

Wireless Antennas - Making Wireless Communications Perform Like Wireline Communications

Yingbo Hua¹, Yan Mei, and Yu Chang

*Department of Electrical Engineering
University of California
Riverside, California, USA, 92521
<http://www.ee.ucr.edu/~yhua>
yhua@ee.ucr.edu*

INVITED PAPER

August 14, 2003

Abstract

We first provide an overview of some of the latest developments in wireless communications using multiple transmitters and multiple receivers. We point out the importance of SNR control in fast random fading environment. For applications where large antenna arrays are not suitable, we introduce the concept of wireless antennas or wireless relays that are distributed between a source and a destination. We propose Hurwitz-Radon space-time code for the wireless relays. Each relay receives a noisy baseband signal simultaneously from the source. The baseband signals (symbols) are not decoded into information bits at the (non-regenerative) relays, but rearranged (i.e., space-time modulated) in their orders, amplitudes and phases according to the Hurwitz-Radon code. The relays do not exchange symbols with each other, but forward the modified sequences of symbols in parallel to the destination. Our study shows that with R relays, a diversity factor around $R/2$ can be achieved, i.e., the averaged bit error rate is in the order of $1/SNR^{R/2}$ as opposed to $1/SNR$ for a single (regenerative) relay system. More than 10 dB power saving, from the baseline of a single relay system, is possible with eight relays. Issues such as channel estimation, symbol synchronization, medium access protocols and signal processing hardware are also discussed.

1. Introduction

1.1 The Challenge of Fast Random Fading

Fast random fading (also called small scale fading) is a key factor that distinguishes mobile wireless communications from wireline communications. In wireline communications, signal-to-noise ratio (SNR) is easy to predict on line and does not change rapidly. The reliability of wireline communications is generally high. But in mobile wireless environment, fast random fading causes the SNR at a receiver difficult to predict by a transmitter. Without a proper control of SNR, the error rate of data transmission is high.

There are several techniques for SNR control in mobile wireless environment. A simple way to control SNR is to force transmitter to transmit more power when fading is deep and less power when fading is not deep. In IS-95, a mobile adjusts its transmission power so that the signal received by base station is relatively constant [RAP]. This technique is effective to handle large scale fading (primarily due to the distance between base station and mobile) but not much so for

¹ This work was supported in part by ARL's CTA program and NSF. This paper was invited by NSF for presentation at IEEE AP-S Topical Conference on Wireless Communication Technology, Honolulu, Hawaii, USA, Oct 15-16, 2003.

small scale fading. A small scale fading factor² varies rapidly in both space and time, and the squared magnitude of the small scale fading factor is typically degree-2 chi-square distributed [PRO].

If we force a mobile to use its transmission power to maintain a constant SNR at the base station for each packet of data (typically, a packet length³ is 0.5 - 1 ms), then the average power consumption by the mobile over a period of many packets is given by the following:

$$E(P_T) = P_0 \int_0^{\infty} \frac{1}{x} p_2(x) dx = \infty \quad (1)$$

where P_T is the transmission power of the mobile for each packet, P_0 is the desired constant power at the base station (up to a squared large scale fading factor), x is the squared magnitude of the small scale fading factor from the mobile to the base station, and $p_n(x)$ is the probability density function of the degree- n chi-square random variable, i.e.,

$$p_n(x) = \frac{1}{\sigma^n 2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2\sigma^2} \quad (1a)$$

and $n\sigma^2$ is the mean of x .

The implication of (1) is that the battery of the mobile becomes flat very soon. On the other hand, if the mobile can not output very large power on average, the SNR at base station will be quite low for some packets, and these packets will be most likely to be lost. (As mobile phone users, have we all experienced broken voice transmissions and even undesired disconnections?)

Another technique of SNR control is a passive one where data transmission happens only when the fading condition is good. Stopping data transmission between base station and mobile when the fading at the mobile is deep is little different from data interruption. However, a recent study [VIS] argues that a base station communicating at any given time with the mobile that has the best fading condition yields a better network capacity. This argument requires several assumptions. One of them is that the base station knows the fast fading condition at each mobile at all time. Another assumption is that the delays caused to any mobile during its bad luck period are acceptable. In [VIS], a randomized beamforming is applied at base station so that each bad luck period for a mobile may be statistically brief and statistically fair to all.

The techniques mentioned above follow a top-down approach that does not enrich the capacity of each individual mobile, but provides a good top-down management of the existing resources.

1.2 SNR Control Using Smart Antennas

Another SNR or fading control technique is known as beamforming through multiple antennas. Beamforming is a physical layer technique while the previously mentioned two methods are networking techniques. Beamforming can be done at transmitting site and/or at receiving site. The beamforming techniques for receivers have been a classic topic in signal processing. Good reviews and some recent advances on this topic can be found in [VAN] and [HUA(a)]. The beamforming techniques for transmitters have been a hot topic in the past a few years, e.g., see [DAH] and [GIA]. This topic has led to an exciting cross-fertilization between information theory and signal processing. Space-time coding is believed to be the key to increase the throughput of a system with multiple transmitters and multiple receivers (also known as MIMO system). In the context of MIMO system, the notion of SNR control is in fact embodied in the notion of diversity. A system is said to have a diversity d if the average error rate (after averaging over random fading) is

² A complex value in general.

³ At the speed of 120 km per hour, we travel 0.034 m during 1 ms, which is only 0.112λ at 1 GHz. So, for each packet of 1 ms or less, even the small scale fading can be more or less constant. In IS-95, the transmission power of a mobile is updated by the base station once every 1.25 ms, which is within the range of small scale fading.

governed by $1/SNR^d$, where SNR is the averaged SNR over random fading. The maximum achievable diversity of M transmitters and N receivers is known to be MN . The maximum possible data rate (bits per second per Hertz) over the same system is $\min(M, N)\log_2 SNR$ for large SNR [TEL], [ZHE]. Many existing space-time codes can achieve the full diversity MN . But none is available to achieve the rate $r\log_2 SNR$ over all values of high SNR where $0 < r \leq \min(M, N)$. An optimal tradeoff between the maximum value of d and the maximum value of r is recently reported in [ZHE].

The current smart antenna techniques focus on transmitters and/or receivers exclusively. This approach is basically confined within a point-to-point link, which therefore can be viewed as a bottom-up approach in the broad context of wireless networks.

1.3 SNR Control Using Wireless Antennas

The conventional “wired” antennas have limitations. The spacing between antennas generally needs to be larger than half a wavelength to avoid fading correlation and antenna coupling. However, for none-line-of-sight and omni-directional communications, the carrier frequency can not be much higher than 1 GHz, or equivalently the wavelength can not be much smaller than 0.3 m. A large number of antennas with many meters in dimension is not often practical for a mobile node. This is particularly true in an “ad hoc mobile networking” (MANET) environment (also known as peer-to-peer mobile networking).

Without multiple antennas at each mobile node, how can we minimize the damage of small scale fading by taking the advantage of the spatially independent nature of small scale fading?

We believe that the logical approach is to design and develop wireless antennas or wireless relays that are distributed between a source and a destination. The wireless relays should be designed in such a way that the virtual point-to-point channel between the source and the destination has an improved quality of data transmission. The wireless relays should not increase either the total power consumption or the total bandwidth usage, which are the two most limited resources in wireless mobile environment.

Among many possible applications, wireless relays may be embedded in all mobile nodes in the MANET environment. In other words, each node may serve as a relay between two other nodes. Now, it may seem that a relay is no different from a node. The real answer is yes or no, depending on how the relays are designed.

The basic idea of our approach is captured by the term: *node-relays-node (NRN)*. The current practice in wireless mobile networks follows the *node-to-node (NTN)* strategy, where the bits are completely decoded (and often stored) at each node. It is important to note that both the NRN and NTN strategies refer to the physical layer. (The routing techniques at the networking level to avoid bad paths are all built upon the physical layer, and the quality of the physical layer underpins the entire network. Networking coding techniques to reduce error rates are all at the expense of additional bits transmitted at the physical layer.) If there is a single relay between the two nodes, this relay may do two different things: decode and forward, or amplify and forward. In the first case, the relay is no different from a node, and hence there is no diversity improvement. In the second case, as shown later, the effective channel between the two nodes actually has a poorer diversity, i.e., suffers more from small scale fading. Therefore, the NRN strategy does not include the single relay case.

If the relays in the NRN mode work in a serial fashion, then it is also obvious that there is no diversity improvement. Hence, the NRN strategy does not include serial single relays either. The relays in the NRN mode must then work in parallel. In a fast random fading environment, the fading condition at each relay may vary significantly for each arriving packet, and the (wireless mobile) relays may not be able to inform each other who has the best fading condition so that all

other relays can simply shut off, giving right to the chosen relay to relay the packet. In such a fast random fading environment, the only choice left for the relays is for them to relay their received packets *simultaneously*. If any of the relays completely decodes its received packet, then the total error rate of the NRN system is no better than the error rate of the NTN system. This once again rules out the idea of using regenerative relays for the NRN system.

What else can we do to make the NRN idea work? By simply amplifying and forwarding the received packets in a straightforward way, it does not work (we actually tried the naïve method when the idea of NRN was just conceived). A good solution must then come from a good design of space-time coding for the relays. Indeed, as we will show, good solutions do exist and yield a great improvement of the quality of data transmission between two nodes through the NRN strategy. In the rest of this paper, we will simply refer to the NRN strategy as wireless relays or wireless antennas.

