic signalling schemes on single MIMO relay channels. More advanced algorithms for single MIMO relay channels, when complete CSI is available at the relays are proposed in section IV. Section V discusses the routing techniques for the multiple relay case. In section VI, applications of single signal beamforming techniques in relay channels are given, and their performance is compared with some spatial multiplexing configurations developed in section III and IV. Concluding remarks are given in Section VII.

A note on notation: We use boldface to denote matrices and vectors and $\mathrm{E}\left\langle \bullet \right\rangle_x$ denotes the expectation with respect to x. $\det(\mathbf{X})$ denotes the determinant and $\mathbf{X}^{\dagger}$ denotes the pseudo-inverse of a matrix $\mathbf{X}$. $\mathbf{X}^H$ denotes the conjugate transpose and $\mathrm{tr}(\mathbf{X})$ denotes the trace. $\mathbf{I}_M$ denotes the $M \times M$ identity matrix. $\|\bullet\|_F$ denotes the Frobenius-norm.

II. System Model

The basic system model for a single user two hop relay network is shown in Fig. 1. We consider a two hop network model with one source, one destination and K relays located randomly and uniformly within the middle region between the source and the destination. Note that many results in the paper are not necessarily restricted to this assumption. We ignore the direct link between the source and the destination due to the larger distance and additional pathloss compared to the relay links. We also assume that the total transmit power for the source and the relays is the same; it is equally distributed among the relays. Each relay processes the received signals independently. For notational simplicity, we assume in the paper that the source and destination have the same number of transmit and receive antennas M, while each relay has $L \geq M$ antennas. The results can be extended to a more general case where different number of antennas are available among each transmitter or receiver.

We restrict our discussion to the case of a slow, frequency-flat block fading model. The data transmission is over two time slots using two hops. In the first transmission time slot, the source broadcasts the signal to all the relay terminals. The input/output relation for the source to the kth relay is given by

$$\mathbf{r}_k = \sqrt{\eta}\mathbf{H}_k\mathbf{s} + \mathbf{n}_k, \tag{1}$$

where $\mathbf{r}_k$ is the $L \times 1$ received signal vector. η denotes the power per transmit antenna at the source. The vector $\mathbf{s}$ is the $M \times 1$ transmit signal vector with covariance matrix $\mathbf{I}_M$ and $\mathbf{n}_k$ is the $L \times 1$ complex circular additive white Gaussian noise vector at relay k which has zero mean and identity covariance matrix $\mathbf{I}_L$. $\mathbf{H}_k$ is the $L \times M$ channel transfer matrix from source to the kth relay. $\mathbf{H}_k$ can be further expressed as $\mathbf{H}_k = \sqrt{\alpha_k}\tilde{\mathbf{H}}_k$, where the entries of $\tilde{\mathbf{H}}_k$ are identically independent distributed (i.i.d) complex Gaussian random variables with unit variance. Each factor α_k contains the pathloss and independent lognormal shadowing terms. It can be written as $\alpha_k = x^{-\gamma}10^{\zeta_k/10}$, where x is the distance between the source and relay k. The scalar γ denotes the path loss exponent (in this paper it is always set to 4). The lognormal shadowing term, ζ_k, is a random variable drawn from a normal distribution with a mean of 0 dB and a standard deviation δ (dB). In our simulations we use $\delta = 8$ dB, which is a value typical of shadowing deviations in urban cellular environments. We normalized the range between the source and destination so that x is 0.5.

Each relay processes their received signals and transmits them to the destination. The signal received at the destination can be written as:

$$\mathbf{y} = \sum_{i=1}^{K}\mathbf{G}_k\mathbf{d}_k + \mathbf{n}_d, \tag{2}$$

where the matrix $\mathbf{G}_k$ is the channel matrix from kth relay to the destination, which might also be written as $\mathbf{G}_k = \sqrt{\beta_k}\tilde{\mathbf{G}}_k$, where each entry of $\tilde{\mathbf{G}}_k$ is an i.i.d. complex Gaussian random variable with unit variance. β_k contains the same pathloss as α_k and independent lognormal shadowing terms with the same mean and deviation as in α_k. The vector $\mathbf{n}_d$ is the $M \times 1$ complex circular additive white Gaussian noise at the

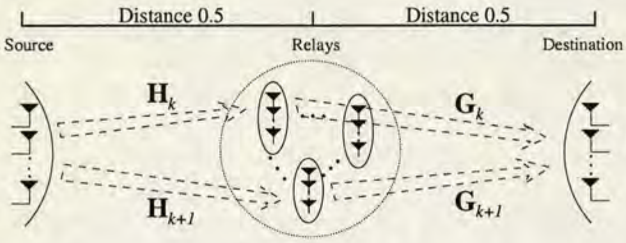


Fig. 1. Basic system model of a MIMO two hop network.

destination with identity covariance matrix. The vector $\mathbf{d}_k$ is the transmit signal vector at relay k, which should meet the total transmit power constraint:

$$\mathrm{E}\left\langle \|\mathbf{d}_k\|_F^2 \right\rangle \leq \frac{\eta M}{K}. \tag{3}$$

Unless specifically stated, we assume that the source does not know the channel information; the destination knows all the channel information. Based on the different CSI available at relays, we place MIMO relay channels into the following two categories [17]:

1) Non-coherent Relay Channels: where the kth relay only has full knowledge of the channel matrix $\mathbf{H}_k$.

2) Coherent Relay Channels: where the kth relay can obtain full knowledge of both $\mathbf{H}_k$ and $\mathbf{G}_k$.

Note that in practice for the signalling or routing schemes to be effectively implemented for rate adaptive transmission, feedback should be allowed from the receiver to the transmitter. We will specifically state the feedback required by the particular techniques when we introduce them later in the paper.

III. Relaying Schemes for Non-coherent Single Relay Channels

In this section we discuss the relaying configurations for non-coherent single relay channels. We replace the notation $\mathbf{H}_k$ and $\mathbf{G}_k$ by $\mathbf{H}$ and $\mathbf{G}$.

A. Digital Relaying (DR)

In this scheme, the message at the source is multiplexed into M different signal streams, each independently encoded and transmitted to the relay. We apply a low-rate feedback V-BLAST MMSE detector [19] at the receiver to decode the signal streams, where the signal to interference plus noise ratios (SINRs) at the receiver are fed back to the transmitter. Specifically, the relay uses L antennas to detect each signal stream through successive interference cancellation, which consists of M iterations, each aimed at decoding one signal stream. It then decodes and re-encodes each signal stream to forward the M signal streams to the destination using *arbitrary* M antennas. The destination can then apply the V-BLAST MMSE detector to detect and decode each signal stream in the same way. The Shannon capacity from the source to the relay can be achieved by the V-BLAST MMSE detector if we assume that each signal is correctly decoded:

$$C = \log_2 \det\left(\mathbf{I} + \eta\mathbf{H}\mathbf{H}^H\right) = \sum_{i=1}^{M} \log_2\left(1 + \eta\psi_{\mathbf{H}}^i\right) \tag{4}$$

in bits per channel use, where $\psi_{\mathbf{H}}^i$ is the output SINR for signal i in the V-BLAST detector at the relay. Note that this capacity (4) can be achieved regardless of the decoding order. For a closed form expression for SINR for different streams for any decoding order, please also refer to equation (4) in section 2.3 of [19]. In this paper, we choose the decoding order as follows: for each iteration, the receiver detects and decodes the *strongest* signal (with the highest SINR in that iteration). The main reason we recommend this decoding order is to reduce possible error propagation (i.e. improve bit error ratio) in the V-BLAST-MMSE detector in a real system. However, with rate allocation and long code block lengths, error propagation can be ignored. Maximum SINR ordering is also useful for coherent relay channels, which will be studied in the next section.

The detection at the destination is performed in the same way, and the capacity from relay to destination can also be expressed as $C = \sum_{i=1}^{M} \log_2\left(1 + \eta\psi_{\mathbf{G}}^i\right)$, where $\psi_{\mathbf{G}}^i$ is the output SINR for signal i at destination. Note that the channel matrix $\mathbf{G}$ is of size $M \times M$ instead of $M \times L$. The network capacity for this relay configuration can be easily observed as:

$$C = 0.5 \times \sum_{i=1}^{M} \log_2\left(1 + \eta \times \min\left(\psi_{\mathbf{H}}^i, \psi_{\mathbf{G}}^i\right)\right), \tag{5}$$

where the factor 0.5 denotes the half multiplexing loss due to relaying compared with a relay-free scenario. We note that in order to achieved the capacity of the system, the SINR information ($\psi_{\mathbf{H}}^i$ or $\psi_{\mathbf{G}}^i$) needs to be fed back to the transmitter for rate adaptive transmission. If the relay knows the SINR values $\psi_{\mathbf{G}}^i$ at the destination, e.g. through overhearing the channel feedback from destination to source, equation (5) can be maximized by ranking the $\psi_{\mathbf{H}}^i, \psi_{\mathbf{G}}^i$ in the same order (e.g. both are monotonically decreasing sequences).

With the power constraint proposed by (3), the mutual information upper bound for the non-coherent relay channel is

$$C = 0.5 \min\left(\log\left(\mathbf{I} + \eta\mathbf{H}\mathbf{H}^H\right), \log\left(\mathbf{I} + \eta\frac{M}{L}\mathbf{G}\mathbf{G}^H\right)\right) \tag{6}$$

where the channel matrix $\mathbf{G}$ has dimension $M \times L$ and the factor M/L represents the power constraint at the relay. From a practical point of view, this upper bound could be achieved by joint coding/interleaving across antennas and maximum likelihood (ML) detection at the relay and destination. However, it is well known that the ML detection complexity increases exponentially with the number of antennas [19].

The capacity result in equation (6) could also be achieved by the V-BLAST MMSE structure, when the relay has knowledge of the channel SINR values $\psi_{\mathbf{G}}^i$. However, unlike ML detector, this may require breaking the M received data packets and reconstructing L different data packets for retransmission to the destination, in order to exploit optimally the relay-destination SINRs $\psi_{\mathbf{G}}^i$. This results in a higher complexity and might raise implementation issues in practice. If an end-to-end Automatic Repeat-reQuest (ARQ) protocol is employed [20], it implies that one ARQ process must be used for all M source packets, rather than M or L separate ARQ processes. An error

$$C = 0.5 \log_2 \det \left(\eta\rho \left(\mathbf{GWH}\right)^H \left(\mathbf{I} + \rho\mathbf{GW}\left(\mathbf{GW}\right)^H\right)^{-1} \mathbf{GWH} + \mathbf{I} \right) \tag{11}$$

in one of the M or L packets means that all M packets must be retransmitted from source, reducing throughput compared to the case where only the errored packet is retransmitted [21]. Alternatively, separate ARQ processes must be set up for the source-relay and relay-destination links. Omitting this breaking/reconstructing process simplifies the task of the relay and the required ARQ process and is therefore assumed in our scheme, which achieves the capacity in (5).

As will be observed in later simulations, our configuration performs almost the same as the upper bound in (6) when $L = M$. This is mainly because with uncorrelated lognormal shadowing considered on both links in the network, one link will usually experience much higher signal power than the other link. This means the minimum SINRs of the two channel matrices in the V-BLAST MMSE detectors are often all for the same link, which means that (5) equals (6) for that channel realization. In the unlikely event that the shadowing coefficients are similar in magnitude, the minimum SINRs for different streams may be chosen from both links. However, $\psi_{\mathbf{G}}^i$ and $\psi_{\mathbf{H}}^i$ are likely to be similar in value as i increases. This is mainly due to the increasing diversity order, which can help average out the fading effect, in each iteration of the V-BLAST detector. The diversity gain obtained by the VBLAST zero forcing (ZF) or MMSE detector without ordering is shown to be $(L - M + i)$ (see Chapter 8.5.1 of [22]), which equals to i when $L = M$. It has also been speculated that changing the detection order will not affect the diversity order [23], [24].

The situation is slightly different when $L > M$ due to increased degrees of freedom for retransmission at the relays. However, with the power constraint at the relay, the performance advantage of $M \times L$ channels over that for $M \times M$ channels is limited, as it still only has M eigenvalues and can only effectively support M data streams.

B. Analogue Relaying (AR)

In this scheme, the relay amplifies its received signal vector by:

$$\sqrt{\rho} = \sqrt{\frac{M \times \eta}{\eta \left\|\mathbf{H}\right\|_F^2 + L}} \tag{7}$$

to meet the power constraint described by (3) and forwards it to the destination.[1] Compared with *DR*, one obvious defect for *AR* is that while the relays amplify the signals, they also amplify the receiver noise. The average total signal to total noise ratio (TSTNR) τ at the relay can be defined as:

$$\tau \triangleq \frac{\eta \mathrm{E}\left\langle \left\|\mathbf{H}\right\|_F^2 \right\rangle_{\tilde{\mathbf{H}}}}{L} = \eta\alpha_k M. \tag{8}$$

However, this scheme requires no decoding at the relay. This means there is no decoding delay at the relay and it requires less processing power at the relay compared with *DR*, which employs full decoding and re-encoding at the relay. As will be described below, *AR* can be regarded as the simplest *hybrid relaying* scheme.

[1]Note that another suboptimal but more convenient way of scaling power is to take expectation of $\left\|\mathbf{H}\right\|_F^2$ regarding $\tilde{\mathbf{H}}$ so that $\sqrt{\rho} = \sqrt{\frac{M \times \eta}{\eta\alpha_k ML + L}}$

C. Hybrid Relaying (HR)

In the *HR* schemes, the relays only decode the training sequence from the source to obtain full CSI, then filter the received signals based on the knowledge of CSI without decoding them. After multiplying the signal vector by the filtering weight matrix $\mathbf{W}$, the relay then amplifies and forwards the filtered signals to the destination. The amplifying factor now can be written as:

$$\sqrt{\rho} = \sqrt{\frac{M \times \eta}{\eta \left\|\mathbf{WH}\right\|_F^2 + \left\|\mathbf{W}\right\|_F^2}}. \tag{9}$$

Note that for *AR*, $\mathbf{W} = \mathbf{I}$. The input/output relation from the source to destination for *HR* can be expressed as:

$$\mathbf{y} = \sqrt{\eta}\left(\sqrt{\rho}\mathbf{GWH}\right)\mathbf{s} + \left(\mathbf{G}\sqrt{\rho}\mathbf{W}\mathbf{n}_1 + \mathbf{n}_d\right), \tag{10}$$

where $\mathbf{n}_1$ is the white Gaussian noise vector at the relay. Thus we can treat the whole system as a point-to-point MIMO link. The network capacity can be written as equation (11) [25].

The capacity can be achieved by perfect decoding using the V-BLAST detector at the destination. One way to choose $\mathbf{W}$ is to enhance the average SNR at the relay. This can be done by applying a matched filter at relay. This relaying scheme is referred to as *Matched Filter based Relaying (MFR)*. By setting $\mathbf{W} = \mathbf{H}^H$, τ can be written as

$$\tau \triangleq \frac{\eta \mathrm{E}\left\langle \left\|\mathbf{H}^H\mathbf{H}\right\|_F^2 \right\rangle_{\tilde{\mathbf{H}}}}{\mathrm{E}\left\langle \left\|\mathbf{H}^H\right\|_F^2 \right\rangle_{\tilde{\mathbf{H}}}} = \eta\alpha_k\left(L + M\right), \tag{12}$$

which is obtained from the identity [26], $\mathrm{E}\left\langle \left\|\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right\|_F^2 \right\rangle_{\tilde{\mathbf{H}}} = LM\left(L+M\right)$. It can be seen that the τ for *MFR* is $(L+M)/M$ times that for *AR*, which implies that *MFR* has a larger value of τ than *AR*. However, it should also be noted that MFR has the defect of correlating the signals at each antenna, which makes it more difficult for the destination to separate the signals compared with *AR*.

One may consequently consider applying a MMSE filter at the relay. However, we have found that unlike a point-to-point MIMO link, applying MMSE filtering at the relay performs worse than matched filtering. More details can be found in [27].

IV. Relaying Schemes for Coherent Single Relay Channels

Compared with the non-coherent scenario, the relay has the freedom to explore and coordinate both source to relay and relay to destination channels in a coherent scenario. In the following we will investigate both digital and hybrid relaying.

$$
\begin{aligned}
C=&0.5\log_2\det\left(\eta\rho\mathbf{V_H}\mathbf{D_H^H}\mathbf{U_H^H}\mathbf{V}\mathbf{D}^H\mathbf{U}^H\left(\mathbf{I}+\rho\mathbf{U}\Lambda\mathbf{U}^H\right)^{-1}\mathbf{U}\mathbf{D}\mathbf{V}^H\mathbf{U_H}\mathbf{D_H}\mathbf{V_H^H}+\mathbf{I}\right)\\
=&0.5\log_2\det\left(\eta\mathbf{V_H}\mathbf{D_H^H}\mathbf{U_H^H}\mathbf{V}\left(\frac{\rho\Lambda}{1+\rho\Lambda}\right)\mathbf{V}^H\mathbf{U_H}\mathbf{D_H}\mathbf{V_H^H}+\mathbf{I}\right)\\
=&0.5\log_2\det\left(\eta\Lambda_{\mathbf{H}}^{1/2}\mathbf{U_H^H}\mathbf{V}\left(\frac{\rho\Lambda}{1+\rho\Lambda}\right)\mathbf{V}^H\mathbf{U_H}\Lambda_{\mathbf{H}}^{1/2}+\mathbf{I}\right)
\end{aligned}
\tag{16}
$$

A. Modified Digital Relaying (MDR)

In this scheme the relay uses all the L antennas to retransmit the M streams to exploit beamforming gain on the relay to destination channels. We use the singular value decomposition of the $M\times L$ channel matrix $\mathbf{G}=\mathbf{U_G}\mathbf{D_G}\mathbf{V_G^H}$. After decoding the signals in the same way as in non-coherent channels, the relay multiplies the decoded signal vector $\mathbf{d}$ by a $L\times M$ unitary matrix $\mathbf{V}_{\mathbf{G},M}$, which contains the M columns of $\mathbf{V_G}$ corresponding to the M nonzero singular values of $\mathbf{G}$. The destination then multiplies the received signal vector by $\mathbf{U_G}$. The source to destination MIMO channels become M parallel channels and each antenna branch can perform detection independently. To achieve the optimal capacity from the relay to destination link, a waterfilling algorithm [28] is applied at the relay transmitter. The capacity of this scheme can therefore be expressed as

$$C=0.5\times\sum_{i=1}^{M}\log_2\left(1+\eta\times\min\left(\psi_{\mathbf{H}}^i,\frac{M}{L}\gamma^i\lambda_{\mathbf{G}}^i\right)\right),\tag{13}$$

where $\lambda_{\mathbf{G}}^i$ is the ith eigenvalue[2] of matrix $\mathbf{G}\mathbf{G}^H$ and γ^i represents the power allocation at the relay transmit antennas for stream i and can be expressed as:

$$\gamma^i=\left(\mu-\frac{L}{M\eta\lambda^i}\right)^+,\sum_{i=1}^{M}\gamma^i=L.$$

Equation (13) can be maximized by ranking the columns of $\mathbf{U_G}$ and $\mathbf{V}_{\mathbf{G},M}$ so that $\psi_{\mathbf{H}}^i$ and $\gamma^i\lambda_{\mathbf{G}}^i$ are ranked in the same order (e.g. both are monotonically decreasing sequences). It is not difficult to see that the decoding order at the relay suggested in Section III. A might help improve the capacity (13) in this scenario, as it can maximize the smallest $\psi_{\mathbf{H}}^i$ by decoding the weakest signal last.

The mutual information upper bound for coherent relay channels is:

$$C=0.5\min\left(\log\left(\mathbf{I}+\eta\mathbf{H}\mathbf{H}^H\right),\log\left(\mathbf{I}+\eta\frac{M}{L}\mathbf{G}\Sigma\mathbf{G}^H\right)\right)\tag{14}$$

where $\Sigma=\operatorname{diag}\left\{\gamma^1,\cdots,\gamma^M\right\}$ denotes the digonal matrix generated from the iterative waterfilling algorithm conducted at the relay before retransmission. Note that this upper bound again requires de-multiplexing and re-multiplexing of the data streams at the relay. It can also be observed that (13) approaches (14) for the same reason as stated in section III.A. for digital relaying for non-coherent relay channels, though this time it is true not only for $L=M$ but also for $L>M$.

[2]In this paper we assume the eigenvalues of matrices are ordered arbitrarily unless specifically stated, as we will discuss the eigenvalue ordering problem in the following section

Note that in a point-to-point MIMO link for $L=M$ the benefit of waterfilling is quite small and even then is only useful for low receive SNR values. In a relay scenario where lognormal shadowing terms are considered, this benefit is even negligible due to the minimum function in capacity calculation (i.e. the average value of (6) is about the same as that of (14)). However, the benefit of waterfilling grows when L becomes larger than M, as energy is only allocated to the non-zero eigenmodes of the channel matrix. As can be seen from (13) and (14), when $L>M$ the capacity performance is mainly constrained by the relay to destination link due to the scaling factor M/L, thus waterfilling can significantly improve the network capacity by increasing the gain of γ^i as L increases.

B. Optimal Hybrid Relaying (OHR)

We now give an information-theoretic study on the optimal configuration for hybrid relaying based on knowledge of both $\mathbf{G}$ and $\mathbf{H}$ at the relay. We first replace $\mathbf{GW}$ with $\mathbf{M}$ in (11). We write the singular value decomposition $\mathbf{M}=\mathbf{U}\mathbf{D}\mathbf{V}^H$ and $\mathbf{H}=\mathbf{U_H}\mathbf{D_H}\mathbf{V_H^H}$. Recall the identity

$$\det\left(\mathbf{I}+\mathbf{AB}\right)=\det\left(\mathbf{I}+\mathbf{BA}\right).\tag{15}$$

The capacity (11) can then be modified to equation (16), where $\Lambda=\mathbf{D}\mathbf{D}^H=\operatorname{diag}\left\{\lambda_1,\lambda_2,\cdots,\lambda_N\right\}$, contains the eigenvalues of $\mathbf{M}\mathbf{M}^H$ and $\Lambda_{\mathbf{H}}=\operatorname{diag}\left\{\lambda_{\mathbf{H}}^1,\cdots,\lambda_{\mathbf{H}}^N\right\}$. Using Hadamard's inequality [29], the capacity is maximized when

$$\mathbf{U_H}=\mathbf{V}.\tag{17}$$

Thus $\mathbf{M}=\mathbf{U}\mathbf{D}\mathbf{U_H^H}$ and (16) can be written as:

$$C=0.5\sum_{i=1}^{M}\log_2\left(1+\eta\lambda_{\mathbf{H}}^i\frac{\rho\lambda_i}{\rho\lambda_i+1}\right)\tag{18}$$

$$=0.5\sum_{i=1}^{N}\log\left(1+\eta\lambda_{\mathbf{H}}^i\left(\frac{M\eta\lambda_i}{J+M\eta\lambda_i}\right)\right),\tag{19}$$

where $J=\eta\left\|\mathbf{WH}\right\|_F^2+\left\|\mathbf{W}\right\|_F^2$ is the signal energy received at the relay. For every fixed λ_i, maximizing C is equivalent to minimizing J. By replacing $\mathbf{W}$ with $\mathbf{W}=\mathbf{G}^\dagger\mathbf{M}$, it follows that

$$
\begin{aligned}
J=&\operatorname{tr}\left(\eta\mathbf{W}\mathbf{H}\mathbf{H}^H\mathbf{W}^H+\mathbf{W}\mathbf{W}^H\right)\\
=&\operatorname{tr}\left(\mathbf{G}^\dagger\mathbf{U}\mathbf{D}\left(\eta\Lambda_{\mathbf{H}}+\mathbf{I}\right)\mathbf{D}^H\mathbf{U}^H\left(\mathbf{G}^\dagger\right)^H\right) &(20)\\
=&\operatorname{tr}\left(\mathbf{U}^H\left(\mathbf{U_G}\Lambda_{\mathbf{G}}^\dagger\mathbf{U_G^H}\right)\mathbf{U}\left(\mathbf{D}\left(\eta\Lambda_{\mathbf{H}}+\mathbf{I}\right)\mathbf{D}^H\right)\right), &(21)
\end{aligned}
$$

where the identity $\operatorname{tr}\left(\mathbf{AB}\right)=\operatorname{tr}\left(\mathbf{BA}\right)$ is used to go from (20) to (21). To minimize (21), note that for a unitary matrix $\mathbf{U}$,

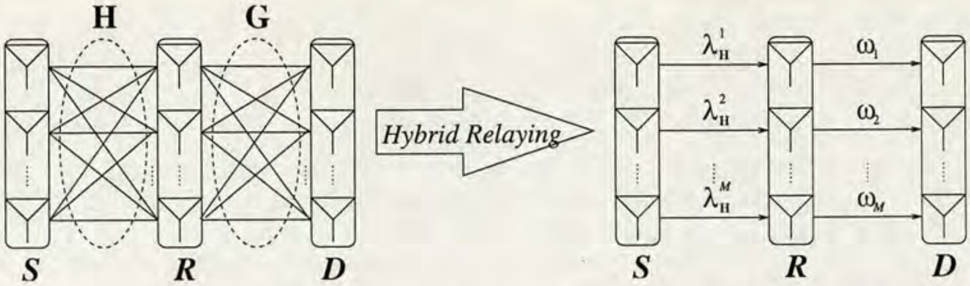


Fig. 2. The process of hybrid relaying for coherent relay channels. The MIMO relay channels are decomposed into several parallel channels each with gain $\lambda_{\mathbf{H}}^i\omega_i$, where $\omega_i = \frac{\rho\lambda_i}{1+\rho\lambda_i}$. For modified analogue relaying, λ_i is $\lambda_{\mathbf{G}}^i$. For modified matched filter relaying, λ_i is $\lambda_{\mathbf{H}}^i\lambda_{\mathbf{G}}^i$.

$N \times N$ Hermitian matrix $\mathbf{S}$ and diagonal matrix $\mathbf{\Sigma}$, we have the following identity [33]:

$$\mathrm{tr}\left(\mathbf{U}^H\mathbf{S}\mathbf{U}\mathbf{\Sigma}\right) \geq \sum_{i=1}^{N} a_i b_{N-i+1}, \tag{22}$$

where $a_1 \leq \cdots \leq a_N$ are eigenvalues of $\mathbf{S}$ and $b_1 \leq \cdots \leq b_N$ are the diagonal entries of Σ. So we see that the minimum J should be in the form:

$$J = \sum_{i=1}^{M} \left(\eta\lambda_{\mathbf{H}}^i + 1\right)\lambda_{\mathbf{G}^\dagger}^i\lambda_i, \tag{23}$$

which can be obtained by choosing

$$\mathbf{U} = \mathbf{U}_{\mathbf{G}}, \tag{24}$$

where $\lambda_{\mathbf{H}}^i$ and $\lambda_{\mathbf{G}^\dagger}^i$ are the M non-zero eigenvalues of $\mathbf{H}\mathbf{H}^H$ and $\left(\mathbf{G}\mathbf{G}^H\right)^\dagger$, respectively. Note that an ordering of $\lambda_{\mathbf{G}^\dagger}^i$ corresponds to the reverse ordering of $\lambda_{\mathbf{G}}^i$.

To obtain the maximum C in (19), we should follow three steps:

Step 1: We calculate J in equation (23) as a function of λ_i for every ordering of $\lambda_{\mathbf{H}}^i$ and $\lambda_{\mathbf{G}^\dagger}^i$[3]. Thus we obtain $M!$ expressions for J.

Step 2: For every expression for J, we evaluate (19). Then the capacity becomes a function of λ_i. We then calculate the maximum value of this function. There are $M!$ maximum values of C and we denote each one of them as $C_{\max}^i$.

Step 3: The final optimal C_{opt} is obtained as $\max\left(C_{\max}^1, \cdots, C_{\max}^{M!}\right)$.

Though complicated in the calculation process, the underlying ideas behind these steps are very simple: (a) to optimally match the source to relay and relay to destination eigenmodes; (b) to find the optimal power allocation at the relay transmitters based on these matched eigenmodes. A closed form solution for each value of $C_{\max}^i$ might be extremely complicated, as we shall first obtain the optimum relation between each λ_i by solving the M differential equations $\partial C/\partial\lambda_i = 0$. In practice we can calculate $C_{\max}^i$ numerically (i.e. by `fminbnd` function in Matlab). The calculation complexity for J is $M!$, which is also extremely high for $M > 2$. However, this approach gives us the theoretical capacity upper bound for MIMO coherent relay channels when using hybrid relaying schemes.

[3] We conjecture from our simulation results that arranging $\lambda_{\mathbf{H}}^i$ and $\lambda_{\mathbf{G}}^i$ both in decreasing order results in the optimal C

We know that in order to maximize C the matrix $\mathbf{GW}$ should be in the form of $\mathbf{U}_{\mathbf{G}}\mathbf{D}\mathbf{U}_{\mathbf{H}}^H$. The MIMO relay channel can thus be decomposed into several parallel channels each with gain $\lambda_{\mathbf{H}}^i\omega_i$, where $\lambda_{\mathbf{H}}^i$ is the source to relay channel gain and $\omega_i = \rho\lambda_i/(1+\rho\lambda_i)$ can be regarded as the relay to destination channel gain. ω_i is optimized by the weight matrix $\mathbf{W}$ under the power constraint at the relay. An visual description for this process is shown in Fig. 2. Based on this discovery, we now propose two practical suboptimal hybrid relaying schemes.

C. Suboptimal Hybrid Relaying Schemes

1) Modified Analogue Relaying (MAR): One simple way to make $\mathbf{GW}$ have the form of $\mathbf{U}_{\mathbf{G}}\mathbf{D}\mathbf{U}_{\mathbf{H}}^H$ is to make

$$\mathbf{W} = \mathbf{V}_{\mathbf{G}}\mathbf{U}_{\mathbf{H}}^H. \tag{25}$$

Then $\mathbf{GW} = \mathbf{U}_{\mathbf{G}}\mathbf{D}_{\mathbf{G}}\mathbf{U}_{\mathbf{H}}^H$ and $\mathbf{GWH} = \mathbf{U}_{\mathbf{G}}\mathbf{D}_{\mathbf{G}}\mathbf{D}_{\mathbf{H}}\mathbf{V}_{\mathbf{H}}^H$. The capacity can be expressed as:

$$C = 0.5\sum_{i=1}^{M}\log_2\left(1+\eta\lambda_{\mathbf{H}}^i\frac{\rho\lambda_{\mathbf{G}}^i}{\rho\lambda_{\mathbf{G}}^i+1}\right). \tag{26}$$

$$= 0.5\sum_{i=1}^{M}\log_2\left(1+\eta\lambda_{\mathbf{H}}^i\frac{\eta M\lambda_{\mathbf{G}}^i}{\eta M\lambda_{\mathbf{G}}^i+\eta\sum_{m=1}^{M}\lambda_{\mathbf{H}}^m+L}\right) \tag{27}$$

Compared with (18), *MAR* simply replaces λ_i with $\lambda_{\mathbf{G}}^i$. To maximize C, the columns of $\mathbf{U}_{\mathbf{H}}$ can be ordered to make $\lambda_{\mathbf{H}}^1 \leq \cdots \leq \lambda_{\mathbf{H}}^i \leq \cdots \leq \lambda_{\mathbf{H}}^M$; the columns of $\mathbf{V}_{\mathbf{G}}$ can be ordered in a similar fashion. The amplifying factor ρ can be written as:

$$\rho = \frac{M\eta}{\eta\left\|\mathbf{V}_{\mathbf{G}}\mathbf{U}_{\mathbf{H}}^H\mathbf{H}\right\|_F^2 + \left\|\mathbf{V}_{\mathbf{G}}\mathbf{U}_{\mathbf{H}}^H\right\|_F^2} = \frac{M\eta}{\eta\left\|\mathbf{H}\right\|_F^2 + L} \tag{28}$$

where we use the identity

$$\|\mathbf{UA}\|_F = \|\mathbf{A}\|_F \tag{29}$$

for any unitary matrix $\mathbf{U}$. This is the same value for ρ as in *AR*. We thus denote this scheme as modified analogue relaying. We also note that J and τ for *MAR* have the same form as in *AR* by employing (29). Compared with *AR*, it can be seen from (26) that the relay is able to decompose the channels and coordinate the backward channels with the forward channels to optimize the sum capacity for M parallel data streams.

2) Modified Matched Filter Relaying (MMFR): As in *MAR*, we design a new $\mathbf{W}$ based on the matched filter weight matrix $\mathbf{H}^H$. If we make $\mathbf{W} = \tilde{\mathbf{W}}\mathbf{H}^H$ and write the following:

$$\mathbf{GW} = \mathbf{U_G D_G V_G^H \tilde{W} V_H D_H^H U_H^H}. \tag{30}$$

we can see that to make $\mathbf{M}$ have the form $\mathbf{U_G D U_H}$, we can make

$$\tilde{\mathbf{W}} = \mathbf{V}_{\mathbf{G},M}\mathbf{V}_{\mathbf{H}}^H, \tag{31}$$

and $\mathbf{D}$ becomes $diag\left(d_{\mathbf{G}}^1\lambda_{\mathbf{H}}^1, \cdots, d_{\mathbf{G}}^i\lambda_{\mathbf{H}}^i, \cdots, d_{\mathbf{G}}^M\lambda_{\mathbf{H}}^M\right)$, where $d_{\mathbf{G}}^i$ are the singular values of $\mathbf{G}$ and $\mathbf{V}_{\mathbf{G},M}$ has been introduced in section IV.A. The capacity can be written as

$$C = 0.5\sum_{i=1}^{M}\log_2\left(1 + \eta\lambda_{\mathbf{H}}^i\frac{\rho\lambda_{\mathbf{H}}^i\lambda_{\mathbf{G}}^i}{\rho\lambda_{\mathbf{H}}^i\lambda_{\mathbf{G}}^i + 1}\right). \tag{32}$$

By comparing (32) with (18), we see that *MMFR* simply replaces λ_i with $\lambda_{\mathbf{H}}^i\lambda_{\mathbf{G}}^i$. To maximize C, the columns of $\mathbf{V_H}$ can be ordered to make $\lambda_{\mathbf{H}}^1 \leq \cdots \leq \lambda_{\mathbf{H}}^i \leq \cdots \leq \lambda_{\mathbf{H}}^M$; the columns of $\mathbf{V}_{\mathbf{G},M}$ can be ordered in a similar fashion. We also note that J, ρ and τ for *MMFR* have the same form as in *MFR* by employing (29). Compared with *MAR*, this scheme has the advantage of enhancing the SNR at the relay.

It should be noted that when M reduces to 1, equation (32) can be rewritten as:

$$C = 0.5 \times \log_2\left(1 + \eta\lambda_{\mathbf{H}}^1\frac{\eta\lambda_{\mathbf{G}}^1}{\eta\lambda_{\mathbf{G}}^1 + \eta\lambda_{\mathbf{H}}^1 + 1}\right). \tag{33}$$

It is not hard to see that (33) equals (19) if we replace J in (19) with the expression in (23) for $M = 1$. This means that *MMFR* becomes the optimal hybrid relaying scheme for $M = 1$. However, for $M \geq 2$, the signals become more correlated due to the matched filter factor $\mathbf{H}^H$ in the weight matrix $\mathbf{W}$, which impairs the sum capacity.

As we mentioned before, compared with analogue relaying, hybrid relaying has to decode some training sequence (or feedback) to obtain the CSI, and needs a filter at the relay to refine the messages. However, it can be seen in this section that by filtering the signals at the relays, the MIMO relay channel can be decomposed to several independent parallel channels, This will significantly reduce the detection complexity at the destination compared with analogue relaying, as each streams can be detected separately in parallel and no non-linear detector such as V-BLAST is required.

D. Comparison of Relaying Schemes for Single Relay Channels

We calculate the average Shannon capacity (in bits per channel use) for 1000 channel realizations and we define $P = M\eta$ as the total transmit power at the source. For simplicity, we neglect the training interval in the capacity calculation, assuming that the maximum channel Doppler frequency is much less than the signalling frequency. Some examples for the impact of training on MIMO capacity performance can be found in [34] and [35]. Its specific impact on MIMO relay channels is beyond the scope of this work. Fig. 3 shows the performance results for different relaying schemes for either coherent or non-coherent relay channels when $L = M$. It can be seen that digital schemes offer the best capacity performance. For either coherent channels or non-coherent channels, *AR (MAR)* outperforms *MFR (MMFR)*, especially for higher SNR values. This implies that weakening the noise at the relays cannot compensate for the disadvantage of signal correlation in *MMFR* for $L = M$. It can also be seen that the relaying schemes designed for coherent relay channels give only a small performance advantage over those for non-coherent relay channels. In particular, the performance of digital relaying schemes (i.e. *DR*, capacity upper bound of *DR* (*DRUB*), capacity upper bound for *MDR* (*MDRUB*)) are all about the same. The reason has been stated in section III.A and IV.A..

However, the situation is different for $L > M$ and Fig. 4 shows the simulation results for this case. It can be seen that unlike the $L = M$ case, relaying schemes for coherent channels offer a significant advantage over schemes for non-coherent channels by exploiting $\lambda_{\mathbf{G}}^i$ on the relay to destination channels. For non-coherent relay channels, *MFR* outperforms *AR* and approaches the performance of *DR*. Since signal correlation becomes less important as the ratio L/M increases, the value of τ for *MFR* in (12) increases compared to that for *AR* in (8). The upper bound for digital relaying (*DRUB*) is still close to the performance of *DR*, the reason for this has been given in section III.A.. For coherent relay channels, digital relaying schemes still perform best; however, their advantage over modified hybrid relaying schemes is smaller than for $L = M$. As previously discussed, *MMFR* is the optimal hybrid scheme for $M = 1$. For $M > 1$, it can be seen that both *MAR* and *MMFR* approaches the optimal hybrid scheme as the ratio L/M increases.

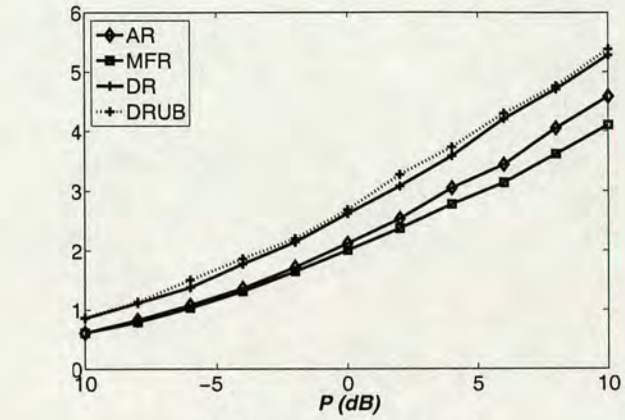
From the above discussion we conclude that by increasing the number of antennas at the relays and obtaining the forward CSI of the relay channels, we can significantly increase the network capacity, especially for the digital and hybrid schemes. The hybrid schemes give a closer performance to the digital schemes when $L > M$ and are attractive if we consider the tradeoff between performance and complexity.

V. Routing for Multiple-Relay Channels

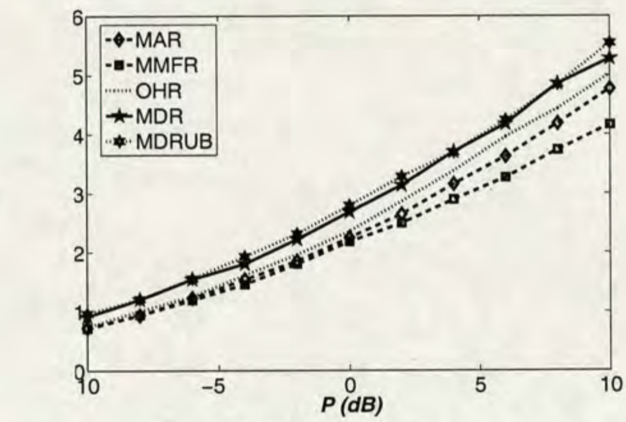
In this section we extend the discussion to multiple relay channels. We assume that the relays do not communicate with each other, which is the most practical case. We will discuss and compare the capacity performance of two routing schemes based on different relay selection criteria. In contrast to *cooperative diversity* as defined in the literature [11], [12], we defined the diversity achieved by these schemes as *selection diversity*. We will compare the the performance of selective routing with that of multi-cast routing where all the relays are used to forward the message, and a relay cooperation protocol exploiting *cooperative diversity* in the networks. We will show the advantage of selective routing over the other two schemes in certain scenarios.

A. Optimal Selective Routing (OSR)

We choose the single best relay through which the highest network capacity can be obtained. We denote the capacity for the kth single relay channel as C_k, where only relay k is used.

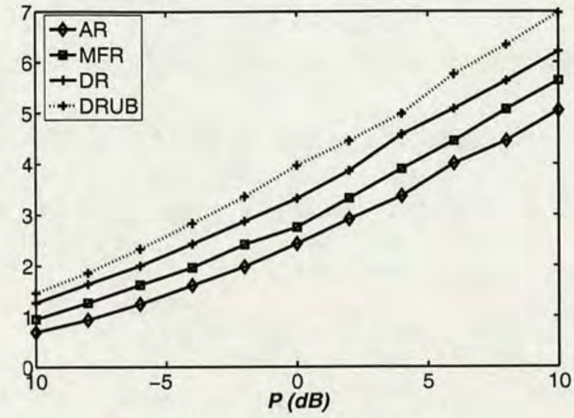


(a) Non-coherent relay channel

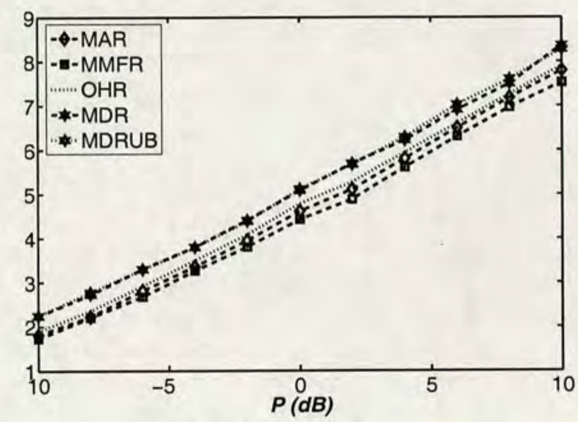


(b) Coherent relay channel

Fig. 3. Average capacity of single MIMO relay channels when $L = M = 2$. (a) Non-coherent relay channel. (b) Coherent relay channel.



(a) Non-coherent relay channel



(b) Coherent relay channel

Fig. 4. Average capacity of single MIMO relay channels when $M = 2$, $L = 8$. (a) Non-coherent relay channel . (b) Coherent relay channel.

The network capacity can be expressed as:

$$C = \max\left(C_1, \cdots, C_k, \cdots, C_K\right). \tag{34}$$

We can apply selective routing for analogue relaying, hybrid relaying and digital relaying, where for each relaying scheme C_k is calculated according to the capacity formula provided in the previous two sections. In practice, the channel capacity of each relay link must be fed back to the source. The source then decides which relay to choose. This scheme should be optimal in the sense that it selects the single best relay channel that maximizes the network capacity. However, this scheme might require extremely high signaling overhead since for every channel realization, all K relay channel capacities have to be tested and compared before the best relay is chosen. This might only be practical for a very slow fading environment.

B. Pathloss and Shadowing based Selective Routing (PSSR)

Instead of calculating C for each single relay route, in this scheme the relay is chosen with respect to only pathloss and shadowing coefficients of the channels, i.e. we neglect $\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{G}}_k$ which contain the Rayleigh fading coefficients. In the following we will give detailed explanation of this routing scheme and show that it approaches the performance upper bound of *OSR* for large values of L or M.

By the law of large numbers [36], we have the following identity for $L \times M$ matrix $\tilde{\mathbf{H}}_k$:

$$\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H \overset{M\to\infty}{\rightarrow} M\mathbf{I}_L, \tilde{\mathbf{H}}^H\tilde{\mathbf{H}} \overset{L\to\infty}{\rightarrow} L\mathbf{I}_M.$$

Therefore, it is not difficult to see that for large values of both M and L the capacity upper bound of each relay channel for both *DR* and *MDR* becomes[4]

$$C = 0.5\min\left(M\log\left(1+\eta M\alpha_k\right), M\log\left(1+\eta M\beta_k\right)\right)$$

Therefore we set the selection criterion as follows:

$$\begin{aligned}\Re_i =& \arg\max_{\Re_i\in\{1,2,\cdots,K\}} \left(M\log_2\left(1+\eta\min\left(M\alpha_{\Re_i}, M\beta_{\Re_i}\right)\right)\right)\\ =& \arg\max_{\Re_i\in\{1,2,\cdots,K\}} \left(\min\left(\alpha_{\Re_i}, \beta_{\Re_i}\right)\right).\end{aligned} \tag{35}$$

For hybrid relaying (analogue relaying) schemes, note that for large value of L, $\lambda^i_{\mathbf{H}}$ and $\lambda^i_{\mathbf{G}}$ tend towards the value L. From the analysis of section IV.B. it is not difficult to see that $\mathbf{W} = \mathbf{I}$ leads to the optimal hybrid relaying solution, which is equal to *MAR* in this scenario. Therefore equation (11) can

[4]For large L only, the capacity for *DR* is $C = 0.5\min\left(M\log\left(1+\eta L\alpha_k\right), M\log\left(1+\eta M\beta_k\right)\right)$ and for *MDR* is $C = 0.5\min\left(M\log\left(1+\eta L\alpha_k\right), M\log\left(1+\eta L\beta_k\right)\right)$.

be modified as follows for large L and M:

$$C = 0.5M\log_2\left(1+\eta M \times \frac{M\eta\alpha\beta}{\eta M\left(\alpha+\beta\right)+1}\right). \quad (36)$$

The selection criterion is thus:

$$\Re_i = \arg\max_{\Re_i\in\{1,2,\cdots,K\}}\left(\frac{\alpha_{\Re_i}\beta_{\Re_i}}{\eta M\left(\alpha_{\Re_i}+\beta_{\Re_i}\right)+1}\right). \quad (37)$$

Compared with the *OSR* scheme, *PSSR* doesn't require knowledge of $\tilde{\mathbf{H}}_k$ and $\tilde{\mathbf{G}}_k$. Since the calculation is based on shadowing coefficients which will change much more slowly than the fading channel realizations, fewer routing updates are required. This will significantly reduce signalling and computation overhead compared to *OSR*.

Fig. 5 gives some examples of the capacity performance for *OSR* and *PSSR* schemes as the number of relays increases, while P is set to 0dB. We can see that PSSR gives very good performance even for small values of M and L. As the values of M and L both increase, the absolute capacity gap between the *OSR* and *PSSR* schemes for the same number of relays becomes smaller, and the percentage gap shrinks. Also, for a fixed value of M, increasing L can also decrease the capacity gap (e.g. $M=2, L=4$). These observations also indicate that MIMO configurations can significantly reduce the variability of the instantaneous channel capacity caused by Rayleigh fading.

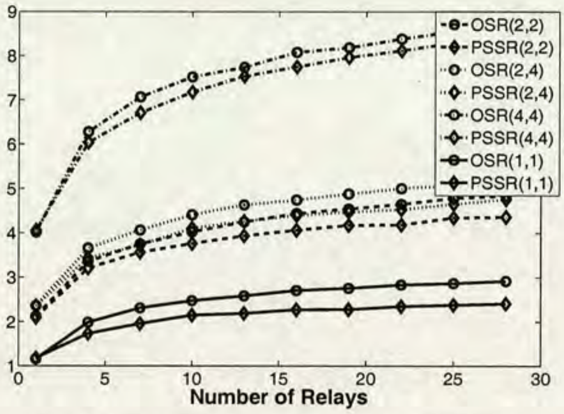
C. Multi-cast Routing (MR)

We also give a short discussion on possible multi-cast routing schemes, where all the relays are used to forward the message. For digital relaying, every relay decodes the signals it received, and forwards them to the destination. The destination receives the signals from all the relays and decodes them, so the effective MIMO channel from relays to destination can be written as $\mathbf{G}_{sum} = \sum_{i=1}^{K}\mathbf{G}_k$. For the destination to decode the signals correctly, each relay has to decode the signals correctly. For each signal s_m, we denote the Shannon capacity from source to relay k by $C_{m,\mathbf{H}_k}$, and from the relays to the destination by $C_{m,\mathbf{G}_{sum}}$. The Shannon capacity of the network for s_m can be expressed by:

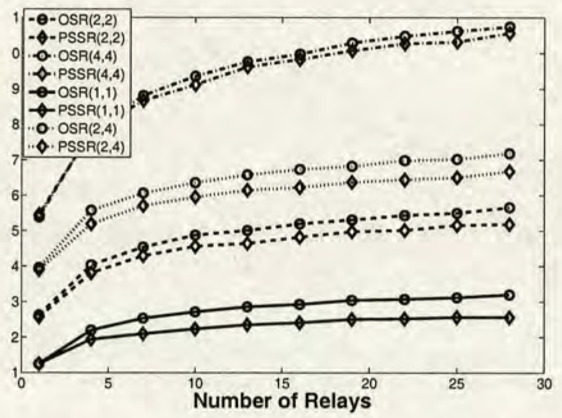
$$C_m = \min\left(\min\left(C_{m,\mathbf{H}_1},\cdots,C_{m,\mathbf{H}_K}\right), C_{m,\mathbf{G}_{sum}}\right). \quad (38)$$

The network capacity for all the M signals are $C=\sum_{m=1}^{M}C_m$. It can be seen from (38) that the capacity for each signal is constrained by the worst channel among the K source to relay channels and the relays to destination channel. As the number of antennas increases, the capacity for each signal will reduce. Therefore, the *MR* scheme turns out to be the poorest routing choice in this scenario. This observation also explains why the source to relay link is often tested before relaying the message in most of the existing literature discussing digital relaying (e.g. [11], [12]).

For hybrid or analogue relaying schemes, the destination receives the filtered or amplified signals from all the relays available in the network. The source to destination channel can be written as:



(a) Analogue relaying for non-coherent relay channels.



(b) Modified digital relaying for coherent relay channels.

Fig. 5. Comparison of average capacity of multiple-relay channels for different selective routing schemes for different antenna number allocations (M,L). The circled marked curves denote the optimal selective routing, and the diamond marked curves denote the pathloss and shadowing based selective routing. $P = 0dB$. (a) non-coherent relay channels. (b) coherent relay channels. Note that similar curves are obtained for digital and hybrid relaying in both cases

$$\mathbf{y} = \sqrt{\eta}\left(\sum_{k=1}^{K}\sqrt{\rho_k}\mathbf{G}_k\mathbf{W}_k\mathbf{H}_k\right)\mathbf{s} + \left(\sum_{k=1}^{K}\sqrt{\rho_k}\mathbf{W}_k\mathbf{G}_k\mathbf{n}_k + \mathbf{n}_d\right) \quad (39)$$

where ρ_k is the power amplifying factor for relay k, which is the $1/K$ times the value for single relay channels expressed by (9).

Note that the relay channels can not be made orthogonal to each other due to the different singular vectors for each relay channel when $M>1$ and $L>1$, unless each relay obtains the knowledge of the other $2\times(K-1)$ backward and forward channel matrices (i.e. the relays cooperate). However, this would involve extremely large signalling overheads. How to choose the proper weight at each relay to suppress the co-channel interference thus remains an open topic. [17] gives an suboptimal scheme which requires no joint detection at the destination. However, when $M=1$, it is possible to combine the signals effectively at the destination, interested readers can refer to [37].

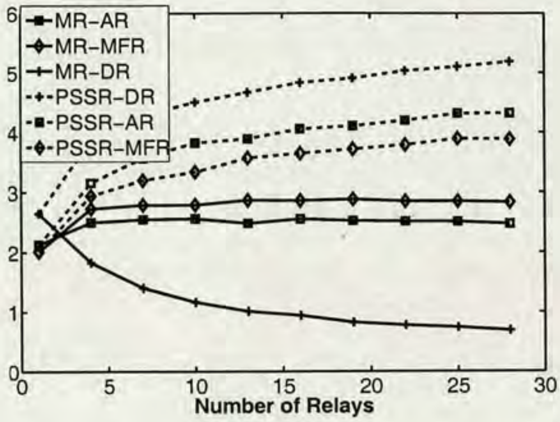


Fig. 6. Average capacity of non coherent multiple-relay channels for multi-cast routing and shadowing based selective routing shemes, $P = 0dB$, $L = M = 2$. Similar curve behavior can be found for modified relaying schemes for coherent multiple relay channels.

If we simply choose $\mathbf{W}_k$ designed for single relay channels as those for multiple relay channels, each product $\mathbf{G}_k\mathbf{W}_k\mathbf{H}_k$ and $\mathbf{G}_k\mathbf{W}_k$ are i.i.d. random matrices. Therefore, for large K the contribution of the equivalent multi-cast MIMO channel to the capacity becomes the average over all single K relay channels. So the capacity of multi-cast routing schemes will be upper bounded by that of the selective routing schemes, for which only the best (near best for *PSSR*) relay channel is chosen. This trend is shown in Fig. 6 where *PSSR* is compared to *MR*. For *MR* schemes, the signals becomes less correlated at the destination for larger number of relays. So matched filter based relaying outperforms analogue relaying due to the reduction in the amplified noise generated at the relays.

D. Selection Diversity vs. Cooperative Diversity

We now compare selective routing with a protocol that exploits the *cooperative diversity* of the relay channels for digital relaying, where two relays are used. This comparison can also be regarded as a specific example of a comparison between the benefits of *cooperative diversity* and *selection diversity*. It has been shown theoretically in [12] that full *cooperative diversity* can be achieved by applying space-time distributed codes in a single antenna scenario. A practical example to achieve full diversity is given in [38], where Alamouti codes [32] are applied at the relays; the same diversity gain as in [12] can be achieved provided that the relays are selected to be able to decode the signals correctly.

We simply extend [38] to a multi-antenna scenario and refer to this as space-time block coded relaying (*STBCR*). The source uses one antenna to transmit signals s_1 and s_2 for the first two symbol slots. The two relays use multiple antennas to decode the two signals (i.e. maximum ratio combining) and re-send them using *STBC* encoding. The destination then decodes the signals by multiplying the received signal vector by the corresponding *STBC* weight matrix. It will be more difficult to apply the STBC encoding to noisy analogue waveforms in analogue or hybrid relaying, so they are not discussed in the paper. Note that unlike the analysis on distributed space-time coding in [12], by applying STBC we only need to apply repetition coding rather than i.i.d. random coding at the relay for the two relay network. This is much more convenient to implement in practice [11], [12].

After decoding and re-encoding, relay 1 uses one antenna to transmit $[s_1, -s_2^*]$ and relay 2 transmits $[s_2, s_1^*]$ over two symbol durations. Assuming the channel stays constant for the two symbol transmission periods, the capacity of the relay to destination channels for each signal can be written as

$$C_{r,d}^{STBC} = 0.5\log_2\left(1 + 0.5\eta\sum_{i=1}^{2}\sum_{m=1}^{M}|g_{m,i}|^2\right), \qquad (40)$$

where $g_{m,i}$ are the channel coefficients from relay i to antenna m at the destination. The factor 0.5 right before η denotes the half power scaling factor for each relay. For optimal selective routing, the destination uses maximum ratio combining to enhance the received signal power. The capacity can then be written as:

$$C^{DSR} = \max\left(\min\left(C_{s,r_1}, C_{r_1,d}\right), \min\left(C_{s,r_2}, C_{r_2,d}\right)\right). \qquad (41)$$

The capacity C_{s,r_i} is the capacity of the channel between the source and relay i. The capacity $C_{r_i,d}$ is the capacity of the channel between relay i and the destination and can be written as:

$$C_{r_i,d} = \log_2\left(1 + \eta Q_i\right), \qquad (42)$$

where $Q_i = \sum_{m=1}^{M}|g_{m,i}|^2$. Note that equations (41) and (42) also hold for $M = 1$, which is the single antenna scenario. Without loss of generality, if we assume that $Q_1 \geq Q_2$, it can be seen that

$$C_{r_1,d} \geq C_{r,d}^{STBC}. \qquad (43)$$

It follows that

$$C^{DSR} \geq \min\left(C_{s,r_1}, C_{r_1,d}\right) \geq \min\left(C_{s,r_1}, C_{s,r_2}, C_{r,d}^{STBC}\right). \qquad (44)$$

The right hand side of this inequality is the capacity of each signal for the *STBC* relaying scheme. Therefore optimal selective relaying actually outperforms *STBC* relaying in this example. Specifically, inequality (43) suggests that selection diversity can offer a better *power gain* over cooperative diversity if there is a power constraint at the relays. A higher capacity can be achieved if we give all the transmission power to the single best relay instead of splitting it equally among different relays, even if full cooperative diversity can be achieved at the relay to destination link. Inequality (44) implies that selection diversity might offer more diversity gain if the relays are randomly chosen. This is because there is only one source to relay link being considered in the minimum function for selective routing; but for space-time coded relaying, all the source to relay links have to be considered in the capacity function due to its multi-casting nature. This also confirms the previous work for cooperative diversity (e.g. [11], [12]), where it is often suggested that the relays need to be chosen to be able to correctly decode the messages so that the source-to-relay capacity C_{s,r_i} can be removed from the network capacity constraint function. In fact, is has been proved in [11] that no diversity gain can be obtained if the relays are chosen totally at random, selection diversity is clearly an advance in this sense. Finally, we note that these two inequalities (43) and (44) can be generalized to the scenarios where direct link is included and the number of relays is more than 2 (e.g. the cooperative diversity schemes discussed in [12] when random coding is applied).

VI. Spatial Multiplexing vs. Single Signal Beamforming

So far we have mainly concentrated on spatial multiplexing systems where the spatial multiplexing gain M is fully achieved for MIMO communications. In this section we discuss another MIMO configuration which exploits the spatial diversity of MIMO channels. We refer to this as *single signal beamforming* [30], [31]. The transmitter uses all of its antennas to transmit one signal instead of multiplexing different signals simultaneously. The weights for transmitter and receiver are chosen as the columns of right and left singular vectors of channel corresponding to the largest singular value of the MIMO channel matrix. The channel gain at the receiver is the square of the largest singular value of channel matrix which can achieve full diversity of $L \times M$. It should be noted that this scheme requires CSI at the transmitter; thus, it can only be applied to coherent channels. Furthermore, the source to relay CSI is required at the source when $M > 1$.

If we apply single signal beamforming to MIMO relay channels, we can also define digital beamforming relaying (*DBR*) and hybrid beamforming relaying (*HBR*). For *DBR*, the relay decodes the signal, re-encodes it, weights it and forwards it to the destination. The capacity of a single relay channel can be written as:

$$C = 0.5 \times \log_2\left(1 + M\eta \times \min\left(\lambda_{\mathbf{H}}^{\max}, \lambda_{\mathbf{G}}^{\max}\right)\right) \tag{45}$$

where $\lambda_{\mathbf{H}}^{\max}$ and $\lambda_{\mathbf{G}}^{\max}$ are the largest eigenvalues of $\mathbf{H}\mathbf{H}^H$ and $\mathbf{G}\mathbf{G}^H$. For *HBR*, after multiplying the received signal vector by the corresponding left singular vector of the channel $\mathbf{H}$, the relay amplifies, weights and forwards the signal to the destination. The capacity for *HBR* scheme can be expressed as:

$$\begin{aligned} C =& 0.5\log_2\left(1 + M\eta\lambda_{\mathbf{H}}^{\max}\frac{\rho\lambda_{\mathbf{G}}^{\max}}{1+\rho\lambda_{\mathbf{G}}^{\max}}\right) & (46)\\ =& 0.5\log_2\left(1 + M\eta\lambda_{\mathbf{H}}^{\max}\frac{\eta M\lambda_{\mathbf{G}}^{\max}}{1+\eta M\lambda_{\mathbf{H}}^{\max}+\eta M\lambda_{\mathbf{G}}^{\max}}\right) & (47)\end{aligned}$$

where ρ is the amplifying factor at the relay:

$$\rho = \frac{M\eta}{1 + M\eta\lambda_{\mathbf{H}}^{\max}}. \tag{48}$$

It should be noted that *HBR* reduces to *MMFR* when $M = 1$.

We first compare *DBR* with *MDR* for spatial multiplexing systems with source to relay CSI available at the source. For *MDR*, the source can apply the same beamforming operations as at the relay. Thus $\psi_{\mathbf{H}}^i$ is replaced by $\lambda_{\mathbf{H}}^i$ in equation (13). We find that *DBR* outperforms *MDR* for low output SNR values at the destination receiver. This is because $\log_2\left(1 + SNR_{receiver}\right) \approx SNR_{receiver}$ for small values of $SNR_{receiver}$. The capacity of *MDR* then becomes $0.5\eta \times \sum_{i=1}^{M} \min\left(\lambda_{\mathbf{H}}^i, \lambda_{\mathbf{G}}^i\right)$ for low output SNR, and this is smaller than $0.5M\eta \times \min\left(\lambda_{\mathbf{H}}^{\max}, \lambda_{\mathbf{G}}^{\max}\right)$, which approximates the capacity of *DBR* expressed in (45) for low output SNR.

For hybrid relaying schemes, we can also find that *HBR* outperforms *MAR* in a similar way at low transmit power levels. It can be proved that (27) is smaller than (47) for low $SNR_{receiver}$ with the inequalities:

$$\begin{aligned} &\sum_{i=1}^{M}\eta\lambda_{\mathbf{H}}^i\frac{\eta M\lambda_{\mathbf{G}}^i}{L+\eta\sum_{m=1}^{M}\lambda_{\mathbf{H}}^i+\eta M\lambda_{\mathbf{G}}^i} & (49)\\ \le&\sum_{i=1}^{M}\eta\lambda_{\mathbf{H}}^i\frac{\eta M\lambda_{\mathbf{G}}^{\max}}{1+\sum_{m=1}^{M}\eta\lambda_{\mathbf{H}}^i+\eta M\lambda_{\mathbf{G}}^{\max}} & (50)\\ \le& M\eta\lambda_{\mathbf{H}}^{\max}\frac{M\times\eta M\lambda_{\mathbf{G}}^{\max}}{1+\eta M\lambda_{\mathbf{H}}^{\max}+\eta M\lambda_{\mathbf{G}}^{\max}}. & (51)\end{aligned}$$

Generally, the performance gain of relaying over non-relaying for either spatial multiplexing or single signal beamforming configurations can be expressed as follows:

$$G \approx \frac{0.5\log_2\left(1+\kappa SNR_{receiver}\right)}{\log_2\left(1+SNR_{receiver}\right)} \tag{52}$$

where $SNR_{receiver}$ is the receive SNR for the direct link and κ denotes the link gain due to relaying (restricted to the worse link between the source-relay and the relay-destination links). It is easy to see that $G \approx \kappa/2$ when $SNR_{receiver} \to 0$; and $G \approx 0.5$ when $SNR_{receiver} \to +\infty$. This means that relaying should be used for low SNR scenarios. When the receive SNR for direct link transmission is high, the benefit for increasing the link reliability by relaying will not compensate for its bandwidth loss of 0.5 due to the half duplex nature of the relay. Therefore, considering the performance advantage of single signal beamforming over spatial multiplexing for low receive SNR, it is better to combine single signal beamforming with relaying when multiple antennas are applied at each node in this sense. However, if we increase the number of antenna at the relays, or apply relay selection in the networks, the receive SNR in every point-to-point link in the relay network (i.e. source to relay link or relay to destination link) will be improved. In this scenario, spatial multiplexing will outperform the single signal beamforming and can still be the preferred choice to be combined with relaying.

The simulation results confirms this discussion. Figure 7 gives the simulation results for single relay channels for different values of (M, L). It can be seen that at low values of η single signal transmission schemes outperform spatial multiplexing schemes. However, this advantage is weakened when a larger number of antennas L are deployed at the relay. This is because larger values of L can enhance the output SNR at the destination receiver by increasing $\lambda_{\mathbf{H}}^i$ and $\lambda_{\mathbf{G}}^i$. Increasing the number of relays can also result in a larger performance improvement for spatial multiplexing compared with single signal transmission. In the multiple relay scenario there is more freedom to obtain large values of $\lambda_{\mathbf{H}}^i$ and $\lambda_{\mathbf{G}}^i$ by selecting the proper relay. This trend is clearly shown in Fig. 8 where *PSSR* schemes are applied.

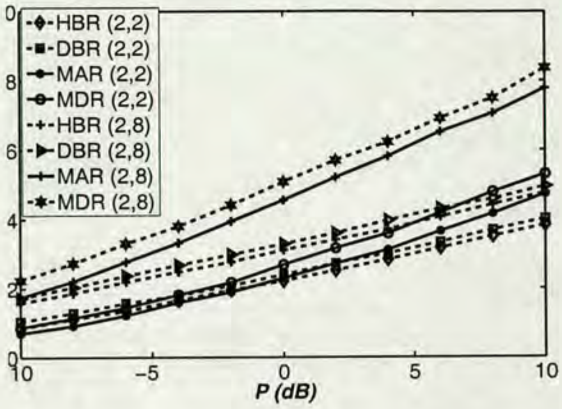


Fig. 7. Average capacity of MIMO single relay channels for single signal beamforming schemes and spatial multiplexing schemes when $M = L = 2$ (2,2) and $M = 2$, $L = 8$ (2,8).

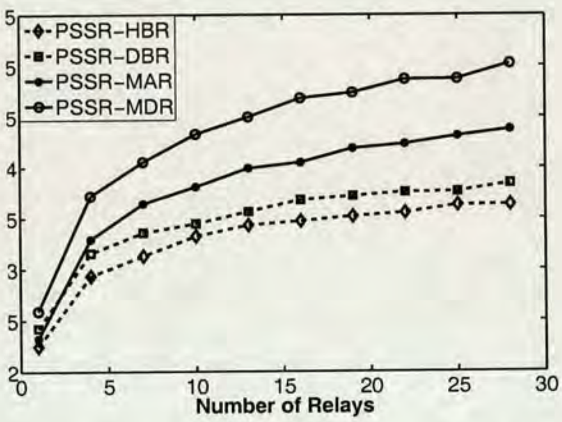


Fig. 8. Average capacity of MIMO multiple relay channels for single signal beamforming and spatial multiplexing systems, where shadowing based selective routing schemes are used. $M = L = 2$, $P = 0dB$.

VII. Conclusions and Future Work

In this paper we study MIMO spatial multiplexing configurations for relay channels. For single relay channels, we show that increasing the number of antennas at the relay can be beneficial. First it will make the proposed hybrid relaying schemes good suboptimal choices for either coherent or non-coherent relay channels, when compared with digital relaying schemes. Secondly, it is shown that in such scenarios knowledge of forward CSI at the relay can help improve the network capacity significantly, if we exploit and coordinate the backward and forward channel eigenmodes at the relay. For multiple relay channels, selective routing schemes will give better performance than multi-cast or cooperative diversity schemes when the total transmit power at the relays is fixed. Then multiple relay channels can be simplified to single relay channels where only one relay is used. In this case, the conclusions for single relay channels in the paper can also be applied to multiple relay channels. We also compare the spatial multiplexing solutions with single signal beamforming solutions which exploiting the diversity of a MIMO link. For low receive SNR values, the single signal beamforming schemes perform better than spatial multiplexing schemes and can be preferred choices to be combined with relaying.

Some interesting topics for future work include: (a) effective signalling for digital and hybrid relaying when only partial forward CSI is available at the relay for single relay channels (e.g. semi-coherent relay channels); (b) effective signal combining at the destination when the effect of the direct link is taken into account; (c) effective signalling for digital and hybrid relaying when multi-cast routing is used.

References

[1] D. Gesbert, M. Shafi, S. Da-shan, P. J.Smith, and A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 281–302, April 2003.

[2] A. Goldsmith *et al.*, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 684–702, June 2003.

[3] W. Rhee and J. M. Cioffi, "On the capacity of multiuser wireless channels with multiple antennas," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, Oct. 2003.

[4] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1691-1706, July 2003.

[5] S. Vishwanath, N. Jindal, and A. J. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels", *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[6] N. Jindal *et al.*, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[7] Q. H. Spencer *et al.*, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, no. 2, Feb. 2004.

[8] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, Sept. 2004.

[9] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks wireless communications," *IEEE Personal Communications*, vol. 9, no. 4, pp. 8–27, Aug. 2002.

[10] R. Pabst *et al.*, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, vol. 42, no. 9, pp. 80–89, Sept. 2004.

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

[12] 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.

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

[14] H. Bolcskei and R. U. Nabar, "Realizing MIMO gains without user cooperation in large single-antenna wireless networks," in *Proc. ISIT 2004*.

[15] V. I. Morgenshtern, H. Bolcskei, and R. U. Nabar, "Distributed orthogonalization in large interference relay networks," in *Proc. ISIT 2005*.

[16] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

[17] H. Bolcskei *et al.*, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, June 2006.

[18] O. Oyman and A. J. Paulraj, "Energy efficiency in MIMO relay networks under processing cost," in *Proc. Conf. Inf. Science and Systems*, Johns Hopkins University, March 16-18, 2005.

[19] S. T. Chung *et al.*, "Approaching the MIMO capacity with a low-rate feedback channel in V-BLAST," *EURASIP J. App. Sig. Proc.*, vol 5, pp. 762–771, 2004.

[20] H. El Gamal, G. Caire, and M. O. Damen, "The MIMO ARQ channel: diversity-multiplexing-delay tradeoff," *IEEE. Trans. Wireless Commun.*, vol. 52, no. 8, pp. 3601–3619, Aug. 2006.

[21] H. Zheng, A. Lozano, and M. Haleem, "Multiple ARQ processes for MIMO systems," *EURASIP J. App. Sig. Proc.*, vol 5, pp. 772-782, 2004.

[22] D. Tse and V. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[23] S. Loyka and F. Gagnon, "Performance analysis of the V-BLAST algorithm: an analytical approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1326–1337, July 2004.

[24] H. Zhang, H. Dai, and B. L. Hughes, "On the diversity-multiplexing tradeoff for ordered SIC receivers over MIMO channels," in *Proc. IEEE ICC 2006*.

[25] F. R. Farrokhi *et al.*, "Link-optimal space-time processing with multiple transmit and receive Antennas," *IEEE Commun. Lett.*, vol. 5, no. 3, March 2001.

[26] A. M. Tulino and S. Verdu, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, 2004.

[27] Y. Fan and J. S. Thompson, "On the outage capacity of MIMO multihop networks," in *Proc. IEEE Globecom 2005*.

[28] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, March 1998.

[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[30] J. Bach Andersen, "Antenna arrays in mobile communications: gain, diversity, and channel capacity," *IEEE Antennas Propagat. Mag.*, vol. 42, no. 2, pp. 12–16, April 2000.

[31] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[32] S. M. Alamouti, "A simple transmit diverstiy technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[33] J. Dattorro, *Convex Optimization and Euclidean Distance Geometry*. Meboo Publishing USA, 2005.

[34] B. Hassibi and B.M. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inf. Theory*, vol. 49, np. 4, pp. 951–963, April 2003.

[35] T.L. Marzetta, "BLAST training: estimating channel characteristics for high capacity space-time wireless," in *Proc. 1999 Allerton Conference*, Monticello, IL, 1999.

[36] A. Papoulis, *Probability, Random Variables, and Stochastic Process*. New York: McGraw-Hill, 1984.

[37] Y. Fan, J. S. Thompson, A. B. Adinoyi, and H. Yanikomeroglu, "On the diversity multiplexing tradeoff for multi-antenna multi-relay channels," submitted to *IEEE Trans. Inf. Theory*, special issue on models, theory and codes for relaying and cooperation in communication networks.

[38] P. A. Anghel, G. Leus, and M. Kavehl, "Multi-user space-time coding in cooperative networks," in *Proc. ICASSP'03*, vol. 4, pp. 73-76.

Yijia(Richard) Fan received his BEng degree in electrical engineering from Shanghai Jiao Tong University (SJTU), Shanghai, P.R. China, in July 2003. Since October 2003, he has been a Ph.D. candidate at the Institute for Digital Communications, University of Edinburgh. His Ph.D. project is fully funded by Engineering and Physical Sciences Research Council (EPSRC), UK. His current research interests include signal processing and information theory for MIMO wireless systems and their applications in future generation wireless networks.

John S Thompson received his BEng and PhD degrees from the University of Edinburgh in 1992 and 1996, respectively. From July 1995 to August 1999, he worked as a postdoctoral researcher at Edinburgh, funded by the UK Engineering and Physical Sciences Research Council (EPSRC) and Nortel Networks. Since September 1999, he has been a lecturer at the School of Engineering and Electronics at the University of Edinburgh. In October 2005, he was promoted to the position of reader. His research interests currently include signal processing algorithms for wireless systems, antenna array techniques and multihop wireless communications. He has published approximately 100 papers to date, including a number of invited papers, book chapters and tutorial talks, as well as co-authoring an undergraduate textbook on digital signal processing. He is currently an editor-in-chief of *IEE Proceedings on Vision, Image and Signal Processing*, and is the technical programme co-chair for the IEEE International Conference on Communications (ICC) 2007, to be held in Glasgow.

Frequenz
59 (2005)
7-8

An Introduction to Multihop Multi-Antenna Communications

By John S Thompson, Peter M Grant and Yijia Fan

Abstract – Recently there has been significant interest in the use of multihop wireless links for use in mobile communications. In such a system, data is forwarded by one or more relays from the source to the destination, rather than simply via a direct wireless link. This paper gives an overview of the benefits of such methods to cellular wireless systems. In particular, the high data rate coverage of a wireless network can be increased by using these techniques. However, the use of such techniques does imply an increased signalling overhead, in order to set up and maintain multihop routes. This paper will also explain how smart antenna concepts can be used to improve further the performance of multihop links.

Index Terms – Wireless Relaying, Multihop, Antenna Array Processing, Multiple Input Multiple Output

1. Introduction

'he use of multiple antennas at both ends of a wireless link, alled multiple-input multiple-output (MIMO) technology, romises significant improvements in performance [1]. MIMO ystems can be used to perform spatial multiplexing [2], where nultiple data streams are sent in parallel from transmit (TX) ntennas to the receiver (RX). However, such a system performs est when the channel is subject to significant multipath scattering, so that the transmit antennas are spatially separable. An lternative scheme is space-time coding [3], where a single data tream is encoded and sent from several TX antennas to provide liversity at the receiver. Diversity means that several independent replicas of the TX signal are available at the RX, which educes variability in the channel due to multipath fading [4].

Recently, the introduction of MIMO techniques into ad hoc ıetwork or multihop cellular networks [5] has been studied [6]–11]. In multihop networks, information sent by a source termiıal is relayed via intermediate terminals to the destination. The ıse of relay links typically reduces the channel attenuation on each link, when compared to that for the source-destination path. This means that in a cellular network, the coverage of high data ate services can be extended or the transmission power reduced. Two main processing techniques for the relays have been studed. These are analogue relaying, where the relays amplify and send the received signals, and digital relaying where the relays decode, re-encode, and re-transmit the signals [6].

One of the first papers to study MIMO relays was [7] which ısed space-time coding concepts to permit several relays to ransmit simultaneously to the destination. In one configuration, he space-time block codes proposed in [12] can be used to ensure that each relay's transmission of the data can be separated from all others to provide diversity gain against fading or shadowing. However, [13] suggests that selecting the single best relay is a better option to maximize capacity – this will be discussed further in Section 4. Reference [8] investigates different protocols for when the source and relay can transmit. Their analysis suggests that permitting interference between the source and relay transmissions can offer better performance than by allocating different channels to the source and relay.

Several authors have considered the combination of spatial multiplexing MIMO systems and multihop links. In [9], bounds on the Shannon capacity of a MIMO link with a single relay are considered. A lower bound on capacity is derived by using either the relay link or the source-destination link separately and an upper bound is derived by summing the SNRs of the source-destination and the relay links. However, this paper makes the rather unrealistic assumption that a relay can receive and transmit on the same frequency at the same time. In [10], a MIMO system is considered in an environment where there is no multipath scattering, so that spatial multiplexing cannot normally be used. It is suggested in that paper that a large number of relays be used as artificial "scatterers", so that spatial multiplexing can then be performed. Finally in [11], another spatial multiplexing configuration using relaying is proposed. In that scheme, a number of co-located transmitters wish to communicate over a large distance to a number of closely spaced receivers. This time, the receivers can collectively act as a MIMO receiver, if they can communicate with each other over a separate short range wireless link to exchange channel and received signal information.

In this paper we will discuss our research on MIMO relaying configurations. We investigate an alternative to analogue and digital relaying called hybrid relaying. In this approach, the relays only decode the training sequence from the source or use feedback from the destination to obtain channel state information (CSI) for the appropriate channel. Then the relay can apply spatial filtering to the received signals, based on the CSI, without decoding them and retransmit the filtered signals to the destination. We also discuss different algorithms for hybrid relaying for single MIMO relay channels. We then extend the discussion to the multiple relay case, taking account of large scale pathloss and shadowing effects. We compare two different strategies for selecting the best relay for a given source and destination.

2. MIMO System Model

We consider the two hop model shown in Fig. 1 with one source, one destination and K relays located randomly and uniformly within the middle region between source and destination. We normalize the range between the source and destination and ignore the direct link between source and destination. We also assume that the total transmit power for both the source and the relays are the same. In the latter case, the transmit power is assumed to be equally distributed among the relays. Each relay processes the received signals independently, without cooperation or exchange of channel state information. We assume in the paper that the source and destination have the same number of transmit and receive antennas M, while each relay has L □ M antennas.

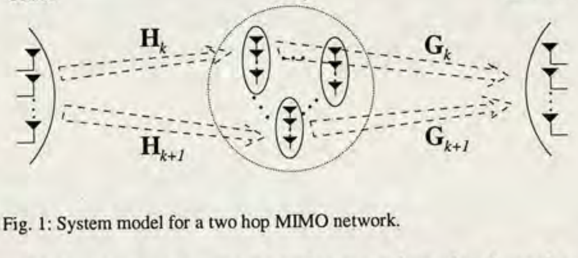


Fig. 1: System model for a two hop MIMO network.

We restrict our discussion to the case where the channels are slow, frequency-flat Rayleigh fading with a block fading model. The data transmission is over two consecutive time slots using two hops. In the first transmission time slot, the source broadcasts its data signals to all the relay terminals. Each relay processes its received signals and then retransmits them to the destination in the second time slot. The $M \times L$ channel matrix $\mathbf{H}_k$ denotes the MIMO channel from the source to the kth relay. Similarly, the $L \times M$ channel matrix $\mathbf{G}_k$ denotes the MIMO channel from the kth relay to the destination.

We assume that source does not know the channel state information, but the destination knows all the channel information. Based on the different channel information available at relays, we separate MIMO relay channels into two categories [10]. The first is *non-coherent relay channels,* where each relay only knows the backward channel matrix $\mathbf{H}_k$. The second is *coherent relay channels,* where each relay has knowledge of both the backward channel matrix $\mathbf{H}_k$ and the forward channel matrix $\mathbf{G}_k$.

3. Performance of Single Relay Systems

In this section, we will discuss algorithms for performing relaying in the case where there is a single relay available. We begin by discussing the non-coherent relay channel case, and then move on to discuss the coherent relay scenario. Finally simulation results are presented to compare the different algorithms that are proposed.

3.1. Non-coherent Relay Channels

For digital relaying (DR), the relay uses L antennas to jointly decode the signals received from the source using the VBLAST algorithm [14]. The relay will then re-encode the M source data streams and use an arbitrary choice of M antennas to forward the encoded signals to the destination. The destination will also apply the VBLAST algorithm to detect the signals.

For analogue relaying (AR), the relay amplifies its received signal vector (of size L) to meet the transmit power constraint and forwards it to the destination. There, the receiver again applies VBLAST detection to the received signals. Compared with digital relaying, one obvious defect for analogue relaying is that while the relays amplify the signals, they also amplify the receiver noise. However, this scheme has the advantage of not requiring decoding at the relay. This means there is no decoding delay at the relay and hence it requires less processing complexity at the relay compared with digital relaying. As discussed below, *AR* can be regarded as the simplest hybrid relaying scheme considered in this paper.

Compared with digital relaying and analogue relaying, hybrid relaying (HR) makes a tradeoff between performance and complexity of the AR and DR methods. It is rather similar to AR, except that an $L \times L$ spatial filtering matrix $\mathbf{W}$ is applied to the received signal, based on knowledge of the channel matrix $\mathbf{H}_k$. However, it does not fully decode and encode the signal, as is required in the DR scheme. After multiplying the signal vector by the filtering weight matrix $\mathbf{W}$, the relay then amplifies and forwards the filtered signals to the destination.