The notion of wireless antennas shown here is related to virtual antenna arrays as shown in [LAN], [ANG] and [SCA] and [DOH] in that multiple spatially distributed mobile nodes are exploited to assist data transmission for each other. But none of them looks at the physical layer issues the way we do here. In [ANG], the space-time code proposed in [ALA] was attempted in a very brief and straightforward fashion.

1.4 Our Recent Key Discovery

Our contribution in this work can be briefly described as follows. We assume that there are R relays between a source and a destination. These relays are all wireless and mobile. Each relay receives the signal transmitted from the source, and constructs a sequence (or packet) of estimated symbols. These estimated symbols are essentially a baseband analog discrete signal, and they are not further quantized (or decoded) into bits at the relays. In other words, each relay performs demodulation and sampling, but no detection. After the demodulation, each relay reshuffles the order of the estimated symbols, with possible amplitude and phase changes, according to a space-time code. There is no exchange of these estimated symbols between the (mobile wireless) relays. Communication of control information between the relays is not required. The new sequence of estimated symbols at each relay is then transmitted to a destination. The relays retransmit their estimated symbols in a parallel and simultaneous fashion, where symbol synchronization of the relays is governed by their (common) input signal from the (common) source. The destination then receives a superimposed version of the symbols transmitted from the relays. With the knowledge of the space-time code used at the relays, the destination is then able to estimate the channel parameters and to detect the original symbol sequence transmitted from the source.

Our space-time code used for the relays is based on Hurwitz-Radon matrices. The space-time code effectively reshapes the SNR distribution of the detection statistics at the destination. The reshaped SNR distribution has a smaller variance than the SNR distribution without the relays. We will show that with one transmitter ($M=1$), one receiver ($N=1$), and R parallel wireless relays between the source and the destination, a diversity factor around $R/2$ can be achieved, i.e., the averaged bit error rate with R relays is in the order of $1/SNR^{R/2}$, which is opposed to $1/SNR$ in the case of no (or single) relay. More than 10 dB power saving from the baseline of a single relay system can be achieved with eight relays.

The Hurwitz-Radon matrices introduced in [GER]⁴ lay the mathematical foundation for our approach. The mathematical subject on Hurwitz-Radon matrices is arcane but however very rich. It was only recently brought to the information theory community [TAR]. We will next review,

⁴ The first author of [GER] graduated with Ph.D. from Syracuse University in 1968 exactly 20 years before the first author of this paper. If this indicates how slow information used to propagate, consider the fact that the Hurwitz-Radon matrices were established 90 years ago. It seems that many golden nuggets are yet to be found from the past.

and provide a new insight into, the Hurwitz-Radon matrices in the context of multiple transmitting antennas. We will then apply the Hurwitz-Radon matrices to wireless antennas.

The main contents of the rest of this paper are summarized by the headings of the sections and subsections. Part of this work was also presented in [HUA(b)-(d)].

2. Hurwitz-Radon Matrices for Multiple Transmitting Antennas

2.1 Hurwitz-Radon Matrices

Theorem 1 [GER]: There are p matrices $\{\mathbf{A}_0 \ \mathbf{A}_1 \ \cdots \ \mathbf{A}_{p-1}\}$ where \mathbf{A}_i is $L \times L$, $\mathbf{A}_0 = \mathbf{I}$, $\mathbf{A}_i^T \mathbf{A}_i = \mathbf{I}_L$ (i.e., $L \times L$ identity matrix), $\mathbf{A}_i^T \mathbf{A}_j = -\mathbf{A}_j^T \mathbf{A}_i$ for $i \neq j$, and furthermore the maximum value p_{\max} of p is given as follows. Let $L = 2^a b$ where b is odd, and $a = 4c + d$ where $0 \leq d < 4$, then $p_{\max} = 8c + 2^d$.

The matrices⁵ defined above are known as the family of Hurwitz-Radon matrices and can be easily generated as shown in [GER]. From the properties of these matrices, it is easy to verify that each element of \mathbf{A}_i is from $\{0, 1, -1\}$, and there is one and only one non-zero element in each row of \mathbf{A}_i .

It follows from Theorem 1 that for any real vector \mathbf{a} , we have the property

$$\mathbf{a}^T \mathbf{A}_i^T \mathbf{A}_j \mathbf{a} = \begin{cases} \|\mathbf{a}\|^2 & i = j \\ 0 & i \neq j \end{cases}.$$

Furthermore, p_{\max} is independent of b . If a is large and $b=1$ then $p_{\max} \cong 2 \log_2 L$. If b is allowed to be any odd number, the relationship between p_{\max} and L is not monotonic. One can also verify that $p_{\max} \leq L$ in general, and $p_{\max} = L$ if and only if $L = 2$, $L = 4$, or $L = 8$. The following table illustrates more details (with $b=1$):

L	a	c	d	p_{\max}
2	1	0	1	2
4	2	0	2	4
8	3	0	3	8
16	4	1	0	9
32	5	1	1	10
64	6	1	2	12
128	7	1	3	16
256	8	2	0	17
512	9	2	1	18
1024	10	2	2	20

Table 1: Examples of the key parameters of the Hurwitz-Radon matrices

2.2 Transmitting Real Symbols

Let \mathbf{s} be a $L \times 1$ real vector (such as binary PSK and M-PAM symbol modulations) and

⁵ The identity matrix is excluded from the family in the original definition.

$$\mathbf{X} = [\mathbf{A}_0 \mathbf{s} \quad \mathbf{A}_1 \mathbf{s} \quad \cdots \quad \mathbf{A}_{M-1} \mathbf{s}] \quad (2)$$

which is $L \times M$ where $M \leq p_{\max}$. Then it follows that $\mathbf{X}^T \mathbf{X} = \|\mathbf{s}\|^2 \mathbf{I}_M$ where $\|\mathbf{s}\|$ denotes the norm of the vector \mathbf{s} , i.e., all columns of the real matrix \mathbf{X} are orthogonal and of equal norm. The property of orthogonality is important. For example, if we let the i th column of \mathbf{X} be the sequence of L symbols from the i th transmitting antenna, then a corresponding sequence of symbols (or baseband signals) received by a receiver can be represented by the following vector: $\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{w}$ where the i th element of \mathbf{h} is the real fading factor between the i th transmitter and the receiver, and \mathbf{w} is the noise vector. Furthermore, we can write $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{w}$ where

$$\mathbf{H} = \sum_{i=0}^{M-1} h_i \mathbf{A}_i \quad (3)$$

which is also an orthogonal matrix, i.e., $\mathbf{H}^T \mathbf{H} = \|\mathbf{h}\|^2 \mathbf{I}_L$. Then, the sufficient statistics for detection of the symbol vector \mathbf{s} is given by $\mathbf{r} \triangleq \mathbf{H}^T \mathbf{y} = \|\mathbf{h}\|^2 \mathbf{s} + \mathbf{H}^T \mathbf{w}$ where all elements of \mathbf{s} are decoupled and the noise vector $\mathbf{H}^T \mathbf{w}$ is white Gaussian provided \mathbf{w} is white Gaussian. Hence, the optimal detection of the symbol vector \mathbf{s} based on \mathbf{r} is very simple⁶. Also, note that for every M symbols that each transmitter takes in, there are M coded symbols that the transmitter spills out. So, the above real valued code matrix \mathbf{X} has a full-rate.

2.3 Transmitting Complex Symbols

However, if we let \mathbf{s}_c be a block of complex symbols (as in QPSK or M-PSK symbol modulation), then the matrix $\mathbf{B} = [\mathbf{A}_0 \mathbf{s}_c \quad \mathbf{A}_1 \mathbf{s}_c \quad \cdots \quad \mathbf{A}_{M-1} \mathbf{s}_c]$ (similar to \mathbf{X}) is no longer an orthogonal matrix in general. But if we define

$$\mathbf{X}_c = \begin{bmatrix} \mathbf{B} \\ \mathbf{B}^* \end{bmatrix},$$

we have a complex orthogonal matrix, i.e., one can verify that $\mathbf{X}_c^H \mathbf{X}_c = \mathbf{B}^H \mathbf{B} + (\mathbf{B}^H \mathbf{B})^* = 2\|\mathbf{s}\|^2 \mathbf{I}_M$. This transformation from \mathbf{s}_c to \mathbf{X}_c is known as half-rate orthogonal complex space-time block code [TAR]. If the i th column of \mathbf{X}_c represents the $2L$ symbols transmitted from the i th transmitter, then a corresponding vector of complex symbols received by a receiver is represented by $\mathbf{y}_c = \mathbf{X}_c \mathbf{h}_c + \mathbf{w}_c$ where the i th element of \mathbf{h}_c is the complex fading between the i th transmitter and the receiver, and \mathbf{w}_c is the complex noise vector. Alternatively, we can write

$$\mathbf{y}_c = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \text{ and } \mathbf{y}_c' \triangleq \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2^* \end{bmatrix}.$$

Then, it follows that $\mathbf{y}_c' = \mathbf{H}_c \mathbf{s}_c + \mathbf{w}_c$ where

$$\mathbf{H}_c = \begin{bmatrix} \sum_{i=0}^{M-1} h_{c,i} \mathbf{A}_i \\ \sum_{i=0}^{M-1} h_{c,i}^* \mathbf{A}_i \end{bmatrix} \quad (4)$$

which is complex and orthogonal, i.e., $\mathbf{H}_c^H \mathbf{H}_c = 2\|\mathbf{h}_c\|^2 \mathbf{I}_L$. Like the real case, the optimal detection of the symbol vector \mathbf{s}_c from the sufficient statistics $\mathbf{H}_c^H \mathbf{y}_c'$ is very simple as all

⁶ In the case of binary PSK, the optimal detection of each symbol is done by thresholding the corresponding element in the vector \mathbf{r} .

symbols in \mathbf{s}_c are decoupled from each other and the noise vector in the sufficient statistics is also white Gaussian.

2.4 Using Two Transmitters

The full-rate orthogonal complex space-time (linear) block codes however do not exist unless $p=2$. The proof took 2 pages in [TAR] and only nine lines in [GAN]. The latter used an existing theorem established in [GER]. When $p=2$, we let $\mathbf{s} = [s_1 \ s_2]^T$. From now on, all symbols are complex unless stated otherwise, and hence we drop the subscript c . Alamouti [ALA] introduced the following code

$$\mathbf{C} = \begin{bmatrix} s_1 & s_2 \\ -s_2^* & s_1^* \end{bmatrix}$$

where the i th column corresponds to the i th transmitter. (The original matrix shown in [ALA] is a transpose of this one, but they are equivalent to each other since the i th row of the matrix in [ALA] corresponds to the i th transmitter.) It is easy to verify that \mathbf{C} is an orthogonal matrix. But the 2×2 complex orthogonal matrix is not unique. Taking transpose and/or complex conjugate of \mathbf{C} yields different 2×2 orthogonal matrices. Multiplying any complex unitary matrix such as $\text{diag}[e^{j\theta_1} \ e^{j\theta_2}]$ to \mathbf{C} from left or right also obviously yields different 2×2 orthogonal matrices. Some examples were given in [JAF]. Furthermore, all of these orthogonal matrices can be expressed in terms of the 2×2 Hurwitz-Radon matrices. With $p=2$, the Hurwitz-Radon matrices $\{\mathbf{A}_0 \ \mathbf{A}_1\}$ are

$$\mathbf{A}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \mathbf{A}_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Then, one can verify that $\mathbf{C}^T = [\mathbf{A}_0 \mathbf{s}_r \ -\mathbf{A}_1 \mathbf{s}_r] + j[\mathbf{A}_0 \mathbf{s}_i \ \mathbf{A}_1 \mathbf{s}_i]$ where $j = \sqrt{-1}$, $\mathbf{s}_r = \text{Re}(\mathbf{s})$ and $\mathbf{s}_i = \text{Im}(\mathbf{s})$.

We now illustrate the importance of the spatial diversity in relation to the discussion of (1). Assuming that we use two transmitters and the space-time code \mathbf{C}^T (the transpose of the Alamouti code). Then, the vector of the two consecutive symbols received by a receiver is⁷

$$\begin{bmatrix} y(1) \\ y(2) \end{bmatrix} = \begin{bmatrix} s(1) & -s^*(2) \\ s(2) & s^*(1) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} w(1) \\ w(2) \end{bmatrix}$$

or equivalently,

$$\begin{bmatrix} y(1) \\ y^*(2) \end{bmatrix} = \begin{bmatrix} h_1 & -h_2 \\ h_2^* & h_1^* \end{bmatrix} \begin{bmatrix} s(1) \\ s^*(2) \end{bmatrix} + \begin{bmatrix} w(1) \\ w^*(2) \end{bmatrix}.$$

We consider the sufficient statistics:

$$\begin{bmatrix} r(1) \\ r^*(2) \end{bmatrix} \hat{=} \begin{bmatrix} h_1 & -h_2 \\ h_2^* & h_1^* \end{bmatrix}^H \begin{bmatrix} y(1) \\ y^*(2) \end{bmatrix} = \|\mathbf{h}\|^2 \begin{bmatrix} s(1) \\ s^*(2) \end{bmatrix} + \begin{bmatrix} h_1 & -h_2 \\ h_2^* & h_1^* \end{bmatrix}^H \begin{bmatrix} w(1) \\ w^*(2) \end{bmatrix}$$

Assuming that the symbols $s(n)$ are white and each of the variance σ_s^2 , and the noise $w(n)$ is also white and each of the variance σ_w^2 , then the SNR for a fixed channel fading is

$$\text{SNR}(\mathbf{h}) = \|\mathbf{h}\|^2 \frac{\sigma_s^2}{\sigma_w^2}.$$

If we want to maintain a constant SNR at the receiver (i.e., a constant probability of errors), the transmitter must consume the power $P_T = P_0 / \|\mathbf{h}\|^2$ for each packet of two symbols. The

⁷ We consider frequency flat fading here. The theory shown in this paper can be applied to each narrowband carrier in a multicarrier system.

probability density function $p_4(x)$ of $x \triangleq \|\mathbf{h}\|^2$ (with two independent complex Gaussian random variables of variance σ_h^2) is known to be chi-square distribution of degree-4 (page 41 in [PRO]). Then, the averaged power consumption can be shown to be

$$E(P_T) = P_0 \int_0^\infty \frac{p_4(x)}{x} dx = \frac{P_0}{\sigma_h^2}$$

which is a finite number in contrast to (1). A similar result can be obtained for one transmitter and two receivers. Therefore, the method of power control as proposed in IS-95 would require at least two transmitters, or two receivers, or equivalently at least two independent channels (in time, frequency or code). Indeed, in IS-95, a rake receiver with 3 taps is used, which is supposed to yield 3 equivalent diversity channels provided that the multipath time spreading is more than 3 chip intervals. But when the multipath time spread is not large (which highly depends on the terrain) little diversity gain may be available from the rake receiver.

2.5 An Asymptotical Property

It is important to note that although the orthogonality of the transmitted symbol matrix from multiple transmitters generally implies the orthogonality of the coefficient matrix (such as (2) and (3)) of the received symbol vector at the receiver. But special situations may happen. For example, if the complex symbols in the vector \mathbf{s} are independent and identically distributed (i.i.d.), which is a reasonable assumption, then the complex matrix \mathbf{X} defined in (2) is asymptotically orthogonal as p becomes large. To prove this fact, we write

$$\begin{aligned} (\mathbf{X}^H \mathbf{X})_{k,l} &= \mathbf{s}^H \mathbf{A}_k^T \mathbf{A}_l \mathbf{s} = \mathbf{s}_r^T \mathbf{A}_k^T \mathbf{A}_l \mathbf{s}_r + \mathbf{s}_i^T \mathbf{A}_k^T \mathbf{A}_l \mathbf{s}_i + j(\mathbf{s}_r^T \mathbf{A}_k^T \mathbf{A}_l \mathbf{s}_i - \mathbf{s}_i^T \mathbf{A}_k^T \mathbf{A}_l \mathbf{s}_r) \\ &= \begin{cases} \|\mathbf{s}\|^2 & k = l \\ j(\mathbf{s}_r^T \mathbf{A}_k^T \mathbf{A}_l \mathbf{s}_i - \mathbf{s}_i^T \mathbf{A}_k^T \mathbf{A}_l \mathbf{s}_r) & k \neq l \end{cases} \end{aligned}$$

Note that $\mathbf{A}_k^T \mathbf{A}_l$ is an orthogonal matrix, i.e., $(\mathbf{A}_k^T \mathbf{A}_l)^T \mathbf{A}_k^T \mathbf{A}_l = \mathbf{I}_L$, and there is only one nonzero (1 or -1) element per each row of $\mathbf{A}_k^T \mathbf{A}_l$. Assuming that all elements in \mathbf{s} are i.i.d. random variables, and each of the variance σ_s^2 , the large sample theorem in statistics implies that

$$\lim_{L \rightarrow \infty} \frac{1}{L} (\mathbf{X}^H \mathbf{X})_{k,l} = \begin{cases} \sigma_s^2 & k = l \\ O\left(\frac{1}{L}\right) \rightarrow 0 & k \neq l \end{cases} \quad (5)$$

In the above limit, the zero is in the order of $O(1/L)$. The number of off-diagonal elements in $\mathbf{X}^H \mathbf{X}$ is in the order of $O(M^2)$. The error of the orthogonality of \mathbf{X} can be measured by

$$E_{orth}(\mathbf{X}) = \frac{\sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M |(\mathbf{X}^H \mathbf{X})_{i,j}|^2}{\sum_{i=1}^M |(\mathbf{X}^H \mathbf{X})_{i,i}|^2}$$

Using (5), we have that for large L , $E_{orth}(\mathbf{X}) = O\left(\frac{M^2/L^2}{M}\right) = O\left(\frac{M}{L^2}\right) = O\left(\frac{2 \log_2 L}{L^2}\right) \rightarrow 0$.

Therefore, despite the growing dimension of $\mathbf{X}^H \mathbf{X}$ as L becomes large, we still have $\lim_{L \rightarrow \infty} \frac{1}{L} (\mathbf{X}^H \mathbf{X}) = \sigma_s^2 \mathbf{I}_M$. The next figure shows ten independent realizations of the orthogonality measure.

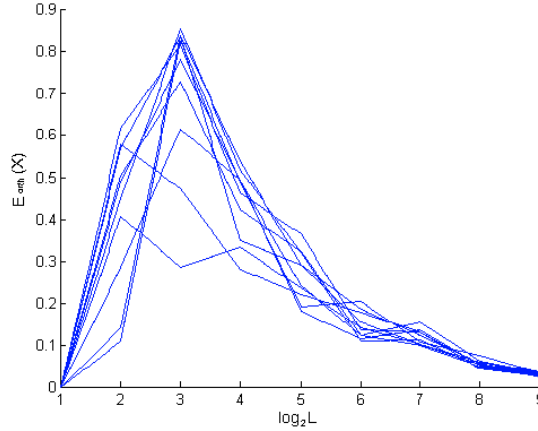


Figure 1: Asymptotically orthogonal property of full rate space-time codes: ten independent realizations.