The key advantage of HR compared to AR is that the filtering matrix $\mathbf{W}$ can enhance the signal quality of the retransmitted signal. One way to choose $\mathbf{W}$ is to enhance the average signal to noise ratio (SNR) at the relay. This can be done by applying a spatial matched filter at the relay. This relaying scheme is referred to as *matched filter based relaying (MFR)* [13]. In the case where the number of relay antennas $L > M$, the source and destination array size, HR can be used to reduce the dimension of the retransmitted signal waveform. It can be seen that the SNR at the relay for MFR is $(L+M)/M$ times that for AR. This implies that as L increases the MFR method should provide better SNR performance than the AR method. However, it should also be noted that MFR has the defect of correlating the signals at each antenna, which makes it more difficult for the destination to separate the signals compared with AR.

3.2. Coherent Relay Channels

Compared with the non-coherent scenario, the relay has the freedom to explore and coordinate both backward and forward MIMO channels in a coherent scenario. For example, in digital relaying, the relay can use all the L antennas to retransmit signals to exploit beamforming gain on the relay to destination channels. In this way, the relay to destination MIMO channels become M orthogonal spatial channels and each antenna branch can perform detection independently. The relay can match the highest backward channel gain with the highest forward channel gain to optimize the sum capacity of the complete relay link. We refer to this digital scheme as *modified digital relaying* (MDR).

In [13], we study different hybrid relaying configurations and derive an optimal hybrid relaying (OHR) scheme, based on the backward and forward channel knowledge at the relay. By choosing the optimal filtering matrix $\mathbf{W}$ at the relay, the MIMO relay channel can be decomposed into M orthogonal spatial channels, according to the eigenvalues and vectors of the channel matrices. The decomposition that is used in coherent channels for hybrid relaying is illustrated in Fig. 2.

Fig. 2: The hybrid relaying transformation for coherent relay channels.

Each spatial channel has gain $\lambda_{\mathbf{H}}^i \omega_i$, where $\lambda_{\mathbf{H}}^i$ is the source to relay channel beamforming gain (namely the ith eigenvalue of the source-to-relay channel matrix $\mathbf{H}_k$). Further, the scalar ω_i can be regarded as the relay to destination channel gain for the ith channel, which is optimized by the weight matrix $\mathbf{W}$ under a transmit power constraint at the relay.

Since the calculation for the optimal choice of $\mathbf{W}$ in the OHR scheme is complicated for the case when the number of source antennas $M>2$, we also propose two suboptimal hybrid relaying schemes. The first scheme is a modification of analogue relaying. We refer to it as *modified analogue relaying (MAR).* The matrix $\mathbf{W}$ is set as the product of the left singular vector of $\mathbf{H}_k$ with a diagonal matrix of scaling factors ω_i and the right singular vector of $\mathbf{G}_k$. The scaling factors are chosen to be

$$\omega_i = \rho\lambda_{\mathrm{G}}^i \Big/ \left(1+\rho\lambda_{\mathrm{G}}^i\right) \quad (1)$$

Where λ_{G}^i is the relay to destination beamforming gain (ith eigenvalue of $\mathbf{G}_k$) and ρ is the amplifying factor at the relay. This choice again decomposes the MIMO relay channels into orthogonal channels as shown in Fig. 2. The MAR scheme means that relay is able to decompose the channels and coordinate the backward channels with the forward channels to try to maximize the throughput for M parallel data streams.

The second scheme is a modification of matched filter based relaying, which we refer to as *modified matched filter based relaying (MMFR).* After multiplying the received signal vector by a matched filter matrix, the relay further filters the signals to excite the channel gain for either backward or forward channels. The filtering matrix is constructed in the same way as for the MAR scheme, but the scaling coefficients ω_i for this scheme become

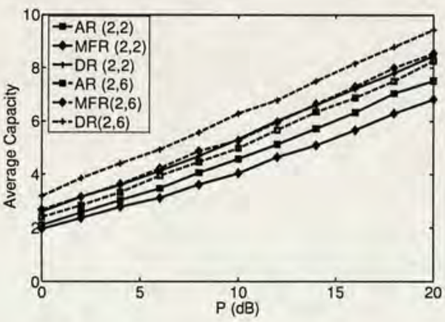
$$\omega_i = \rho\lambda_{\mathrm{H}}^i\lambda_{\mathrm{G}}^i \Big/ \left(1+\rho\lambda_{\mathrm{H}}^i\lambda_{\mathrm{G}}^i\right) \quad (2)$$

Compared with the MAR scheme, this approach has the advantage of enhancing the SNR at the relay. This is achieved by including the eigenvalue λ_{H}^i in the equation for the scaling factor ω_i. For the simple case when $M = 1$, MMFR becomes the optimal hybrid relaying scheme. However, for M □ 2, the signals become more correlated with each other due to the matched filter factor in the weight matrix $\mathbf{W}$, which impairs the throughput of this scheme.

3.3. Algorithm Simulation Results

In this section, we will present results for the relaying techniques described above using the single relay model of section 2. Fig. 3(a) shows performance results for different relaying schemes for non-coherent relay channels and part (b) is for coherent relay channels. We calculate and plot the average Shannon capacity (in bits per channel use) for 1000 channel realizations. The x-axis P in Fig. 3 is defined as total transmit power at the source (and hence also at the relay). When $L=M=2$, it can be seen that digital schemes perform best. For the non-coherent channel results in part (a), AR outperforms MFR especially for higher SNR. This indicates that reducing the noise at the relay cannot compensate for the disadvantage of correlating the signals for the case where $L = M$. Comparing parts (a) and (b) shows that the relaying schemes designed for coherent relay channels give only a small performance advantage over those for non-coherent relay channels. In particular, it can be observed that the performance of MDR is about the same as DR. OHR performs almost as well as MDR and can be a good suboptimal scheme for the tradeoff between performance and complexity.

(a)



(b)

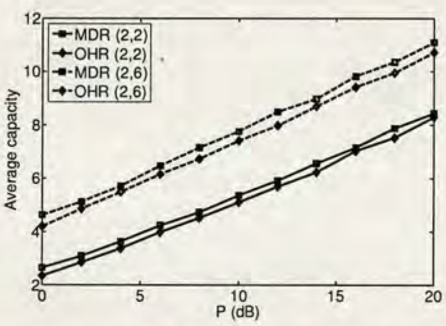


Fig. 3: Average capacity (in bits/channel use) of single relay schemes for (a) non-coherent and (b) coherent relay channels. The number of source/destination antennas M=2 and the number of relay antennas L=2 or 6.

When the number of relay antennas L is increased compared to the number of source/destination antennas M, relaying schemes for coherent channels (Fig 3(b)) perform significantly better than schemes for non-coherent channels (Fig 3(a)). This is because coherent schemes can exploit channel gain on the relay to destination channels when $L>M$. This difference between non-coherent and coherent schemes can be clearly seen by comparing Figs. 3(a) and 3(b). For example, when L=6 and P=20 dB, MDR achieves an average capacity of 11 bits/channel use, which is 1.5 bits better than the result for DR. Another difference from the $L=M=2$ case in Fig. 3(a) is that MFR outperforms AR and achieves performance closer to DR. Since correlating the signals becomes less important as the ratio L/M increases, the SNR at the relay for MFR increases over that for AR. The MDR scheme in Fig 3(b) still performs best overall, though its performance gain over OHR is smaller than for $L=M$=2.

Fig. 4 shows the performance results for different hybrid relaying schemes for coherent relay channels for M=2 and L=2 or 6. It can be seen that OHR outperforms MAR for L=2 and MMFR provides the poorest performance. This is due to the signal correlation effect of matched filtering at relay, described in Section 3.2. When the relay has L=6 antennas, the performance of all three techniques is very similar, but again OHR slightly outperforms MAR in this case.

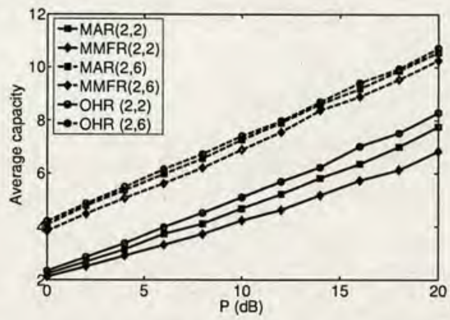


Fig. 4: Average capacity (in bits/channel use) for coherent single MIMO relay channels when hybrid relaying schemes are applied.

4. Multiple Relay Channels

Our work has also investigated the performance of multiple relay channels. The source has two options for routing information to the destination. Firstly, *selective routing* is where a single relay is chosen to forward the source signals to the destination, in a similar manner to the single relay channels discussed above. Secondly, *multicast routing* is where multiple relays can forward the source signals simultaneously to the destination.

For multiple relay channels where selective routing is applied, we will now discuss and compare the capacity performance of two selective routing schemes. In *pathloss and shadowing based selective routing* (*PSSR*), the best relay is selected based on the pathloss and shadowing coefficients of the relay channels only. This means that the effect of fast Rayleigh fading is neglected in the route selection process. Since the calculation is based on the shadowing coefficients, which will change much more slowly than the fading channel realizations, much slower routing updates are required. This approach is suboptimal compared to the *optimal selective routing (OSR)* case where Rayleigh fading is incorporated in the route selection process [13]. Typical simulation results comparing PSSR and OSR are shown in Fig. 5, where the DR method is used in all cases with transmit power P=0dB. Fig. 5 shows some degradation for PSSR, for the single antenna case M=L=1. This is because the Rayleigh fading on each link has a significant impact on capacity. However as M and L are increased to 2 or 4, the capacity of each link becomes much more stable and fading has less effect. In this case, the performance loss of PSSR becomes negligible compared to OSR. It should be noted that the OSR scheme requires a much higher signalling overhead than PSSR and may give rise to routes that change rapidly with time.

Fig. 6 shows the performance of multicast routing (MR) and selective routing based on the PSSR scheme for the noncoherent relay channel scenario. The average capacity of the relay link is plotted against the number of relays that are available. Techniques for non-coherent channels are shown in this figure. The number of antennas at all terminals, including relays, is set to $L=M=2$. The transmit power is set to P=0dB.

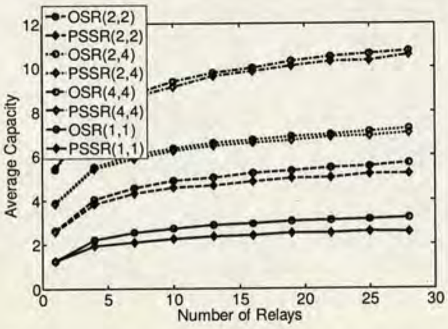


Fig. 5. Average capacity results (in bits/channel use) vs No of relays for the DR method using OSR and PSSR selective routing schemes; The legend shows the values (M,L) and the transmit power P=0dB.

The results in Fig. 6 show that the best scheme is DR combined with PSSR route selection, while the worst scheme is DR with MR. The reason that the MR-DR combination performs so poorly is that all relays must be able to decode the source signal, which means that the poorest relay channel limits performance. This is clearly not a problem when DR is combined with PSSR based selective routing. The hybrid MFR scheme performs the best amongst the multicast routing schemes, but these are generally inferior in performance to the relay selection schemes, where AR outperforms the MFR scheme. These results show it is much better to use all the transmit power on the best relay channel, than to split it among several relay channels. Our simulation studies of the coherent relay channel case are similar to those in Fig. 6, where MDR with selective routing performs the best.

Our studies of these two options described in [15] have shown that, in general, selective routing performs better than multicast routing. This result is valid for our system model where each relay experiences shadowing coefficients that are uncorrelated with all other relays. In this case, the effect of shadowing is typically that one relay will experience significantly better channel conditions than other candidate relays. Therefore, significant performance gains are obtained by only using the best relay to forward information to the destination.

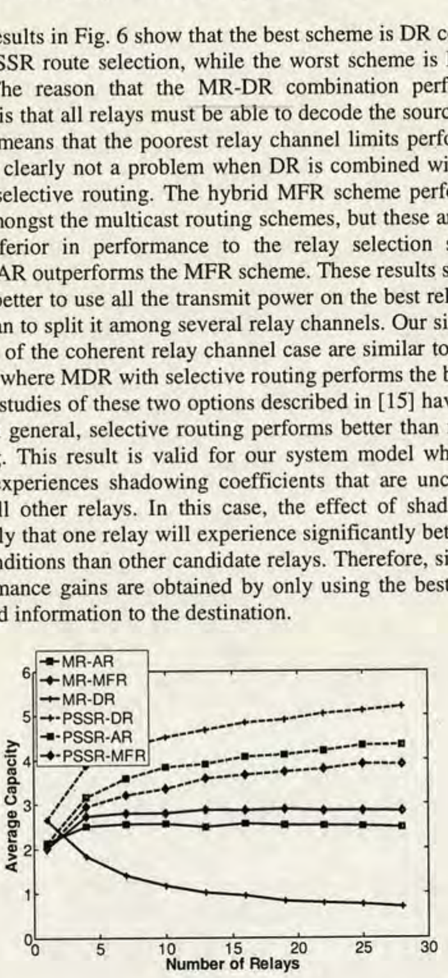


Fig. 6. Average capacity (in bits/channel use) for noncoherent MIMO relay channels using selective routing (PSSR) and multicast routing (MR). The number of antennas $L=M=2$ and transmit power P=0dB.

5. Conclusions

In this paper we studied MIMO spatial multiplexing configurations for relay channels. For single relay channels, we show that increasing the number of antennas at the relay can bring benefits. Firstly it will make proposed hybrid relaying schemes good suboptimal choices for either coherent or non-coherent relay channels compared with digital relying schemes. Secondly, it is shown that in such scenarios, knowledge of forward channel state information at the relay can further improve the network capacity.

For multiple MIMO relay channels, a simple relay selection scheme based on only shadowing/pathloss was found to perform almost as well as using fading effects. If we apply the selective routing schemes discussed in the paper to exploit spatial diversity in the network, multiple relay channels can be simplified to single relay channels. In this case, the conclusions for single relay channels in this paper can also be applied to multiple relay channels.

6. References

[1] D. Gesbert et al, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE Jnl Sel Areas Comm.*, Vol. 21, pp. 281 - 302, 2003.

[2] G.J. Foschini, "Layered space--time architecture for wireless communication in a fading environment when using multi--element antennas", *Bell Labs Tech Jnl*, pp 41-59, Autumn 1996.

[3] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction", *IEEE Trans Info Theory*, Vol 44, pp 744-765, 1998.

[4] J.G. Proakis, *Digital Communications*, New York:McGraw-Hill, 4th Edition, Chapter 14, 2000.

[5] R. Pabst et al, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Comms Mag*, Vol 42, pp 80-89, 2004.

[6] J. N. Laneman et al, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans Info Theory*, Vol 50, pp3062 – 3080, 2004.

[7] J.N. Laneman and G.W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans Info Theory*, Vol. 49, pp.2415-2425, 2003.

[8] R. U. Nabar et al, "Fading relay channels: Performance limits and space-time signal design," *IEEE Jnl Sel Areas Comms*, Vol. 22, pp. 1099-1109, 2004.

[9] B. Wang et al, "On the capacity of MIMO relay channels," *IEEE Trans Info Theory*, Volume 51, No 1, Jan. 2005, pp29 – 43.

[10] H. Bolcskei et al, "Capacity Scaling Laws in MIMO Relay Networks," *IEEE Trans Wireless Comms*, 2004, submitted.

[11] P. Gupta and P.R. Kumar, "Towards an Information Theory of Large Networks: An Achievable Rate Region", IEEE Transactions on Information Theory, Vol 49(8), pp 1877-1894, August 2003.

[12] V. Tarokh, H. Jafarkhani and A.R. Calderbank, "Space-time block codes from orthogonal designs", IEEE Trans Info Theory, Vol 45, pp 1456-1467, 1999.

[13] Y.Fan and J.S. Thompson, "MIMO Configurations for Relay Channels: Theory and Practice", *IEEE Trans Wireless Comms*, 2005, submitted.

[14] G.D. Golden et al, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture", *Elec Lett*, Vol. 35, pp14 – 16, 1999.

[15] Y. Fan and J. Thompson, "On the Outage Capacity of MIMO Multihop Networks", IEEE Globecom Conference, St. Louis, USA, pp 2204-2208, November 2005.

The work of Y Fan was funded by EPSRC Grant GR/S58782.

John S Thompson
School of Engineering and Electronics, University of Edinburgh,
Edinburgh, EH9 3JL
United Kingdom
Fax: +44 131 650 6554
E-mail: John.Thompson@ed.ac.uk

Peter M Grant
School of Engineering and Electronics, University of Edinburgh,
Edinburgh, EH9 3JL
United Kingdom
Fax: +44 131 650 6554
E-mail: Peter.Grant@ee.ed.ac.uk

Yijia Fan
School of Engineering and Electronics, University of Edinburgh,
Edinburgh, EH9 3JL
United Kingdom
Fax: +44 131 650 6554
E-mail: Y.Fan@ed.ac.uk

On the Diversity-Multiplexing Tradeoff for Multi-antenna Multi-relay Channels

Yijia Fan, Abdulkareem Adinoyi, John S Thompson, Halim Yanikomeroglu

Abstract

In this letter we analyze the performance of multiple relay channels when multiple antennas are deployed only at relays. Specifically, we investigate the simple repetition-coded decode-and-forward protocol and apply two antenna combining techniques at relays, namely maximum ratio combining (MRC) on receive and transmit beamforming (TB). We assume that the total number of antennas at all relays is fixed to N. With a reasonable power constraint at the relays, we show that the antenna combining techniques can exploit the full spatial diversity of the relay channels and can achieve the same diversity multiplexing tradeoff as achieved by more complex distributed space-time random coding techniques proposed earlier.

I. Introduction

Cooperative diversity schemes, which exploit the spatial diversity of the relay channels, have been an active research area in the last few years. The performance limits of distributed space-time codes, which can exploit cooperative diversity, are discussed in [1]–[3] for single-antenna

This work was supported by the EPSRC Grant GR/S58782/01, UK.

Y. Fan and J. S. Thompson are with the Institute for Digital Communications, Joint Research Institute for Signal and Image Processing/Integrated Systems/Energy, University of Edinburgh, Edinburgh, EH9 3JL, UK. (e-mail: y.fan@ed.ac.uk, john.thompson@ed.ac.uk)

A. Adinoyi and H. Yanikomeroglu are with Broadband Communications and Wireless Systems (BCWS) centre, Department of Systems and Computer Engineering, Carleton University, Ottawa, K1S 5B6, Canada (Email: adinoyi@sce.carleton.ca)

relay networks. However, the design and implementation of practical codes that approach these limits are a challenging open research area. For example, [2] and [3] use *new independent random codebooks* at the relays to achieve spatial diversity. Though those approaches have theoretical values, they face stiff implementation challenges. It is suggested in [2] that the existing space-time codes for the point-to-point multiple-input multiple-output (MIMO) link might be implemented into relay networks. However, the system becomes more complicated as antennas in relay networks are distributed rather than centralized. For example, each relay may need to know all of the uncoded data, before sending only one part of the codeword to the destination. This might cause additional time delay and processing complexity. An easier way to obtain full diversity for a multiple relay network is to implement repetition coding [2], where the relays decode and re-encode the information using the *same* codebook as applied at the source. In this scenario, instead of using ideal random coding, any capacity achieving AWGN channel codes (e.g, turbo codes or LDPC codes) alone are sufficient to offer excellent performance. However, this scheme requires the relays to transmit in orthogonal time slots. This will result in a significant *multiplexing loss* compared with the ideal space-time coding approach.

In this letter we exploit the spatial diversity of the relay channels in a different way from the space-time codes-based approach. We concentrate on the repetition-coding decode-and-forward technique while applying two kinds of antenna combining techniques at the relay, namely maximum ratio combining (MRC) [4] for reception and transmit beamforming (TB) [5] for transmission. Those techniques are often used in point-to-point single-input multiple-output (SIMO) or multiple-input single-output (MISO) wireless links, where either the transmitter or receiver is equipped with multiple antennas. It has been shown that MRC (TB) is able to achieve the optimal diversity multiplexing tradeoff of SIMO (MISO) systems [6]. In a relay context, we move the multiple antennas to the relays, while the source and the destination are only equipped with a single antenna. It appears that this network consists of a SIMO and MISO channel, for which MRC and TB are the best schemes to use. However, unlike the point-to-point link, the antennas might be deployed in a distributed fashion. Studying the tradeoff between the number of relays and the number of antennas at each relay is a important topic and the main concern of the letter. We analyze the performance of this system based on a slow fading scenario. More specifically, we examine the outage probability and the diversity multiplexing tradeoff of the network.

Note that unlike [2], we allow multiple antennas to be deployed at the relays. However, it will be shown later that the diversity gain that can be achieved by our approach is the same as that of the space-time coded approach proposed by Laneman and Wornell in [2]. This holds as long as the total number of antennas at all relays is fixed, regardless of the number of antennas at each relay and the number of relays. Thus, the same diversity gain can be achieved even when each relay is deployed with a single antenna. Therefore, the application of our scheme is quite general. Furthermore, we note that some recent work on single antenna network has also considered beamforming approach, although they mainly focused on the energy efficiency or capacity scaling behavior (e.g. [9], [10]).

II. SYSTEM DESCRIPTION

We consider a two hop network model with one source, one destination and K relays. For simplicity we ignore the direct link between source and destination. The extension of all the results to include the direct link is straightforward. We assume that the source and destination are deployed with single antenna, while relay k is deployed with m_k antennas; the total number of antennas at all relays is fixed to N. This can be expressed as

$$\sum_{k=1}^{K} m_k = N. \tag{1}$$

We restrict the discussions to the case where the channels are slow and frequency-flat fading. We assume a coherent relay channel configuration context where the kth relay can obtain full knowledge of both the backward channel vector $\mathbf{h}_k$ and the forward channel vector $\mathbf{g}_k$. Note that the forward channel knowledge can be obtained easily if the relay-destination link operates in a Time-Division-Duplex (TDD) mode. One example where the relays obtain the required channel information can be found in [10]. Otherwise, obtaining forward channel knowledge might require additional signalling overhead. In slow fading channel, which is the focus of this letter, this overhead is negligible. For fair comparison, we also assume that for each channel realization, all the backward and forward channel coefficients for all N antennas remain the same regardless of the number of relays K. Fig. 1 shows the system model.

The data transmission is over two time slots using two hops. In the first transmission time slot, the source broadcasts the signal to all the relay terminals. The input/output relation for the source to the kth relay is given by

$$\mathbf{r}_k = \sqrt{\eta}\mathbf{h}_k s + \mathbf{n}_k, \tag{2}$$

where $\mathbf{r}_k$ is the $m_k \times 1$ receive signal vector, and η denotes the transmit power at the source. The scalar s is the unit mean power transmit signal and $\mathbf{n}_k$ is the $m_k \times 1$ complex circular additive white Gaussian noise vector at relay k with identity covariance matrix $\mathbf{I}_{m_k}$. The entries of the channel vector $\mathbf{h}_k$ are identically and independently distributed (i.i.d) complex Gaussian random variables with zero mean and unit variance. We assume that each relay performs MRC of the received signals, by multiplying the received signal vector by the vector $\mathbf{h}_k^H \big/ \|\mathbf{h}_k\|_F$. The signal at the output of the relay receiver is given by

$$\tilde{r}_k = \sqrt{\eta \sum_{i=1}^{m_k} |h_{i,k}|^2} s + \frac{\sum_{i=1}^{m_k} h_{i,k}^* n_{i,k}}{\sqrt{\sum_{i=1}^{m_k} |h_{i,k}|^2}}, \tag{3}$$

where $h_{i,k}$ denotes the channel coefficient from the source to the ith antenna at relay k, and $n_{i,k}$ denotes the noise factor for ith receiver input branch. The SNR at the output of the receiver can be written as:

$$\rho_k^{m_k} = \eta \sum_{i=1}^{m_k} |h_{i,k}|^2. \tag{4}$$

After the relays decode the signals, each relay re-encodes the signal using the same codebook as used at the source, then performs TB of the decoded waveform. If we denote the transmitted signals as t_k with unit variance, the transmitted signal vector $\mathbf{d}_k$ for relay k can be written as

$$\mathbf{d}_k = \sqrt{\frac{\eta m_k}{N}} \frac{\mathbf{g}_k^H}{\|\mathbf{g}_k\|_F} t_k, \tag{5}$$

where the vector $\mathbf{g}_k$ is the $1 \times m_k$ channel vector from the kth relay to the destination, of which each entry is an i.i.d. complex Gaussian random variable with unit variance. The vector $\mathbf{d}_k$ in (5) is designed to meet the total transmit power constraint:

$$\mathrm{E}\left[\|\mathbf{d}_k\|_F^2\right] \le \frac{\eta m_k}{N}, \tag{6}$$

where $\|\bullet\|_F$ denotes the Frobenius norm. Here we assume that the total power at all relays is fixed to be η, i.e. the same as at the source. However, all the conclusions in the paper also hold when the total power at all relays is fixed to *an arbitrary constant*. We note that this power assumption has a meaningful practical implications: in reality the users with larger number of antennas can often transmit with a higher power (in proportional to the number of transmit antennas in this paper).

The destination receiver simply detects the combined signals from all K relays. If the signals are correctly decoded at all the relays (i.e. $t_k = s$), the output signal at the destination can be written as:

$$y = s \sum_{k=1}^{K} \sqrt{\frac{\eta m_k}{N} \sum_{i=1}^{m_k} |g_{i,k}|^2} + n_d, \tag{7}$$

where the scalar n_d is the complex additive white Gaussian noise at the destination with unit variance. The output SNR at the destination receiver can therefore be written as:

$$\rho_d^{m_k} = \left(\sum_{k=1}^{K} \sqrt{\frac{\eta m_k}{N} \sum_{i=1}^{m_k} |g_{i,k}|^2} \right)^2. \tag{8}$$

When all the relays are deployed with a single antenna, there is no maximum ratio combining gain at the relays and the destination. However, the destination still observes a set of equal gain combined [7] amplitude signals from all relays.[1] Since we assume that the backward and

[1]Unlike [7], the equal gain combining for relay channels is applied at the transmitter instead of the receiver.

forward channel coefficients for each antenna are kept the same for different values of K and m_i, the output SNR at the destination can be rewritten as $\rho_d^1 = \frac{\eta}{N}\left(\sum\limits_{k=1}^{K}\sum\limits_{i=1}^{m_i}|g_{i,k}|\right)^2$; when all the antennas are deployed in one relay (i.e. $K = 1$ and $m_1 = N$), full diversity gain is achieved among all the N antennas at the relay and also at the destination. The SNR for this case can be rewritten as $\rho_d^N = \eta \sum\limits_{k=1}^{K}\sum\limits_{i=1}^{m_i}|g_{i,k}|^2$.

III. Outage Analysis

In a slow fading environment, the outage probability can be defined as

$$P_{out} \triangleq P\left[C < R\right], \tag{9}$$

where C denotes the capacity for the particular channel realization and R denotes the transmission rate. The diversity-multiplexing tradeoff can be defined as [8]:

Definition 1: **(Diversity-Multiplexing Tradeoff)** Consider a family of codes C_η operating at SNR η and having rates R bits per channel use. The multiplexing gain and diversity order are defined as[2]

$$r \triangleq \lim_{\eta\to\infty}\frac{R}{\log_2\eta}, \quad d \triangleq -\lim_{\eta\to\infty}\frac{\log_2 P_{out}\left(R\right)}{\log_2\eta}. \tag{10}$$

We first study a simple protocol, in which all the relays participate in the decoding and forwarding process. We refer to this protocol as multi-cast decoding. An outage occurs whenever any relay or the destination fails to decode the signals. We firstly introduce a lemma on the bounds of the value of $\rho_d^{m_k}$, i.e. the output SNR at the destination given that the signal is correctly decoded at all relays. This lemma has been shown in our earlier work [11]:

Lemma 1: For any m_k, $\rho_d^1 \leq \rho_d^{m_k} \leq \rho_d^N$.

We omit the proof due to limited space, the interested reader can refer to [11] for details. This lemma implies that the increased "equal gain combining" gain at the destination can not compensate for the loss of maximum ratio combining gain at the relay and the destination when

[2]we assume that the block length of the code is large enough, so that the detection error is arbitrarily small and the main error event is due to outage.

the number of relays K is increased and the numbers of antennas at each relay are reduced. Based on *Lemma 1*, we now begin our outage analysis with the following lemma:

Lemma 2: Conditioned on all the relays correctly decoding the messages, the outage probability for the relay channels is bounded by:

$$\frac{1}{N!}\left(\frac{N\left(2^{2R}-1\right)}{\eta}\right)^{N} \geq P_{out}^{m_k} \geq \frac{1}{N!}\left(\frac{2^{2R}-1}{\eta}\right)^{N} \tag{11}$$

Proof: See Appendix A.

Lemma 2 indicates that the full diversity of N can be achieved regardless of the number of relays K, provided that the signals are correctly decoded at the relays. However, the diversity of the network might decrease if certain outage occurs at the relays. This is especially true for the multi-cast decoding protocol, for which we have the following theorem:

Theorem 1: For large η, the outage probability for the multi-cast decoding is bounded by:

$$N\left(\frac{2^{2R}-1}{\eta}\right) \geq P_{out} \geq \frac{2}{N!}\left(\frac{2^{2R}-1}{\eta}\right)^{N} \tag{12}$$

with equality to the right-hand side if $K = 1$, to the left-hand side if $K = N$.

Proof: See Appendix B. ∎

This theorem implies that for multi-cast decoding, having more relays and less antennas per relay actually loses diversity. Since requiring all the relays to fully decode the source information limits the performance of the decode and forward to that of the poorest source to relay link. Specifically, it can be seen that for $K = N$ no diversity gain is offered by relaying i.e. the SNR exponent is -1, as no diversity gain can be obtained from the source to relay links in this case. We note that similar observations regarding this point have been made in [1], [2]. *However*, for $K = 1$ the full diversity of N can be achieved, as the diversity gain for the source to relay link is also N. This observation implies that diversity gain can still be obtained for multi-cast decoding, if multiple antennas are deployed at the relays. In terms of the diversity multiplexing tradeoff, we have the following theorem.

Theorem 2: The diversity-multiplexing tradeoff curve for the multi-cast decoding scheme is bounded by

$$1 - 2r \leq d \leq N\left(1 - 2r\right), 0 \leq r \leq 0.5 \tag{13}$$

with equality to right-hand side if $K = 1$, to left-hand side if $K = N$.

Proof: For large η, replace R with $r \log_2 \eta$ in (12), the proof is straightforward.

It can be seen from Theorem 3 that when $K = N$, the diversity-multiplexing tradeoff for multi-cast decoding is strictly worse than that for direct transmission, which is $d = 1 - r$ [8]. When $K = 1$, however, the diversity-multiplexing tradeoff is the same as the space-time distributed coding schemes proposed in [2]. In fact, we can combine the antenna diversity schemes with a protocol *similar to* the one proposed by [2], which exploit further the diversity of source to relay channels by selecting the qualified relays that meet the transmission rate R, to improve the network performance when $K > 1$. Specifically, the protocol for the antenna diversity schemes is proposed as follows:

Protocol 1: **(Selection Decoding)** Select $\tilde{K}$ relays with a total number of antennas $\tilde{N}$, denoted as $\Re\left(\tilde{N}, \tilde{K}\right)$, that could successfully decode the source message at a transmission rate R, to decode and forward the messages.

We can obtain outage probability for the selection decoding by the following theorem:

Theorem 3: For large η, the outage probability for the selection decoding scheme is bounded by:

$$\left(\frac{2^{2R}-1}{\eta}\right)^N \times \sum_{\tilde{N}=1}^{N} \binom{N}{\tilde{N}} \frac{1}{\tilde{N}!}\tilde{N}^{\tilde{N}} \geq P_{out}^{m_k} \geq \left(\frac{2^{2R}-1}{\eta}\right)^N \sum_{\tilde{N}=1}^{N} \binom{N}{\tilde{N}} \frac{1}{\tilde{N}!\left(N-\tilde{N}\right)!}, \tag{14}$$

while the upper bound is met when the selected $\tilde{N}$ antennas are all within one relay.

Proof: See Appendix C.

It can be seen from Theorem 3 that for selection decoding full diversity can always be achieved regardless of the number of relays K and selected relays $\tilde{K}$. This is clearly an advantage over the multi-cast decoding scheme. Replacing R with $r \log_2 \eta$ in (14), we can directly obtain the diversity-multiplexing tradeoff for selection decoding:

Theorem 4: The diversity-multiplexing tradeoff curve for selection decoding is

$$d = N(1-2r), 0 \le r \le 0.5, \tag{15}$$

which is the same as that for the space-time distributed coding protocol proposed in [2][3]. Fig. 2 shows the diversity multiplexing tradeoff for different protocols discussed in the paper, when $N = 5$.

IV. Conclusions

Three conclusions can be drawn: (a) provided the messages are successfully decoded at the relays, having less relays will offer better performance due to increased combining (power) gain at the destination, although the full diversity order N of the network can always be achieved regardless of the number of antennas; (b) if all the relays participate in the decoding and forwarding process, the network performance will degrade as the number of relays increases, as the performance is always restricted to the worst source to relay link. In this sense, deploying all the antennas at single relay is the optimal choice; (c) however, full diversity can be achieved if we apply the relay selection schemes to choose the potential relays. More specifically, the diversity-multiplexing tradeoff achieved by the antenna combining techniques is the same as that achieved by more complicated distributed space-time coding schemes such as [2]. In this scenario, deploying more antennas at fewer relays is still a better choice due to improved combining (power) gain.

Appendix

A. *Proof of* Lemma 1

Based on Lemma 1, it is clear that

$$P_{out}^{1} \ge P_{out}^{m_k} \ge P_{out}^{N}, \tag{16}$$

[3]Note that $d = (N+1)(1-2r)$ if direct link is included.

where P_{out}^1 denotes the outage probability for N relay case and P_{out}^N for 1 relay case, given that the signals are correctly decoded at all the relays. Note that

$$\rho_d^1 \geq \frac{\eta}{N} \sum_{k=1}^{K} \sum_{i=1}^{m_i} |g_{i,k}|^2 = \frac{\rho_d^N}{N}, \tag{17}$$

inequality (16) can be extended and modified as:

$$P\left[\sum_{k=1}^{K} \sum_{i=1}^{m_i} |g_{i,k}|^2 \leq \frac{N\left(2^{2R}-1\right)}{\eta}\right] \geq P_{out}^{m_k} \geq P\left[\sum_{k=1}^{K} \sum_{i=1}^{m_i} |g_{i,k}|^2 \leq \frac{2^{2R}-1}{\eta}\right]. \tag{18}$$

Since $\sum\limits_{k=1}^{K} \sum\limits_{i=1}^{m_i} |g_{i,k}|^2$ is chi-square distributed with dimension $2N$, for small ε it is easy to show that [6]

$$P\left(\sum_{k=1}^{K} \sum_{i=1}^{m_i} |g_{i,k}|^2 \leq \varepsilon\right) \approx \frac{1}{N!}\varepsilon^N. \tag{19}$$

Put (19) to (18) finishes the proof of *Lemma 2*.

B. *Proof of Theorem 1*

If we denote $C_r^{k,m_k} = 0.5\log_2\left(1+\rho_k^{m_k}\right)$ as the Shannon capacity from source to relay k channel for each channel realization. The outage probability is given by:

$$\begin{aligned}
P_{out} &= P\left[\min\left(C_r^{k,m_k}\right) < R\right] + P\left[\min\left(C_r^{k,m_k}\right) > R\right] P_{out}^{m_k} \qquad (20)\\
&= 1 - \prod_{k=1}^{K}\left(1 - P\left[\sum_{i=1}^{m_i} |h_{i,k}|^2 < \frac{2^{2R}-1}{\eta}\right]\right) + P_{out}^{m_k} \prod_{k=1}^{K}\left(1 - P\left[\sum_{i=1}^{m_i} |h_{i,k}|^2 < \frac{2^{2R}-1}{\eta}\right]\right)\\
&\overset{\eta\to\infty}{\approx} 1 - \prod_{k=1}^{K}\left(1 - \frac{1}{m_k!}\left(\frac{2^{2R}-1}{\eta}\right)^{m_k}\right) + P_{out}^{m_k}, \qquad (21)
\end{aligned}$$

where $P_{out}^{m_k}$ is bounded by (11). For large η, retaining only the term containing the lowest exponent of $1/\eta$ in the first term, (21) can be further modified as

$$P_{out} \approx \sum_{k=1}^{K} \frac{1}{m_k!}\left(\frac{2^{2R}-1}{\eta}\right)^{m_k} + P_{out}^{m_k}. \tag{22}$$

Observing that $\frac{1}{a!\eta^a} < \frac{1}{b!\eta^b}$ when $a > b$, P_{out} is maximized when $m_k = 1$,$K = N$. Therefore for large η

$$P_{out} \leq N\left(\frac{2^{2R}-1}{\eta}\right), \tag{23}$$

where $P_{out}^{m_k}$ is omitted due to its higher exponent. P_{out} is minimized when $m_k = N$,$K = 1$ and $P_{out}^{m_k} = P_{out}^{N}$. We obtain the lower bound

$$P_{out} \geq \frac{2}{N!}\left(\frac{2^{2R}-1}{\eta}\right)^{N} \tag{24}$$

and thus complete the proof.

C. Proof of Theorem 3

Since $\Re\left(\tilde{N},\tilde{K}\right)$ is a random set, we utilize the total probability law and write

$$P_{out} = \sum_{\Re\left(\tilde{N},\tilde{K}\right)} P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] P_{out}^{m_k|\Re\left(\tilde{N},\tilde{K}\right)}, \tag{25}$$

where $P_{out}^{m_k|\Re\left(\tilde{N},\tilde{K}\right)}$ denotes the outage probability conditioned on $\Re\left(\tilde{N},\tilde{K}\right)$ is chosen, and can be bounded by (11) by replacing N with $\tilde{N}$. The probability for any relay to be chosen can be expressed as:

$$P\left[r \in \Re\left(\tilde{N},\tilde{K}\right)\right] = P\left[\sum_{i=1}^{m_k}\left|h_{i,k}\right|^2 \geq \frac{2^{2R}-1}{\eta}\right] = 1 - P\left[\sum_{i=1}^{m_k}\left|h_{i,k}\right|^2 \leq \frac{2^{2R}-1}{\eta}\right]. \tag{26}$$

Therefore any $\Re\left(\tilde{N},\tilde{K}\right)$ exists with a probability that can be written as:

$$\begin{aligned} P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] = & \prod_{r\in\Re\left(\tilde{N},\tilde{K}\right)} \left(1 - P\left[\sum_{i=1}^{m_k}\left|h_{i,k}\right|^2 \leq \frac{2^{2R}-1}{\eta}\right]\right) \\ & \times \prod_{r\in\Re\left(N-\tilde{N},K-\tilde{K}\right)} P\left[\sum_{i=1}^{m_k}\left|h_{i,k}\right|^2 \leq \frac{2^{2R}-1}{\eta}\right]. \end{aligned} \tag{27}$$

Based on (19), at high SNR, $P\left[\Re\left(\tilde{N},\tilde{k}\right)\right]$ can be approximated as

$$P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] \approx \left(\frac{2^{2R}-1}{\eta}\right)^{N-\tilde{N}} \prod_{r\in\Re\left(N-\tilde{N},K-\tilde{K}\right)} \frac{1}{m_k!}, \tag{28}$$

Which can be bounded by:

$$\frac{1}{\left(N-\tilde{N}\right)!}\left(\frac{2^{2R}-1}{\eta}\right)^{N-\tilde{N}} \leq P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] \leq \left(\frac{2^{2R}-1}{\eta}\right)^{N-\tilde{N}} \tag{29}$$

Note that the bounds are independent of K. Putting (29) and (11) into (25), we obtain the bounds (14) and thus complete the proof.

REFERENCES

[1] 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, pp. 3062-3080, December 2004.

[2] 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.

[3] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol 51, no. 12 pp. 4152-4172, Dec. 2005.

[4] J. G. Proakis, *Digital Communications: Fourth Edition*, 2001.

[5] J. Bach Andersen, "Antenna arrays in mobile communications: gain, diversity, and channel capacity", *IEEE Antennas Propagat. Mag.*, vol. 42, no. 2, pp. 12 - 16, April 2000.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005.

[7] Y. Chen, C. Tellambura, "Performance analysis of L-branch equal gain combiners in equally correlated Rayleigh fading channels", *IEEE Commun. Letters*, vol. 8, no. 3, pp. 150 - 152, March 2004.

[8] 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.

[9] A. F. Dana and B. Hassibi, "On the power efficiency of sensory and adhoc wireless networks", IEEE Trans. Inf. Theory, vol. 52, no. 7, pp. 2890 - 2914, Jul. 2006.

[10] B. Wang, J. Zhang and L. Zheng, "Achievable Rates and Scaling Laws of Wideband Sensory Relay Networks", *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4084- 4104, Sept. 2006.

[11] Y. Fan, A. Adinoyi, J. S. Thompson and H. Yanikomeroglu, "Antenna combining for multi-antenna multi-relay channels", *European Trans. Telecom., special issue on European wireless conference 2006*, to appear, 2007. Available on web:http://www.see.ed.ac.uk/~s0345317/relay_ETT.pdf.

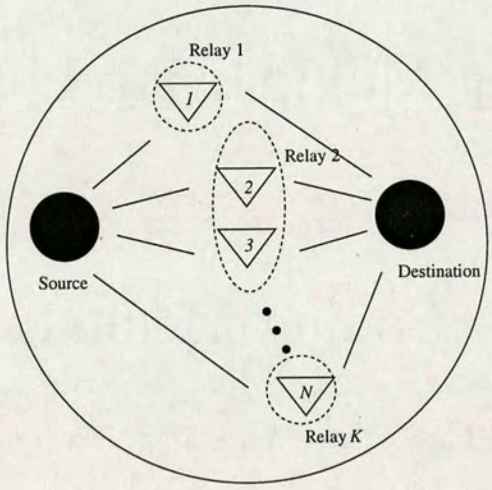


Fig. 1. System model for a two hop network: Source and destination are each deployed with 1 antenna. Totally N antennas are deployed at K relays. For each channel realization, either backward or forward channel coefficients for all N antennas remains the same regardless of the number of relays K.

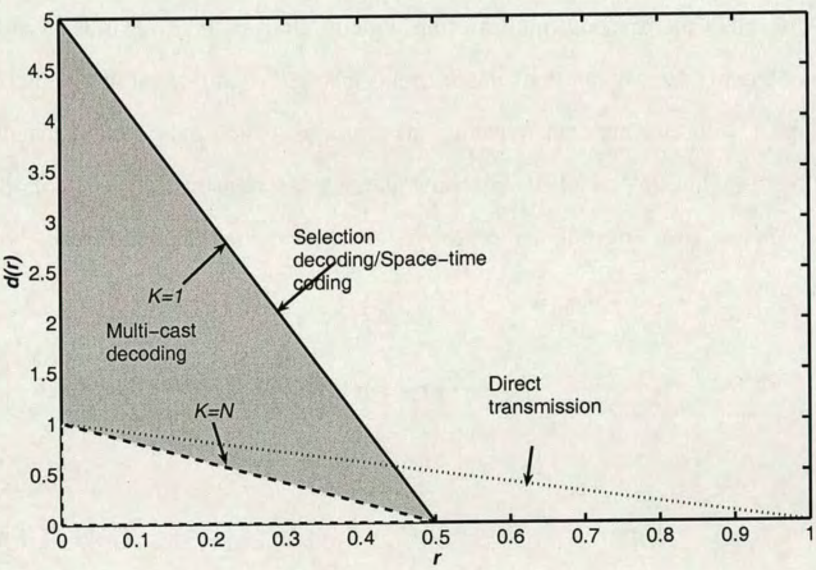


Fig. 2. The diversity multiplexing tradeoff for different protocols, when $N = 5$.

Recovering Multiplexing Loss through Successive Relaying Using Simple Repetition Coding Method

Yijia Fan, Chao Wang, John Thompson

Abstract

In this paper, we study a transmission protocol for a two relay wireless network where simple repetition coding is applied at the relays. We give information-theoretic achievable rates for this transmission scheme, and develop a space-time V-BLAST signalling and detection method which can approach them. We show through the diversity multiplexing tradeoff analysis that our transmission scheme can recover the multiplexing loss of the half-duplex relay network, while retaining some diversity gain. We also compare it with conventional transmission protocols which only exploit the diversity of the network at the cost of a multiplexing loss. We show that our 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

Manuscript submitted June 2006, revised Feb 2007

Y. Fan's work was supported by the EPSRC Grant GR/S58782/01, UK and C. Wang's work was supported by the UK Mobile Virtual Centre of Excellence (VCE) Core 4 Project (www.mobilevce.com).

The authors are with the Institute for Digital Communications, University of Edinburgh, Edinburgh, EH9 3JL, UK. (e-mail: y.fan@ed.ac.uk, john.thompson@ed.ac.uk)

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 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 to be used 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 0.5. 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 if assuming 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 [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 time 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. We make the following observations in this paper for the proposed protocol:

- 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 decoding 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 protocol as a function of SNR, as well as the achievable rates for different channel conditions. We believe these analysis offer strong insights for *adaptive protocol design*, where relaying and direct transmission can be combined together. 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.

- 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 firstly appeared in [30]. It was focused on amplify-and-forward relaying and did not offer insight on 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], [32]. In [6] capacity analysis was made while direct link is ignored. Hence no diversity can be obtained. At about the *same* time as we submitted this manuscript, references [31], [32] were published. These papers included the direct link and offered diversity multiplexing tradeoff analysis under the assumption that the two relays were isolated. Here we note that in our work we assume both interference and direct link exists.

Just before submitting this manuscript, we found [5] which also analyzes the capacity for such schemes when digital relaying is used. One major difference between [5] and our work is that direct link is ignored in [5] while it is considered in this paper. Therefore the analysis becomes different for the following reasons. Firstly, the scheme in [5] does not offer any *cooperative diversity* gain, which is a very important benefit that the relay could offer. We will show in this paper a diversity gain of 2 can be obtained by considering the direct link. Secondly, this difference 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* analyse 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] might be obtained if assuming the channel coefficient for direct link is zero in our model. Therefore our analysis is more general, and the adaptive protocols introduced by us 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_l. 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 till *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 codewords, 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). How to suppress the interference thus becomes a major problem. We will discuss this problem further in the next two sections.

III. Achievable Rates

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 $h_{a,b}$ as the channel coefficient between node a and b, which may contain path-loss, Rayleigh fading, and lognormal shadowing. For simplicity, we denote $C\left(x\right)$ as the capacity function $\log_2\left(1+x\right)$, where 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

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

where r_i is the ith element in the L dimensional relay index vector

$$\mathrm{r} = \left[\begin{array}{ccccc} R1 & R2 & R1 & R2 & R1\cdots \end{array}\right], \tag{2}$$

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 only concern ourselves 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 $R1$ transmits s_1 while $R2$ is receiving s_2, if $|h_{R1,R2}| \succ |h_{S,R2}|$, $R2$ firstly 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 $R2$, which can be expressed as follows:

$$R_1 \leq C\left(\frac{|h_{R1,R2}|^2 SNR}{1+|h_{S,R2}|^2 SNR}\right). \tag{3}$$

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

$$R_2 \leq C\left(\frac{|h_{S,R2}|^2 SNR}{1+|h_{R1,R2}|^2 SNR}\right). \tag{4}$$

Note that this decoding criterion applies from the second time slot to the Lth time slot when transmitting each frame. In slot i, equation (1) can be adapted to a constraint on R_{i-1} and equation (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:

$$\mathbf{y} = \sqrt{SNR}\underbrace{\begin{bmatrix} h_{S,D} & 0 & 0 & 0 & 0 \\ h_{r_1,D} & h_{S,D} & 0 & 0 & 0 \\ 0 & h_{r_2,D} & h_{S,D} & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & h_{r_{L-1},D} & h_{S,D} \\ 0 & 0 & 0 & 0 & h_{r_L,D} \end{bmatrix}}_{\mathbf{H}} \mathbf{s} + \mathbf{n}, \tag{5}$$

where $\mathbf{y}$ is the $(L+1)\times 1$ receive signal vector, $\mathbf{s}$ is the $L \times 1$ transmit signal vector and $\mathbf{n}$ is the $(L+1)\times 1$ complex circular additive white Gaussian noise vector at the destination. Unlike conventional multiple access MIMO channels, the dimensions of $\mathbf{y}$, $\mathbf{s}$ and $\mathbf{n}$ 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]:

$$R_k \leq \log_2\left(\det\left(\mathbf{I} + \mathbf{h}_k\mathbf{h}_k^H SNR\right)\right), \tag{6}$$

$$R_{k_1} + R_{k_2} \leq \log_2\left(\det\left(\mathbf{I} + SNR\left(\mathbf{h}_{k_1}\mathbf{h}_{k_1}^H + \mathbf{h}_{k_2}\mathbf{h}_{k_2}^H\right)\right)\right), \tag{7}$$

$$\cdots$$

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

where $\mathbf{h}_k$ denotes the kth column of $\mathbf{H}$. As it is extremely complicated to give an exact description for the rate region of each signal when $L > 2$, we will only concentrate 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 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:

$$R_{S,r_1} \leq C\left(|h_{S,r_1}|^2 SNR\right) \tag{9}$$

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` $h_{R1,R2} \succ h_{S,r_{i+1}}$,

$$R_i \leq \min\left(C\left(\frac{|h_{R1,R2}|^2 SNR}{1+|h_{S,r_{i+1}}|^2 SNR}\right), R_{S,r_i}, C\left(|h_{S,D}|^2 SNR + |h_{r_i,D}|^2 SNR\right)\right),$$
$$R_{S,r_{i+1}} \leq C\left(|h_{S,r_{i+1}}|^2 SNR\right); \tag{10}$$

`else`

$$R_i \leq \min\left(R_{S,r_i}, C\left(|h_{S,D}|^2 SNR + |h_{r_i,D}|^2 SNR\right)\right),$$
$$R_{S,r_{i+1}} \leq C\left(\frac{|h_{S,r_{i+1}}|^2 SNR}{1+|h_{R1,R2}|^2 SNR}\right); \tag{11}$$

`end`.

Note that the term $C\left(|h_{S,D}|^2 SNR + |h_{r_i,D}|^2 SNR\right)$ 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:

$$R_L \leq \min\left(R_{S,r_L}, C\left(|h_{S,D}|^2 SNR + |h_{r_L,D}|^2 SNR\right)\right). \tag{12}$$

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

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

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 on 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

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

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

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

It is obvious that the rate bounds provided by (15) is significantly larger than that 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:

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

The capacity expressed by (13) can be simplified to

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

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

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

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

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

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.

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 develop 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 only used 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 utilizing 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 a 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]:

$$C = \log_2 \det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^H SNR\right) = \sum_{i=1}^{M} \log_2\left(1 + SINR_i\right), \tag{20}$$

where $SINR_i$ 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 is suggested to feed the value of $SINR_i$ back to the transmitter. Adaptive modulation and coding should be applied to make the transmission rate for s_i lower than $\log_2\left(1 + SINR_i\right)$.

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 *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 $1 \le i \le L-1$), based on the same interference cancellation criterion as in Section III, the rate calculation can be performed as:

`if` $h_{R1,R2} \succ h_{S,r_{i+1}}$

$$R_i \leq \min\left(C\left(\frac{|h_{R1,R2}|^2 SNR}{1+|h_{S,r_{i+1}}|^2 SNR}\right), R_{S,r_i}, \log_2\left(1+SINR_{r_i}\right)\right),$$
$$R_{S,r_{i+1}} \leq C\left(|h_{S,r_{i+1}}|^2 SNR\right); \tag{21}$$

`else`

$$R_i \leq \min\left(R_{S,r_i}, \log_2\left(1+SINR_{r_i}\right)\right), R_{S,r_{i+1}} \leq C\left(\frac{|h_{S,r_{i+1}}|^2 SNR}{1+|h_{R1,R2}|^2 SNR}\right); \tag{22}$$

`end.`

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

$$R_L \leq \min\left(R_{S,r_L}, \log_2\left(1+SINR_{r_L}\right)\right). \tag{23}$$

The $SINR_{r_i}$ denotes the SINR for s_i, which is decoded, encoded and forwarded by relay r_i. The network capacity is therefore

$$C_{ach_{BLAST}} = \frac{1}{L+1}\max_{R_1\cdots R_L}\left\{\sum_{i=1}^{L} R_i\right\}. \tag{24}$$

The condition for interference free transmission discussed in Section III-E and can be expressed as:

$$C\left(\frac{|h_{R1,R2}|^2 SNR}{1+|h_{S,r_{i+1}}|^2 SNR}\right) \geq \min\left(C\left(|h_{S,r_i}|^2 SNR\right), \log_2\left(1+SINR_{r_i}\right)\right). \tag{25}$$

The rate for the ith $(1 \leq i \leq L)$ signal under this condition can be expressed as:

$$R_i \leq \min\left(C\left(|h_{S,r_i}|^2 SNR\right), \log_2\left(1+SINR_{r_i}\right)\right). \tag{26}$$

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:

$$C\left(|h_{S,r_i}|^2 SNR\right) \geq \log_2\left(1 + SINR_{r_i}\right), \tag{27}$$

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:

$$SINR_{r_i} \leq \left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right) SNR. \tag{28}$$

This further implies that the conditions in (25) and (27) are better suited to assist the VBLAST algorithm to achieve the rate in (19), than those in (14) and (16). However, these conditions also imply an increased signalling overhead between the source, relays and destination in order to obtain the required SINR information. Furthermore, we note that V-BLAST might only be applied to a slow fading scenario where 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:

$$\begin{aligned} C = \frac{1}{3} \times \min\Big(& C\left(|h_{S,R1}|^2 SNR\right), C\left(|h_{S,R2}|^2 SNR\right), \\ & C\left(\left(|h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2\right) SNR\right)\Big), \end{aligned} \tag{29}$$

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 maximum ratio combining to combine the signals received from all the four time slots in order to detect and decode them. The capacity achieved by this protocol can be written as:

$$\begin{aligned} C = \frac{1}{2} \times \min\Big(\quad & C\left(|h_{S,R1}|^2 SNR\right), C\left(|h_{S,R2}|^2 SNR\right), \\ & C\left(\left(|h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2\right) SNR\right)\Big), \end{aligned} \tag{30}$$

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:

$$\min\left(C\left(|h_{S,R1}|^2 SNR\right), C\left(|h_{S,R2}|^2 SNR\right)\right) \geq C\left(\left(|h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2\right) SNR\right) \tag{31}$$

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, 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 procotol II:

$$G \triangleq \frac{\frac{1}{L+1} \mathrm{E}\left[\log_2\left(\det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^H SNR\right)\right)\right]}{0.5 \times \mathrm{E}\left[C\left(\left(|h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2\right) SNR\right)\right]}, \tag{32}$$

where $\mathrm{E}[\bullet]$ denotes the expectation and we assume each $h_{a,b}$ is an identically, independent distributed (i.i.d), complex, zero mean Gaussian random variable with unit variance. 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

In this section we study further the diversity multiplexing tradeoff [33] for such 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 the results as follows:

Theorem 1: 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:

$$d(r) = 2(1 - \frac{L+1}{L}r)^{+} \tag{33}$$

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 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. However, a higher diversity gain up to $2L$ might be achievable by coding across $L+1$ time slots in this scenario.

V. Simulation results

In this section we make further comparison of the protocols for different network geometries in terms of achievable rates. We only compare classic protocol II with the proposed protocol. As mentioned previously, to achieve a 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 develop adaptive protocols, we provide three examples here, which might be the easiest to occur to in terms of link capacity: (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 to the analysis in this paper, condition (a) appears the *simplest* since it does not require knowledge of the relays to destination links. In the following we will only adopt (a) in the simulations. 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 case 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. It can be written as $h_{a,b} = v{x_{a,b}}^{-\gamma}10^{\zeta_k/10}$, where v is an i.i.d complex Gaussian random variable with unit variance, $x_{a,b}$ is the distance between the nodes a and b. The scalar γ denotes the path loss exponent (in this paper it is always set to 4). The lognormal shadowing term ζ_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 can be negligible compared with the source to relays links (i.e. near 0). 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 10000 channel realizations. It can be clearly seen from all three figures that the V-BLAST algorithm approaches the capacity bounds introduced for the proposed protocol in the 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

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

Note that one very important factor that impairs the capacity performance of the proposed protocol is the 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 alterative 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 topics will remain as interesting future work.

APPENDIX

PROOF OF *Theorem 1*

As mentioned in Section III.E, conditioned on relays correctly decoding 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 a $(m+1)\times m$ MIMO channel matrix $\mathbf{H}_m$ in the same form as in (5). Define v_0 as the exponential orders [9] of $1/|h_{S,D}|^2$ and v_k as the exponential orders of $1/|h_{r_k,D}|^2$. Furthermore, Let $\mathbf{M}_{m+1} = \mathbf{I} + \frac{1}{2}\Sigma_S\Sigma_n^{-1}$, where Σ_S and Σ_n denote the covariance matrices of the observed signal and noise components *at the receiver*, respectively. We assume 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

$$\begin{aligned} P_{PE|v_0,v_1} &\leq \det\left(\mathbf{I} + \frac{1}{2}\Sigma_{S_1}\Sigma_n^{-1}\right)^{-l} \\ &= \left(1 + \frac{1}{2}\rho\left|h_{S,D}\right|^2 + \frac{1}{2}\rho\left|h_{r_1,D}\right|^2\right)^{-l} \\ &\doteq \rho^{-l(\max\{1-v_0,1-v_1\})^+} \end{aligned} \tag{34}$$

where $\doteq$ denotes the exponential equality [33] and SNR is replaced by ρ for notation simplicity. We assume each s_i is transmitted with data rate R (bits per transmission time slot). Since the successive relaying protocol uses $(L+1)$ time slots to transmit L different signal symbols, the average transmission rate is $\overline{R} = \frac{L}{L+1}R$. We assume the average transmission rate changes as $\overline{R} = r\log\rho$ with respect to ρ, then it is easy to see $R = \frac{L+1}{L}r\log\rho$. Therefore, we have a total of $\rho^{\frac{L+1}{L}rl}$ codewords. Thus, the error probability can be bounded by

$$P_{E|v_0,v_1} \dot{\leq} \rho^{-l((\max\{1-v_0,1-v_1\})^{+}-\frac{L+1}{L}r)}. \tag{35}$$

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. The set should be defined as

$$O^{+} = \{(v_0,v_1) \in R^{2+} | (\max\{1-v_0, 1-v_1\})^{+} \leq \frac{L+1}{L}r\} \tag{36}$$

Then, for any error event belongs to non-outage set, we can choose l to make the probability for it 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

$$d_o(r) = 2(1-\frac{L+1}{L}r)^{+} \tag{37}$$

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

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

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

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

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

$$\det(\mathbf{I}+\Sigma_{S_m}\Sigma_n^{-1})\dot{\geq}\rho^{\max\{m(1-v_0)^+, m(1-v)^+\}}. \quad (39)$$

If we define $\det(\mathbf{I}+\Sigma_{S_m}\Sigma_n^{-1})\dot{=}\rho^{f(v_0,v_1,v_2)}$ and

$$\rho^{\max\{m(1-v_0)^+, m(1-v)^+\}}\dot{=}\rho^{g(v_0,v_1,v_2)} \quad (40)$$

we have

$$f(v_0,v_1,v_2)\dot{\geq}g(v_0,v_1,v_2), \qquad \forall(v_0,v_1,v_2)\in R^{3+} \quad (41)$$

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

$$O_f^+ = \{(v_0,v_1,v_2)\in R^{3+}|f(v_0,v_1,v_2)\leq \frac{L+1}{L}mr\} \quad (42)$$

where m denotes that m symbols are transmitted and the equivalent data rate $R=\frac{L+1}{L}mr\log\rho$. We define

$$O_g^+ = \{(v_0,v_1,v_2)\in R^{3+}|g(v_0,v_1,v_2)\leq \frac{L+1}{L}mr\} \quad (43)$$

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

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

$$\inf_{(v_0,v_1,v_2)\in O_g^+}(v_0+v_1+v_2) = 2(1-\frac{L+1}{L}r)^+ \quad (44)$$

Comparing (37) and (44), we can see the diversity gain achieved by a multiple access MIMO channel with channel matrix $\mathbf{H}_m$ $(m > 1)$is always larger than that in $\mathbf{H}_1$.

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

$$\prod_{i=1}^{n} \det(\mathbf{I} + \frac{1}{2}\Sigma_{S_{m_i}}\Sigma_n^{-1}) \dot{\geq} \rho^{\max\{(\sum_{i=1}^{n} m_i)(1-v_0)^+, (\sum_{i=1}^{n} m_i)(1-v)^+\}}$$

Define $\rho^{f_n(v_0,v_1,v_2)} \doteq \prod_{i=1}^{n} \det(\mathbf{I} + \frac{1}{2}\Sigma_{S_{m_i}}\Sigma_n^{-1})$ and $\rho^{g_n(v_0,v_1,v_2)} \doteq \rho^{\max\{(\sum_{i=1}^{n} m_i)(1-v_0)^+, (\sum_{i=1}^{n} m_i)(1-v)^+\}}$. It can be seen

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

Similarly, applying $O_{g_n}^+ = \{(v_0, v_1, v_2) \in R^{3+} | g(v_0, v_1, v_2) \leq \frac{L+1}{L}(\sum_{i=1}^{n} m_i) r\}$, we can have

$$\inf_{v_0, v_1, v_2 \in O_{f_n}^+} (v_0 + v_1 + v_2) \geq 2(1 - \frac{L+1}{L} r)^+$$

The determinant of matrix $(\mathbf{I} + \frac{1}{2}\Sigma_S\Sigma_n^{-1})$ can always be decomposed to the product of the determinants of several submatrices $(\mathbf{I} + \frac{1}{2}\Sigma_{S_{mi}}\Sigma_n^{-1})$. Therefore the error exponent is always larger than or equal to $2(1 - \frac{L+1}{L} r)^+$ and the proof is completed.

REFERENCES

[1] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperative diversity-Part I and II," *IEEE Trans. Commun*, vol. 51, 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, 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. Allerton Conf. Comm., Contr. and Comp.*, (Monticello, IL), Oct. 2005.

[7] E Zimmermann, P. Herhold, G. Fettweis, "On the performance of cooperative diversity protocols in practical wireless systems," IEEE VTC 2003-Fall.

[8] P.A. Anghel, G. Leus, M. Kavehl, "Multi-user space-time coding in cooperative networks," ICASSP'03, Apr. 2003.

[9] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol 51, no. 12 pp. 4152-4172, Dec. 2005.

[10] Y. Zhao, R. Adve and T. J. Lim, "Improving amplify-and-forward relay networks: Optimal power allocation versus selection", submitted to *IEEE. Trans. Wireless. Commun.*, Feb. 2006.

[11] Y. Zhao, R. Adve and T. J. Lim, "Outage probability at arbitrary SNR with cooperative diversity," *IEEE Commun. Lett.*, vol 9, no. 8, pp. 700-702, Aug. 2005.

[12] M. Janani et al, "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE. Trans. Signal. Process.*, vol. 52, no. 2, pp. 362-371, Feb. 2004.

[13] M. Costa, "On the Gaussian interference channel", *IEEE. Trans. Inf. Theory*, vol. 31, no. 5, pp. 607-615, Sep. 1985.

[14] B. Suard et al, "Uplink channel capacity of space-division-multiple-access schemes," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, Jul. 1998.

[15] T. Cover and J. A. Thomas, *Elements of information theory*, Wiley, 1991.

[16] R. Pabst et al, "Relay-based deployment concepts for wireless and mobile broadband radio," , vol. 42, no. 9, pp. 80-89, Sep. 2004.

[17] Y. Fan and J. S. Thompson, "MIMO configurations for relay channels: Theory and practice", submitted to *IEEE Trans. Wireless. Commun.*, Aug. 2005.

[18] A Bletsas, A Lippnian, DP Reed, "A simple distributed method for relay selection in cooperative diversity wireless networks, based on reciprocity and channel measurements", IEEE VTC 2005-Spring, 2005.

[19] J. Cho, Z. J. Haas, "On the throughput enhancement of the downstream channel in cellular radio networks through multihop relaying," *IEEE. J. Sel. Commun.*, vol. 22, no. 7, pp. 1206-1219, Sep. 2004.

[20] A. Bletsas et al, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Comm.* vol. 24, no. 3, Mar. 2006.

[21] G. J. Foschini et al, "Analysis and performance of some basic space-time architectures," *IEEE Journal on Selected Area in Communications*, Vol. 21, No. 3, April 2003.

[22] G. J. Foschini et al, "Simplified processing for high spectral efficiency wireless communications employing multi-element arrays," *IEEE J. Sel. Commun.*, vol. 17, no.11, pp. 1841-1852, 1999.

[23] S. Verdu and S. Shamai,"Spectral Efficiency of CDMA with Random Spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, Mar. 1999.

[24] P. A. Anghel and M. Kaveh, "On the performance of distributed space-time coding systems with one or two non-regenerative relays," *IEEE Trans. Wireless Comm.*, vol. 5, no. 3, pp. 682?C692, Mar. 2006.

[25] M. Janani et al., "Coded cooperation in wireless communications: space-time transmission and iterative decoding," *IEEE. Trans. Signal. Process.*, vol. 52, no. 2, pp. 362?C371, Feb. 2004.

[26] A. Stefanov and E. Erkip, "Cooperative coding for wireless networks," *IEEE Trans. Comm.*, vol. 52, no. 9, pp. 1470?C1476, Sept. 2005.

[27] T. E. Hunter and A. Nosratinia, "Diversity through coded cooperation," *IEEE Trans. wireless Comm.*, vol. 5, no. 2, pp. 283?C 289, Feb. 2006.

[28] Peyman Razaghi and Wei Yu: "Bilayer low-density parity-check codes for decode-and-forward in relay channels", submitted to *IEEE Trans. Inf. Thy*, September, 2006.

[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 Comm.*, 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, Mar. 2004.

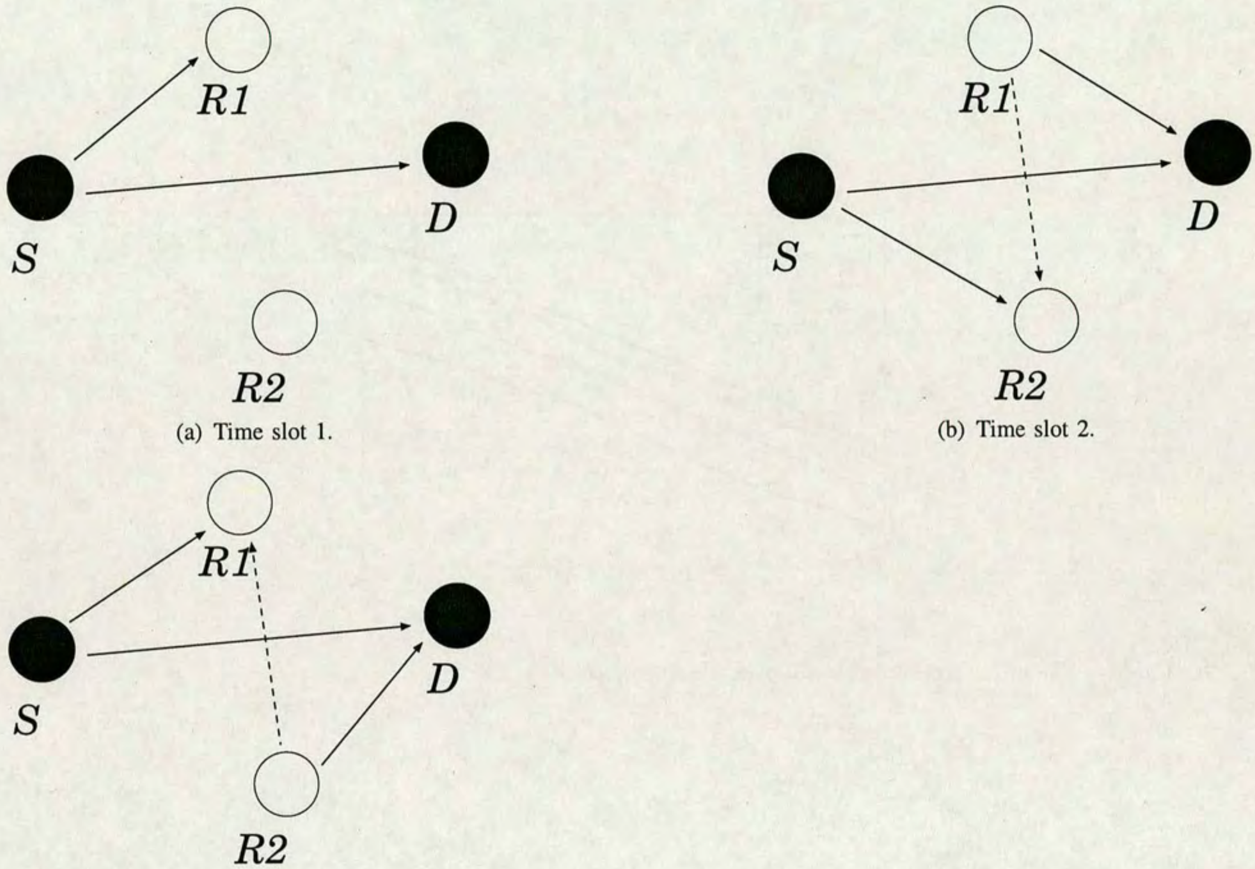
[31] Sheng Yang, Jeau-Claude Belfiore, "Towards the optimal Amplify and Forward Cooperative Diversity Scheme," *35th IEEE Comm. Thy. Workshop*, Dorado, Puerto Rico, May 2006.

[32] Sheng Yang, Jeau-Claude Belfiore, "On Slotted Amplify-and-Forward Cooperative Diversity Schemes", *ISIT*, 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, 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.



(a) Time slot 1.

(b) Time slot 2.

(c) Time slot 3.

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

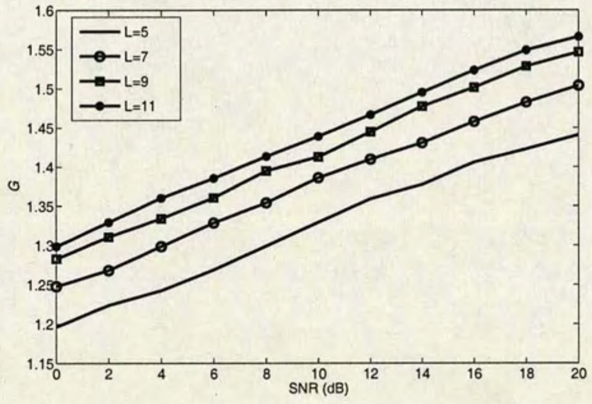


Fig. 2. Capacity gain of the proposed protocol over classic protocol II.

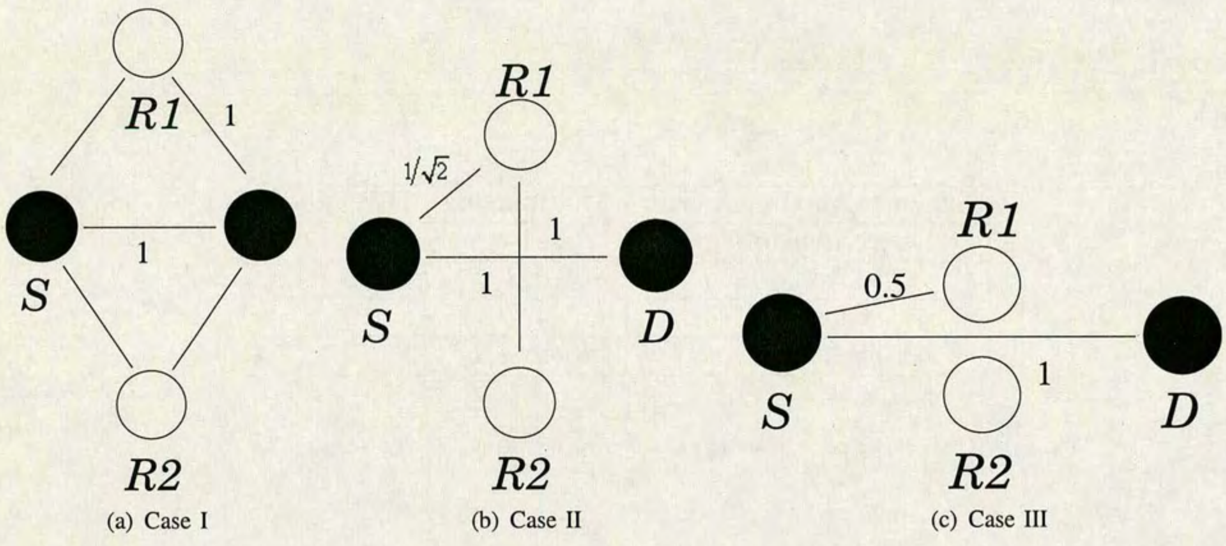
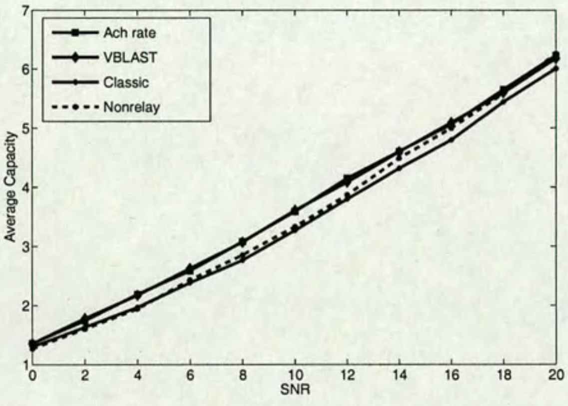
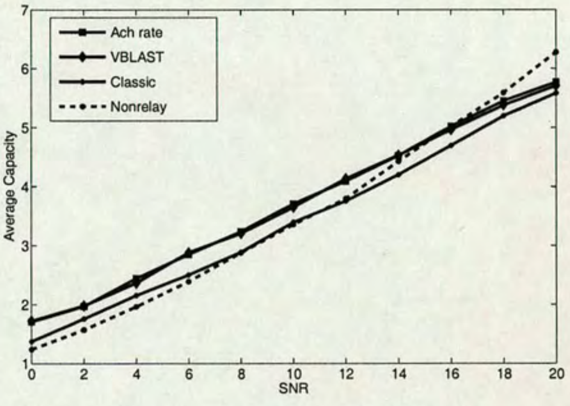


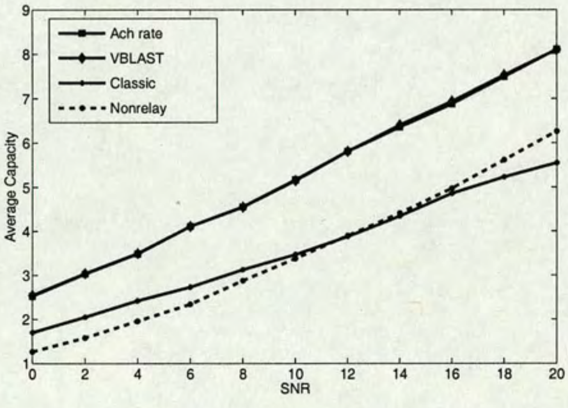
Fig. 3. Network models for different geometries.



(a) Case I



(b) Case II



(c) Case III

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

On the Optimum Threshold of Digital Cooperative Relaying Schemes

Furuzan Atay Onat[1], Abdulkareem Adinoyi[1], Yijia Fan [2], Halim Yanikomeroglu [1], and John S. Thompson[2]

[1]Broadband Communications and Wireless Systems (BCWS) Centre,
Department of Systems and Computer Engineering Carleton University, Ottawa, Canada
e-mail: {furuzan, adinoyi, halim}@sce.carleton.ca

[2]Institute for Digital Communications, School of Engineering and Electronics
University of Edinburgh, Edinburgh, UK
e-mail: {y.fan, john.thompson}@ed.ac.uk

Abstract— **In this paper we study the performance of threshold based digital relaying. We derive the end-to-end bit error rate minimizing threshold for a two-hop relaying network. In symmetric networks (or balanced networks) the average signal-to-noise ratio (SNRs) of all the relay links are assumed to be equal. Hence, previous literature on this topic tied the threshold selection to the average SNR of the source-relay channel in digital relaying. This paper reveals that in realistic relay network scenarios where, in general, the channel statistics are not identical, the optimal threshold is not a function of the source-relay average SNR. Instead, the optimal threshold depends only on average SNR of the relay-destination and source-destination channels. In addition to these findings, we perform end-to-end system simulations which collaborate the analytical developments in the paper.**

I. Introduction

There are two types of processing in the relay terminals. In analog relaying, relay does not make a hard detection about the data symbols or blocks. Instead, it amplifies the signal possibly along with some phase rotation. In digital relaying, symbols or blocks received from the source are detected and then regenerated possibly after decoding. If the relay detection is correct, the destination receives the signal through two branches and obtains diversity. However, if the relay has detection error, the effective SNR at the destination after combining is reduced enormously. This phenomenon is called *error propagation*.

The error propagation limits the performance of digital fixed protocol relaying. One way of reducing error propagation is by using embedded error detection codes such that once any errors are detected at the relay, the relay remains silent. While this idea is very effective in minimizing error propagation, it requires the relay to perform channel estimation, demodulation and then error detection for each data block before making a forwarding decision. These operations cause additional delay of the data even if the relay eventually decides not to transmit. Moreover, the relay node must be in either transmit or receive mode continuously during the data transfer. In wireless terminals, the amount of power consumed for receiving is negligible in comparison to the transmit power. However, this amount is significant in very low power devices such as sensor nodes.

A simpler way of reducing error propagation is to make forwarding decisions based on the received SNR at the relay. If the received SNR is large, the data is less likely to have errors. Then, a relay node can go to "sleep mode" for durations in the scale of the coherence time of the channels between the relay and its neighbors.

The choice of the threshold has considerable impact on the end-to-end performance of the cooperative diversity schemes. For instance, consider a relay detection threshold value of zero. This is akin to fixed protocol digital relaying. It is known that the error propagation limits the diversity gains and the diversity order of such a single cooperation relay network is limited to one [1]. On the other hand, for a very high value of threshold setting, the system degenerates to one path channel, which is the source to destination channel and the dual diversity is not realized. The threshold detection strategy prevents relay from forwarding erroneous signal to the destination, ensuring that when the relay forwards, reliable and independent signal is obtained at the destination thereby providing dual diversity in a single cooperation relay networks. It is apparent that the threshold has to be chosen carefully to balance between creating the required diversity branches to the destination and at the same time minimizing the risk of error propagation.

In [2], the authors use a SNR threshold to determine if the source-relay channel is reliable and propose a heuristic formula to calculate the value of this threshold. The work in [3] studies the performance of threshold relaying in a multi-antenna multi-relay architecture and shows the importance of the threshold especially when the relay has a small number of antennas. Recently, the problem of deciding whether using a particular relay is advantageous over direct transmission has been studied in the context of coded cooperation [4]. In contrast to the work in [4], we consider a setup where the relay is not aware of the channel coding and performs only symbol-by-symbol detection.

Most previous work consider symmetric networks where the average SNR of all the relay links are assumed to be equal. This assumption, though convenient for analysis, is not

This work was supported by Wireless Tech. Labs, Nortel Networks.

realistic in some cases, especially in multi-relay networks. In such symmetric networks SNR of all source-relay, relay-destination and source-destination are identically distributed. Hence, the threshold is usually presented as if it is determined by the source-relay average SNR. Our study indicates that, in contrast to what is commonly assumed in the literature, for arbitrary network configurations, the optimal threshold value is not determined by the source-relay channel rather, it is a function of the relay-destination and source-destination channels. Although the importance of threshold based relaying has been acknowledged in the literature, the issue of how to choose the optimal threshold has not been investigated comprehensively. Hence, the treatment presented in this paper fill an important gap by deriving a closed-form expression for the optimal threshold value to minimize average end-to-end performance.

II. System Model

In this paper, we consider a source S, a destination D and a relay R, assisting S-D communication. We assume a block fading channel model, where each channel stays the same during each block and the channel states at different blocks are i.i.d. Rayleigh distributed. γ_{sd}, γ_{sr} and γ_{rd} denote the instantaneous SNRs of S-D, S-R and R-D links. Due to Rayleigh fading, γ_{sd}, γ_{sr} and γ_{rd} are exponential r.v.s with expected values σ_{sd}^2, σ_{sr}^2 and σ_{rd}^2, respectively.