In fact, a more general fact is that the following matrix is asymptotically orthogonal under the i.i.d. condition of \mathbf{s} :

$$\mathbf{X}(\mathbf{s}) = [\mathbf{A}_0 \mathbf{s}_r \quad \mathbf{A}_1 \mathbf{s}_r \quad \cdots \quad \mathbf{A}_{M-1} \mathbf{s}_r] + j[\mathbf{B}_0 \mathbf{s}_i \quad \mathbf{B}_1 \mathbf{s}_i \quad \cdots \quad \mathbf{B}_{M-1} \mathbf{s}_i]$$

where the two sets of matrices $\{\mathbf{A}_k\}$ and $\{\mathbf{B}_l\}$ are each chosen from the Hurwitz-Radon matrices. When the two sets of matrices are equal, $\mathbf{X}(\mathbf{s})$ reduces to (2).

However, the asymptotical orthogonality of \mathbf{X} does not imply that the corresponding complex coefficient matrix \mathbf{H} defined in (3) is also asymptotically orthogonal. In fact, \mathbf{H} is not asymptotically orthogonal even if the complex elements in \mathbf{h} are i.i.d. (though a reasonable assumption of small scale fading factors). This is because that the size $L \times L$ of the matrix \mathbf{H} grows faster than the number M of the elements in \mathbf{h} (i.e., $M \leq p_{\max} \approx 2 \log_2 L$) and hence $E_{orth}(\mathbf{H}) \rightarrow 0$ even as $M \rightarrow \infty$.

3. Hurwitz-Radon Matrices for Multiple Wireless Antennas

The diversity techniques using multiple conventional antennas may be suitable for base stations. For a mobile user, even two antennas (of, say, half meter apart) can be too cumbersome. In a peer-to-peer mobile network, the conventional antenna arrays may have very limited use. For this reason, we have explored the idea of space-time coding for multiple wireless antennas or wireless relays. Among many possible applications, these relays can be ad hoc and embedded in the mobile nodes in a peer-to-peer mobile network.

3.1 Using Two Wireless Antennas

Let us start with two wireless relays ($R=2$), one transmitter at the source ($M=1$) and one receiver at the destination ($N=1$). The source produces two consecutive symbols, $s(1)$ and $s(2)$. Each of the relays receives two corrupted versions, i.e., the two corresponding complex symbols received by the i th relay are

$$\begin{bmatrix} s_i(1) \\ s_i(2) \end{bmatrix} = \eta h_i \begin{bmatrix} s(1) \\ s(2) \end{bmatrix} + \begin{bmatrix} w_i(1) \\ w_i(2) \end{bmatrix} \quad (6)$$

where η is the real valued large scale (slow) fading factor between the source and the group of two relays, and h_i is the complex valued small scale (fast) fading factor (of unit variance)⁸ between the source and the i th relay, and $w_i(n)$ is the complex white noise at the i th relay and of the variance σ_w^2 . We now let the output symbols of the two relays be

$$\begin{bmatrix} x_1(1) & x_2(1) \\ x_1(2) & x_2(2) \end{bmatrix} \triangleq \sqrt{G} \begin{bmatrix} s_1(1) & -s_2^*(2) \\ s_1(2) & s_2^*(1) \end{bmatrix} \quad (7)$$

where $x_i(n)$ is the output from the i th relay, and G is the power gain factor. This structure resembles the matrix \mathbf{C}^T (the transpose of Alamouti code but not the original as shown in [ALA]). The receiver at the destination then receives the following corresponding two symbols:

$$\begin{bmatrix} y(1) \\ y(2) \end{bmatrix} = \gamma \begin{bmatrix} x_1(1) & x_2(1) \\ x_1(2) & x_2(2) \end{bmatrix} \cdot \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} v(1) \\ v(2) \end{bmatrix} \quad (8)$$

where γ is the real valued large scale fading factor between the relays and the destination, and g_i is the complex valued small scale fading factor (of unit variance) between the i th relay and the destination, and $v(n)$ is the noise (additional) at the destination and of the variance σ_v^2 . A simple analysis of (6)-(8) shows that

$$\begin{bmatrix} y(1) \\ y^*(2) \end{bmatrix} = \sqrt{G\gamma\eta} \begin{bmatrix} g_1 h_1 & -g_2 h_2^* \\ g_2^* h_2 & g_1^* h_1^* \end{bmatrix} \cdot \begin{bmatrix} s(1) \\ s^*(2) \end{bmatrix} + \sqrt{G\gamma} \begin{bmatrix} g_1 w_1(1) - g_2 w_2^*(2) \\ g_1^* w_1^*(2) + g_2^* w_2(1) \end{bmatrix} + \begin{bmatrix} v(1) \\ v^*(2) \end{bmatrix}. \quad (9)$$

We see that the coefficient matrix of the original symbol vector is complex orthogonal, and the last two noise terms are still white and of the variance $G\gamma^2 \|\mathbf{g}\|^2 \sigma_w^2 + \sigma_v^2$. Therefore, the optimal detection of the original symbols can be carried out on each individual component of the following sufficient statistics:

$$\begin{bmatrix} r(1) \\ r^*(2) \end{bmatrix} \triangleq \begin{bmatrix} g_1 h_1 & -g_2 h_2^* \\ g_2^* h_2 & g_1^* h_1^* \end{bmatrix}^H \begin{bmatrix} y(1) \\ y^*(2) \end{bmatrix} \quad (10)$$

and the probability of detection error is governed by the SNR of this statistics. The SNR of (10) can be shown to be

$$SNR_2(\mathbf{h}; \mathbf{g}) = \frac{G\gamma^2 \eta^2 \left(\sum_{i=1}^2 |g_i|^2 |h_i|^2 \right) \sigma_s^2}{G\gamma^2 \sigma_w^2 \sum_{i=1}^2 |g_i|^2 + \sigma_v^2}. \quad (11)$$

With this expression of SNR and a set of fixed fading factors, the bit error rate (BER) is known for a variety of symbol modulation schemes. For example, if the symbol modulation is QPSK, the BER conditional upon the fadings is given by $Q(\sqrt{SNR_2(\mathbf{h}; \mathbf{g})})$ [PRO]. However, the averaged BER (averaged over the random fading factors) $E(Q(\sqrt{SNR_2(\mathbf{h}; \mathbf{g})}))$ depends on the distribution of $SNR_2(\mathbf{h}; \mathbf{g})$.

We next discuss several special cases to examine the structures of $SNR_2(\mathbf{h}; \mathbf{g})$.

3.1.1 When the relays are near the destination

⁸ We have not been quite consistent about this variance. Two choices are used, i.e., the variance of the complex small scale fading factor is either one or two. But we hope that the choice is clear in the context and causes no confusion.

If the relays are much closer to the destination than the source, then the relays may consume a much smaller amount of power than the source and ensure that $G\gamma^2\sigma_w^2\sum_{i=1}^2|g_i|^2 \gg \sigma_v^2$. In this special case, we denote $SNR_2(\mathbf{h};\mathbf{g})$ by $SNR_{2,destination}(\mathbf{h};\mathbf{g})$, and we have from (11) that

$$SNR_{2,destination}(\mathbf{h};\mathbf{g}) = \frac{\eta^2 \left(\sum_{i=1}^2 |g_i|^2 |h_i|^2 \right) \sigma_s^2}{\sigma_w^2 \sum_{i=1}^2 |g_i|^2} = SNR^* \left(\sum_{i=1}^2 \left(\frac{|g_i|^2}{\sum_{i=1}^2 |g_i|^2} \right) |h_i|^2 \right) \quad (12)$$

where

$$SNR^* = \frac{\eta^2 \sigma_s^2}{\sigma_w^2} \quad (12a)$$

is the average SNR of the signal received by each relay (in the neighborhood of the destination). We see that (12) is a weighted average of the degree-2 chi-square random variables $|h_i|^2$, and the weights themselves depend on the degree-2 chi-square random variables $|g_i|^2$.

Without the relays, the SNR in the neighborhood of the destination is simply

$$SNR_0(h) = SNR^* \cdot |h|^2 \quad (13)$$

where $|h|^2$ has the same distribution as $|h_1|^2$ and $|h_2|^2$, and $E(SNR_0(h)) = SNR^*$. SNR^* is also the mean of $SNR_{2,destination}(\mathbf{h};\mathbf{g})$, i.e.,

$$\begin{aligned} E(SNR_{2,destination}(\mathbf{h};\mathbf{g})) &= E(E(SNR_{2,destination}(\mathbf{h};\mathbf{g}|\mathbf{g}))) \\ &= E \left(\frac{\eta^2 \sigma_s^2}{\sigma_w^2} \left(\sum_{i=1}^2 \left(\frac{|g_i|^2}{\sum_{i=1}^2 |g_i|^2} \right) \sigma_h^2 \right) \right) = \frac{\eta^2 \sigma_s^2}{\sigma_w^2} = SNR^* \end{aligned} \quad (14)$$

We see that with or without the relays that are near the destination, the average SNR is not affected.