To facilitate the explanation, we assume a protocol with two time slots. In the first time slot the source transmits while the relay and the destination listens. In the second time slot, according to its decision, the relay either transmits or remains silent causing an idle time slot. As we mentioned earlier in Section I, our protocols make forwarding decisions based on the channel state information CSI (exact and/or statistics). Hence, relying on the assumption that the channel states remain the same for some time, one can skip the second time slot when the relay decides not to transmit.

We assume that exact CSI is available at the receiver side for all S-D, S-R and R-D channels. Besides, the relay has average CSI of S-D and R-D channels (i.e. σ_{sr}^2 and σ_{rd}^2).

For simplicity, we assume a BPSK modulation for all the transmissions. However, the derivations can be extended to M-PSK modulation with modest effort.

III. Static Relaying

In this protocol the decision of whether to use the relay or not is made based on the long-term channel statistics, which can include path loss and shadow fading. This protocol represents the case where the relaying decisions are made at a higher layer for a time-scale much larger than the time scale of Rayleigh fading. In other words, the relay is either on or off as long as there are no major changes in the environment. If the relay always retransmits, we call this protocol as *fixed relaying*. The end-to-end BER of fixed relaying is given by:

$$\mathrm{BER}_{e2e}^{fix} = \mathrm{BER}^{s\to r} P_{prop} + (1 - \mathrm{BER}^{s\to r})\, \mathrm{BER}_{coop} \quad (1)$$

The $\mathrm{BER}^{s\to r}$ and $\mathrm{BER}^{s\to d}$ are the bit error rates of S-R and S-D links and given by:

$$\mathrm{BER}^{s\to r} = \mathrm{BER}_{p2p}(\sigma_{sr}^2), \quad \mathrm{BER}^{s\to d} = \mathrm{BER}_{p2p}(\sigma_{sd}^2) \quad (2)$$

where $\mathrm{BER}_{p2p}(\sigma^2)$ denotes the bit error rate of point-to-point link using BPSK modulation under Rayleigh channel with an average SNR of σ^2, which is equal to [5]:

$$\mathrm{BER}_{p2p}(\sigma^2) = \frac{1}{2}\left(1 - \sqrt{\frac{\sigma^2}{1+\sigma^2}}\right) \quad (3)$$

The BER_{coop} denotes the probability of error at the destination after the maximal ratio combining (MRC) of the signals received from S and R, given the relay has correct detection. This is the well-known 2 branch diversity BER performance for Rayleigh fading [5]:

$$\mathrm{BER}_{coop}(\sigma_{sd}^2, \sigma_{rd}^2) = \begin{cases} \frac{1}{2}\left(1 - \sqrt{\frac{\sigma_{sd}^2}{1+\sigma_{sd}^2}}\right)^2 \left(1 + \frac{1}{2}\sqrt{\frac{\sigma_{sd}^2}{1+\sigma_{sd}^2}}\right), & \sigma_{sd}^2 = \sigma_{rd}^2; \\ \frac{1}{2}\left[1 - \frac{1}{\sigma_{rd}^2 - \sigma_{sd}^2}\left(\sigma_{rd}^2\sqrt{\frac{\sigma_{rd}^2}{1+\sigma_{rd}^2}} - \sigma_{sd}^2\sqrt{\frac{\sigma_{sd}^2}{1+\sigma_{sd}^2}}\right)\right], & \text{o.w.} \end{cases} \quad (4)$$

The P_{prop} denotes the *error propagation probability*, which the probability of bit error at the destination after MRC given that the relay has bit error. We use the following approximate formula to calculate P_{prop} for BPSK modulation:

$$P_{prop}(\sigma_{sd}^2, \sigma_{rd}^2) \cong \frac{\sigma_{rd}^2}{\sigma_{rd}^2 + \sigma_{sd}^2} \quad (5)$$

The derivation of (5) is given in the Appendix. Throughout this paper, this approximation is used for all analytical developments.

Let I represent the action chosen by the relay. If the relay transmits $I = 1$ and 0, otherwise. Let $\mathrm{BER}_{e2e}^{st}(\sigma_{sr}^2, \sigma_{sd}^2, \sigma_{rd}^2, I)$ denote end-to-end bit error rate of static relaying protocol for given channel statistics and forwarding decision I.

$$\begin{aligned} &\mathrm{BER}_{e2e}^{st}(\sigma_{sr}^2, \sigma_{sd}^2, \sigma_{rd}^2, I) \\ &= I\,\mathrm{BER}_{e2e}^{fix} + (1 - I)\mathrm{BER}^{s\to d} \end{aligned} \quad (6)$$

from which we can see that the minimum end-to-end error probability achieved by static relaying is:

$$\begin{aligned} \mathrm{BER}^{e2e,st}(\sigma_{sr}^2, \sigma_{sd}^2, \sigma_{rd}^2) = \min\Big\{ &\mathrm{BER}^{s\to d}, \\ &\mathrm{BER}^{s\to r} P_{prop} + (1 - \mathrm{BER}^{s\to r})\, \mathrm{BER}_{coop} \Big\} \end{aligned} \quad (7)$$

and the optimal strategy is to choose relaying if relaying results in a BER lower than the direct transmission.

IV. Dynamic Relaying

Since the relay knows the exact SNR realization of S-R channel, for each data block it can choose to retransmit or not, based on this information. We denote the action taken by the relay as I and $I(\gamma_{sr})$ represents relaying policy, which can be a function of instantaneous S-R SNR. Our objective is to minimize average end-to-end bit error rate of dynamic relaying, which is denoted by BER_{e2e}^{dyn}. We denote the end-to-end bit error rate of a given γ_{sr} realization and policy $I(\gamma_{sr})$ as $\mathrm{BER}_{e2e}^{dyn,I}(\gamma_{sr})$. Then the average end-to-end BER of a given policy $I(.)$ is:

$$\mathrm{BER}_{e2e}^{dyn,I} = \int p_{\gamma_{sr}}(\gamma_{sr}) \mathrm{BER}_{e2e}^{dyn,I}(\gamma_{sr}) \mathrm{d}\gamma_{sr} \quad (8)$$

where $p_{\gamma_{sr}}(.)$ represents the pdf of γ_{sr}. The minimum error achieved by all possible policies is:

$$\begin{aligned}\mathrm{BER}_{e2e}^{dyn} &= \min_{I(.)} \mathrm{BER}_{e2e}^{dyn,I} \\ &= \int p_{\gamma_{sr}}(t) \min_{I(.)} \mathrm{BER}_{e2e}^{dyn,I}(\gamma_{sr})\mathrm{d}t \qquad (9)\end{aligned}$$

From (9), we conclude that minimizing average end-to-end bit error performance is equivalent to minimizing the end-to-end bit error performance for each realization of γ_{sr}.

$$\begin{aligned}&\mathrm{BER}_{e2e}^{dyn,I}(\gamma_{sr}) = \\ &I(\gamma_{sr})Pr\{\text{e2e bit error } |\gamma_{sr}, I(\gamma_{sr}) = 1\} \\ &+(1 - I(\gamma_{sr}))Pr\{\text{e2e bit error } |I(\gamma_{sr}) = 0\}\end{aligned}$$

Then, the optimal strategy for the relay is:

$$I^*(\gamma_{sr}) = \begin{cases} 1, & Pr\{\text{e2e bit error } |\gamma_{sr}, I(\gamma_{sr}) = 1\} \\ & < Pr\{\text{e2e bit error } |I(\gamma_{sr}) = 0\}; \\ 0, & \text{o.w.} \end{cases} \qquad (10)$$

If the relay does not transmit, the end-to-end bit error probability for the block depends only on the direct channel:

$$Pr\{\text{bit error } |I(\gamma_{sr}) = 0\} = \mathrm{BER}^{s\to d} = \mathrm{BER}_{p2p}(\sigma_{sd}^2) \qquad (11)$$

If the relay transmits,

$$\begin{aligned}&Pr\{\text{e2e bit error } |\gamma_{sr}, I(\gamma_{sr}) = 1\} = \\ &\quad Pr\{\text{bit error at S-R link}|\gamma_{sr}\}P_{prop} \\ &\quad +Pr\{\text{no bit error at S-R link}|\gamma_{sr}\} \times \mathrm{BER}_{coop} \qquad (12)\end{aligned}$$

The probability of error of BPSK modulation given the signal to noise ratio is [5]:

$$Pr\{\text{bit error at S-R link}|\gamma_{sr}\} = Q(\sqrt{2\gamma_{sr}})$$

By substituting (11) and (12) in (10), we obtain:

$$I^*(\gamma_{sr}) = \begin{cases} 1, & \gamma_{sr} > \gamma_t; \\ 0, & \text{o.w.} \end{cases} \qquad (13)$$

where

$$\begin{aligned}\gamma_t &= \frac{1}{2}\left(Q^{-1}(\delta(\sigma_{sd}^2, \sigma_{rd}^2))\right)^2 \\ &= \left(\mathrm{erf}^{-1}(1 - 2\delta(\sigma_{sd}^2, \sigma_{rd}^2))\right)^2 \qquad (14)\end{aligned}$$

and Q^{-1} denotes the inverse Q function, erf^{-1} denotes the inverse error function and δ is defined as

$$\delta(\sigma_{sd}^2, \sigma_{rd}^2) \triangleq \frac{\mathrm{BER}_{p2p}(\sigma_{sd}^2) - \mathrm{BER}_{coop}(\sigma_{rd}^2, \sigma_{sd}^2)}{P_{prop}(\sigma_{rd}^2, \sigma_{sd}^2) - \mathrm{BER}_{coop}(\sigma_{rd}^2, \sigma_{sd}^2)}$$

From (13) and (14), it clear that the optimal relaying strategy is a simple threshold rule. The optimal threshold is a function of only two parameters: average SNR of S-D and R-D channels. We note that for any $\sigma_{sd}^2, \sigma_{rd}^2 > 0$

$$\mathrm{BER}_{coop}(\sigma_{sd}^2, \sigma_{rd}^2) < \mathrm{BER}_{p2p}(\sigma_{sd}^2) < P_{prop}(\sigma_{sd}^2, \sigma_{rd}^2)$$

Hence, $\delta(\sigma_{sd}^2, \sigma_{rd}^2) < 1$ and $Q^{-1}(\delta(\sigma_{sd}^2, \sigma_{rd}^2))$ is defined for any positive σ_{sd}^2 and σ_{rd}^2. Moreover, for positive σ_{sd}^2 and σ_{rd}^2 values (14) always gives positive threshold values.

1) End-to-end Bit Error Rate (BER) Calculation: For any threshold value we can compute the average end-to-end bit error probability.

$$\begin{aligned}\mathrm{BER}^{e2e} &= Pr\{\gamma_{sr} < \gamma_t\}\mathrm{BER}^{s\to d} \\ &\quad +Pr\{\gamma_{sr} \geq \gamma_t\}\Big(\mathrm{BER}_t\, P_{prop} \\ &\quad (1 - \mathrm{BER}_t)\mathrm{BER}_{coop}\Big) \qquad (15)\end{aligned}$$

where BER_t is the bit error rate in S-R link given that $\gamma_{sr} > \gamma_t$. From the formulation of [2]:

$$\begin{aligned}\mathrm{BER}_t(\gamma_t, \sigma_{sr}^2) &= Q(\sqrt{2\gamma_t}) \\ &\quad -e^{\gamma_t/\sigma_{sr}^2}\sqrt{\frac{\sigma_{sr}^2}{1+\sigma_{sr}^2}}\, Q\left(\sqrt{2\gamma_t(1 + \frac{1}{\sigma_{sr}^2})}\right) \qquad (16)\end{aligned}$$

By combining (2)-(5) and (14)-(16), one can analytically compute the average error performance of threshold based digital relaying.

V. Lower Bound for the BER of Digital Relaying Protocols

In this section, we study a hypothetical system where the relay can detect all the bit errors and transmits each symbol only if it is detected correctly. We call this protocol as the *genie-aided relaying* protocol. The end-to-end BER of the genie-aided relaying protocol is:

$$\mathrm{BER}_{e2e}^{gen} = \mathrm{BER}^{s\to r}\mathrm{BER}^{s\to d} + (1 - \mathrm{BER}^{s\to r})\mathrm{BER}_{coop},$$

where $\mathrm{BER}^{s\to r}$ and $\mathrm{BER}^{s\to d}$ are as given in (2).

VI. Numerical Results

In Fig. 1, we plot the optimal threshold values given by (14) in closed form and threshold values obtained through the numerical minimization of (15). We observe that the closed form results derived in this paper fully agree with the numerical results. It is also seen that the optimal threshold value is independent of the average SNR of the S-R link. Fig. 2 shows end-to-end bit error rates obtained from (15) using (20). We note that although the optimal threshold is independent of σ_{sr}^2, the performance achieved by this threshold is a function of σ_{sr}^2.

In order to understand the effect of relative channel qualities in the performance of static and dynamic relaying, we plot BER of these protocols as a function of the average SNR of S-R and R-D links. For this study, we keep the S-D link SNR constant at 10 dB. Fig. 3 shows the performance of various protocols including fixed relaying and no relaying (i.e. only S-D transmission) as a function of σ_{sr}^2 for a weak and a strong R-D link. We observe that if R-D link is relatively poor ($\sigma_{rd}^2 = 0$ dB), the performance is mainly constrained by this link and all relaying protocols approach to the fixed relaying performance. However, when R-D link is strong ($\sigma_{rd}^2 = 20$ dB) performance is determined by S-R link. Dynamic relaying is more efficient than static relaying since it can benefit from S-R link when it is reliable and prevent error propagation when it is unreliable.

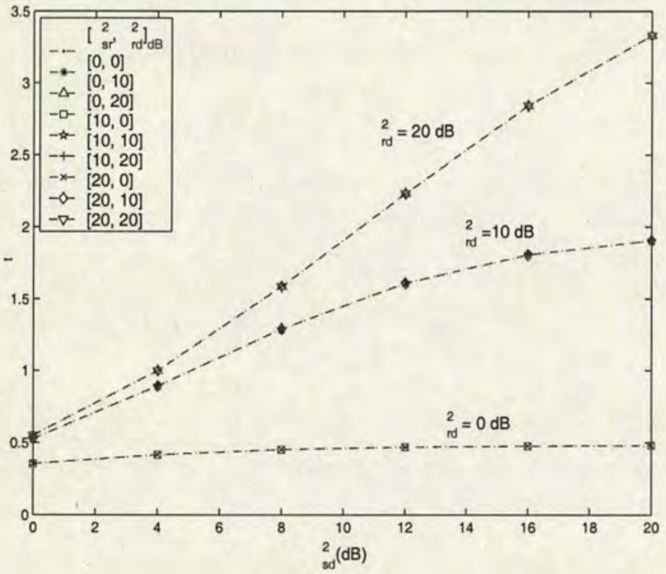


Fig. 1. Threshold values that minimize end-to-end BER given in (14) (dotted curves) and the thresholds obtained through numerical minimization of (15)(dashed curves) as a function of S-D SNR, for different S-R and R-D SNR values

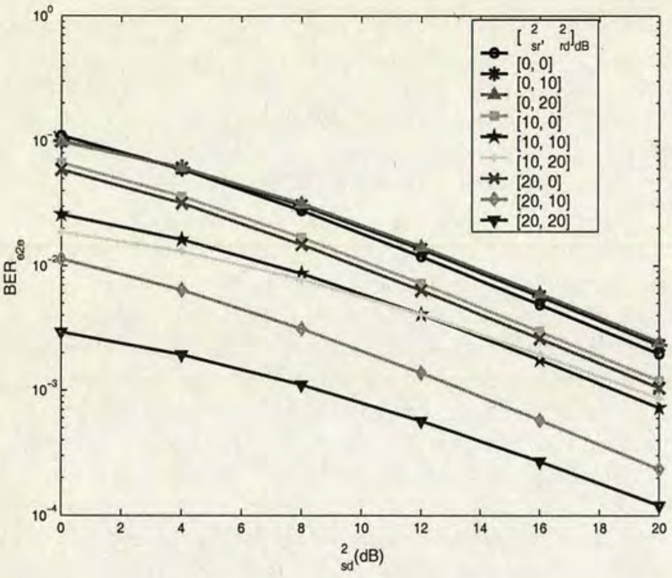


Fig. 2. End-to-end BER of threshold relaying from 15) using the optimal threshold given in (14), as a function of S-D SNR, for different S-R and R-D SNR values

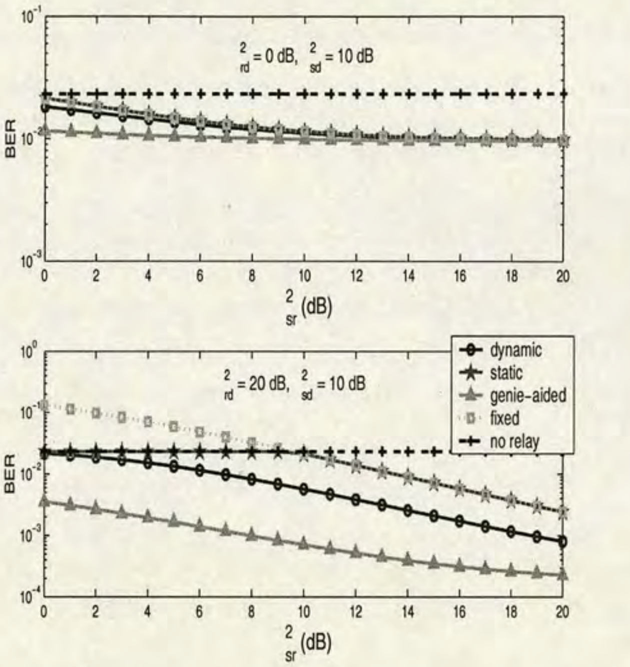


Fig. 3. End-to-end BER comparison of all protocols for fixed $\sigma_{sd}^2 = 10$ dB at two different σ_{rd}^2 values (0 and 20 dB) as a function of σ_{sr}^2.

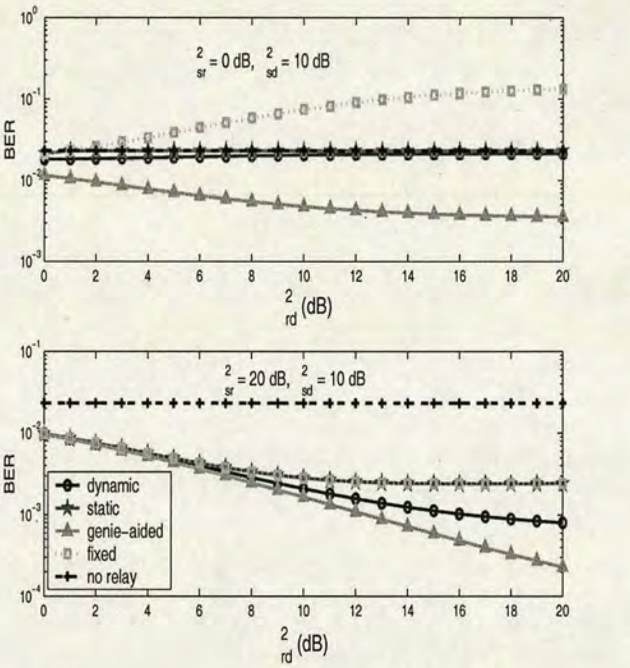


Fig. 4. End-to-end BER comparison of all protocols for fixed $\sigma_{sd}^2 = 10$ dB at two different σ_{sr}^2 values (0 and 20 dB) as a function of σ_{rd}^2.

In Fig. 4, we study the effect of σ_{rd}^2 when the S-R channel is relatively weak (0 dB) and strong (20 dB). If S-R link is poor, both static and dynamic relaying protocols keep the relay silent most of the time. Hence, the performance is independent of R-D link and close to the performance of no relay case. An interesting observation from Fig. 4 is that the performance of fixed relaying gets worse as R-D link becomes stronger due to linear increase in P_{prop}. In the second case, where S-R link is strong, error events must be rare at the relay. However, as R-D link gets stronger, they are dominant at the destination. Dynamic relaying can further decrease error propagation by increasing the threshold as R-D link becomes stronger, which provides a performance advantage over static relaying.

VII. Remarks and Conclusions

In this paper, we studied dynamic digital relaying for a source-destination node pair and a relay helping their communication. Relaying decisions are made based on the state of the wireless links among the three nodes. We assumed that the exact channel state information is available only at the receiver side for all three links while the relay has channel statistics of all the links.

Given that the relay has two possible actions, to retransmit the data or to remain silent, we showed that the optimal policy is a threshold rule on the received SNR at the relay. We derived the optimal threshold value in closed-form for i.i.d. Rayleigh channels. In contrast to the tendency in the literature to tie the threshold selection to the average SNR of the source-relay channel, we showed that the optimal threshold is a function of the average SNR of source-destination and relay-destination links.

We evaluated the BER performance of dynamic digital relaying and compared it to a static relaying protocol and a general performance upper bound for all digital relaying protocols. We showed that dynamic relaying using the optimal threshold value can take advantage of the source-relay channel adaptively providing significant performance advantage over static relaying.

APPENDIX

APPROXIMATION FOR P_{prop}

Since we assume BPSK modulation, without loss of generality, we assume that the source sends $+1$ symbol, the relay sends -1 symbol and an error occurs if the destination decides that -1 is send. The received signals from the source and the relay are denoted by y_1 and y_2, respectively.

$$y_1 = \alpha_{sd}\, x + n_1$$
$$y_2 = -\alpha_{rd}\, x + n_2$$

where α_{sd} and α_{rd} are the fading coefficients and n_1 and n_2 are i.i.d. Gaussian r.v.'s with zero mean and $N_0/2$ variance. After MRC, the decision variable is

$$\begin{aligned} y &= \alpha_{sd}^*\, y_1 + \alpha_{rd}^*\, y_2 \\ &= (|\alpha_{sd}|^2 - |\alpha_{rd}|^2)\, x + \alpha_{sd}^*\, n_1 + \alpha_{rd}^*\, n_2 \\ &= (|\alpha_{sd}|^2 - |\alpha_{rd}|^2) x + \tilde{n} \end{aligned}$$

where $\tilde{n}$ is the total noise whose power is equal to $E[|\tilde{n}|^2] = (N_0/2)(|\alpha_{sd}|^2 + |\alpha_{rd}|^2)$

Since the destination assumes that both S and R send the same symbol, the optimal decision rule at the destination is:

$$\hat{x} = \begin{cases} 1, & y \geq 0; \\ -1, & y < 0. \end{cases}$$

The probability of error given the fading coefficients is

$$\begin{aligned} Pr\{\text{bit error}|\alpha_{sd}, \alpha_{rd}\} &= Q\left(\frac{|\alpha_{sd}|^2 - |\alpha_{rd}|^2}{\sqrt{(N_0/2)(|\alpha_{sd}|^2 + |\alpha_{rd}|^2)}}\right) \\ &= Q\left(\frac{\gamma_{sd} - \gamma_{rd}}{\sqrt{(\gamma_{sd} + \gamma_{rd})/2}}\right) \end{aligned} \quad (17)$$

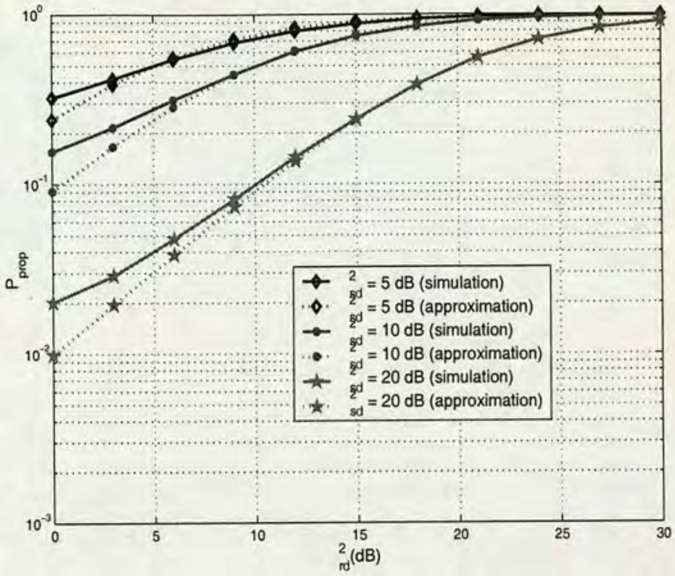


Fig. 5. Comparison of P_{prop} values obtained from the approximation in (20) and from simulation as a function of σ_{rd}^2 for different σ_{sd}^2 values.

where we substituted $\gamma_{sd} = |\alpha_{sd}|^2/N_0$ and $\gamma_{rd} = |\alpha_{rd}|^2/N_0$. Then, P_{prop} is given by

$$\begin{aligned} P_{prop} = \int\int & \left[Q\left(\frac{\gamma_{sd} - \gamma_{rd}}{\sqrt{(\gamma_{sd} + \gamma_{rd})/2}}\right) p_{\gamma_{rd}}(\gamma_{rd}) \right. \\ & \left. \times p_{\gamma_{sd}}(\gamma_{sd})\right] \mathrm{d}\gamma_{rd} \mathrm{d}\gamma_{sd} \end{aligned} \quad (18)$$

Due to the complexity of the exact expression given in (18), we use a high SNR approximation for P_{prop}.

If we assume $\gamma_{sd} >> 1$ and $\gamma_{rd} >> 1$ and $\gamma_{rd}/\gamma_{sd} = k$ $(0 < k < \infty)$, then

$$\begin{aligned} Q\left(\frac{\gamma_{sd} - \gamma_{rd}}{\sqrt{(\gamma_{sd} + \gamma_{rd})/2}}\right) &= Q\left(\sqrt{\gamma_{sd}}\left(\frac{1-k}{\sqrt{(1+k)/2}}\right)\right) \\ &\approx \begin{cases} 1, & k \geq 1; \\ 0, & k < 1. \end{cases} \end{aligned} \quad (19)$$

where, we use

$$Q(x) \approx \begin{cases} 1, & x >> 1; \\ 0, & x << 1. \end{cases}$$

Substituting (19) in (18), we obtain

$$\begin{aligned} P_{prop} &\approx Pr\{k > 1\} = Pr\{\gamma_{sd} < \gamma_{rd}\} \\ &= \int_0^\infty \int_0^{\gamma_{rd}} \frac{1}{\sigma_{sd}^2} \exp(-\gamma_{sd}/\sigma_{sd}^2) \\ &\qquad \times \frac{1}{\sigma_{rd}^2} \exp(-\gamma_{rd}/\sigma_{rd}^2) \mathrm{d}\gamma_{sd}\, \mathrm{d}\gamma_{rd} \\ &= \frac{\sigma_{rd}^2}{\sigma_{sd}^2 + \sigma_{rd}^2} \end{aligned} \quad (20)$$

To check the accuracy of this approximation at different σ_{sd}^2 and σ_{rd}^2 values, in Fig. 5 we plot probability of error propagation obtained both from simulations and from (20). We observe that the analytical approximation is for the σ_{rd}^2 and σ_{sd}^2 range used, which cover most scenarios in practice.

Acknowledgment

The authors would like to thank Dr. Ian Marsland, Carleton University, for the useful discussions during the initial phase of this work.

References

[1] N. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. on Information Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[2] P. Herhold, E. Zimmermann, and G. Fettweis, "On the performance of cooperative amplify-and-forward relay networks," in *Proceedings of ITG Conference on Source and Channel Coding (SCC)*, Jan. 2004.

[3] A. Adinoyi and H. Yanikomeroglu, "Multi-antenna aspects of wireless fixed relays," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC 2006)*, 2006.

[4] Z. Lin, E. Erkip, and A. Stefanov, "Cooperative regions and partner choice in coded cooperative systems," *IEEE Transactions on Communications*, vol. 54, July 2006.

[5] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2000.

On the Diversity-Multiplexing Tradeoff for Multi-antenna Multi-relay Channels

Yijia Fan , John S. Thompson
Institute for Digital Communications,
Joint Research Institute for Signal and Image Processing
School of Engineering and Electronics
The University of Edinburgh, Edinburgh, EH9 3JL, UK.
Email: y.fan@ed.ac.uk

Abdulkareem Adinoyi, Halim Yanikomeroglu
Broadband Communications and Wireless Systems centre
Department of Systems and Computer Engineering
Carleton University, Ottawa, K1S 5B6, Canada
Email: adinoyi@sce.carleton.ca

Abstract— **In this paper we analyze the performance of multiple relay channels when multiple antennas are deployed only at relays. Specifically, we investigate the simple repetition-coded decode-and-forward protocol and apply two antenna combining techniques at relays, namely maximum ratio combining (MRC) on receive and transmit beamforming (TB). We assume that the total number of antennas at all relays is fixed to N. With a reasonable power constraint at the relays, we show that the antenna combining techniques can exploit the full spatial diversity of the relay channels and can achieve the same diversity multiplexing tradeoff as achieved by more complex space-time distributed coding techniques, such as those proposed by Laneman and Wornell (2003).**

I. Introduction

It is widely believed that ad hoc networking [1] or multi-hop cellular networks [2] are important new concepts for future generation wireless systems, where either mobile or fixed nodes (often referred to as relays) are used to help forward the information to the desired user. One advantage of these structures is that it is possible to unite multiple relays in the network as a "virtual antenna array" to forward the information cooperatively, while appropriate combining at the destination realizes diversity gain. The diversity achieved in this way is often named as *user cooperation diversity* or *cooperative diversity* [4], as it mimics the performance advantages of multiple-input multiple-output (MIMO) systems [3] by exploiting the spatial diversity of the relay channels. The performance limits of distributed space-time codes, which can exploit cooperative diversity, are discussed in [4]–[7] for single-antenna relay networks. However, the design and implementation of practical codes that approach these limits is difficult and a challenging open area of research.

In this paper we exploit the spatial diversity of the relay channels in a way different from the space-time codes-based approach. We apply two kinds of antenna combining techniques at the relay, namely maximum ratio combining (MRC) [8] for reception and transmit beamforming (TB) [9] for transmission. Those techniques were often used in point-to-point single-input multiple-output (SIMO) or multiple-input single-output (MISO) wireless links, where either the transmitter or receiver is equipped with multiple antennas. It has been shown that MRC (TB) is able to achieve the information theoretic upper bound and optimal diversity multiplexing tradeoff of SIMO (MISO) systems [10]. In a relay context, we move the multiple antennas to the relays, while the source and the destination are only equipped with a single antenna. Our investigation is based on the repetition-coded decode-and-forward transmission [4], where each relay simply fully decodes the source message, re-encodes it with the same codebook as the source and forwards it to the destination. This method avoids any form of space-time coding or network coding and is easy to implement in practice. We analyze the performance of this system based on a slow fading scenario. More specifically, we exam the outage probability and the diversity multiplexing tradeoff of the network.

Note that one different assumption between our approach and the space-time coded approach is that we allow multiple antennas to be deployed at the relays. However, it will be shown later that the diversity gain that can be achieved by our approach is the same as that of the space-time coded approach proposed by Laneman and Wornell in [5], as long as the total number of antennas at all relays is fixed, regardless of the number of antennas at each relay and the number of relays. Thus, the same diversity gain can be achieved even when each relay is deployed with a single antenna. Therefore the application of our scheme is quite general.

The rest of this paper is organized as follows. In Section II, the basic system model and assumptions are introduced. Section III introduces the antenna combining techniques. The outage analysis is presented in section IV. Finally, conclusions are drawn in Section V.

II. System Model

We consider a two hop network model with one source, one destination and K relays. For simplicity we ignore the direct link between source and destination. The extension of all the results to include the direct link is straightforward. We assume that the source and destination are deployed with single antennas, while relay k is deployed with m_k antennas; the total number of antennas at all relays is fixed to N. This can be expressed as

$$\sum_{k=1}^{K} m_k = N. \tag{1}$$

We restrict our discussion to the case where the channels are slow, frequency-flat fading. The data transmission is over two time slots using two hops. In the first transmission time slot, the source broadcasts the signal to all the relay terminals. The input/output relation for the source to the kth relay is given by

$$\mathbf{r}_k = \sqrt{\eta}\mathbf{h}_k s + \mathbf{n}_k, \tag{2}$$

where $\mathbf{r}_k$ is the $m_k \times 1$ receive signal vector, and η denotes the transmit power at the source. The scalar s is the unit mean power transmit signal and $\mathbf{n}_k$ is the $m_k \times 1$ complex circular additive white Gaussian noise vector at relay k with identity covariance matrix $\mathbf{I}_{m_k}$. The vector $\mathbf{h}_k$ is the $m_k \times 1$ channel transfer matrix from source to the kth relay. The entries of $\mathbf{h}_k$ are identically independent distributed (i.i.d) complex Gaussian random variables with zero mean and unit variance. In the second hop, each relay processes its received signals and re-transmits them to the destination. The signal received at the destination can be written as:

$$y = \sum_{i=1}^{K} \mathbf{g}_k \mathbf{d}_k + n_d, \tag{3}$$

where the vector $\mathbf{g}_k$ is the channel matrix from kth relay to the destination, of which each entry is an i.i.d. complex Gaussian random variable with unit variance. The scalar n_d is the complex additive white Gaussian noise at the destination with unit variance. The vector $\mathbf{d}_k$ is the transmit signal vector at relay k, which should meet the total transmit power constraint:

$$\mathrm{E}\left[\|\mathbf{d}_k\|_F^2\right] \leq \frac{\eta m_k}{N}, \tag{4}$$

where $\|\bullet\|_F$ denotes the Frobenius norm. This power constraint means that the power is allocated at each relay in proportion to its number of antennas. For presentation simplicity we assume here that the total power at all relays is fixed to be η, i.e. the same as at the source. However, all the conclusions in the paper also hold when the total power at all relays is fixed to an arbitrary constant. We assume a coherent relay channel configuration context where the kth relay can obtain full knowledge of both the backward channel vector $\mathbf{h}_k$ and the forward channel vector $\mathbf{g}_k$. Note that the forward channel knowledge can be obtained if the relay-destination link operates in a Time-Division-Duplex (TDD) mode. For fair comparison, we also assume that for each channel realization, all the backward and forward channel coefficients for all N antennas remain the same regardless of the number of relays K. Fig. 1 shows the system model.

III. Antenna Diversity Techniques in Relay Channels

In this section we apply MRC and TB techniques to the system model described in section II. We assume that each

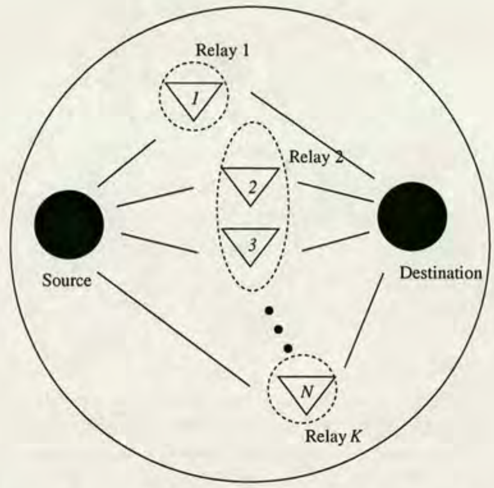


Fig. 1. System model for a two hop network: Source and destination are each deployed with 1 antenna. Totally N antennas are deployed at K relays. For each channel realization, either backward or forward channel coefficients for all N antennas remains the same regardless of the number of relays K.

relay performs MRC of the received signals, by multiplying the received signal vector by the vector $\mathbf{h}_k^H/\|\mathbf{h}_k\|_F$. The signal at the output of the relay receiver is given by

$$\tilde{r}_k = \sqrt{\eta \sum_{i=1}^{m_k} |h_{i,k}|^2} s + \frac{\sum_{i=1}^{m_k} h_{i,k}^* n_{i,k}}{\sqrt{\sum_{i=1}^{m_k} |h_{i,k}|^2}} \tag{5}$$

where $h_{i,k}$ denotes the ith antenna at relay k, and $n_{i,k}$ denotes the noise factor for ith receiver input branch. The SNR at the output of the receiver can be written as:

$$\rho_k^{m_k} = \eta \sum_{i=1}^{m_k} |h_{i,k}|^2. \tag{6}$$

After the relays decode the signals, each relay then performs TB of the decoded waveform. If we denote the transmitted signals as t_k with unit variance, the transmitted signal vector $\mathbf{d}_k$ for relay k can be written as

$$\mathbf{d}_k = \sqrt{\frac{\eta m_k}{N}} \frac{\mathbf{g}_k^H}{\|\mathbf{g}_k\|_F} t_k. \tag{7}$$

The destination receiver simply detects the combined signals from all K relays. If the signals are correctly decoded at all the relays (i.e. $t_k = s$), the output signal at the destination can be written as:

$$y = s \sum_{k=1}^{K} \sqrt{\frac{\eta m_k}{N} \sum_{i=1}^{m_k} |g_{i,k}|^2} + n_d = s \sum_{k=1}^{K} \tilde{g}_k + n_d \tag{8}$$

It can be seen from (8) that by applying antenna diversity schemes at the relays, the networks can be decomposed to K diversity channels, each with channel gain $\tilde{g}_k$. The output SNR at the destination receiver can therefore be written as:

$$\rho_d^{m_k} = \left(\sum_{k=1}^{K} \sqrt{\frac{\eta m_k}{N} \sum_{i=1}^{m_k} |g_{i,k}|^2} \right)^2 . \tag{9}$$

When all the relays are deployed with a single antenna, there is no traditional maximum ratio combining gain at the relays and the destination. However, the destination still observes a set of equal gain combined [12] amplitude signals from all relays.[1] Since we assume that the backward and forward channel coefficients for each antenna are kept the same for different values of K and m_i, the output SNR at the destination can be rewritten as

$$\rho_d^1 = \frac{\eta}{N} \left(\sum_{k=1}^{K} \sum_{i=1}^{m_i} |g_{i,k}| \right)^2 ; \tag{10}$$

when all the antennas are deployed in one relay (i.e. $K = 1$ and $m_1 = N$), full diversity gain is achieved among all the N antennas at the relay and also at the destination. The SNR for this case can be rewritten as

$$\rho_d^N = \eta \sum_{k=1}^{K} \sum_{i=1}^{m_i} |g_{i,k}|^2. \tag{11}$$

IV. Outage Analysis

When the channel fading is slow, i.e. codewords span less than one channel block, the Shannon capacity for the Rayleigh fading channels is zero. Therefore a certain outage probability must be allowed for communicating at any finite data rate. The outage probability can be defined as

$$P_{out} \triangleq P\left[C < R\right]. \tag{12}$$

This further allow us to exploit and investigate the diversity-multiplexing tradeoff of the systems, which is defined as follows [13]:

Definition 1: **(Diversity-Multiplexing Tradeoff)** Consider a family of codes C_η operating at SNR η and having rates R bits per channel use. The multiplexing gain and diversity order are defined as[2]

$$r \triangleq \lim_{\eta \to \infty} \frac{R}{\log_2 \eta}, \quad d \triangleq - \lim_{\eta \to \infty} \frac{\log_2 P_{out}(R)}{\log_2 \eta}. \tag{13}$$

We first study a simple protocol, in which all the relays participate in the decoding and forwarding process. We refer to this protocol as multi-cast decoding. An outage occurs whenever any relay or the destination fails to decode the signals. Before starting the outage analysis, we firstly introduce a lemma on the bounds of the value of $\rho_d^{m_k}$, i.e. the output SNR at the destination given that the signal is correctly decoded at all relays. This lemma has been shown in our earlier work [14]:

Lemma 1: For any m_k, $\rho_d^1 \le \rho_d^{m_k} \le \rho_d^N$.

We omit the proof, which can be found in Appendix of [14]. This Lemma is important throughout the analysis in the paper, as it implies that the increased "equal gain combining" gain at the destination can not compensate for the loss of maximum ratio combining gain at the relay and the destination when the number of relays K is increased and the numbers of antennas at each relay are reduced. Based on *Lemma 1*, we now begin our outage analysis with the following lemma:

Lemma 2: Conditioned on all the relays correctly decoding the messages, the outage probability for the relay channels is bounded by:

$$\frac{1}{N!} \left(\frac{N\left(2^{2R}-1\right)}{\eta} \right)^N \ge P_{out}^{m_k} \ge \frac{1}{N!} \left(\frac{2^{2R}-1}{\eta} \right)^N \tag{14}$$

Proof: See Appendix I. ∎

Lemma 2 indicates that the full diversity of N can be achieved regardless of the number of relays K, provided that the signals are correctly decoded at the relays. However, the diversity of the network might decrease if certain detection error happens at the relays. This is especially true for the multi-cast decoding protocol, for which we have the following theorem:

Theorem 1: For large η, the outage probability for the multi-cast decoding is bounded by:

$$N \left(\frac{2^{2R}-1}{\eta} \right) \ge P_{out} \ge \frac{2}{N!} \left(\frac{2^{2R}-1}{\eta} \right)^N \tag{15}$$

with equality to the right-hand side if $K = 1$, to the left-hand side if $K = N$.

Proof: See Appendix II. ∎

This theorem implies that for multi-cast decoding, having more relays and less antennas per relay actually loses diversity. Since requiring all the relays to fully decode the source information limits the performance of the decode and forward to that of the poorest source to relay link. Specifically, it can be seen that for $K = N$ no diversity gain is offered by relaying i.e. the SNR exponent is -1, as no diversity gain can be obtained from the source to relay links in this case. However, for $K = 1$ the full diversity of N can be achieved, as the diversity gain for the source to relay link is also N. In terms of the diversity multiplexing multiplexing tradeoff, we have the following theorem.

Theorem 2: The diversity-multiplexing tradeoff curve for the multi-cast decoding scheme is bounded by

$$1 - 2r \le d \le N\left(1 - 2r\right), 0 \le r \le 0.5 \tag{16}$$

with equality to right-hand side if $K = 1$, to left-hand side if $K = N$.

Proof: For large η, replace R with $r \log_2 \eta$ in (15), the proof is straightforward. ∎

It can be seen from Theorem 3 that when $K = N$, the diversity-multiplexing tradeoff for multi-cast decoding is strictly worse than that for direct transmission, which is $d = 1 - r$ [13]. When $K = 1$, however, the diversity-multiplexing tradeoff is the same as the space-time distributed coding schemes proposed in [5]. In fact, we can combine the antenna diversity schemes with a protocol *similar to* the one proposed by [5], which exploit further the diversity of source to relay channels by selecting the qualified relays

[1] Unlike [12], the equal gain combining for relay channels is applied at the transmitter instead of the receiver.

[2] we assume that the block length of the code is large enough, so that the detection error is arbitrarily small and the main error event is due to outage

that meet the transmission rate R, to improve the network performance when $K > 1$. Specifically, the protocol for the antenna diversity schemes is proposed as follows:

Protocol 1: **(Selection Decoding)** Select $\tilde{K}$ relays with a total number of antennas $\tilde{N}$, denoted as $\Re\left(\tilde{N},\tilde{K}\right)$, that could successfully decode the source message at a transmission rate R, to decode and forward the messages.

We can obtain the outage probability for the selection decoding by the following theorem:

Theorem 3: For large η, the outage probability for the selection decoding scheme is bounded by:

$$\left(\frac{2^{2R}-1}{\eta}\right)^N \times \sum_{\tilde{N}=1}^{N} \binom{N}{\tilde{N}} \frac{1}{\tilde{N}!} \tilde{N}^{\tilde{N}} \geq P_{out}^{m_k} \geq \left(\frac{2^{2R}-1}{\eta}\right)^N \sum_{\tilde{N}=1}^{N} \binom{N}{\tilde{N}} \frac{1}{\tilde{N}!\left(N-\tilde{N}\right)!}, \quad (17)$$

while the upper bound is met when the selected $\tilde{N}$ antennas are all within one relay.

Proof: See appendix III. ■

It can be seen from Theorem 4 that for selection decoding full diversity can always be achieved regardless the number of relays K. This is clearly an advantage over the multi-cast decoding scheme. Replacing R with $r\log_2\eta$ in (17), we can directly obtain the diversity-multiplexing tradeoff for selection decoding:

Theorem 4: The diversity-multiplexing tradeoff curve for selection decoding is

$$d = N(1-2r), 0 \leq r \leq 0.5, \quad (18)$$

which is the same as that for the space-time distributed coding protocol proposed in [5].

We claim that compared with distributed space-time coding, the messages for antenna combining techniques are simply repetition coded. Therefore it is much easier to implement than space-time coding in practice, provided that each relay antenna can obtain its forward (relay to destination) CSI. Fig. 2 shows the diversity multiplexing tradeoff for different protocols discussed in the paper, when $N = 5$.

V. Conclusions

From the above analysis, we can draw several conclusions regarding the antenna combining techniques introduced in the paper: (a) provided the messages are successfully decoded at the relays, having less relays will offer better performance due to increased combining (power) gain at the destination, though the full diversity N of the network can be achieved regardless the number of antennas; (b) if all the relays participate in the decoding and forwarding process, the network performance will degrade as the number of relays increases, as the performance is always restricted to the worst source to relay link. In this sense, deploying all the antennas at single relay is the optimal choice; (c) however, full diversity can be achieved if we apply the relay selection schemes to choose the potential relays. More specifically, the diversity-multiplexing tradeoff achieved by the antenna combining techniques is the same as that achieved by more complicated space-time distributed coding schemes. In this scenario, deploying more antennas at fewer relays is still a better choice due to improved combining (power) gain.

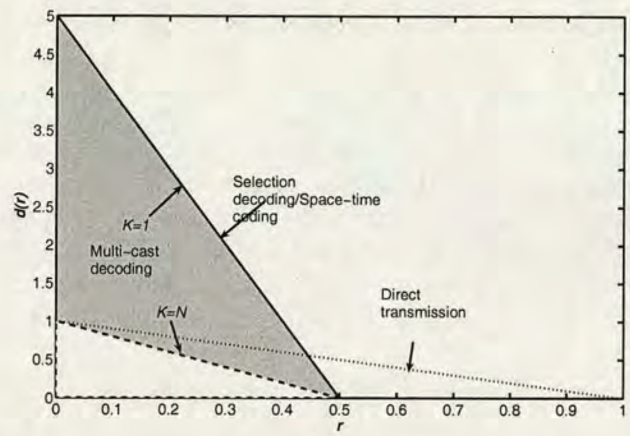


Fig. 2. The diversity multiplexing tradeoff for different protocols, when $N = 5$.

The analysis in the paper also implies that given a certain amount of available antennas in the network, the wired cooperation (i.e. all the antennas belong to one terminal) outperforms the wireless cooperation (i.e. each antenna belongs to different terminals). We further note that the recently proposed fixed relay concept [2] in mesh networks allows the possibility to deploy large number of antennas at the relay. This provide a good application for the antenna combining techniques discussed in the paper.

Appendix I
Proof of *Lemma 2*

Based on Lemma 1, it is clear that

$$P_{out}^1 \geq P_{out}^{m_k} \geq P_{out}^N, \quad (19)$$

where P_{out}^1 denotes the outage probability for N relay case and P_{out}^N for 1 relay case, given that the signals are correctly decoded at all the relays. Note that

$$\rho_d^1 \geq \frac{\eta}{N}\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2 = \frac{\rho_d^N}{N}, \quad (20)$$

inequality (19) can be extended as:

$$P\left[\frac{1}{2}\log_2\left(1+\frac{\rho_d^N}{N}\right) < R\right] \geq P_{out}^{m_k} \geq P\left[\frac{1}{2}\log_2\left(1+\rho_d^N\right) < R\right] \quad (21)$$

After some modification, (21) can be rewritten as:

$$P\left[\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2 \le \frac{N\left(2^{2R}-1\right)}{\eta}\right] \ge P_{out}^{m_k}$$
$$\ge P\left[\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2 \le \frac{2^{2R}-1}{\eta}\right]. \quad (22)$$

Since $\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2$ is chi-square distributed with dimension $2N$, for small ε it is easy to show that

$$P\left(\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2 \le \varepsilon\right) \approx \frac{1}{N!}\varepsilon^N. \quad (23)$$

Put (23) to (22) finishes the proof of *Lemma* 2.

Appendix II
Proof of Theorem 1

If we denote $C_r^{k,m_k} = 0.5\log_2\left(1+\rho_k^{m_k}\right)$ as the Shannon capacity from source to relay k channel for each channel realization. The outage probability is given by:

$$\begin{aligned} P_{out} &= P\left[\min\left(C_r^{k,m_k}\right) < R\right] + P\left[\min\left(C_r^{k,m_k}\right) > R\right]P_{out}^{m_k} \\ &= 1-\prod_{k=1}^{K}\left(1-P\left[\sum_{i=1}^{m_i}|h_{i,k}|^2 < \frac{2^{2R}-1}{\eta}\right]\right) \\ &\quad +P_{out}^{m_k}\prod_{k=1}^{K}\left(1-P\left[\sum_{i=1}^{m_i}|h_{i,k}|^2 < \frac{2^{2R}-1}{\eta}\right]\right) \\ &\overset{\eta\to\infty}{\approx} 1-\prod_{k=1}^{K}\left(1-\frac{1}{m_k!}\left(\frac{2^{2R}-1}{\eta}\right)^{m_k}\right) + P_{out}^{m_k}, \end{aligned} \quad (24)$$

where $P_{out}^{m_k}$ is bounded by (14). For large η, retaining only the term containing the lowest exponent of $1/\eta$ in the first term. (24) can be further modified as

$$P_{out} \approx \sum_{k=1}^{K}\frac{1}{m_k!}\left(\frac{2^{2R}-1}{\eta}\right)^{m_k} + P_{out}^{m_k}. \quad (25)$$

Observing that $\frac{1}{a!\eta^a} < \frac{1}{b!\eta^b}$ when $a > b$, P_{out} is maximized when $m_k = 1, K = N$. Therefore for large η

$$P_{out} \le N\left(\frac{2^{2R}-1}{\eta}\right), \quad (26)$$

where $P_{out}^{m_k}$ is omitted due to its higher exponent. P_{out} is minimized when $m_k = N, K = 1$ and $P_{out}^{m_k} = P_{out}^N$. We obtain the lower bound

$$P_{out} \ge \frac{2}{N!}\left(\frac{2^{2R}-1}{\eta}\right)^N \quad (27)$$

and thus complete the proof.

Appendix III
Proof of Theorem 2

Since $\Re\left(\tilde{N},\tilde{K}\right)$ is a random set, we utilize the total probability law and write

$$P_{out} = \sum_{\Re\left(\tilde{N},\tilde{K}\right)} P\left[\Re\left(\tilde{N},\tilde{K}\right)\right]P_{out}^{m_k|\Re\left(\tilde{N},\tilde{K}\right)}, \quad (28)$$

where $P_{out}^{m_k|\Re\left(\tilde{N},\tilde{K}\right)}$ denotes the outage probability conditioned on $\Re\left(\tilde{N},\tilde{K}\right)$ is chosen, and can be bounded by (14) by replacing N with $\tilde{N}$. The probability for any relay to be chosen can be expressed as:

$$\begin{aligned} P\left[r\in\Re\left(\tilde{N},\tilde{K}\right)\right] &= P\left[\sum_{i=1}^{m_k}|h_{i,k}|^2 \ge \frac{2^{2R}-1}{\eta}\right] \\ &= 1-P\left[\sum_{i=1}^{m_k}|h_{i,k}|^2 \le \frac{2^{2R}-1}{\eta}\right]. \end{aligned} \quad (29)$$

Therefore any $\Re\left(\tilde{N},\tilde{K}\right)$ exists with a probability that can be written as:

$$\begin{aligned} P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] &= \prod_{r\in\Re\left(\tilde{N},\tilde{K}\right)}\left(1-P\left[\sum_{i=1}^{m_k}|h_{i,k}|^2 \le \frac{2^{2R}-1}{\eta}\right]\right) \\ &\quad \times \prod_{r\in\Re\left(N-\tilde{N},K-\tilde{K}\right)} P\left[\sum_{i=1}^{m_k}|h_{i,k}|^2 \le \frac{2^{2R}-1}{\eta}\right] \end{aligned} \quad (30)$$

Based on (23), at high SNR, $P\left[\Re\left(\tilde{N},\tilde{k}\right)\right]$ can be approximated as

$$P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] \approx \left(\frac{2^{2R}-1}{\eta}\right)^{N-\tilde{N}} \prod_{r\in\Re\left(N-\tilde{N},K-\tilde{K}\right)}\frac{1}{m_k!}, \quad (31)$$

which can be bounded by:

$$\begin{aligned} \frac{1}{\left(N-\tilde{N}\right)!}\left(\frac{2^{2R}-1}{\eta}\right)^{N-\tilde{N}} &\le P\left[\Re\left(\tilde{N},\tilde{K}\right)\right] \\ &\le \left(\frac{2^{2R}-1}{\eta}\right)^{N-\tilde{N}}. \end{aligned} \quad (32)$$

Note that the bounds are independent of K. Putting (32) and (14) into (28), we obtain the bounds (17) and thus complete the proof.

References

[1] A. J. Goldsmith, S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks Wireless Communications", *IEEE Pers. Commun.*, Vol. 9, no. 4, pp. 8 - 27, Aug. 2002.

[2] R. Pabst et al, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, Sept. 2004.

[3] D. Gesbert, M. Shafi, Shiu Da-shan, P. J. Smith, A. Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 281 - 302, April 2003.

[4] 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*, to appear.

[5] 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.

[6] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol 51, no. 12 pp. 4152-4172, Dec. 2005.

[7] 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.

[8] J. G. Proakis, *Digital Communications: Fourth Edition*, 2001.

[9] J. Bach Andersen, "Antenna arrays in mobile communications: gain, diversity, and channel capacity", *IEEE Antennas Propagat. Mag.*, vol. 42, no. 2, pp. 12 - 16, April 2000.

[10] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005.

[11] W. Rhee, J. M. Cioffi, "On the capacity of multiuser wireless channels with multiple antennas", *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2580 - 2595, October 2003.

[12] Y. Chen, C. Tellambura, "Performance analysis of L-branch equal gain combiners in equally correlated Rayleigh fading channels", *IEEE Commun. Letters*, vol. 8, no. 3, pp. 150 - 152, March 2004.

[13] 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.

[14] Y. Fan, J. S. Thompson, A. B. Adinoyi and H. Yanikomeroglu, "Space diversity for multi-antenna multi-relay Channels", European Wireless Conference 2006, April 2 - 5 2006, Athens, Greece.

Recovering Multiplexing Loss through Successive Relaying

Yijia Fan , John S. Thompson
Institute for Digital Communications,
Joint Research Institute for Signal and Image Processing/ Integrated Systems/Energy
School of Engineering and Electronics
The University of Edinburgh, Edinburgh, EH9 3JL, UK.
Email: y.fan@ed.ac.uk

***Abstract*— In this paper, we develop a novel transmission protocol for a two relay wireless network to improve one of the fundamental tradeoffs in half-duplex relay networks, between multi-path diversity gain and multiplexing loss. We give an information-theoretic upper bound for this transmission scheme, and develop a space-time V-BLAST signalling and detection method which can approach this capacity upper bound. We show that our transmission scheme can recover the multiplexing loss of the half-duplex relay network, while retaining some diversity gain. We also compare it with conventional transmission protocols which only exploit the diversity of the network at the cost of a multiplexing loss. We show that our new transmission protocol offers significant performance advantages over conventional protocols, especially when the interference between the two relays is sufficiently strong.**

I. Introduction

In the past few years, cooperative diversity protocols [1]–[3] have been studied intensively to improve the diversity of relay networks, where nodes help each other by relaying transmissions. In most of the prior work, a time-division-multiple-access (TDMA) half-duplex transmission is assumed and relaying transmission protocols often use two TDMA time slots. The most popular transmission protocol (e.g. [1]) can be described as follows: for the first time slot, the source broadcasts the message to both the relays and the destination, the relays then forward the signals to the destination in the second time slot. Finally, the destination combines the message received from both time slots and performs decoding. We refer to this protocol as the *classic protocol* throughout the paper.

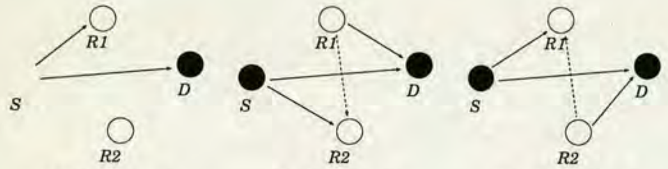
For digital relaying, where the relay decodes, re-encodes and forwards the message, repetition coding is often applied and discussed due to its low complexity compared with other coding methods. Diversity gain can be obtained to enhance the link reliability if the *classic protocol* is applied. However, when compared with direct transmission, repetition coded relaying will lose spectral efficiency for the high signal to noise ratio (SNR) region due to its bandwidth inefficiency. We can write the capacity for direct transmission as $\log_2(1+SNR)$ (bits per channel use), where SNR denotes the receive SNR at the destination for direct transmission. Then, the capacity for repetition coded relaying based on *classic protocols* can be expressed as $0.5\times\log_2(1+\kappa SNR)$, where κ denotes the link gain achieved due to relaying and 0.5 denotes the multiplexing loss due to the use of two time slots in the classic protocol. The capacity ratio of relaying to direct transmission can be expressed as:

$$G \triangleq \frac{0.5\log_2(1+\kappa SNR)}{\log_2(1+SNR)} \tag{1}$$

It is obvious to see that $G \approx \kappa/2$ when $SNR \rightarrow 0$; and $G \approx 0.5$ when $SNR \rightarrow +\infty$. This means that the classic protocol can improve link capacity only for the low SNR region (with a gain of $\kappa/2$). When the receive SNR for direct link transmission is high, the benefit for increasing the link reliability by relaying will not compensate for its multiplexing loss of 0.5.

In this paper, we propose a novel transmission protocol 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/array gain can still be obtained. Specifically, for L symbols transmitted in $(L+1)$ time slots with joint decoding at the destination, the system can be modeled as a Multiple-Input Multiple-Output (MIMO) system with L inputs and $L+1$ outputs if the signals are successfully decoded at the relays. It can offer multiplexing gain of at most $L/(L+1)$ and a diversity/array gain at most 2 compared with direct transmission. Thus it can make relaying more beneficial for the high SNR region while retaining diversity gain. We give a mutual information upper bound for this protocol, and also propose a practical low-rate feedback V-BLAST detection algorithm which approaches this upper bound. Based on our practical network models we show that the proposed protocol can give a significant performance advantage over the classic protocol, especially when the two relays are located close to each other.

Two recent conference papers [4] [5] present a *similar* relaying strategy to the one proposed in this paper. We would like to point out four major differences between our work and theirs: a) They ignore the direct (source to destination) link. In our paper we consider the impact of both the relay and direct links. b) In our model signal collision happens not only at the relay, but also at the destination. c) In our work we specifically study strong interference relay links and give specific conditions for an interference-free transmission scenario. d) By considering



(a) Time slot 1. (b) Time slot 2. (c) Time slot 3.

Fig. 1. Transmission schedule for the proposed protocol.

the effect of direct link, our protocol can be modified to become an adaptive protocol which can select between direct transmission and relay transmission strategies, depending on the current channel conditions.

The rest of the paper is organized as follows. In Section II, we propose the new transmission protocol. We analyze the capacity upper bound of this protocol in section III, and design a V-BLAST signalling and detection method to approach this upper bound in section IV. In section IV, we also introduce the classic protocol and compare it with the proposed protocol in some representative cases. More comprehensive comparisons based on different network geometries can be found in section V. Finally, we provide conclusions to the paper in section VI.

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 (packages), each containing L symbols denoted as s_l. These L symbols are transmitted continuously by the source, and decoded and forwarded by two relays successively in turn. Before decoding L symbols, the destination waits for $L+1$ transmission time slots until all L symbols are received, from both the direct link and the relay links. It then performs joint decoding of all L symbols. The specific steps for each transmission (reception) time slot for every frame are described as follows:

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

Time slot 2: S transmits symbol 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 symbol 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 process repeats till *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 symbols 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.

III. Capacity Upper Bound

We assume a slow, flat, block fading environment, where the channel remains static for each message frame transmission (i.e. $L+1$ time slots). We also assume that each transmitter transmits with equal power (i.e. no power allocation or saving among the source and relays). We denote $h_{a,b}$ as the channel coefficient between node a and b. For simplicity, we denote $C(\rho)$ the capacity function $\log_2(1+\rho SNR)$, where SNR denotes the ratio of signal power to the noise variance at the receiver.

A. Source-Relay Link

It is clear that 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

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

where r_i is the ith element in the L dimensional relay index vector

$$\mathbf{r} = \begin{bmatrix} R1 & R2 & R1 & R2 & R1 & \cdots \end{bmatrix}, \tag{3}$$

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 [7], 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 only concern ourselves 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.

For example, when $R1$ transmits s_1 while $R2$ is receiving s_2, if $|h_{R1,R2}| \succ |h_{S,R2}|$, $R2$ firstly 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 additional rate constraint for s_1 to be correctly decoded at $R2$, which can be expressed as follows:

$$R_1 \leq C\left(\frac{|h_{R1,R2}|^2}{1+|h_{S,R2}|^2}\right). \tag{4}$$

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

$$R_2 \leq C\left(\frac{|h_{S,R2}|^2}{1+|h_{R1,R2}|^2}\right). \tag{5}$$

Note that this decoding criterion applies from the second time slot to the Lth time slot when transmitting each frame. In slot i, equation (2) can be adapted to a constraint on R_{i-1}, equation (5) 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:

$$\mathbf{y} = \underbrace{\begin{bmatrix} h_{S,D} & 0 & 0 & 0 & 0 \\ h_{r_1,D} & h_{S,D} & 0 & 0 & 0 \\ 0 & h_{r_2,D} & h_{S,D} & 0 & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & h_{r_{L-1},D} & h_{S,D} \\ 0 & 0 & 0 & 0 & h_{r_L,D} \end{bmatrix}}_{\mathbf{H}} \mathbf{s} + \mathbf{n}, \quad (6)$$

where $\mathbf{y}$ is the $(L+1)\times 1$ receive signal vector, $\mathbf{s}$ is the $L \times 1$ transmit signal vector and $\mathbf{n}$ is the $(L+1)\times 1$ complex circular additive white Gaussian noise vector at the destination. Unlike conventional multiple access MIMO channels, the dimension of $\mathbf{y}$, $\mathbf{s}$ and $\mathbf{n}$ is expanded in the time domain rather than the space domain. However, the capacity region should be the same, which can be expressed as follows [8]:

$$R_k \le \log_2\left(\det\left(\mathbf{I} + \mathbf{h}_k \mathbf{h}_k^H SNR\right)\right), \quad (7)$$

$$R_{k_1} + R_{k_2} \le \log_2\left(\det\left(\mathbf{I} + SNR\left(\mathbf{h}_{k_1}\mathbf{h}_{k_1}^H + \mathbf{h}_{k_2}\mathbf{h}_{k_2}^H\right)\right)\right) \quad (8)$$

$$\cdots$$

$$\sum_{k=1}^{L} R_k \le \log_2\left(\det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^H SNR\right)\right), \quad (9)$$

where $\mathbf{h}_k$ denotes the kth column of $\mathbf{H}$. As it is extremely complicated to give an exact description for the rate region of each signal when $L > 2$, we will only concentrate on inequalities (7) and (9) 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 achievable when a space-time V-BLAST algorithm is applied at the destination to decode the signals.

D. Network Capacity Upper Bound

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. The specific steps are described as follows:

First, we impose a rate constraint R_i for each transmitted symbol s_i.

In the first time slot (initialization), we write:

$$R_{S,r_1} \le C\left(|h_{S,r_1}|^2\right) \quad (10)$$

For the $(i+1)$th time slot (for $1 \le i \le L-1$), 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` $h_{R1,R2} \succ h_{S,r_{i+1}}$,

$$R_i \le \min\left(C\left(\frac{|h_{R1,R2}|^2}{1+|h_{S,r_{i+1}}|^2}\right), R_{S,r_i}, C\left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right)\right), \quad (11)$$

$$R_{S,r_{i+1}} \le C\left(|h_{S,r_{i+1}}|^2\right); \quad (12)$$

`else`

$$R_i \le \min\left(R_{S,r_i}, C\left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right)\right), \quad (13)$$

$$R_{S,r_{i+1}} \le C\left(\frac{|h_{S,r_{i+1}}|^2}{1+|h_{R1,R2}|^2}\right); \quad (14)$$

`end`.

Note that the term $C\left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right)$ represents the constraint expressed by (7). The purpose of the `if` statement is to select the decoding order at the relay and to decide whether equation (4) or (5) is the correct constraint to apply.

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

$$R_L \le \min\left(R_{S,r_L}, C\left(|h_{S,D}|^2 + |h_{r_L,D}|^2\right)\right). \quad (15)$$

Combining these constraints with the sum capacity constraint expressed by (9), a network capacity upper bound for $L+1$ time slots can then be written as

$$C_{upper} = \min\left(\max_{R_1 \cdots R_L}\left\{\sum_{i=1}^{L} R_i\right\}, C_{MIMO}\right), \quad (16)$$

where $C_{MIMO} = \log_2\left(\det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^H SNR\right)\right)$. The first term in the min function comes from the calculation described above, the second one comes from equation (9). If the signal is correctly decoded and transmitted by the relays, the first term in (16) is omitted and the system mimics a MIMO system with L transmit antennas and $L+1$ receive antennas, for which the maximum multiplexing order is L. Compared with direct transmission over $L+1$ time slots which has multiplexing order of $L+1$, the multiplexing gain of relaying over direct transmission is $L/(L+1)$, which approaches 1 for large L. We also expect the diversity/array gain achieved by this protocol to be 2, since each signal transmission involves two independent fading channels (the direct link and a relay link).

E. Interference Free Transmission

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 [7]. We mentioned in the previous subsection that the interference channel between the relays mimics a Gaussian interference channel, so the same conclusion can be made if we use the interference cancellation criterion developed in the previous subsection. Specifically, if the interference between relays (or the value of $|h_{R1,R2}|$) is so large that the following inequality holds

$$\frac{|h_{R1,R2}|^2}{1+|h_{S,r_{i+1}}|^2} \geq \min\left(|h_{S,r_i}|, |h_{S,D}|^2 + |h_{r_i,D}|^2\right), i = 1 \cdots L \quad (17)$$

the relay can always correctly decode the the interference and subtract it before decoding the desired message, without affecting the whole network capacity. In this situation, the capacity analysis for ith ($1 \leq i \leq L$) transmitted signal as expressed by (11)-(15) can be simplified to

$$R_i \leq \min\left(C\left(|h_{S,r_i}|^2\right), C\left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right)\right). \quad (18)$$

It is obvious that the rate bounds provided by (18) is significantly larger than that provided by (11)-(15).

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. [1], [2]). 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 use relaying or not. For the interference free scenario discussed here, the constraint becomes:

$$|h_{S,r_i}|^2 \geq |h_{S,D}|^2 + |h_{r_i,D}|^2, 1 \leq i \leq L. \quad (19)$$

The capacity upper bound expressed by (16) can be simplified to

$$C_{upper} \leq \min\left(\sum_{i=1}^{L} C\left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right), C_{MIMO}\right). \quad (20)$$

By Jensen's inequality it is clear that

$$\sum_{i=1}^{L} C\left(|h_{S,D}|^2 + |h_{r_i,D}|^2\right) \geq C_{MIMO}. \quad (21)$$

Therefore the upper bound is equal to the MIMO channel capacity equation:

$$C_{upper} \leq C_{MIMO}. \quad (22)$$

This result shows that the proposed protocol can offer the best capacity performance conditioned on (17) and (19), which guarantees that the relays will correctly decode the message without affecting the network capacity.

It should be noted that this large interference scenario (i.e. condition (17)) is *not uncommon* in reality. When the two relays (e.g. mobiles) are located close to each other, if the routing techniques are develop to choose these relays, the capacity performance can be significantly improved by applying the proposed protocol. In a fixed relay network scenario [9], the source to relay links are often assumed to be significantly better than the corresponding relay to destination links and the direct link. Both (17) and (19) can be met by choosing the two nearby fixed relays.

IV. THE V-BLAST-MMSE ALGORITHM

The V-BLAST-MMSE algorithm was initially designed for spatial multiplexing MIMO systems [10]. For a system with M transmit antennas, the message at the transmitter is multiplexed into M different signal streams, each independently encoded and transmitted to the receiver. The receiver uses L antennas to detect and decode each signal stream by a V-BLAST MMSE detector. The detection consists of M iterations, each aimed at decoding one signal stream. The Shannon capacity of this system can be achieved if we assume that each signal is correctly decoded:

$$\begin{aligned} C &= \log_2 \det\left(\mathbf{I} + \mathbf{H}\mathbf{H}^H SNR\right) \\ &= \sum_{i=1}^{M} \log_2\left(1 + SINR_i\right), \end{aligned} \quad (23)$$

where $SINR_i$ is the output signal to interference plus noise ratio (SINR) for signal s_i in the V-BLAST-MMSE detector. In order for each signal to be correctly decoded, a low-rate feedback channel is suggested to feed the value of $SINR_i$ back to the transmitter [10]. Adaptive modulation and coding should be applied to make the transmission rate for s_i lower than $\log_2(1 + SINR_i)$.

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 time dimension rather than the space dimension. When considering the rate bound R_i, the same analysis should be made as in Section III. We summarize the capacity calculation process as follows:

The initialization step is the same as (10).

For $(i+1)$th time slot (for $1 \leq i \leq L-1$), the rate calculation can be performed as:

`if` $h_{R1,R2} \succ h_{S,r_{i+1}}$

$$\begin{aligned} R_i \leq \min\Bigg(& C\left(\frac{|h_{R1,R2}|^2}{1+|h_{S,r_{i+1}}|^2}\right), R_{S,r_i}, \\ & \log_2\left(1 + SINR_{r_i}\right)\Bigg), \end{aligned} \quad (24)$$

$$R_{S,r_{i+1}} \leq C\left(|h_{S,r_{i+1}}|^2\right); \quad (25)$$

`else`

$$R_i \leq \min\left(R_{S,r_i}, \log_2\left(1 + SINR_{r_i}\right)\right), \quad (26)$$

$$R_{S,r_{i+1}} \leq C\left(\frac{|h_{S,r_{i+1}}|^2}{1+|h_{R1,R2}|^2}\right); \quad (27)$$

`end.`

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

$$R_L \leq \min\left(R_{S,r_L}, \log_2\left(1 + SINR_{r_L}\right)\right). \quad (28)$$

The $SINR_{r_i}$ denotes the SINR for s_i, which is decoded, encoded and forwarded by relay r_i. The network capacity is therefore upper bounded by

$$C_{upper} = \max_{R_1 \cdots R_L} \left\{ \sum_{i=1}^{L} R_i \right\}. \quad (29)$$

The condition for interference free transmission discussed in Section III-E can be expressed as:

$$C\left(\frac{|h_{R1,R2}|^2}{1+|h_{S,r_{i+1}}|^2}\right) \geq \min\Big(C\left(|h_{S,r_i}|^2\right), \log_2\left(1+SINR_{r_i}\right)\Big). \quad (30)$$

The rate for the ith ($1 \leq i \leq L$) signal under this condition can be expressed as:

$$R_i \leq \min\left(C\left(|h_{S,r_i}|^2\right), \log_2\left(1+SINR_{r_i}\right)\right). \quad (31)$$

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:

$$C\left(|h_{S,r_i}|^2\right) \geq \log_2\left(1+SINR_{r_i}\right), \quad (32)$$

it is clear from (23) that (29) equals (22) under conditions (32) and (30). This implies that the V-BLAST algorithm can achieve the network capacity upper bound (22) for the protocol if the interference channel between relays and source to relay channels are sufficiently strong.

V. Comparison with Classic Protocols

A. Classic Protocol I

The first classic protocol was presented by Laneman and Wornell [2], 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:

$$C = \frac{1}{3} \times \min\Big(C\left(|h_{S,R1}|^2\right), C\left(|h_{S,R2}|^2\right), C\left(|h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2\right)\Big), \quad (33)$$

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

B. Classic Protocol II

An simple improvement of Classic protocol I is to apply distributed Alamouti codes at the relays [6]. 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 maximum ratio combining to combine the signals received from all the four time slots in order to detect and decode them. The capacity achieved by this protocol can be written as:

$$C = \frac{1}{2} \times \min\Big(C\left(|h_{S,R1}|^2\right), C\left(|h_{S,R2}|^2\right), C\left(|h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2\right)\Big), \quad (34)$$

it is clear that (34) outperforms (33) as it has less multiplexing loss compared with direct transmission.

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

$$\min\Big(C\left(|h_{S,R1}|^2\right), C\left(|h_{S,R2}|^2\right)\Big) \geq C\left(|h_{S,D}|^2 + |h_{R1,D}|^2 + |h_{R2,D}|^2\right) \quad (35)$$

when relaying is used. The network under this condition can achieve the best capacity performance (i.e. the third term in (33) and (34)). This result clearly mimics the performance of a 3×1 single-input multiple-output (SIMO) or multiple-input single-output (MISO) system, which obtains a diversity/array gain of 3 compared with direct transmission. The maximum diversity and multiplexing gains can be achieved by the two classic relaying schemes are compared with these of the proposed transmission scheme and direct transmission in Table I for the two relay case.

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

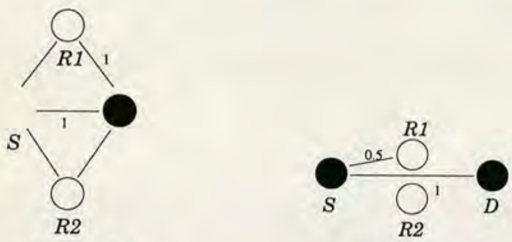
TABLE I

Comparison of the different transmission schemes for the two relay case

VI. Simulation results and Comparison

We compare classic protocol II with the proposed protocol. As mentioned previously, to achieve a 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 develop adaptive protocols, we provide three examples here. They are ordered by increasing implementation complexity: (a) $\min\left(|h_{S,R1}|, |h_{S,R2}|\right) \geq |h_{S,D}|$, i.e. the source to relay link is better than the direct link; (b) condition (19) holds or (c) condition (35) holds. Although (b) and (c) fits better to the analysis in this paper, condition (a) appears the *simplest* since it doesn't require knowledge of the relay to destination links. In the following we will only adopt (a) in the simulations. However, *similar* curve behaviors can be found if condition (b) or (c) is adopted.

Our simulations are based on two network geometries, which are shown in Fig.2. We assume that each channel coefficient $h_{a,b}$ contains Rayleigh fading (i.i.d complex Gaussian random variable with unit variance), pathloss (with exponent 4) and independent lognormal shadowing terms (with mean of



(a) Case I (b) Case II

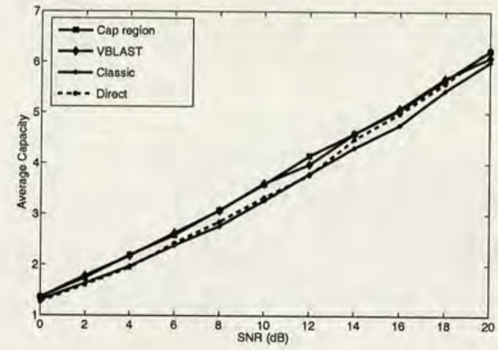
Fig. 2. Network models for different geometries.

0 dB and a standard deviation 8 dB). In case I, the distance between the source to relays and relays to destination are all normalized, the distance between the two relays is therefore $\sqrt{3}$. In case II, the relays are located in the middle region between the source and destination, so that the distance between the source and relays is 0.5 while the distance between the relays can be negligible compared with the source to relays links (i.e. near 0). For the proposed protocol, these two cases represent a tradeoff between the strength of source to relay channels and the interference channel between the two relays.

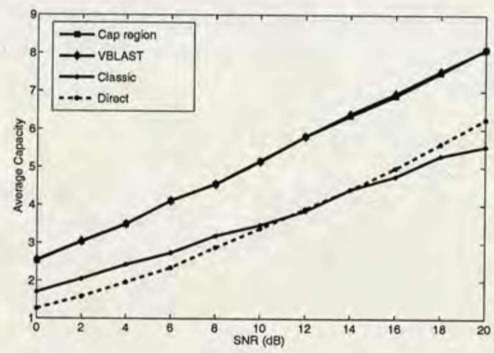
We assume $L = 7$ in the simulation, and the performance for the proposed protocol will certainly increase as L increases. Fig.3 shows the capacity bounds for the proposed protocols (cap region), the capacity achieved by V-BLAST MMSE detection (VBLAST), the classic protocols (classic) and direct transmission (direct), all averaged over 1000 channel realizations. It can be clearly seen from all three figures that the V-BLAST algorithm approaches the capacity bounds introduced for the proposed protocol in the paper.

Fig.3(a) implies that it is not helpful to implement relaying protocols when the source to relay link is about the same quality as the source to destination link, as the link gain due to relaying is small in this case. However, the proposed protocol still offers a performance gain over the direct transmission for both high and low SNR region in this case.

Compared with case I, in case II the source to relay links are much stronger, and the relays become very close to each other so that the interference is sufficiently strong to allow interference free transmission, which has been discussed previously in section III and IV. It can be clearly seen in Fig. 3(b) that the proposed protocol gives a significant performance advantage over direct transmission for both low and high SNR regions due to its array gain and negligible multiplexing loss. The classic protocol performs poorly at high SNR due to its significant multiplexing loss compared with direct transmission, although it achieves some performance gain over direct transmission for the low SNR region.



(a) Case I



(b) Case III

Fig. 3. Average capacity of the network for different network geometries.

VII. Conclusions

From the simulation study based on different geometries in the paper, we can draw two main conclusions: (a) For both the proposed and classic protocols, selecting the relays in the middle region between the source and destination seems a good choice to improve the average network capacity; (b) in this scenario, while the classic protocol still loses its performance advantage for high SNR region, the proposed protocol scheme can give significant performance advantages for both the low and high SNR region.

References

[1] 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, pp. 3062-3080, December 2004.

[2] 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.

[3] 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.

[4] 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.

[5] B. Rankov and A. Wittneben, "Spectral efficient protocols for non-regenerative half-duplex relaying," in *Proc. Allerton Conf. Comm., Contr. and Comp.*, (Monticello, IL), Oct. 2005.

[6] P.A. Anghel, G. Leus, M. Kavehl, "Multi-user space-time coding in cooperative networks," ICASSP'03, Apr. 2003.

[7] M. Costa, "On the Gaussian interference channel", *IEEE. Trans. Inf. Theory*, vol. 31, no. 5, pp. 607-615, Sep. 1985.

[8] B. Suard et al, "Uplink channel capacity of space-division-multiple-access schemes," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, Jul. 1998.

[9] R. Pabst et al, "Relay-based deployment concepts for wireless and mobile broadband radio," , vol. 42, no. 9, pp. 80-89, Sep. 2004.

[10] S. T. Chung *et al* "Approaching the MIMO Capacity with a Low-Rate Feedback Channel in V-BLAST", *Eurasip Jnl App Sig Proc*, Vol 5, pp. 762-771, 2004.

On the Performance of MIMO Spatial Multiplexing Relay Channels

Yijia Fan , John S. Thompson
Institute for Digital Communications,
Joint Research Institute for Signal and Image Processing/ Integrated Systems/Energy
School of Engineering and Electronics
The University of Edinburgh, Edinburgh, EH9 3JL, UK.
Email: y.fan@ed.ac.uk

Abstract—**In this paper we discuss and compare different signalling and relaying methods for multiple-input multiple-output (MIMO) relay networks in terms of network capacity, where every terminal is equipped with multiple antennas. We propose a new relaying mode called *hybrid relaying*, by which the MIMO relay channels can be decomposed into several point-to-point parallel channels without decoding the desired signals at the relay. We show that the proposed hybrid relaying algorithms outperform the conventional analogue relaying scheme. They can be good suboptimal choices compared with digital relying schemes, providing an attractive tradeoff between performance and complexity, especially when larger numbers of antennas are deployed at the relay than at the source and destination.**

I. Introduction

The use of multiple antennas at both ends of a wireless link, called multiple-input multiple-output (MIMO) technology, promises significant improvements in terms of spectral efficiency and link reliability. A large number of research projects has been conducted in the last decade on the point-to-point MIMO link (see [1] and references therein).