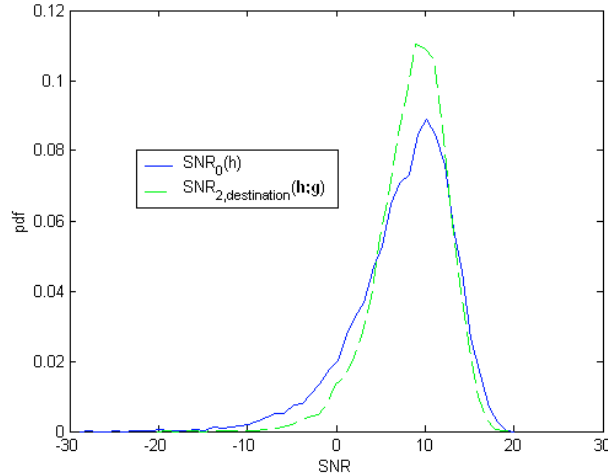


Figure 2: pdf of $SNR_0(h)$ and $SNR_{2,destination}(\mathbf{h};\mathbf{g})$ in dB with $SNR^* = 10dB$.

But because of the average effect shown in (12), the probability density function (pdf) of $SNR_{2, \text{near}}(\|\mathbf{h}\|^2)$ is more compressed towards SNR^* than the pdf of $SNR_0(h)$. Reducing the pdf value of SNR in its lower value region has a major impact on reducing the averaged BER at the destination. Simulations based on (14) show that the diversity is about 1.6, i.e., $E\left(Q\left(\sqrt{SNR_{2, \text{destination}}(\mathbf{h}; \mathbf{g})}\right)\right) \propto 1/SNR^{1.6}$. Figure 2 shows the effect of space-time coding on the SNR distribution at the destination, i.e., the SNR distribution with the relays is more compressed to its mean than that without the relays. Figure 3 shows the average BER assuming QPSK. We see that more than 5 dB SNR gain is achieved at an average BER of 10^{-3} or lower. There is some overhead associated with this number, which includes the transmission power by the relays and the communication channel required between the relays and the destination. The transmission power by the relays may be negligible since the relays here are assumed to be near the destination. The communication channel between the relays and the destination is additional with respect to a zero-relay system. Later on, we will consider more than two relays, where the SNR gain is much more significant.

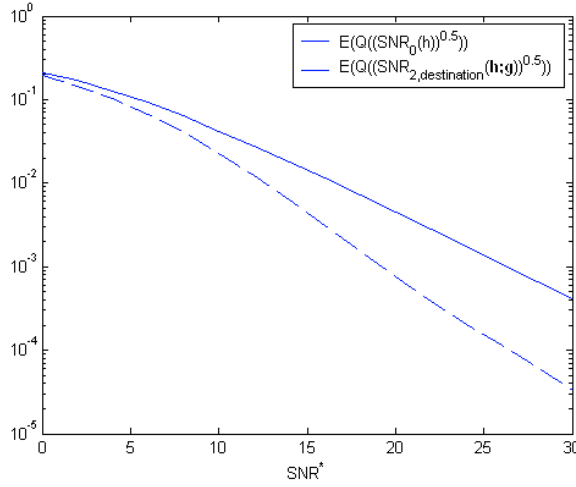


Figure 3: Average BER, i.e., $E\left(Q\left(\sqrt{SNR_{2, \text{destination}}(\mathbf{h}; \mathbf{g})}\right)\right)$ and $E\left(Q\left(\sqrt{SNR_0(\mathbf{h})}\right)\right)$, versus SNR^* in dB.

3.1.2 When the relays are near the source

The previous analysis also holds when the relays are near the source provided that $G\gamma^2\sigma_w^2\sum_{i=1}^2|g_i|^2 \gg \sigma_v^2$. But this assumption requires a very large power consumption at the relays. However, the power consumption at the source in this case can be very small. In fact, even with a relatively small power consumption at the source, the SNR of the signals received by the relays can be very high (i.e., $\eta^2\sigma_s^2 \gg \sigma_w^2$), and then it is even possible to have $G\gamma^2\sigma_w^2\sum_{i=1}^2|g_i|^2 \ll \sigma_v^2$. If this is true, then (11) becomes

$$SNR_{2, \text{source}}(\mathbf{h}; \mathbf{g}) = SNR^* \left(\frac{1}{2} \sum_{i=1}^2 |g_i|^2 |h_i|^2 \right) \quad (15)$$

where

$$SNR^* = \frac{2G\gamma^2\eta^2\sigma_s^2}{\sigma_v^2} \quad (15a)$$

This SNR^* is equivalent to (12a) in the sense that if the total power consumption by the source (negligible for (15)) and the relays (negligible for (12)) remains constant, SNR^* is the average SNR of the signal in the neighborhood of the destination. We see that (15) represents a different average of small scale fading from (12). But their general impact on the average BER is similar although our simulations show that (15) has a slightly smaller diversity than (12).

3.1.3 Optimal placement of the relays

We now consider how to optimally place the relays in a large scale. The large scale distance between the relays and the destination is denoted by d , and the large scale distance between the source and the destination is denoted by D . The large scale fading factors can be modelled as:

$$\eta^2 = \left(\frac{d_0}{D-d}\right)^n \quad \text{and} \quad \gamma^2 = \left(\frac{d_0}{d}\right)^n \quad (16)$$

where d_0 is a small constant in comparison to D , and n is the exponent of the large scale fading. The exponent may range from 2 to 4 or even higher depending on the environment [RAP]. Assume that $\sigma_w^2 = \sigma_v^2 = \sigma^2$. We denote by P_T the power transmitted by the source, and P_R the power by each relay. That is, $P_T = \sigma_s^2$ and

$$P_R = G(\eta^2 P_T + \sigma^2). \quad (16a)$$

Then, we have from (11), (16) and (16a) that

$$SNR_2(\mathbf{h}; \mathbf{g}) \propto \frac{P_T}{\sigma^2} \frac{GR \left(\frac{d_0}{d}\right)^n \left(\frac{d_0}{D-d}\right)^n}{GR \left(\frac{d_0}{d}\right)^n + 1} = \frac{P_T}{\sigma^2} \frac{\left(\frac{d_0}{d}\right)^n \left(\frac{d_0}{D-d}\right)^n}{\left(\frac{d_0}{d}\right)^n + \left(\frac{d_0}{D-d}\right)^n \frac{P_T}{RP_R} + \frac{\sigma^2}{RP_R}} \quad (17)$$

where $R = 2$ for the two relays case. To find the optimal d , we must maximize (17). Assume that

$P_T = RP_R$ and $P_T \left(\frac{d_0}{D}\right)^n \gg \sigma^2$, then the last term in the denominator of (17) is negligible, and

furthermore the maximizer of (17) is $d = \frac{D}{2}$. In particular, with $D - d_0 \approx D$, we have

$$SNR_2(\mathbf{h}; \mathbf{g}) \propto \begin{cases} \frac{P_T}{\sigma^2} \left(\frac{d_0}{D}\right)^n; & d = d_0, \quad d = D - d_0 \\ \frac{P_T}{\sigma^2} \left(\frac{d_0}{D}\right)^n 2^{n-1}; & d = D/2 \end{cases}. \quad (18)$$

Therefore, we see that by placing the relays at half way between the source and the destination (instead of nearing the source or the destination), we have a SNR gain in the order of 2^{n-1} (which is about 9 dB when $n=4$). We next provide a more accurate evaluation of the half way relays.

Assuming $d = \frac{D}{2}$, $P_T = RP_R$ and $P_T \left(\frac{d_0}{D}\right)^n \gg \sigma^2$, (16a) yields $G = \frac{P_R}{P_T} \left(\frac{D/2}{d_0}\right)^n$. Substituting this expression in (11) yields the following equivalent form:

$$SNR_{2,mid}(\mathbf{h}; \mathbf{g}) = SNR^* \cdot 2^n \cdot \frac{\left(\frac{1}{2} \sum_{i=1}^2 |g_i|^2 |h_i|^2 \right)}{\frac{1}{2} \sum_{i=1}^2 |g_i|^2 + 1} \quad (19)$$

where

$$SNR^* = \frac{P_T}{\sigma^2} \left(\frac{d_0}{D} \right)^n \quad (20)$$

which is the mean SNR at the destination when no relay is used and a total power of P_T is used at the source. This is in fact the same SNR^* defined in (12a). Note that from the denominator of (19), we can see that the optimal placement of the relays effectively makes the variance of the noise term due to the relays equal to the variance of the noise term due to the destination.

We also see that the averaging of the small scale fading magnitudes in (12) or (15) is not exactly the same but similar to that in (19). For (12) or (15), the total power consumption is slightly higher than P_T , but for (19), the total power consumption is $2P_T$. Considering the above two observations, the net SNR gain from (12) or (15) to (19) is between 2^{n-2} and 2^{n-1} (which is about 6-9 dB when $n=4$).

3.1.4 More comparison examples

To further evaluate the significance of (19), we now consider a few different alternatives.

Example 1: we consider a single relay placed at half way between the source and the destination, the same power constraint (16a) is imposed, and we let $P_T = P_R$ (since $R=1$). Following the same analysis leading to (19), we have

$$SNR_1(h; g) = SNR^* \cdot 2^n \cdot \frac{|g|^2 |h|^2}{|g|^2 + 1} \quad (21)$$

where $|h|$ and $|g|$ have the same statistical distribution as those in (19). We see that in (19), there is an averaging, but in (21), there is none. However, unlike the zero relay case, (19) and (21) benefit equally from the large scale fading. Simulations based on (21) show that the diversity factor of the single relay method is less than one.

Example 2: we consider a single regenerative relay. This relay decodes symbols into bits and then retransmit the bits to the destination. The BER at the relay is governed by this SNR:

$$SNR_{1,reg}(h) = SNR^* \cdot 2^n \cdot |h|^2 \quad (22)$$

where 2^n is due to the half way distance of the relay. Then, assuming QPSK, the BER at the relay conditioned upon h is

$$BER_h(h) = Q\left(\sqrt{SNR_{1,reg}(h)}\right). \quad (22a)$$

The additional BER from the relay to the destination is given by the same expression $BER_g(g)$.