More recently, applying MIMO techniques into more advanced networks has also come under consideration. It is widely believed that the ad hoc networking [2] or multi-hop cellular networks [3] are important new concepts for future generation wireless systems. Either mobile or fixed nodes (often referred to as relays) are used to help forward the information to the desired user. One advantage of these structures is that it is possible to unite multiple relays in the network as a "virtual antenna array" to forward the information cooperatively, while appropriate signal combining at the destination realizes diversity gain. The diversity achieved in this way is often called *user cooperation diversity* or *cooperative diversity* [4], as it mimics the performance advantages of MIMO systems in exploiting the spatial diversity of the relay channels. The performance limits of space-time codes, which can exploit cooperative diversity, are discussed in [5]–[7] for single-antenna relay networks. For multiple-antenna or MIMO relay channels where every terminal in the network can be deployed with multiple antennas, studies are mainly concentrated on spatial multiplexing systems. Capacity bounds for single relay MIMO channels are presented in [8]. Quantitative capacity results for a multiple MIMO relay network has been reported in [9], where diversity is achieved again through cooperation among all the relays available in the network. The discussion is extended to multiple source-destination scenarios in [10], where the energy efficiency of the MIMO multiple relay network considering multiple source-destination pairs are further discussed.

Generally there are three kinds of relaying modes, the first two kinds are well known: *analogue relaying*, where the relays simply amplify the signals, and *digital relaying* where the relays decode, re-encode, and re-transmit the signals. We propose a new one for MIMO relay channels, which is called *hybrid relaying* (or *modified analogue relaying*), where the relays only decode the training sequence from the source or the feedback from the destination to obtain full channel state information (CSI) for either the source to relay (backward) or the relay to destination (forward) channels. Then the relay applies a special filter to the received signals based on the CSI without decoding them and retransmits the filter outputs. In this paper we propose an effective algorithm for *hybrid relaying* for single MIMO relay channels, We focus on MIMO spatial multiplexing systems where streams of independent data are transmitted over different antennas. We also assume that the relay can obtain the full CSI of both backward and forward channels.

The rest of this paper is organized as follows. In Section II, the basic system model and assumptions for MIMO relay channels is introduced. Section III introduces conventional signalling processing schemes (i.e. *analogue* and *digital* relaying) on single MIMO relay channels. The hybrid relaying schemes are discussed in section IV. Simulation results and performance comparisons can be found in section V. Concluding remarks will be found in Section VI.

A note on notation: We use boldface to denote matrices and vectors and $\mathrm{E}\left\langle \bullet \right\rangle_x$ for expectation regarding to x. $\det(\mathbf{X})$ denotes the determinant and $\mathbf{X}^{\dagger}$ denotes the pseudo-inverse of a matrix $\mathbf{X}$. $\mathbf{X}^H$ denotes the conjugate transpose and $\mathrm{tr}(\mathbf{X})$ denotes the trace. $\mathbf{I}$ denotes the identity matrix and $\|\bullet\|_F$ denotes the Frobenius-norm.

II. System Model

We consider a two hop network model with one source, one destination and one relay located in the middle region between source and destination. Note that the results in the paper are not necessarily restricted to this assumption. We ignore the direct link between source and destination due to large distance. We also assume that total transmit power for the source and relay are the same; for notational simplicity, we assume in the paper that the source and destination have the same number of transmit and receive antennas M, while the relay has $L \geq M$ antennas. The results can be somewhat extended to a more general case where different numbers of antennas are deployed at each transmitter or receiver.

We restrict our discussion to the case where the channels are slow, frequency-flat fading with a block fading model. The data transmission uses two hops over two equal duration time slots during which the channels all remain constant. In the first transmission time slot, the source transmits the signals to the relay. The input/output relation for source to relay link is given by

$$\mathbf{r} = \sqrt{\eta}\mathbf{H}\mathbf{s} + \mathbf{n}_r, \tag{1}$$

where $\mathbf{r}$ is $L \times 1$ receive signal vector. η denotes the power per transmit antenna at the source. The vector s is the $M \times 1$ transmit signal vector with covariance matrix $\mathbf{I}$ and $\mathbf{n}_r$ is the $L \times 1$ complex circular additive white Gaussian noise vector at the relay with identity covariance matrix $\mathbf{I}$. The matrix $\mathbf{H}$ is the $L \times M$ channel transfer matrix from the source to the relay, which can be further expressed as

$$\mathbf{H} = \sqrt{\alpha}\tilde{\mathbf{H}}, \tag{2}$$

where each entry of $\tilde{\mathbf{H}}$ is an identically independent distributed (i.i.d) complex Gaussian random variable with zero mean and unit variance. The factor α contains the pathloss and independent lognormal shadowing terms. It can be written as:

$$\alpha = x^{-\gamma}10^{\zeta/10}, \tag{3}$$

where x is the distance between the source and relay. The scalar γ denotes the path loss exponent (in this paper, always set to 4). The lognormal shadowing term, ζ, is a random variable with a normal distribution, mean of 0 dB and standard deviation δ dB. A value typical of shadowing deviations in urban cellular environments, $\delta = 8$ dB, is used for simulation studies. We normalized the range between the source and destination so that x is 0.5. Each relay processes their received signals and retransmits them to the destination. The signals received at the destination can be written as:

$$\mathbf{y} = \mathbf{G}\mathbf{d} + \mathbf{n}_d, \tag{4}$$

where the matrix $\mathbf{G}$ is the channel matrix from the relay to the destination, which might also be written as:

$$\mathbf{G} = \sqrt{\beta}\tilde{\mathbf{G}} \tag{5}$$

where each entry of $\tilde{\mathbf{G}}$ is an i.i.d. complex Gaussian random variable with zero mean and unit variance. β contains the same pathloss as α and independent lognormal shadowing terms with the same mean and standard deviation as for α. The vector $\mathbf{n}_d$ is the $M \times 1$ complex, circular additive white Gaussian noise at the destination with unit variance. The vector

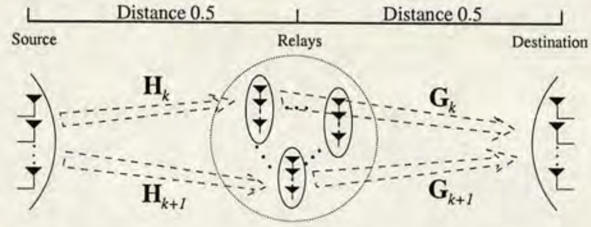


Fig. 1. Basic system model of a MIMO two hop network

$\mathbf{d}$ is the transmit signal vector at the relay, which should meet the transmit power constraint:

$$\mathrm{E}\left[\|\mathbf{d}\|_F^2\right] \leq \eta M. \tag{6}$$

Unless specifically stated, we assume that source does not know the channel information, but the destination knows all the channel information. Both knowledge of $\mathbf{H}$ and $\mathbf{G}$ are available at the relay.

III. Conventional Relaying Schemes

A. Digital Relaying (DR)

This scheme is also called *decode and forward relaying*, where the relay first uses L antennas to jointly decode the signals; it then demuxes and remixes the decoded message to L streams, and uses all L antennas to re-encode and retransmit them to the destination. The network capacity can therefore be expressed as:

$$C = 0.5 \min\left(\log\left(\mathbf{I} + \eta\mathbf{H}\mathbf{H}^H\right), \log\left(\mathbf{I} + \eta\frac{M}{L}\mathbf{G}\Sigma\mathbf{G}^H\right)\right) \tag{7}$$

where $\Sigma = \mathrm{Diag}\left\{\gamma^1, \cdots, \gamma^M\right\}$ denotes the digonal matrix generated from the iterative waterfilling algorithm conducted at the relay before retransmission [8].

$$\gamma^i = \left(\mu - \frac{L}{M\eta\lambda^i}\right)^+, \sum_{i=1}^{L}\gamma^i = M.$$

B. Analogue Relaying (AR)

In this scheme, the relay amplifies its received signal vector by:

$$\sqrt{\rho} = \sqrt{\frac{M \times \eta}{\eta\|\mathbf{H}\|_F^2 + L}} \tag{8}$$

to meet the power constraint described by (6) and forward it to the destination. The input-output relation from the source to destination can be expressed as

$$\mathbf{y} = \sqrt{\eta}\left(\sqrt{\rho}\mathbf{G}\mathbf{H}\right)\mathbf{s} + \left(\mathbf{G}\sqrt{\rho}\mathbf{n}_r + \mathbf{n}_d\right). \tag{9}$$

Thus we can treat the whole system as a point-to-point MIMO link. The network capacity can be written as:

$$C = 0.5\log_2\det\left(\mathbf{I} + \eta\rho\left(\mathbf{G}\mathbf{H}\right)^H\left(\mathbf{I} + \rho\mathbf{G}\mathbf{G}^H\right)^{-1}\mathbf{G}\mathbf{H}\right) \tag{10}$$

Practically this capacity can be achieved by using the VBLAST-MMSE algorithm at the destination to decode the signals [14]. Compared with digital relaying, one obvious defect for *AR* is that while the relays amplify the signals, they also amplify the receiver noise.

IV. Hybrid Relaying

In the *HR* schemes, the relays only decode the training sequence from source to obtain the full CSI, then filter the received signals based on knowledge of the CSI without decoding them. After multiplying the signal vector by the filtering weight matrix $\mathbf{W}$, the relay then amplifies and forwards the filtered signals to the destination. The amplifying factor now can be written as:

$$\sqrt{\rho}=\sqrt{\frac{M\times\eta}{\eta\left\|\mathbf{WH}\right\|_F^2+\left\|\mathbf{W}\right\|_F^2}}. \tag{11}$$

Noted that for *AR*, $\mathbf{W}=\mathbf{I}$. The input/output relation from the source to destination by *HR* can be expressed as:

$$\mathbf{y}=\sqrt{\eta}\left(\sqrt{\rho}\mathbf{GWH}\right)\mathbf{s}+\left(\mathbf{G}\sqrt{\rho}\mathbf{W}\mathbf{n}_1+\mathbf{n}_d\right), \tag{12}$$

where $\mathbf{n}_1$ is the white Gaussian noise vector at the relay. Thus we can treat the whole system as a point-to-point MIMO link. In the next subsections we propose some hybrid relaying schemes for MIMO relay channels.

A. Optimal Hybrid Relaying (OHR)

We now give a information-theoretic study on the optimal configuration for hybrid relaying based on the knowledge of both $\mathbf{G}$ and $\mathbf{H}$ at the relay.

The network capacity of hybrid relaying can be written as (13) shown on the top of next page.

We firstly replace $\mathbf{GW}$ with $\mathbf{M}$ in (13). we use singular value decomposition $\mathbf{M}=\mathbf{UDV}^H$ and $\mathbf{H}=\mathbf{U_H D_H V_H^H}$. Recall the identity

$$\det\left(\mathbf{I}+\mathbf{AB}\right)=\det\left(\mathbf{I}+\mathbf{BA}\right). \tag{15}$$

After some modification, the capacity can then be simplified to (14) shown on the top of next page, where $\Lambda=\mathbf{DD}^H=diag\left\{\lambda_1,\lambda_2,\cdots,\lambda_N\right\}$, contains the eigenvalues of $\mathbf{MM}^H$ and $\Lambda_{\mathbf{H}}=diag\left\{\lambda_{\mathbf{H}}^1,\cdots,\lambda_{\mathbf{H}}^N\right\}$ contains the eigenvalues of $\mathbf{HH}^H$. Using Hadamard's inequality, the capacity is maximized when

$$\mathbf{U_H}=\mathbf{V}. \tag{16}$$

Thus $\mathbf{M}=\mathbf{UDU_H^H}$. And (14) can be written as:

$$C{=}0.5\sum_{i=1}^{M}\log_2\left(1+\eta\lambda_{\mathbf{H}}^i\frac{\rho\lambda_i}{\rho\lambda_i+1}\right) \tag{17}$$

$$={0.5}\sum_{i=1}^{N}\log\left(1+\eta\lambda_{\mathbf{H}}^i\left(\frac{M\eta\lambda_i}{J+M\eta\lambda_i}\right)\right), \tag{18}$$

where $J=\eta\left\|\mathbf{WH}\right\|_F^2+\left\|\mathbf{W}\right\|_F^2$ is the signal energy received at the relay. For every fixed λ_i, maximizing C results from minimizing J. By replacing $\mathbf{W}$ with $\mathbf{W}=\mathbf{G}^\dagger\mathbf{M}$, it follows that

$$\begin{aligned}J&=\mathrm{tr}\left(\eta\mathbf{WHH}^H\mathbf{W}^H+\mathbf{WW}^H\right)\\&=\mathrm{tr}\left(\mathbf{G}^\dagger\mathbf{UD}\left(\eta\Lambda_{\mathbf{H}}+\mathbf{I}\right)\mathbf{D}^H\mathbf{U}^H\left(\mathbf{G}^\dagger\right)^H\right)\\&=\mathrm{tr}\left(\mathbf{U}^H\left(\mathbf{U_G}\Lambda_{\mathbf{G}}^\dagger\mathbf{U_G^H}\right)\mathbf{U}\left(\mathbf{D}\left(\eta\Lambda_{\mathbf{H}}+\mathbf{I}\right)\mathbf{D}^H\right)\right),\end{aligned} \tag{19}$$

while the identity $\mathrm{tr}\left(\mathbf{AB}\right)=\mathrm{tr}\left(\mathbf{BA}\right)$ is used. To minimize (19), for a unitary matrix $\mathbf{U}$, $N\times N$ Hermitian matrix $\mathbf{S}$ and diagonal matrix $\mathbf{\Sigma}$, we have the following identity [15]:

$$\mathrm{tr}\left(\mathbf{U}^H\mathbf{SU\Sigma}\right)\geq\sum_{i=1}^{N}a_ib_{N-i+1}, \tag{20}$$

where $a_1\leq\cdots\leq a_N$ are eigenvalues of $\mathbf{S}$ and $b_1\leq\cdots\leq b_N$ are the diagonal entries of Σ. So it is not hard to see that the minimum J should be in the form:

$$J=\sum_{i=1}^{M}\left(\eta\lambda_{\mathbf{H}}^i+1\right)\lambda_{\mathbf{G}^\dagger}^i\lambda_i, \tag{21}$$

which can be obtained by choosing

$$\mathbf{U}=\mathbf{U_G}, \tag{22}$$

where $\lambda_{\mathbf{H}}^i$ and $\lambda_{\mathbf{G}^\dagger}^i$ are the M non-zero eigenvalues of $\mathbf{HH}^H$ and $\left(\mathbf{GG}^H\right)^\dagger$. Note that ordering the values of $\lambda_{\mathbf{G}^\dagger}^i$ would give the opposite order to $\lambda_{\mathbf{G}}^i$, which are the eigenvalues of $\mathbf{GG}^H$.

To obtain the maximum C in (18), We should follow three steps:

Step 1: We calculate J in equation (21) as a function of λ_i for every ordering of $\lambda_{\mathbf{H}}^i$ and $\lambda_{\mathbf{G}^\dagger}^i$[1]. Thus totally $M!$ expressons for J are obtained.

Step 2: For every expression for J, we evaluate (18). Then the capacity becomes a function of λ_i. We then calculate the maximum value of this function. Totally there are $M!$ maximum values. We denote each one of them as $C_{\max}^i$.

Step 3: The final optimal C_{opt} is obtained as $\max\left(C_{\max}^1,\cdots,C_{\max}^{M!}\right)$.

A closed form solution for each value of $C_{\max}^i$ might be extremely complicated, as we shall first obtain the optimum relation between each λ_i by solving the M differential equations $\partial C/\partial\lambda_i=0$. Instead we might calculate $C_{\max}^i$ numerically (e.g. by *fminbnd* function in Matlab). The calculation complexity for J is $M!$, which is also extremely high for $M>2$. However, this method gives us the theoretical upper bound for MIMO coherent relay channels when using hybrid relaying schemes.

Another point is that in order to optimize C the matrix $\mathbf{GW}$ should be in the form of $\mathbf{U_G D U_H^H}$. The MIMO relay channel can thus be decomposed to several parallel channels each with gain $\lambda_{\mathbf{H}}^i\omega_i$, where $\lambda_{\mathbf{H}}^i$ is the source to relay channel gain and $\omega_i=\rho\lambda_i/(1+\rho\lambda_i)$ can be regarded as the relay to

[1] We conjecture from simulation results that ordering both $\lambda_{\mathbf{H}}^i$ and $\lambda_{\mathbf{G}}^i$ by decreasing size might result in the optimal value of C

$$C=0.5\log_2\det\left(\eta\rho\left(\mathbf{GWH}\right)^H\left(\mathbf{I}+\rho\mathbf{GW}\left(\mathbf{GW}\right)^H\right)^{-1}\mathbf{GWH}+\mathbf{I}\right) \tag{13}$$

$$=0.5\log_2\det\left(\eta\rho\mathbf{V_H}\mathbf{D_H^H}\mathbf{U_H^H}\mathbf{V}\mathbf{D}^H\mathbf{U}^H\left(\mathbf{I}+\rho\mathbf{U}\Lambda\mathbf{U}^H\right)^{-1}\mathbf{UDV}^H\mathbf{U_H}\mathbf{D_H}\mathbf{V_H^H}+\mathbf{I}\right)$$

$$=0.5\log_2\det\left(\eta\mathbf{V_H}\mathbf{D_H^H}\mathbf{U_H^H}\mathbf{V}\left(\frac{\rho\Lambda}{1+\rho\Lambda}\right)\mathbf{V}^H\mathbf{U_H}\mathbf{D_H}\mathbf{V_H^H}+\mathbf{I}\right)$$

$$=0.5\log_2\det\left(\eta\Lambda_{\mathbf{H}}^{1/2}\mathbf{U_H^H}\mathbf{V}\left(\frac{\rho\Lambda}{1+\rho\Lambda}\right)\mathbf{V}^H\mathbf{U_H}\Lambda_{\mathbf{H}}^{1/2}+\mathbf{I}\right) \tag{14}$$

---

destination channel gain. The value ω_i is optimized by the weight matrix $\mathbf{W}$ under a power constraint at the relay. In the following we now propose two practical suboptimal hybrid relaying schemes.

B. Suboptimal Hybrid Relaying Schemes

1) Modified Analogue Relaying (MAR): One simple way to make $\mathbf{GW}$ have the form of $\mathbf{U_G}\mathbf{D}\mathbf{U_H^H}$ is to make

$$\mathbf{W}=\mathbf{V_G}\mathbf{U_H^H}. \tag{23}$$

Then $\mathbf{GW}=\mathbf{U_G}\mathbf{D_G}\mathbf{U_H^H}$ and $\mathbf{GWH}=\mathbf{U_G}\mathbf{D_G}\mathbf{D_H}\mathbf{V_H^H}$. The capacity can be expressed as:

$$C=0.5\sum_{i=1}^{M}\log_2\left(1+\eta\lambda_{\mathbf{H}}^i\frac{\rho\lambda_{\mathbf{G}}^i}{\rho\lambda_{\mathbf{G}}^i+1}\right). \tag{24}$$

$$=0.5\sum_{i=1}^{M}\log_2\left(1+\eta\lambda_{\mathbf{H}}^i\frac{\eta M\lambda_{\mathbf{G}}^i}{\eta M\lambda_{\mathbf{G}}^i+\eta\sum\limits_{m=1}^{M}\lambda_{\mathbf{H}}^m+L}\right) \tag{25}$$

Compared with (17), *MAR* simply replace λ_i with $\lambda_{\mathbf{G}}^i$. To maximize C, the columns of $\mathbf{U_H}$ can be ordered to make $\lambda_{\mathbf{H}}^1\le\cdots\le\lambda_{\mathbf{H}}^i\le\cdots\le\lambda_{\mathbf{H}}^M$, with the same ordering used for the columns of $\mathbf{V_G}$. The amplifying factor of ρ can be written as:

$$\rho=\frac{M\eta}{\eta\left\|\mathbf{V_G}\mathbf{U_H^H}\mathbf{H}\right\|_F^2+\left\|\mathbf{V_G}\mathbf{U_H^H}\right\|_F^2}=\frac{M\eta}{\eta\left\|\mathbf{H}\right\|_F^2+L} \tag{26}$$

by the identity that

$$\left\|\mathbf{UA}\right\|_F=\left\|\mathbf{A}\right\|_F \tag{27}$$

for any unitary matrix $\mathbf{U}$. This is the same value for ρ as in *AR*. We thus name this scheme as modified analogue relaying. Compared with *AR*, it can be seen from (24) that the relay is able to decompose the channels and coordinate the backward channels with the forward channels to optimize the sum capacity for M parallel data streams.

2) Modified Matched Filter Relaying (MMFR): We have developed an algorithm called matched filtered relaying (MFR) [16] which uses $\mathbf{W}=\mathbf{H}^H$ when $\mathbf{G}$ is not known at the relay. Here we present a modified version of *MFR*, i.e. we design a new $\mathbf{W}$ based on the matched filter weight matrix $\mathbf{H}^H$. If we make $\mathbf{W}=\tilde{\mathbf{W}}\mathbf{H}^H$ and write the following:

$$\mathbf{GW}=\mathbf{U_G}\mathbf{D_G}\mathbf{V_G^H}\tilde{\mathbf{W}}\mathbf{V_H}\mathbf{D_H^H}\mathbf{U_H^H}. \tag{28}$$

we can see that to make $\mathbf{M}$ have the form $\mathbf{U_G}\mathbf{D}\mathbf{U_H}$, we can make

$$\tilde{\mathbf{W}}=\mathbf{V}_{\mathbf{G},M}\mathbf{V_H^H}, \tag{29}$$

and $\mathbf{D}$ becomes $\mathrm{Diag}\left(d_{\mathbf{G}}^1\lambda_{\mathbf{H}}^1,\cdots,d_{\mathbf{G}}^i\lambda_{\mathbf{H}}^i,\cdots,d_{\mathbf{G}}^M\lambda_{\mathbf{H}}^M\right)$, where $d_{\mathbf{G}}^i$ are the singular values of $\mathbf{G}$. The capacity can be written as

$$C=0.5\sum_{i=1}^{M}\log_2\left(1+\eta\lambda_{\mathbf{H}}^i\frac{\rho\lambda_{\mathbf{H}}^i\lambda_{\mathbf{G}}^i}{\rho\lambda_{\mathbf{H}}^i\lambda_{\mathbf{G}}^i+1}\right). \tag{30}$$

Also compared with (17), *MMFR* simply replaces λ_i with $\lambda_{\mathbf{H}}^i\lambda_{\mathbf{G}}^i$. To maximize C, the columns of $\mathbf{V_H}$ can be ordered to make $\lambda_{\mathbf{H}}^1\le\cdots\le\lambda_{\mathbf{H}}^i\le\cdots\le\lambda_{\mathbf{H}}^M$, with the same ordering used for the columns of $\mathbf{V}_{\mathbf{G},M}$. Compared with *MAR*, this scheme has the advantage of enhancing the SNR at the relay.

It should be noted that when M reduce to 1, equation (30) can be rewritten as:

$$C=0.5\times\log_2\left(1+\eta\lambda_{\mathbf{H}}^1\frac{\eta\lambda_{\mathbf{G}}^1}{\eta\lambda_{\mathbf{G}}^1+\eta\lambda_{\mathbf{H}}^1+1}\right). \tag{31}$$

It is not hard to see that (31) equals (18) if we replace J in (18) with the expression in (21) for $M=1$. This means that *MMFR* becomes the optimal hybrid relaying scheme for $M=1$. However, for $M\ge2$, the signals becomes more correlated to each other due to the matched filter factor $\mathbf{H}^H$ in the weight matrix $\mathbf{W}$, which impairs the sum capacity.

C. Waterfilling with CSI at the Source

For suboptimal hybrid relaying schemes, it is not difficult to see that waterfilling can be applied at the source if the source can obtain the CSI of $\mathbf{H}$ and $\mathbf{G}$. Here we give an analysis for *MAR* with waterfilling at the source. The capacity can be expressed as:

$$C=0.5\sum_{i=1}^{M}\log_2\left(1+\eta\gamma^i\lambda^i\right), \tag{32}$$

where

$$\gamma^i=\left(\mu-\frac{1}{\eta\lambda^i}\right)^+,\sum_{i=1}^{M}\gamma^i=M$$

and

$$\lambda^i=\lambda_{\mathbf{H}}^i\frac{\eta M\lambda_{\mathbf{G}}^i}{\eta M\lambda_{\mathbf{G}}^i+\eta\sum\limits_{m=1}^{M}\lambda_{\mathbf{H}}^m+L} \tag{33}$$

	CSI	Filtering	Non-linear detection	demux/remux
AR	No	No	No	No
HR	Yes	Yes	No	No
DR	Yes	Yes	Yes	Yes

TABLE I

COMPLEXITY COMPARISON AT THE RELAY FOR DIFFERENT RELAYING METHODS

V. COMPLEXITY COMPARISON

We have shown that compared with digital relaying, analogue relaying and hybrid relaying does not require any form of decoding of the desired messages at the relay. In practice this can significantly reduce the complexity of the relay system, especially for multiple antenna relay networks. Specifically, for digital relaying, to achieve the network capacity (7), the relay has to obtain full CSI and perform non-linear detection to detect and decode the messages, such as V-BLAST-MMSE detector. It also needs to apply a filter at each stage of the non-linear detection process. The processing complexity is very high when the multiplexing gain (i.e. the number of multiplexing streams) of the system is large, and will result in additional power and time latency at the relay. After decoding the multiple streams, the relay has to de-multiplex and re-multiplex the data stream for re-transmission, which might require breaking and reconstructing the data packets into a different layout. In practical data communications, this not only increases the complexity but also raises certain implementation issues (e.g. how to apply Automatic Repeat reQuest (ARQ) protocols to reconstructed packets). On the contrary, hybrid or analogue relaying does not require this process and is much simple to implement.

Compared with analogue relaying, hybrid relaying has to decode some training sequence (or feedback) to obtain the CSI, and needs a filter at the relay to refine the messages. However, it can be seen that by filtering the signals at the relays, the MIMO relay channel can be decomposed to several independent parallel channels. This will significantly reduce the detection complexity at the destination compared with analogue relaying, as each stream can be detected separately in parallel and no non-linear detector is required. A brief comparison of the complexity at the relay for different relaying methods is shown in Table 1.

VI. SIMULATION RESULTS AND COMPARISON

We calculate the average Shannon capacity (in bits per channel use) for 1000 channel realizations and define the total transmit power $P = M\eta$ at the source. Fig. 2(a) shows the performance results for different relaying schemes as a function of P when $L = M$. It can be seen that digital relaying performs best, while *MAR* outperforms *MMFR* specially for higher SNR. This might tell us that weakening the noise at relays can not compensate for the disadvantage of correlating the signals by *MMFR*. The hybrid relaying schemes give only small performance advantage over *AR*. In particular, the performance of *MMFR* is even worse than that of *AR* for high

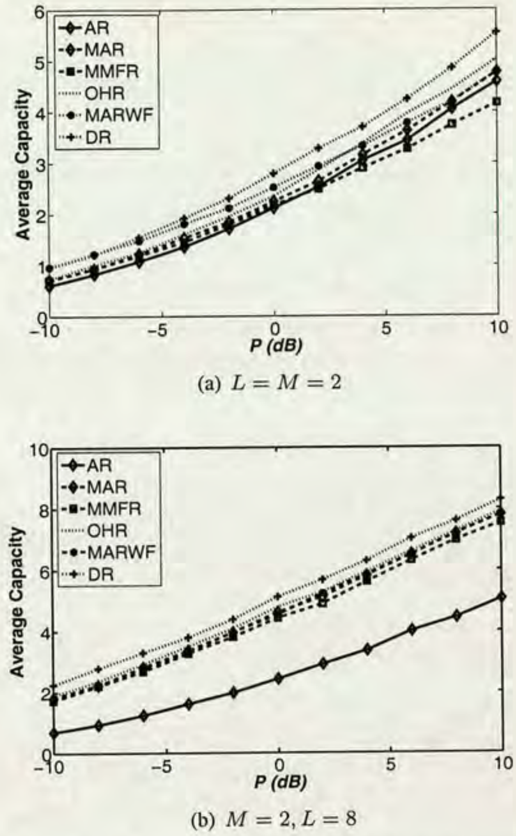


Fig. 2. Average capacity of single MIMO relay channels as a function of transmit power P. (a) $L = M = 2$. (b) $M = 2, L = 8$

SNR values due to signal correlation at the relay. It can also be seen that with the channel knowledge at the source, *MAR* with waterfilling (*MARWF*) can offer quite similar performance to digital relaying for low SNR values, because the waterfilling method accounts for both the $\mathbf{H}$ and $\mathbf{G}$ channels.

Fig. 2(b) shows the simulation results when $L > M$. It can be seen that as the value of L increases, digital and hybrid relaying schemes provide significant capacity advantages over *AR* by exploiting the eigenvalues $\lambda^i_{\mathbf{G}}$ on the relay to destination channels. Digital relaying still performs best. However, the its advantage over hybrid relying schemes is much smaller than for $L = M = 2$. The performance advantage of *MARWF* is not obvious compared with Fig. 2(a). This is mainly because the receive SNR at the destination increases as the value of L increases, so the benefit of waterfilling algorithm becomes negligible [17].

VII. CONCLUSIONS

In this paper we study MIMO spatial multiplexing configurations for relay channels when both source to relay and relay to destination CSI are available at the relay. We show that the proposed hybrid relaying schemes outperform the conventional analogue relaying scheme, and can be good suboptimal choices compared with digital relaying schemes. They avoid non-linear detection, decoding and re-encoding of the desired message at the relay and thus providing an attractive tradeoff between performance and complexity. This is particularly true when larger numbers of antennas are deployed at the relay than at the source and destination.

ACKNOWLEDGMENTS

This work is kindly sponsored by Joint Research Institute for Signal and Image Processing, Edinburgh, and UK EPSRC Grant GR/S58782/01.

REFERENCES

[1] D. Gesbert, M. Shafi, Shiu Da-shan, P.J.Smith, A.Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Comm.*, vol. 21, no. 3, pp. 281 - 302, April 2003.

[2] A. J. Goldsmith, S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks Wireless Communications", *IEEE Pers. Commun.*, Vol. 9, no. 4, pp. 8 - 27, Aug. 2002.

[3] R. Pabst et al, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, Sept. 2004.

[4] J. N. Laneman, G. W. Wornell and D. N. C. Tse, "An efficient protocol for realizing cooperative diversity in wireless networks," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001.

[5] 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*, to appear.

[6] 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.

[7] 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.

[8] B. Wang, J. Zhang, and A. Host-Madsen, "On capacity of MIMO relay channel," *IEEE Trans. Inf. Theory*, 2004, submitted.

[9] H. Bolcskei et al, "Capacity Scaling Laws in MIMO Relay Networks," *IEEE Trans. Wireless Comm.*, Apr. 2004, submitted.

[10] O. Oyman and A. J. Paulraj, "Energy efficiency in MIMO relay networks under processing cost", in *Conf. Inf. Science and Systems*, The Johns Hopkins University, March 16-18, 2005.

[11] A. M. Tulino and S. Verdu, "Random Matrix Theory and Wireless Communications," *Foundations and Trends in Communications and Information Theory*, vol.1, no. 1, 2004.

[12] J. Bach Andersen, "Antenna arrays in mobile communications: gain, diversity, and channel capacity", *IEEE Antennas Propagat. Mag.*, vol. 42, no. 2, pp. 12 - 16, April 2000.

[13] S. M. Alamouti, "A simple transmit diverstiy technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451-1458, Oct. 1998.

[14] S. Verdu and S. Shamai,"Spectral Efficiency of CDMA with Random Spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, Mar. 1999.

[15] J. Dattorro, *Convex Optimization and Euclidean Distance Geometry*, Meboo Publishing USA, 2005.

[16] Y. Fan and J. S. Thompson, "On the outage Capacity of MIMO Multihop Networks", *IEEE Globecom 2005*, 28 November - 2 December 2005, St. Louis, MO, USA.

[17] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication", IEEE Transactions on Communications, Volume 46, Issue 3, pp. 357 - 366, March 1998.

Space Diversity for Multi-antenna Multi-relay Channels

Yijia Fan and John Thompson
Institute for Digital Communications
School of Engineering and Electronics
University of Edinburgh
Edinburgh, EH9 3JL, UK
Email: Y.fan@ed.ac.uk

Abdulkareem Adinoyi
and Halim Yanikomeroglu
Broadband Communications
and Wireless Systems (BCWS) centre
Department of Systems and Computer Engineering
Carleton University
Ottawa, K1S 5B6, Canada
Email: adinoyi@sce.carleton.ca

***Abstract*— In this paper we analyze the performance of multiple relay channels when multiple antennas are deployed only at relays. We apply two antenna diversity techniques at relays, namely maximum ratio combining (MRC) on receive and transmit beamforming (TB). We show that with K relays the network can be decomposed into K diversity channels each with a different channel gain, and that the signals can be effectively combined at the destination. We assume that the total number of antennas at all relays is fixed at N. If the total transmit power for all relays are the same as for the source and equally distributed among all the relays, the network capacity will be lower bounded by that of N relay channels each with single antenna, and upper bounded by that of single relay channels with N antennas.**

I. Introduction

It is widely believed that ad hoc networking [1] or multi-hop cellular networks [2] are important new concepts for future generation wireless systems [3], where either mobile or fixed nodes (often referred to as relays) are used to help forward the information to the desired user. One advantage of these structures are that it is possible to unite multiple relays in the network as a "virtual antenna array" to forward the information cooperatively, while appropriate combining at the destination realizes diversity gain. The diversity achieved in this way is often named as *user cooperation diversity* or *cooperative diversity* [4], as it mimics the performance advantages of multiple-input multiple-output (MIMO) systems [5] in exploiting the spatial diversity of the relay channels. The performance limits of space-time codes, which can exploit cooperative diversity, are discussed in [6]–[8] for single-antenna relay networks. For multiple-antenna relay channels where every terminal in the network can be deployed with multiple antennas, studies are mainly concentrated on spatial multiplexing systems [9]–[11].

In this paper we exploit the spatial diversity of the relay channels in a different way from space-time codes based approach. We apply two kinds of antenna combining techniques at the relay, namely maximum ratio combining (MRC) [12] for reception and transmit beamforming (TB) [13] for transmission. Those techniques were often used in point-to-point wireless links to enhance signal-to-noise ratio (SNR) at the output of the receiver by deploying multiple antennas at either transmitter or receiver. In a relay context, we move the multiple antennas to the relays, while the source and the destination are only deployed with a single antenna. Our investigation is based on *digital relaying*, where the relays decode, re-encode and re-transmit the signals. We show that the network with K relays can be decomposed into K diversity channels each with different channel gain, and the signals from all K branch can be effectively combined at the destination. We derive the capacity bounds for this signal combining techniques. Our analysis results can be applied to both ergodic capacity and outage capacity [14] performance.

The rest of this paper is organized as follows. In Section II, the basic system model and assumptions are introduced. Section III introduces the signal combining techniques. The capacity performance analysis are made in section IV. Section V presents and discusses simulation results and finally, conclusions are drawn in Section VI.

II. System Model

We consider a two hop network model with one source, one destination and K relays. We ignore the direct link between source and destination. We also assume that total transmit power of the source and relays are the same and that it is equally distributed among the relays. Each relay processes the received signals independently. We assume that the source and destination are deployed with single antennas, while relay k is deployed with m_k antennas. We assume that the total number of antennas at all relays is fixed to N. This can be expressed as

$$\sum_{k=1}^{K} m_k = N. \tag{1}$$

We restrict our discussion to the case where the channels are slow, frequency-flat fading. The data transmission is over two times slots using two hops. In the first transmission time slot, the source broadcasts the signal to all the relay terminals. The input/output relation for the source to the kth relay is given by

$$\mathbf{r}_k = \sqrt{\eta}\mathbf{h}_k s + \mathbf{n}_k, \tag{2}$$

where $\mathbf{r}_k$ is $m_k \times 1$ receive signal vector. η denotes the transmit power at the source. The s is the transmit signal with covariance 1 and $\mathbf{n}_k$ is the $m_k \times 1$ complex circular additive white Gaussian noise vector at relay k with identity covariance matrix $\mathbf{I}_{m_k}$. The vector $\mathbf{h}_k$ is the $m_k \times 1$ channel transfer matrix from source to the kth relay and can be further expressed as

$$\mathbf{h}_k = \sqrt{\alpha_k}\tilde{\mathbf{h}}_k, \tag{3}$$

where each entry of $\tilde{\mathbf{h}}_k$ are identically independent distributed (i.i.d) complex Gaussian random variables with unit variance. Each factor α_k contains the pathloss and can be written as $\alpha_k = x_k^{-\gamma}$, where x_k is the distance between the source and relay k. The scalar γ denotes the path loss exponent. In the second hop, each relay processes its received signals and re-transmits them to the destination. The signal received at the destination can be written as:

$$y = \sum_{i=1}^{K} \mathbf{g}_k \mathbf{d}_k + n_d, \tag{4}$$

where the vector $\mathbf{g}_k$ is the channel matrix from kth relay to the destination, which might also be written as:

$$\mathbf{g}_k = \sqrt{\beta_k}\tilde{\mathbf{g}}_k, \tag{5}$$

where each entry of $\tilde{\mathbf{g}}_k$ is an i.i.d. complex Gaussian random variables with unit variance. The scalar β_k also contains the pathloss from the kth relay to the destination. The scalar n_d is the complex additive white Gaussian noise at the destination with unit variance. The vector $\mathbf{d}_k$ is the transmit signal vector at relay k, which should meet the total transmit power constraint:

$$\mathrm{E}\left[\|\mathbf{d}_k\|_F^2\right] \leq \frac{\eta m_k}{N}, \tag{6}$$

where $\|\bullet\|_F$ denotes the Frobenius norm. We assume a coherent relay channel configuration context where the kth relay can obtain full knowledge of both backward channel vector $\mathbf{h}_k$ and forward channel vector $\mathbf{g}_k$. For fair comparison, we also assume that for each channel realization, either backward or forward channel coefficients for all N antennas remains the same regardless of the number of relays K. It will not be difficult to see that the conclusions on either Ergodic capacity or outage capacity in this paper also hold if we extend the discuss to a more general case where each antenna is fixed in the network. Fig. 1 gives a description for the system model.

III. Antenna Diversity Techniques in Relay Channels

In this section we apply MRC and TB techniques to the system model described in section II. We assume each relay performs MRC of the received signals, by multiplying the received signal vectors by the vector $\mathbf{h}_k^H/\|\mathbf{h}_k\|_F$. The signals at output of the relay receiver is given by

$$\tilde{r}_k = \sqrt{\eta \sum_{i=1}^{m_k} |h_{i,k}|^2} s + \frac{\sum_{i=1}^{m_k} h_{i,k}^* n_{i,k}}{\sqrt{\sum_{i=1}^{m_k} |h_{i,k}|^2}} \tag{7}$$

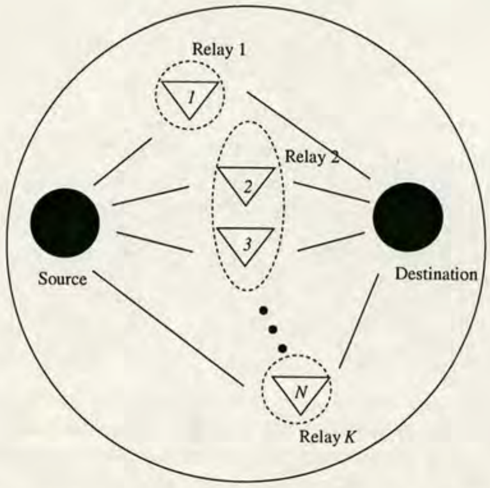


Fig. 1. System model for a two hop network: Source and destination are each deployed with 1 antenna. Totally N antennas are deployed at K relays. For each channel realization, either backward or forward channel coefficients for all N antennas remains the same regardless of the number of relays K.

where $h_{i,k}$ denotes the ith antenna at relay k, and $n_{i,k}$ denotes the noise factor for ith receiver input branch. The SNR at the output of the receiver can be written as:

$$\rho_k^{m_k} = \eta \sum_{i=1}^{m_k} |h_{i,k}|^2. \tag{8}$$

After the relays decode the signals, each relay then performs TB of the transmitted signals. If we denote the transmitted signals as d_k with unit variance, the transmitted signal vector $\mathbf{d}_k$ for relay k can be written as

$$\mathbf{d}_k = \sqrt{\frac{\eta m_k}{N}} \frac{\mathbf{g}_k^H}{\|\mathbf{g}_k\|_F} d_k. \tag{9}$$

The destination receiver simply detects the combined signals from all K relays. If we adjust the transmission data rate so that the signals are correctly decoded at all the relays (i.e. $d_k = s$), the output signal at the destination can be written as:

$$y = s\sum_{k=1}^{K} \sqrt{\frac{\eta m_k}{N} \sum_{i=1}^{m_k} |g_{i,k}|^2} + n_d = s\sum_{k=1}^{K} \tilde{g}_k + n_d \tag{10}$$

It can be seen from (10) that by applying antenna diversity schemes at relays, the networks can be decomposed to K diversity channels each with channel gain $\tilde{g}_k$. The output SNR at the destination receiver can therefore be written as:

$$\rho_d^{m_k} = \left(\sum_{k=1}^{K} \sqrt{\frac{\eta m_k}{N} \sum_{i=1}^{m_k} |g_{i,k}|^2}\right)^2. \tag{11}$$

When all the relays are deployed with a single antenna, there is no traditional maximum ratio combining gain at the relays and the destination. However, the destination still observes a set of equal gain combined [15] amplitude signals from all relays.[1]

[1] Different from [15], the equal gain combining for relay channels is applied at the transmitter instead of the receiver.