The total averaged BER of a single regenerative relay system is simply

$$BER_{reg}(1) = E\left(1 - (1 - BER_h(h))(1 - BER_g(g))\right). \quad (22b)$$

Under a small BER condition, we can write

$$BER_{reg}(1) \approx 2 \cdot BER_h = 2E\left(Q\left(\sqrt{SNR_{1,reg}(h)}\right)\right) \quad (22c)$$

where $BER_h = E(BER_h(h)) = E(BER_g(g))$. Except for the benefit from the large scale fading, the average BER of using a single regenerative relay is similar to the case of no relay. It is obvious that the diversity factor of a single regenerative relay system is one (same as the case of no relay).

Example 3: we consider two regenerative (parallel) relays. With two regenerative relays, we know that the BER at the i th relay is given by $BER_{h_i}(h_i)$. The additional BER from the relays to the destination is governed by the following SNR (like the case of two ideal transmitters):

$$SNR_{2,reg}(\mathbf{g}) = SNR_0 \cdot 2^n \cdot \left(\frac{1}{2} \sum_{i=1}^2 |g_i|^2 \right). \quad (23)$$

Then, the additional BER⁹ between the relays and the destination is

$$BER_{\mathbf{g}}(\mathbf{g}) = Q\left(\sqrt{SNR_{2,reg}(\mathbf{g})}\right). \quad (24)$$

Finally, the total averaged BER of the two regenerative relays system is

$$BER_{reg}(2) = E\left(1 - (1 - BER_{h_1}(h_1))(1 - BER_{h_2}(h_2))(1 - BER_{\mathbf{g}}(\mathbf{g}))\right). \quad (25)$$

Under a small BER condition, it is easy to show that

$$BER_{reg}(2) = 2 \cdot BER_h + BER_{\mathbf{g}} > BER_{reg}(1) \quad (26)$$

where $BER_h = E(BER_{h_i}(h_i))$ and $BER_{\mathbf{g}} = E(BER_{\mathbf{g}}(\mathbf{g}))$. With (26), we conclude that if the bits are regenerated at the relays, we should keep the number of relays at the minimum which is one, i.e., there is no benefit by considering multiple regenerative (parallel) relays.

3.1.5 Other useful variations

We have assumed that $\max\{|h_1|, |h_2|\}$ is unknown to the relays although $|h_1|$ is available at the relay 1, and $|h_2|$ is available at the relay 2. In a very fast fading environment, the relays may not have the time to communicate with each other to find out which of them has less channel attenuation. In this case, all relays should be forced to relay the signals (unless the SNR of some relays is clearly below a threshold). Under the parallel relaying arrangement, the coding scheme (7) applied at the relays yields a better averaged BER than a single-path channel.

But if the fading is not very fast, or in other words, the values of $|h_1|$ and $|h_2|$ remain unchanged for a long period of time, then the relays may be able to communicate with each other to maximize $SNR_2(\mathbf{h}; \mathbf{g})$ by choosing some weightings on $|g_1|$ and $|g_2|$. In fact, under a symmetry condition, the best SNR is achieved by simply using the better relay of the two at any given time. The better relay is the relay whose fading magnitude is given by $\max\{|h_1|, |h_2|\}$. In this paper, we will not further consider the switching scheme.

Somewhere between the switching scheme and the averaging scheme, we can apply some soft switching on the relays. For example, each relay can match (multiplying) its received signal by its (conjugate) small scale fading factor. For example, we can define the following space-time code:

$$\begin{bmatrix} x_1(1) & x_2(1) \\ x_1(2) & x_2(2) \end{bmatrix} \triangleq \sqrt{G\eta} \begin{bmatrix} h_1^* s_1(1) & -h_2 s_2^*(2) \\ h_1^* s_1(2) & h_2 s_2^*(1) \end{bmatrix}. \quad (27)$$

Using this code in (8) yields

$$\begin{bmatrix} y(1) \\ y^*(2) \end{bmatrix} = \sqrt{G\eta\gamma} \begin{bmatrix} g_1 |h_1|^2 & -g_2 |h_2|^2 \\ g_2^* |h_2|^2 & g_1^* |h_1|^2 \end{bmatrix} \cdot \begin{bmatrix} s(1) \\ s^*(2) \end{bmatrix} + \sqrt{G}\gamma \begin{bmatrix} g_1 h_1^* w_1(1) - g_2 h_2 w_2^*(2) \\ g_1^* h_1 w_1^*(2) + g_2^* h_2^* w_2(1) \end{bmatrix} + \begin{bmatrix} v(1) \\ v^*(2) \end{bmatrix} \quad (28).$$

Then, the SNR in the sufficient statistics can be shown to be

⁹ While QPSK is assumed here for simplicity, the final conclusion at the end of this example is independent of this assumption.

$$SNR_{2,match-1}(\mathbf{h}; \mathbf{g}) = \frac{G\gamma^2\eta^2 \left(\sum_{i=1}^2 |g_i|^2 |h_i|^4 \right) \sigma_s^2}{G\gamma^2\sigma_w^2 \sum_{i=1}^2 |g_i|^2 |h_i|^2 + \sigma_v^2} \quad (29).$$

If we assume that $G\gamma^2\sigma_w^2 \sum_{i=1}^2 |g_i|^2 |h_i|^2 \gg \sigma_v^2$ (which can be made valid when the relays are close to the destination), then it follows from (29) that

$$SNR_{2,match-1,destination}(\mathbf{h}; \mathbf{g}) = SNR^* \frac{\sum_{i=1}^2 |g_i|^2 |h_i|^4}{\sum_{i=1}^2 |g_i|^2 |h_i|^2} \geq SNR^* \frac{\sum_{i=1}^2 |g_i|^2 |h_i|^2}{\sum_{i=1}^2 |g_i|^2} = SNR_{2,destination}(\mathbf{h}; \mathbf{g}) \quad (30)$$

where we applied the inequality:

$$\sum_{i=1}^n a_i b_i = \sum_{i=1}^n a_i^{1/2} (a_i^{1/2} b_i) \leq \sqrt{\left(\sum_{i=1}^n a_i \right) \left(\sum_{i=1}^n a_i b_i^2 \right)}$$

with positive a_i and b_i . Therefore, the fading coefficient matching method can improve the SNR at the destination.

The coefficient matching method can be further generalized by using a large exponent of the fading factor. But the exponent may increase or decrease the actual power gain at an array, which depends on whether the fast fading magnitude is larger or smaller than one. (Note that when the fast fading factor has a unit magnitude, the SNR of the signal received by the relay equals its mean value, i.e., SNR^* .) One choice is that the exponent is chosen to be a large value when the fading magnitude is larger than one, or otherwise chosen to be one. In this case, (27) is generalized into

$$\begin{bmatrix} x_1(1) & x_2(1) \\ x_1(2) & x_2(2) \end{bmatrix} \triangleq \sqrt{G\eta} \begin{bmatrix} |h_1|^{(k-1)} h_1^* s_1(1) & -|h_2|^{(k-1)} h_2 s_2^*(2) \\ |h_1|^{(k-1)} h_1^* s_1(2) & |h_2|^{(k-1)} h_2 s_2^*(1) \end{bmatrix} \quad (31)$$

where

$$|h_i|^{(k-1)} = \begin{cases} |h_i|^{k-1} & |h_i| > 1 \\ 1 & |h_i| \leq 1 \end{cases}. \quad (31a)$$

If the relays are at the half way between the source and the destination, and the conditions used for (19) also hold, then an variation of (29) is

$$SNR_{2,match-k,mid}(\mathbf{h}; \mathbf{g}) = SNR^* \cdot 2^n \cdot \frac{\frac{1}{2} \sum_{i=1}^2 |g_i|^2 |h_i|^4 |h_i|^{(2k-2)}}{\frac{1}{2} \sum_{i=1}^2 |g_i|^2 |h_i|^2 |h_i|^{(2k-2)} + 1} \quad (31b)$$

In theory, when at least one of the fading magnitudes is larger than one and k is very large, the above scheme automatically picks up the best fading factor that governs the final SNR at the destination. But this SNR gain quickly saturates (in the form of $\frac{x}{x+1} \rightarrow 1$). Furthermore, in practice, the power amplifier in each relay has a limited dynamic range.

3.1.6 Useful cautions

As discussed in section 2.4, the 2×2 orthogonal complex code is not unique. A natural question is whether any of these orthogonal matrices can be used in the way we construct (7) from \mathbf{C}^T for wireless relays. The answer is that it is not always true. For example, if we simply copy Alamouti's code \mathbf{C} for the wireless relays (as in [ANG]), then the signals received at the destination do not have the desired orthogonality property as shown in (9). To prove this fact, let

$$\begin{bmatrix} x_1(1) & x_2(1) \\ x_1(2) & x_2(2) \end{bmatrix} \triangleq \begin{bmatrix} s_1(1) & s_1(2) \\ -s_2^*(2) & s_2^*(1) \end{bmatrix}.$$

Then, the signals received at the destination are:

$$\begin{bmatrix} y(1) \\ y(2) \end{bmatrix} = \begin{bmatrix} s_1(1) & s_1(2) \\ -s_2^*(2) & s_2^*(1) \end{bmatrix} \cdot \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} + \begin{bmatrix} v(1) \\ v(2) \end{bmatrix}.$$

Using (6) in the above equation yields

$$\begin{bmatrix} y(1) \\ y^*(2) \end{bmatrix} = \begin{bmatrix} h_1 g_1 & h_1 g_2 \\ h_2 g_2^* & -h_2 g_1^* \end{bmatrix} \cdot \begin{bmatrix} s(1) \\ s(2) \end{bmatrix} + \begin{bmatrix} w_1(1)g_1 + w_1(2)g_2 \\ -w_2(2)g_1^* + w_2(1)g_2^* \end{bmatrix} + \begin{bmatrix} v(1) \\ v^*(2) \end{bmatrix}$$

where the noise term is white, but the coefficient matrix of the original symbols is not complex orthogonal in general, unless h_1 and h_2 are real valued. Therefore, the space-time coding for the wireless relays is not quite the same as for multiple transmitters.