Since we assume that the backward and forward channel coefficients for each antennas are kept same for different number of K and m_i. The output SNR at the destination can be rewritten as

$$\rho_d^1 = \frac{\eta}{N}\left(\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|\right)^2 ; \tag{12}$$

when all the antenna are deployed in one relay (i.e. $K = 1$ and $m_1 = N$), full diversity gain is achieved among all the N antennas at the relay and also at the destination. The SNR can be rewritten as

$$\rho_d^N = \eta\sum_{k=1}^{K}\sum_{i=1}^{m_i}|g_{i,k}|^2 \tag{13}$$

IV. Capacity Performance

In this section we derive the capacity bounds for the scheme proposed in previous section. The network capacity for digital relaying can be written as

$$C_D^{m_k} = \min\left(C_r^{1,m_1}, C_r^{2,m_2}, \cdots, C_r^{K,m_k}, C_d^{m_k}\right) \tag{14}$$

where $C_r^{k,m_k} = 0.5\log_2\left(1+\rho_k^{m_k}\right)$ denoting the Shannon capacity from source to relay k channel, and $C_d^{m_k} = 0.5\log_2\left(1+\rho_d^{m_k}\right)$ denoting the Shannon capacity from relays to destination channels. The factor 0.5 denotes the half bandwidth compared with non-relay channels.

We firstly analyze channel capacity from the relays to destination link by bounding the $\rho_d^{m_k}$,i.e. the output SNR at the destination.

Lemma 1: For any m_k, $\rho_d^1 \le \rho_d^{m_k} \le \rho_d^N$.

Proof: See Appendix. ■

From *Lemma 1*, we can see that

$$C_d^1 \le C_d^{m_k} \le C_d^N, \tag{15}$$

where C_d^1 denotes the capacity for relays to destination channels when $K = N$, and C_d^N denotes the capacity for relay to destination channels when $K = 1$. Now also considering the capacity from the source to relays link and extending the analysis to the whole network scenario, we have the following theorem:

Theorem 1: If we denote the network capacity for $K = N$ as C_D^1 and for $K = 1$ as C_D^N, for any m_k, $C_D^1 \le C_D^{m_k} \le C_D^N$.

Proof: Considering the SNR $\rho_k^{m_k}$ for the source to the kth relay link, if we denote it as ρ_n^1 for $K = N$ and ρ_1^N for $K = 1$, it can be shown that

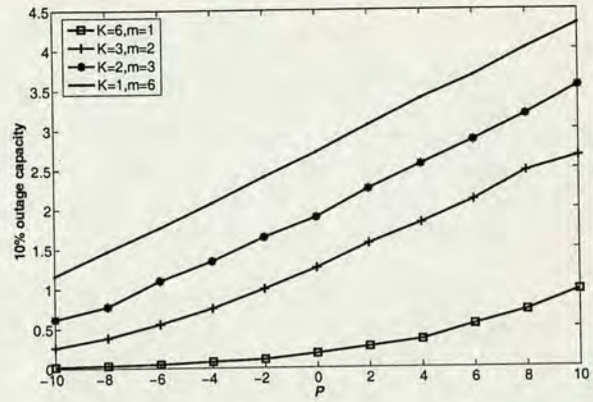
$$\min\left(\rho_n^1\right) \le \min\left(\rho_k^{m_k}\right) \le \rho_1^N. \tag{16}$$

Therefore, we have the following:

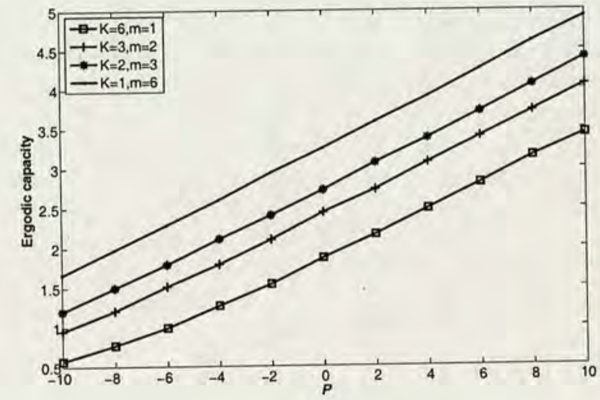
$$\min\left(C_r^{1,1}, \cdots, C_r^{N,1}\right) \le \min\left(C_r^{1,m_1}, \cdots, C_r^{K,m_k}\right) \le C_r^{1,N}. \tag{17}$$

Combining (17) and (15), we thus complete the proof. ■

From the above analysis we have shown that for the signal combining techniques discussed in the paper, the network capacity will be lower bounded by that of N relay channels each with single antenna, and upper bounded by that of a single relay channel with N antennas. This means that even there are more relays, the increased "equal gain combining" gain at the destination can not compensate for the loss of maximum ratio combining gain at the relay and the destination when numbers of antennas at each relay are reduced.



(a) 10% outage capacity



(b) Ergodic capacity

Fig. 2. Capacity of single MIMO relay channels for different number of relays K, while each relay is deployed with m antennas. (a) 10% outage capacity. (b) Ergodic capacity.

V. Simulation Results

We calculate the both ergodic capacity and 10% outage capacity (in bits per channel use) for 1000 channel realizations regarding different values of η, denoted as *P* in the figures. In this simulation example we assume that the distance between source and destination is normalized. The relays are uniformly and randomly located in the middle region between the source and relays. Therefore x_k is set to 0.5. We assume the total number of antennas at relays (N) is 6 and we also assume that all K relays have the same number of antennas m. Fig. 2 shows the capacity performance. We can see that for different (K, m), the capacity is always upper bounded by $(1, 6)$ and lower bounded by $(6, 1)$. These results verify the analysis made in this paper. Furthermore, we can see through the simulation that larger m and small K might give larger benefit, since larger m allows more freedom of cooperation among the antennas at each relay. Therefore when m reaches N (K reduces to 1), full cooperation are made among all the antennas to give rise to the best performance.

VI. CONCLUSIONS

In this paper we analyze the performance of multiple relay channels when multiple antennas are deployed only at relays. We apply antenna diversity techniques at relays which are known as maximum ratio combining and transmit beamforming. We show that with K relays the network can be decomposed into K diversity channels each with a different channel gain, while the signals can be effectively combined at the destination. If we assume that the total number of antennas at all relays are fixed to N and total transmit power at all relays are normalized, the network capacity will be lower bounded by that of N relay channels each with single antenna, and upper bounded by that of single relay channels with N antennas.

APPENDIX
PROOF OF *Lemma 1*

We firstly prove that $\rho_d^1 \le \rho_d^{m_k}$. We write the following

$$\sqrt{\rho_d^{m_k}} - \sqrt{\rho_d^1} = \sum_{k=1}^{K}\left(\underbrace{\sqrt{\frac{\eta m_k}{N}\sum_{i=1}^{m_k}|g_{i,k}|^2}}_{A_k} - \underbrace{\sqrt{\frac{\eta}{N}}\sum_{i=1}^{m_k}|g_{i,k}|}_{B_k}\right). \quad (18)$$

To compare A_k with B_k, we write

$$A_k^2 - B_k^2 = \frac{\eta}{N}\left(m_k\sum_{i=1}^{m_k}|g_{i,k}|^2 - \left(\sum_{i=1}^{m_k}|g_{i,k}|\right)^2\right) \quad (19)$$

$$= \frac{\eta}{N}\left((m_k-1)\sum_{i=1}^{m_k}|g_{i,k}|^2\right) - \frac{\eta}{N}\left(\sum_{i,j=1;i\neq j}^{m_k}|g_{i,k}|\,|g_{j,k}|\right). \quad (20)$$

Note that

$$(m_k-1)\sum_{i=1}^{m_k}|g_{i,k}|^2 = \sum_{i=1}^{m_k}\sum_{j=1,j\neq i}^{m_k}|g_{j,k}|^2 \quad (21)$$

$$= 0.5\sum_{i,j=1;i\neq j}^{m_k}\left(|g_{i,k}|^2 + |g_{j,k}|^2\right). \quad (22)$$

So (20) can be further written as:

$$A_k^2 - B_k^2 = \frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{m_k}\left(|g_{i,k}|^2 - 2|g_{i,k}|\,|g_{j,k}| + |g_{j,k}|^2\right)$$

$$= \frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{m_k}\left(|g_{i,k}| - |g_{j,k}|\right)^2 \ge 0.$$

So $A_k \ge B_k$ and therefore $\rho_d^1 \le \rho_d^{m_k}$.

Next we prove that $\rho_d^N \ge \rho_d^{m_i}$. For simplicity, we denote

$$a_k = \sum_{i=1}^{m_k}|g_{i,k}|^2 \quad (23)$$

in equation (11) and (13). Then $\rho_d^N - \rho_d^{m_i}$ can be written as

$$\rho_d^N - \rho_d^{m_i} = \frac{\eta}{N}\left(\sum_{k=1}^{K}(N-m_k)\,a_k - \sum_{i,j=1;i\neq j}^{K}\sqrt{m_i m_j a_i a_j}\right). \quad (24)$$

Note the constraint by (1) in section II, we have the following:

$$(N-m_k) = \sum_{i=1,i\neq k}^{K} m_i. \quad (25)$$

Putting (25) into (24), we have the following:

$$\rho_d^N - \rho_d^{m_i} = \frac{\eta}{N}\left(\sum_{k=1}^{K}\sum_{i=1,i\neq k}^{K} m_i a_k - \sum_{i,j=1;i\neq j}^{K}\sqrt{m_i a_i m_j a_j}\right). \quad (26)$$

Note the following:

$$\sum_{k=1}^{K}\sum_{i=1,i\neq k}^{K} m_i a_k = \sum_{i,j=1;i\neq j}^{K}(m_i a_j)$$

$$= 0.5\left(\sum_{i,j=1;i\neq j}^{K}(m_i a_j) + \sum_{i,j=1;i\neq j}^{K}(m_j a_i)\right).$$

We can further write (26) as follows:

$$\rho_d^N - \rho_d^{m_i} = \frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{K}(m_i a_j) - \frac{\eta}{N}\sum_{i,j=1;i\neq j}^{K}\sqrt{m_i a_j m_j a_i} + \frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{K}(m_j a_i)$$

$$= \frac{\eta}{2N}\sum_{i,j=1;i\neq j}^{K}\left(\sqrt{m_i a_j} - \sqrt{m_j a_i}\right)^2.$$

Therefore $\rho_d^N \ge \rho_d^{m_i}$ and $\rho_d^1 \le \rho_d^{m_k} \le \rho_d^N$.

REFERENCES

[1] A. J. Goldsmith, S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks Wireless Communications", *IEEE Pers. Commun.*, Vol. 9, no. 4, pp. 8 - 27, Aug. 2002.

[2] R. Pabst et al, "Relay-based deployment concepts for wireless and mobile broadband radio," *IEEE Commun. Mag.*, Sept. 2004.

[3] M. Frodigh, S. Parkvall, C. Roobol, P.Johansson, Larsson, "Future-generation wireless networks" *IEEE Pers. Commun.*, vol. 8, no. 5, pp. 10 - 17, Oct. 2001.

[4] J. N. Laneman, G. W. Wornell and D. N. C. Tse, "An efficient protocol for realizing cooperative diversity in wireless networks," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001.

[5] D. Gesbert, M. Shafi, Shiu Da-shan, P.J.Smith, A.Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 281 - 302, April 2003.

[6] J. N. Laneman, D. N. C. Tse and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, to appear.

[7] 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.

[8] 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.

[9] B. Wang, J. Zhang, and A. Host-Madsen, "On capacity of MIMO relay channel," *IEEE Trans. Inf. Theory*, 2004, submitted.

[10] H. Bolcskei et al, "Capacity Scaling Laws in MIMO Relay Networks," *IEEE Trans. Wireless Comm.*, Apr. 2004, submitted.

[11] O. Oyman and A. J. Paulraj, "Energy efficiency in MIMO relay networks under processing cost", in *Conf. Inf. Science and Systems*, The Johns Hopkins University, March 16-18, 2005.

[12] J. G. Proakis, *Digital Communications: Fourth Edition*, 2001.

[13] J. Bach Andersen, "Antenna arrays in mobile communications: gain, diversity, and channel capacity", *IEEE Antennas Propagat. Mag.*, vol. 42, no. 2, pp. 12 - 16, April 2000.

[14] W. Rhee, J. M. Cioffi, "On the Capacity of Multiuser Wireless Channels With Multiple Antennas", *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2580 - 2595, October 2003.

[15] Y. Chen, C. Tellambura, "Performance analysis of L-Branch equal gain combiners in Equally correlated Rayleigh fading channels", *IEEE Commun. Letters*, vol. 8, no. 3, pp. 150 - 152, March 2004.

On the Outage Capacity of MIMO Multihop Networks

Yijia Fan , John S. Thompson
Institute for Digital Communications, School of Engineering and Electronics
The University of Edinburgh, Edinburgh, EH9 3JL, UK.
Email: y.fan@ed.ac.uk

Abstract— **In this paper we discuss the outage capacity of multihop networks employing terminals with multiple antennas. We discuss the performance of several novel relaying configurations and signalling algorithms for either limited feedback or non-feedback channels. Three relaying types are considered, namely *analogue relaying, digital relaying* and *hybrid relaying*. We find that *digital selective relaying* outperforms all other relaying configurations considered here while requiring the highest signalling overhead. The *matched filter based hybrid relaying* scheme appears to be a good suboptimum choice for its good performance and low signalling overhead in the multiple relay scenario. We also find that for hybrid relaying schemes, the MMSE algorithm, which is usually applied in conventional MIMO systems, might not give effective performance improvement when it is applied at relays.**

I. Introduction

The use of multiple antennas at both ends of a wireless link, called multiple-input multiple-output (MIMO) technology, promises significant improvements in terms of spectral efficiency and link reliability through spatial multiplexing [1] and space-time coding [2], respectively. A lot of research has been conducted on point-to-point MIMO links in the last decade [3]. Recently, capacity limits of MIMO channels in a multi-user or network context has also been theoretically studied. Some important results on MIMO broadcast and multiple-access channels can be found in [4] [5] [6]. In the mean time, deploying MIMO techniques into ad hoc [7] or multihop cellular networks (MCNs) [8] has also come under consideration. In [9],[10], the impact of so-called virtual antenna arrays or cooperative diversity schemes, which mimic the performance advantages of MIMO systems by exploiting the spatial diversity of the wireless channel, have been studied. More recently, the first quantitative capacity results for a Time-Division-Multiple-Access (TDMA) based MIMO relay network has been reported in [11], where every terminal in the network model is equipped with multiple antennas.

So far the work on the capacity of MIMO multi-user or relay channels is mainly done based on an information-theoretic point of view. However, in wireless networks, a desirable goal is to develop suboptimal, but more practical, approaches for routing and signal processing. In this paper we discuss the practical capacity performance of MIMO relay channels using different routing and relaying schemes. We focus on MIMO spatial multiplexing systems where streams of independent data are transmitted over different antennas. We develop and compare different algorithms based on either limited feedback or non-feedback of channel state information (CSI). Three relay types are considered in the paper. The first two kinds are the well known: *analogue relaying*, where the relays simply amplify the signals, and *digital relaying* where the relays decode, re-encode, and re-transmit the signals. We also propose a relaying type called *hybrid relaying*, in which the information from the source is filtered at each relay before retransmission. We will discuss the performance of several novel relaying and signalling algorithms based on these different channel conditions and relaying types.

The reminder of the report is organized as follows. In Section II, the basic system model is introduced. Section III discusses the impact of different relay configurations on MIMO multihop channels where limited feedback exists. The non-feedback case is also briefly discussed in section IV. Section V analyzes the simulation results of all the algorithms developed in this paper. Concluding remarks and future topics will be found in Section VI.

II. Basic System Model

We develop a system model based on [11]. We consider a multihop network model with one source, one destination and K relays located randomly and independently in the middle region between source and destination. We ignore the direct link between the source and destination due to the large range. The data transmission is over two time slots using two hops. In the first time slot the relays communicate with the source. After processing the received signals, the relays then transmit the processed data to the destination during the second time slot while the source is silent. We also assume that total transmit power for source and relays are the same; it is equally distributed among the relays.

We assume that each node in the network is equipped with M antennas. We restrict our discussion to the case where the channels are slow, frequency-flat fading with a block fading model. In the first transmission time slot, the source broadcasts the signals to all the relay terminals. The input/output relation for the source to the kth relay is given by

$$\mathbf{r}_k = \sqrt{\alpha_k}\mathbf{H}_k\mathbf{s} + \mathbf{n}_k = \tilde{\mathbf{H}}_k\mathbf{s} + \mathbf{n}_k \qquad (1)$$

where $\mathbf{r}_k$ is $M \times 1$ receive signal vector, $\mathbf{H}_k$ is the $M \times M$ channel transfer matrix from source to the kth relay, with each entry set to identically independent distributed (i.i.d) complex

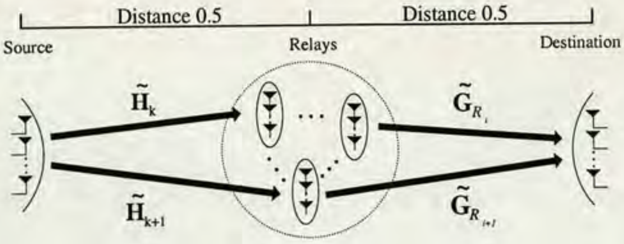


Fig. 1. Basic system model of a MIMO two hop network

Gaussian random variables with unit variance. The vector s is the $M \times 1$ transmit signal vector with covariance matrix $\mathbf{I}_M$ and $\mathbf{n}_k$ is the $M \times 1$ complex circular additive white Gaussian noise vector at relay k with covariance matrix $\sigma^2\mathbf{I}_M$, where $\mathbf{I}_M$ is the $M \times M$ identity matrix.

The factor α_k is the average energy received at one antenna for the kth relay terminal over one symbol period. It may be written as:

$$\alpha_k = P_T x^{-\gamma} 10^{\zeta_k/10} \tag{2}$$

where P_T denotes the transmit signal energy per transmit antenna from the source and x is the distance between the source and relay k. The scalar γ denotes the path loss exponent (in this paper, always set to 4). The lognormal shadowing term, ζ_k, is a random variable with a normal distribution, mean of 0 dB and standard deviation δ (dB). A value typical of shadowing deviations in urban cellular environments, $\delta = 8$ dB is used. We assume that the relay terminals are located randomly and uniformly within the middle region between the source and destination so that α_k will be i.i.d.. If we normalized the range between the source and destination, x will be 0.5.

All the relays process their received vector signal $\mathbf{r}_k$. As will be seen later in the paper, each relay k might send the post-processed signals to the destination using N_k out of M antennas ($0 \leq N_k \leq M$). In the situation where limited feedback exists between the source and relays, the source might choose the best P relays $\{\Re_1, \Re_2, \cdots, \Re_P\}$ out of all K relays to transmit ($0 < P \leq K$) while switching off the others. For all the above cases, the signal received at the destination terminal can be written as:

$$\mathbf{y} = \sum_{i=1}^{P} \sqrt{\beta_{\Re_i}} \mathbf{G}_{\Re_i} \mathbf{t}_{\Re_i} + \mathbf{z} = \sum_{i=1}^{P} \tilde{\mathbf{G}}_{\Re_i} \mathbf{t}_{\Re_i} + \mathbf{z} \tag{3}$$

where $\beta_{\Re_i}$ is the average signal energy over one symbol period from relay $\Re_i$ to the destination, which is given by

$$\beta_{\Re_i} = P_T (1-x)^{-\gamma} 10^{\zeta_{\Re_i}/10}. \tag{4}$$

The matrix $\mathbf{G}_{\Re_i}$ is the corresponding $M \times N_{\Re_i}$ channel matrix with each entry set to i.i.d. complex Gaussian with unit variance, while $N_{\Re_i}$ is the number of antennas transmitting during the second time slot from relay $\Re_i$. The transmit signal vectors $\mathbf{t}_{\Re_i}$ are processed to meet the constraint

$$\mathrm{E}\left[\|\mathbf{t}_{\Re_i}\|^2\right] \leq N_{\Re_i} \frac{M}{\sum_{i=1}^{P} N_{\Re_i}} \tag{5}$$

so that the total transmit power for the relays is the same as for the source. The expression $\mathrm{E}[x]$ represents the expectation operator and $\|x\|$ denotes the Frobenius-norm of x. The vector $\mathbf{z}$ is the $M \times 1$ complex circular additive white Gaussian noise at the destination with covariance matrix $\sigma^2\mathbf{I}_M$. This constraint implies that the more relays are used, the less transmit power will be allocated to each relay. If we set the bandwidth to 1 in a single hop network (e.g. a network without relay), then each hop has bandwidth 0.5 for a two hop network. Fig 1 summarizes the basic system model described above.

III. MIMO Multihop Networks with Limited Feedback

In this section we discuss the relaying configurations in the multihop network when limited feedback is allowed from the receivers to the transmitters. We assume that the receivers obtain the CSI through the training sequence, and can feed the output signal to interference plus noise ratio (SINR) of the signals back to the transmitters periodically. We develop relaying algorithms based on *analogue relaying*, *digital relaying* and *hybrid relaying*.

A. Analogue Full Relaying (AFR)

In this simple analogue relaying scheme. the kth relay scales its received signal by:

$$\rho_k = \sqrt{\frac{M}{K\left(\left\|\tilde{\mathbf{H}}_k\right\|^2 + M\sigma^2\right)}} \tag{6}$$

to meet the constraint described by (5), then all K relays forward the amplified signals to the destination. The input/output relation from the source to destination will be

$$\mathbf{y} = \left(\sum_{k=1}^{K} \rho_k \tilde{\mathbf{G}}_k \tilde{\mathbf{H}}_k\right)\mathbf{s} + \left(\sum_{k=1}^{K} \rho_k \tilde{\mathbf{G}}_k \mathbf{n}_k + \mathbf{z}\right) = \tilde{\mathbf{V}}\mathbf{s} + \tilde{\mathbf{n}} \tag{7}$$

Thus we can treat the whole system as a point-to-point MIMO spatial multiplexing system with equivalent channel matrix $\tilde{\mathbf{V}}$ and noise vector $\tilde{\mathbf{n}}$. We then apply the VBLAST-MMSE [1] algorithm to decode the signals at the destination. The MMSE weight matrix can be written as:

$$\begin{aligned} \mathbf{W}_{MMSE} &= \tilde{\mathbf{V}}^H \left(\mathrm{E}\left(\tilde{\mathbf{n}}\tilde{\mathbf{n}}^H\right) + \tilde{\mathbf{V}}\tilde{\mathbf{V}}^H\right)^{-1} \\ &= \tilde{\mathbf{V}}^H \left(\sigma_n^2 \left(\sum_{i=1}^{K} \rho_i^2 \tilde{\mathbf{G}}_i \tilde{\mathbf{G}}_i^H + \mathbf{I}_M\right) + \tilde{\mathbf{V}}\tilde{\mathbf{V}}^H\right)^{-1} \end{aligned} \tag{8}$$

We assume that the source can encode each signal to be sent at a rate equal to the Shannon capacity for the corresponding SINR fed back from the destination through the relays. So for each channel realization, the capacity (in bits per channel use) of the $M \times M$ MIMO link with the VBLAST-MMSE detector is now given by:

$$C_{BLAST} = \sum_{i=1}^{M} C_i = 0.5 \times \sum_{i=1}^{M} \log_2 (1 + SINR_i). \quad (9)$$

Where C_i denotes the Shannon capacity for each transmitted signal, $SINR_i$ denotes the post-processing SINR obtained by the VBLAST-MMSE algorithm and the factor 0.5 denotes the bandwidth compared with that of single hop network. A more generalized form [12] based on the input/output relation can be written as:

$$C = 0.5 \times \log_2 \left| \mathbf{I}_M + \frac{1}{\sigma^2} \tilde{\mathbf{V}} \tilde{\mathbf{V}}^H \left(\sum_{k=1}^{K} \rho_k^2 \tilde{\mathbf{G}}_k \tilde{\mathbf{G}}_k^H + \mathbf{I}_M \right)^{-1} \right| \quad (10)$$

where $||$ denotes the matrix determinant.

It can be seen from (10) that while the relays amplify the signals, they also amplify the receiver noise. The elements of the equivalent channel matrix might give rise to a low channel capacity when only a small number of relays are used because of the keyhole effect [13]. However, as the number of relays increase, this issue is likely to be overcome. It can be observed from equation (10) that for larger number of relays, the contribution of the equivalent MIMO channel $\tilde{\mathbf{V}}$ to the capacity becomes that of the average over all single K relay channels, where either "good" or "bad" relay channels are used. So we speculate that the capacity under this scheme will become constant for large number of relays.

B. Digital Relaying (DR)

In the following we develop three digital relaying schemes which exploit additional feedback information.

1) Digital Selective Relaying (DSR): In the DSR scheme, the source selects the single best relay for which the highest throughput can be achieved. The capacity for each transmitted signal s_m through relay route k can be written as:

$$C_{m,k} = \min (C_{m,sr_k}, C_{m,r_k d}) \quad (11)$$

where C_{m,sr_k} is the capacity for s_m from source to the relay k and $C_{m,r_k d}$ from relay k to the destination. So the capacity of the $M \times M$ MIMO spatial multiplexing relay link through relay k might be written as:

$$C_k = \sum_{m=1}^{M} C_{m,k} \quad (12)$$

The network capacity for DSR scheme is given by

$$C = \max (C_1, \cdots, C_K) \quad (13)$$

This scheme exploits the spatial diversity of the MIMO relay network by selecting the optimum relay for transmission. However, it should be noted that in order to use this algorithm, the source must know not only the output SINRs at each relay, but also the output SINRs at the destination through each single relay route. So the signalling overhead for this configuration is very high as the number of relays increases. It might only be practical for a very slow fading environment.

2) Digital MMSE Selective Relaying (DMSR): Unlike DSR, where relaying is restricted to a single relay, DMSR might use different relays to decode different signals sent by the source (i.e. the signals are decoded distributively among the relays). In DMSR, each relay first calculates the output SINR of each transmitted signal after performing the Linear-MMSE algorithm [14]. Then each relay feeds the SINRs of all the signals they received back to the source. For signal s_m, the source chooses relay $\Re_i$ to decode and retransmit it using an arbitrary transmit antenna. $\Re_i$ is chosen by the following criterion.

$$\Re_i = \arg \max_{\Re_i \in \{1,2,\cdots,K\}} (SINR_{m,\Re_i}). \quad (14)$$

where $SINR_{m,\Re_i}$ is the SINR for signal s_m at relay $\Re_i$. The destination then performs the VBLAST-MMSE algorithm to decode each signal.

The complexity of this scheme is much lower than DSR because the source does not require knowledge of the output SINRs at the destination through each single relay route. The destination only has to feed the output SINRs back to the relays to perform the VBLAST-MMSE algorithm. Although the linear-MMSE detector does not perform as well as VBLAST detector, the cooperative diversity of source to relay links is further exploited in DMSR comparing with DSR. However, the drawback of DMSR is that the spatial diversity from the relay to destination link is not exploited. It can be seen from (11) that the capacity of DMSR will always be upper bounded by the $M \times M$ point-to-point MIMO link formed by the relays and the destination. Considering this factor, we develop another digital relaying scheme using the transmit antenna selection technique at the relays to exploit the spatial diversity of the relay to destination channels.

3) DMSR with Antenna Selection (DMSR-AS): We modify the DMSR technique so that the relays use the transmit antenna selection algorithm developed in [14]. For each relay chosen by DMSR algorithm, we select $N_{\Re_i}$ of the total M transmit antennas that maximize the minimum column norm of the selected channel $\mathbf{G}_{\Re_i}$, to retransmit the signals, while $\sum N_{\Re_i} = M$. This means that we feed back to the relays the selected antennas and their corresponding SINRs.

We can see that the advantage of this scheme is mainly decided by the number of antennas M at each relay. So for a small M, DMSR-AS might only result in a limited performance gain over DMSR schemes even when the number of relays is large.

C. Hybrid Relaying (HR)

In the hybrid relaying schemes, the relays only decode the training sequence from the source to obtain the full CSI, then filter the received signals based on the CSI without decoding them. After multiplying the signal vector by the filtering weight matrix, each relay then amplifies and forwards the filtered signals to the destination. The amplifying factor ρ_k

for relay k is

$$\rho_k = \sqrt{\frac{M}{\left\|\mathbf{W}_k\tilde{\mathbf{H}}_k\right\|^2 + \|\mathbf{W}_k\|^2\sigma^2}} \tag{15}$$

where $\mathbf{W}_k$ is the weight matrix used for the kth relay. The system model and capacity for this scheme can be described by (7) and (10) replacing $\tilde{\mathbf{G}}_k$ by $\tilde{\mathbf{G}}_k\mathbf{W}_k$. In the following we propose two hybrid relaying configurations, named MMSE relaying and matched filter based relaying.

1) MMSE Relaying (HR-MR): In this scheme, each relay performs MMSE filtering. The MMSE weight for the kth relay might be written as:

$$\mathbf{W}_k = \tilde{\mathbf{H}}_k^H\left(\tilde{\mathbf{H}}_k\tilde{\mathbf{H}}_k^H + \sigma^2\mathbf{I}_M\right)^{-1} \tag{16}$$

The same approach as described in subsection III-A is used at the destination to detect the signals. This scheme offers the possibility of reconstructing the signals at the relays by making a tradeoff between noise amplification and interference suppression.

2) Matched Filter Based Relaying (HR-MFR): In this scheme, each relay performs matched filtering. The weight matrix is given by:

$$\mathbf{W}_k = \tilde{\mathbf{H}}_k^H \tag{17}$$

Comparing with the MR-HR scheme, the transmit signals at the relays might be less separated. However, this scheme has the advantage of reducing the noise at the relays.

IV. MIMO Multihop Network with No Feedback

All the analogue relaying and hybrid relaying schemes discussed in section III can be applied to non-feedback relay channels. However, the network capacity should be calculated differently for the VBLAST detector. As discussed in [14], when the CSI is not known at the transmitter, the Shannon capacity of the MIMO link for each symbol period is restricted by the worst stream during VBLAST detection. So the maximum achievable throughput for the multihop network is given by:

$$C_{BLAST} = 0.5 \times M \times \log_2\left(1 + SINR_{\min}\right), \tag{18}$$

while $SINR_{\min}$ is the minimum post-processing SINR obtained by the VBLAST-MMSE algorithm at the destination. In fact, the linear-MMSE detector might be a better choice than VBLAST detector considering the outage capacity for the low or medium SNR level [14]. However, we believe that for each relay scheme, the behavior of outage capacity for non-feedback channels considering different SNR or number of active relays will be almost identical to that for limited feedback channels. So the non-feedback case will not be discussed further in this paper.

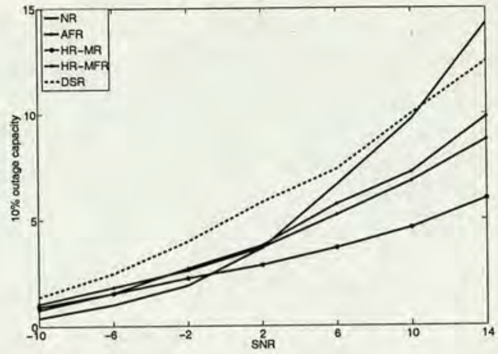


Fig. 2. Outage capacity of the MIMO two hop network with limited feedback, while every terminal is deployed with 4 antennas, the number of relays is 1

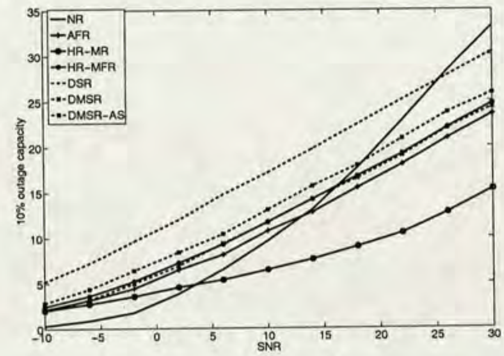


Fig. 3. Outage capacity of the MIMO two hop network with limited feedback, while every terminal is deployed with 4 antennas, the number of relays is 10

V. Simulation Results and Performance Comparison

We consider a system model where the source, the relays and the destination each employ 4 antennas. We assume a slow fading environment where each channel realization lasts for a number of symbol periods so that the transmission and processing time for periodic training sequences and feedback can be ignored. We simulate and compare the 10% outage capacity of the MIMO multihop network. The SNR is defined as $\frac{P_T}{\sigma^2}$. Due to limited space we only give the simulation results for the limited feedback MIMO relay channels.

It has been shown in [10] that relaying in single-input single-output (SISO) multihop networks will only bring benefits given a low or medium level SNR. For a high SNR level, direct transmission will always be preferred due to the advantage of doubled bandwith comparing with relaying. The same conclusion can be made for the MIMO relay networks. Fig 2 shows the performance of different relay configurations for limited feedback relay channels with a single relay available. The NR scheme represents the no relay case, while at the destination the VBLAST-MMSE detector is used to decode the signal directly from the source. We can see that at the high SNR level, relaying does not bring

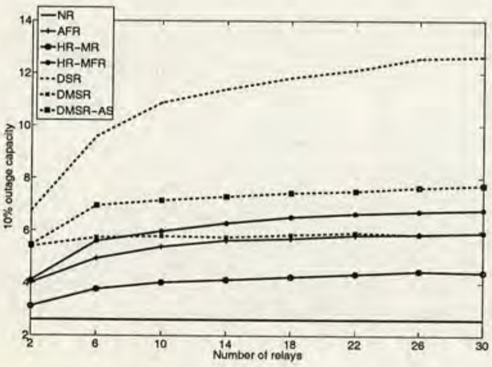


Fig. 4. Outage capacity of the MIMO two hop network with limited feedback, while every terminal is deployed with 4 antennas, the SNR is set to 0dB

any capacity benefit for the MIMO network. When comparing the performance of different relaying configurations, we can see that digital relaying (dashed curve) outperforms all hybrid and analogue relaying schemes at the cost of the highest signalling overhead. Furthermore, it is interesting to see that hybrid relaying schemes do not give a desirable performance advantage over analogue relaying. For matched filter based relaying, it outperforms analogue relaying only for low SNR levels due to its capability of reducing the noise at the relay. However, when the SNR increases, its defect in correlating the signals becomes a dominant factor that limits the capacity performance. So it is not as good as analogue relaying. For MMSE relaying, although it can separate the signals at the relay, the signals undergo another MIMO channel from the relay to the destination, which make the signal suppression at the relay less effective than in a non-relay scenario. It performs the worst due to its noise amplification at the relay. At low SNR levels, the MMSE filter approximates the matched filter, so the two hybrid relaying schemes give the similar performance.

Fig 3 shows the simulation results for 10 relay case. It can be seen that the relaying schemes outperforms non-relaying for a larger SNR region. Unlike the single relay case, the matched filter based relaying outperforms analogue relaying. This is because each relay might be considered as a scatterer in the network, as the number of relays increases the quality of effective point to point MIMO channels are improved, so the signals become less correlated at the destination. The advantage of noise reduction dominates the capacity performance. MMSE relaying still performs worst. Thus, we conjecture that separating the co-channel interference might be not as effective as reducing the noise at the relays to increase the MIMO multihop network capacity.

Fig 4 shows the capacity performance as the number of relays increase. We set the SNR to 0dB. It can be seen that except for DSR, all other relaying schemes give consistent capacity performance for large number of relays. Matched filter based relaying gives comparable performance to the two digital relaying schemes (DMSR and DMSR-AS) while requiring less signalling overhead.

VI. Conclusions

This paper has investigated different relay configurations for multihop networks where all terminals are equipped with multiple antennas. We find that no matter whether limited feedback exists or not, digital selective relaying performs the best. Matched filter based hybrid relaying can be a good suboptimal choice achieving similar performance to two other digital relaying schemes. For hybrid relaying schemes, reducing noise at the relays appears to be more effective than spatially separating the signals in order to improve the whole network capacity.

In this work, we assume only partial CSI at each transmitter. In the future, relaying configurations with full CSI at the transmitter will be considered. More specifically we plan to study further hybrid relaying and dirty paper coding [6] schemes to exploit the spatial nature of the relay channels.

Acknowledgments

This work is kindly sponsored by UK EPSRC Grant GR/S58782/01.

References

[1] P. W. Wolniansky, G. J. Foschini, G. D. Golden, R. A. Valenzuela, "V-BLAST: An Architecture for Realizing Very High Data Rates Over the Rich-Scattering Wireless Channel", invited paper, *Proc. ISSSE-98*, Pisa, Italy, Sept. 29, 1998.

[2] A.F. Naguib et al, "Increasing Data Rate over Wireless Channels", *IEEE Signal Processing Magazine*, Vol 17(3), pp 76-92, May 2000.

[3] D. Gesbert, M. Shafi, Shiu Da-shan, P.J.Smith, A.Naguib, "From theory to practice: an overview of MIMO space-time coded wireless systems", *IEEE Journal on Selected Areas in Communications*, Volume: 21, Issue: 3, Pages:281 - 302, April 2003.

[4] A. Goldsmith et al, "Capacity Limits of MIMO Channels", *IEEE Journal on Selected Areas in Communications*, Volume: 21, Issue: 5, Pages:684 - 702, June 2003.

[5] S. Vishwanath, N. Jindal, A. J. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels", *IEEE Transactions on Information Theory*, Volume: 49, Issue: 10, Pages:2658 - 2668 Oct. 2003.

[6] N. Jindal, A. Goldsmith, "Dirty Paper coding vs. TDMA for MIMO broadcast channels", *IEEE International Conference on Communications*, Volume: 2, 20-24, June 2004

[7] A.J. Goldsmith, S.B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks Wireless Communications", *IEEE Personal Communications*, Volume: 9, Issue: 4, Pages: 8 - 27, Aug. 2002.

[8] Ying-Dar Lin and Yu-Ching Hsu, "Multihop cellular: A New Architecture for wireless Communications", in *INFOCOM 2000.IEEE*, 2000.

[9] J.N.Laneman, G.W.Wornell, "Energy-efficient antenna sharing and relaying for wireless networks", *IEEE Wireless Communications and Networking Conference*, Volume: 1, 23-28, Pages: 7 - 12, Sept. 2000.

[10] J. Laneman, G.W.Wornell, "Distributed Space-Time-Coded Protocols for Exploiting Cooperative Diversity in Wireless Networks", *IEEE Trans. Inform. Theory*. Volume: 49, No. 10, pages:2415 - 2425, Oct. 2003

[11] H. Bolcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity Scaling Laws in MIMO Relay Networks", in *Allerton Conference on Communications, Control, and Computing*, Monticello, IL, Oct. 2003.

[12] F. R. Farrokhi et al, "Link-optimal Space-Time Processing with Multiple Transmit and Receive Antennas", *IEEE Comm. Letters*, Volume: 5, No.3, March 2001.

[13] D. Chizhik et al, "Keyholes, Correlations, and Capacitys of Multielement Transmit and Receive Antennas", *IEEE Trans. Wireless comm.* Volume: 1, No.2, April 2002.

[14] D. Gore, A. Gorokhov, A. Paulraj, "Joint MMSE versus V-BLAST and antenna selection", in *Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, Volume: 1, Pages:505 - 509, Nov. 2002.