3.2 Using Multiple Wireless Antennas

We have shown that with two wireless relays, the quality of data transmission can be improved using 2×2 orthogonal space-time code. To extend this result to more than two relays, we need higher dimensional orthogonal space-time codes. Unfortunately, full rate orthogonal complex space-time block codes only exist for the 2×2 case. With more than two relays, an alternative is to use full rate orthogonal real space-time block code or half rate orthogonal complex space-time block code, both of which are effectively half rate. The loss of the coding rate at the relays can be compensated by a larger constellation per symbol throughout the system. Since the relays stabilize the effective SNR (shown later), large constellations no longer be a major difficulty. The orthogonal codes for the case of multiple transmitters were shown earlier. We now apply these codes to multiple wireless relays.

3.2.1 Relaying real symbols

Let \mathbf{s} be a string of L real symbols transmitted from the source. Then, a corresponding string of real symbols received by the i th relay can be represented by

$$\mathbf{s}_i = \eta h_i \mathbf{s} + \mathbf{w}_i \quad (32)$$

where η is the large scale fading factor between the source and the group of relays, h_i is the (real) small scale fading factor between the source and the i th relay. As discussed before, it is reasonable to assume that h_i is constant for a short packet of, say, 500 μ s or even more. But h_i may vary rapidly between packets and over the index i . We will assume that h_i is real and has zero mean and unit variance. The corresponding string of the symbols transmitted by the i th relay is defined to be

$$\mathbf{x}_i = \sqrt{G} \mathbf{A}_i \mathbf{s}_i \quad (33)$$

where G is the power gain at the relays, and \mathbf{A}_i is a member of the Hurwitz-Radon matrices. These matrices can be stored or regenerated in each relay. Each relay however must know its assigned number in order to use a correct \mathbf{A}_i (i.e., distinct from those used by other relays). The number R of relays should be not greater than $p_{\max}(L)$. Special circuits may be designed to implement (33) efficiently. The string of the corresponding symbols received by the destination is then

$$\mathbf{y} = \gamma \sum_{i=1}^R g_i \mathbf{x}_i + \mathbf{v} \quad (34)$$

where γ is the large scale fading factor between the group of relays and the destination, g_i is the small fading factor between the i th relay and the destination, which is constant during each packet of data but may vary rapidly between packets and over the index i . We will assume that g_i is real and has zero mean and unit variance. Then, it follows from (32)-(34) that

$$\mathbf{y} = \sqrt{G\gamma\eta} \left(\sum_{i=1}^R g_i h_i \mathbf{A}_i \right) \mathbf{s} + \sqrt{G\gamma} \sum_{i=1}^R g_i \mathbf{A}_i \mathbf{w}_i + \mathbf{v} \quad (35)$$

or equivalently,

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad (35a)$$

where the definitions of \mathbf{H} and \mathbf{n} follow obviously from (35). It is easy to verify that

$$\mathbf{H}^T \mathbf{H} = G\gamma^2 \eta^2 \left(\sum_{i=1}^R |g_i|^2 |h_i|^2 \right) \mathbf{I}_L \quad (36)$$

$$E(\mathbf{n}\mathbf{n}^T) = \left(G\gamma^2 \left(\sum_{i=1}^R |g_i|^2 \right) \sigma_w^2 + \sigma_v^2 \right) \mathbf{I}_L \quad (37)$$

where we see that \mathbf{H} is orthogonal, and the noise vector \mathbf{n} is white. The sufficient statistics $\mathbf{r} \triangleq \mathbf{H}^T \mathbf{y}$ has the exactly same property as we saw in the two relays case. In fact,

$$\mathbf{r} \triangleq \mathbf{H}^T \mathbf{y} = \alpha \mathbf{s} + \beta \mathbf{e} \quad (38)$$

where

$$\alpha = RG\gamma^2 \eta^2 \left(\frac{1}{R} \sum_{i=1}^R |g_i|^2 |h_i|^2 \right) \quad (39)$$

$$\beta = \sqrt{\left(RG\gamma^2 \left(\frac{1}{R} \sum_{i=1}^R |g_i|^2 \right) \sigma_w^2 + \sigma_v^2 \right) \left(RG\gamma^2 \eta^2 \left(\frac{1}{R} \sum_{i=1}^R |g_i|^2 |h_i|^2 \right) \right)} \quad (40)$$

$$E(\mathbf{e}\mathbf{e}^T) = \mathbf{I}_L. \quad (41)$$

It is important to note that as R becomes large, both α and β become deterministic (i.e., independent of small scale fading), and hence the effective channel between the original symbol sequence \mathbf{s} and the sufficient statistics \mathbf{r} is a virtual additive white Gaussian noise (AWGN) channel. The SNR of the signal received at the destination remains virtually constant and can be made known to the transmitter at the source. For such a channel, a larger data rate can be achieved by using a larger constellation size [FOR].

We now examine the SNR of (38). It is easy to verify that the SNR is given by

$$SNR_{R,real}(\mathbf{h}; \mathbf{g}) = \sigma_s^2 \frac{\alpha^2}{\beta^2} = \sigma_s^2 \frac{RG\gamma^2 \eta^2 \left(\frac{1}{R} \sum_{i=1}^R |g_i|^2 |h_i|^2 \right)}{RG\gamma^2 \left(\frac{1}{R} \sum_{i=1}^R |g_i|^2 \right) \sigma_w^2 + \sigma_v^2}. \quad (41)$$

With the same assumptions made about (19), we can show that

$$SNR_{R,real,mid}(\mathbf{h}; \mathbf{g}) = SNR^* \cdot 2^n \cdot \frac{\left(\frac{1}{R} \sum_{i=1}^R |g_i|^2 |h_i|^2 \right)}{\frac{1}{R} \sum_{i=1}^R |g_i|^2 + 1} \quad (42)$$

Note that this SNR is achieved with the relays at half way between the source and the destination, and the total power consumption by the relays is the same as that by the source. The power consumed by each relay becomes very small when the number of relays becomes large. That is, each assistant only pays a small price. Furthermore, (42) differs from (19) in that g_i and h_i are *real valued* here instead of being complex. Since each small scale fading factor has two independent components, real part and imaginary part, about half of the channel diversity is hence lost in the real data domain. Our simulations based on (42) produced the following table. The diversity for the real symbol case is about $R/3$.

R	2	3	4	5	6	7	8
d	0.7	0.99	1.30	1.59	1.86	2.11	2.51

Table 2: Relaying real symbols with relays at half way: Diversity d as function of the number R of relays

3.2.2 Relaying complex symbols

If the source transmits a sequence of complex symbols, the relays have two choices. The first choice is to convert the corresponding (estimated) complex sequence into two real sequences, and then relay each real sequence as described previously. The second choice is to use a half-rate orthogonal complex space-time block code. For both choices, the source must stop data transmission to the destination after each packet for at least as long as the length of the packet.

To show the details of the second choice, we now rewrite (32) as follows:

$$\mathbf{s}_i = \eta h_i \mathbf{s} + \mathbf{w}_i \quad (44)$$

where all notations are the same as for (32) except that \mathbf{s}_i , \mathbf{s} , \mathbf{w}_i and h_i are now complex. The space-time code for the i th relay is now defined as

$$\mathbf{x}_i = \sqrt{G} \begin{bmatrix} \mathbf{A}_i h_i^* \mathbf{s}_i \\ \mathbf{A}_i h_i \mathbf{s}_i^* \end{bmatrix} \quad (45)$$

where \mathbf{x}_i is a $2L \times 1$ vector. As shown later, this weighting h_i ensures the orthogonality of the coefficient matrix of the original symbol vector in the packet received at the destination. Alternatively, we can replace $h_i = |h_i| \exp(j\theta_i)$ in (45) by $\exp(j\theta_i)$. It is the phase conjugation that matters the most here. The complex version of (34) is now

$$\mathbf{y} = \gamma \sum_{i=1}^R g_i \mathbf{x}_i + \mathbf{v} \triangleq \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}. \quad (46)$$

Using (45) in (46) yields

$$\mathbf{y}_c \triangleq \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2^* \end{bmatrix} = \sqrt{G} \gamma \eta \begin{pmatrix} \sum_{i=1}^R g_i |h_i|^2 \mathbf{A}_i \\ \sum_{i=1}^R g_i^* |h_i|^2 \mathbf{A}_i \end{pmatrix} \mathbf{s} + \sqrt{G} \gamma \begin{pmatrix} \sum_{i=1}^R g_i h_i^* \mathbf{A}_i \mathbf{w}_i \\ \sum_{i=1}^R g_i^* h_i \mathbf{A}_i \mathbf{w}_i \end{pmatrix} + \mathbf{v}_c \quad (47)$$

or equivalently, we write

$$\mathbf{y}_c = \mathbf{H}_c \mathbf{s} + \mathbf{n}_c. \quad (47a)$$

It is easy to verify that

$$\mathbf{H}_c^H \mathbf{H}_c = 2G\gamma^2 \eta^2 \sum_{i=1}^R |g_i|^2 |h_i|^4 \mathbf{I}_L \quad (48)$$

$$\mathbf{R}_n \triangleq E(\mathbf{n}_c \mathbf{n}_c^H) = G\gamma^2 \sigma_w^2 \begin{bmatrix} \sum_{i=1}^R |g_i|^2 |h_i|^2 \mathbf{I}_L & \sum_{i=1}^R g_i^2 |h_i|^2 \mathbf{I}_L \\ \sum_{i=1}^R g_i^{*2} |h_i|^2 \mathbf{I}_L & \sum_{i=1}^R |g_i|^2 |h_i|^2 \mathbf{I}_L \end{bmatrix} + \sigma_v^2 \mathbf{I}_{2L}. \quad (49)$$

The noise vector in (47a) is not white although the coefficient matrix in (47a) is complex orthogonal. This is a bit unfortunate because the sufficient statistics of \mathbf{s} is no longer given by the simple expression:

$$\mathbf{r}_c \triangleq \mathbf{H}_c^H \mathbf{y}_c \quad (50)$$

which is then a sub-sufficient statistics. With the above definition, we can show with a tedious process that

$$\mathbf{r}_c \triangleq \mathbf{H}_c^H \mathbf{y}_c = \alpha_c \mathbf{s} + \beta_c \mathbf{e}_c \quad (50a)$$

where

$$\begin{aligned} E(\mathbf{e}\mathbf{e}^H) &= \mathbf{I}_L \\ \alpha_c &= 2RG\gamma^2 \eta^2 \lambda_2 \\ \beta_c &= \sqrt{R^2 G^2 \gamma^4 \eta^2 (2\lambda_1 \lambda_2 + 2 \operatorname{Re}(\tau_1 \tau_2)) \sigma_w^2 + 2RG\gamma^2 \eta^2 \lambda_2 \sigma_v^2} \\ \lambda_1 &= \frac{1}{R} \sum_{i=1}^R |g_i|^2 |h_i|^2 \\ \lambda_2 &= \frac{1}{R} \sum_{i=1}^R |g_i|^2 |h_i|^4 \\ \tau_1 &= \frac{1}{R} \sum_{i=1}^R g_i^2 |h_i|^2 \\ \tau_2 &= \frac{1}{R} \sum_{i=1}^R g_i^{*2} |h_i|^4 \end{aligned}$$

Then, we can show that the SNR of \mathbf{r}_c is

$$SNR_{R, \text{complex}, \text{match-1}}(\mathbf{h}; \mathbf{g}) \geq \frac{\sigma_s^2 RG\gamma^2 \eta^2 \lambda_2}{RG\gamma^2 \lambda_1 \sigma_w^2 + \frac{\sigma_v^2}{2}} \quad (53)$$

With the same assumptions made about (19), we have

$$SNR_{R, \text{complex}, \text{match-1}, \text{mid}}(\mathbf{h}; \mathbf{g}) \geq SNR^* \cdot 2^n \cdot \frac{\lambda_2}{\lambda_1 + \frac{1}{2}}. \quad (54)$$

Alternatively, we can consider the sufficient (optimal) statistics

$$\mathbf{r}_{c, \text{opt}} \triangleq \mathbf{H}_c^H \mathbf{R}_n^{-1} \mathbf{y}_c.$$

The structure shown in (49) can be used to provide a more explicit expression of this statistics. But we will not pursue it in this paper.

Our simulations based on the lower bound of (54) produced the following table. The diversity is about $R/2$.

R	2	3	4	5	6	7	8
d	0.93	1.34	1.93	2.35	2.64	3.00	3.59

Table 3: Relaying complex symbols with relays at half way: diversity d as function of the number R of relays

Comparing Table 3 with Table 2 suggests that relaying complex symbols has a higher diversity gain.

3.2.3 Comparison of relaying complex symbols and relaying real symbols

For a fair comparison of relaying two real sequences and relaying one complex sequence, we need to remove the matching amplitude used in (45) as fading magnitude matching at the relays always improves SNR as discussed previously. If we replace $h_i = |h_i| \exp(j\theta_i)$ in (45) by $\exp(j\theta_i)$, then (54) becomes

$$SNR_{R,complex,mid}(\mathbf{h}; \mathbf{g}) \geq SNR^* \cdot 2^n \cdot \frac{\left(\frac{1}{R} \sum_{i=1}^R |g_i|^2 |h_i|^2 \right)}{\frac{1}{R} \sum_{i=1}^R |g_i|^2 + \frac{1}{2}} \quad (55)$$

Let us now compare (42) and (55). First, we note that for the same SNR^* , the source transmission power¹⁰ used for (42) is half the source transmission power used for (55). In (42), SNR^* is the signal power divided by the noise variance in one dimension (i.e., the real part of noise¹¹). But in (55), SNR^* is the signal power divided by the noise variance in two dimensions. However, for an identical fading environment, the small scale fading factors in (42) are simply the real parts of the small scale fading factors in (55). Therefore, $E(|g_i|^2)$ and $E(|h_i|^2)$ used in (55) are each twice the values of $E(|g_i|^2)$ and $E(|h_i|^2)$ used in (42). Furthermore, each of $E(|g_i|^2)$ and $E(|h_i|^2)$ used in (55) is degree-2 chi-squares but each of $E(|g_i|^2)$ and $E(|h_i|^2)$ used in (42) is degree-1 chi-squares. Therefore, relaying complex symbols provides a better diversity than relaying real symbols. One advantage of using the real symbols over using the complex symbols is that the former has a delay of L symbols instead of $2L$ symbols.

Our simulations show that we have more than 10 dB SNR gain from 2 relays to 8 relays under a constant total power consumption. With the optimal statistics $\mathbf{r}_{c,opt} \hat{=} \mathbf{H}_c^H \mathbf{R}_n^{-1} \mathbf{y}_c$, an even higher SNR gain can be achieved.

3.2.5 Diversity Gain

We have showed that using multiple wireless relays yield a high diversity. The tables 2-3 provide the diversity as function of the number of relays for real and complex cases where the relays are at half way between the source and the destination. When the relays are near the source or the destination, the diversity gain is actually even higher. The next table shows the diversity for the complex case where the relays are near the destination and no fading coefficient matching is used at the relays. The diversity is near $R/2$ for large R .

R	1	2	3	4	5	6	7	8
d	0.91	1.54	2.14	2.61	3.00	3.30	3.47	3.86

¹⁰ The source transmission power is half the total transmission of this system, the latter of which includes the relay transmission power.

¹¹ It is assumed that the real part of noise has the same variance as the imaginary part of the noise.

Table 4¹²: Relaying complex symbols with relays near destination: diversity d as function of the number R of relays

The table 5 shows the diversity for the complex case where the relays are near the destination and the fading coefficient matching is used at the relays. The diversity in this case is slightly larger than $R/2$ for large R .

R	1	2	3	4	5	6	7	8
d	0.91	1.72	2.27	2.77	3.27	3.89	4.57	4.90

Table 5: Relaying complex symbols with relays near destination and fading coefficient matching: diversity d as function of the number R of relays

When the relays are near the source and no fading coefficient matching is applied, we have the following table:

R	1	2	3	4	5	6	7	8
d	0.68	1.23	1.71	2.16	2.53	2.88	3.09	3.46

Table 6: Relaying complex symbols with relays near the source and no fading coefficient matching: : diversity d as function of the number R of relays

Comparing Table 6 with Table 4, we see that the diversity is slightly higher when the relays are near the destination than when the relays are near the source.

The above tables clearly suggest that a large diversity gain is possible with a large number of relays. Depending on the location of the relays, the diversity varies around $R/2$ for large R .

To provide further illustrations, we now define the following where the common SNR gain from large scale fading is removed:

$$SNR_{1,mid} = SNR^* \cdot \frac{|g|^2 |h|^2}{|g|^2 + 1}$$

$$SNR_{2,mid} = SNR^* \cdot \frac{\left(\frac{1}{2} \sum_{i=1}^2 |g_i|^2 |h_i|^2 \right)}{\frac{1}{2} \sum_{i=1}^2 |g_i|^2 + 1}$$

$$SNR_{R,mid,bound} = SNR^* \cdot \frac{\lambda_2}{\lambda_1 + \frac{1}{2}}$$

with

$$\lambda_1 = \frac{1}{R} \sum_{i=1}^R |g_i|^2 |h_i|^2$$

$$\lambda_2 = \frac{1}{R} \sum_{i=1}^R |g_i|^2 |h_i|^4$$

¹² The theoretical diversity for $R=1$ in Table 4 and Table 5 is known to be 1. Clearly, our estimates are conservative. This is partly because the reading was done with the mean SNR between 10 dB and 20 dB. The curves generally continue to bend downwards slightly beyond this interval.

where g , h , g_i and h_i are *complex* Gaussian of zero mean and unit variance (i.e., the variance of real or imaginary part is 0.5). The figure 4 shows the pdf of $SNR_{1,mid}$, $SNR_{2,mid}$ and $SNR_{R,mid,bound}$ for $R=2,3,4,5,6,7,8$. The figure 5 shows $E(Q(\sqrt{SNR_{1,mid}}))$, $E(Q(\sqrt{SNR_{2,mid}}))$ and $E(Q(\sqrt{SNR_{R,mid,bound}}))$ for $R=2,3,4,5,6,7,8$. A simple calculation can show that more than 10 dB SNR gain from the baseline of single (regenerative) relay is achieved with eight relays at an average BER of 10^{-3} or lower.

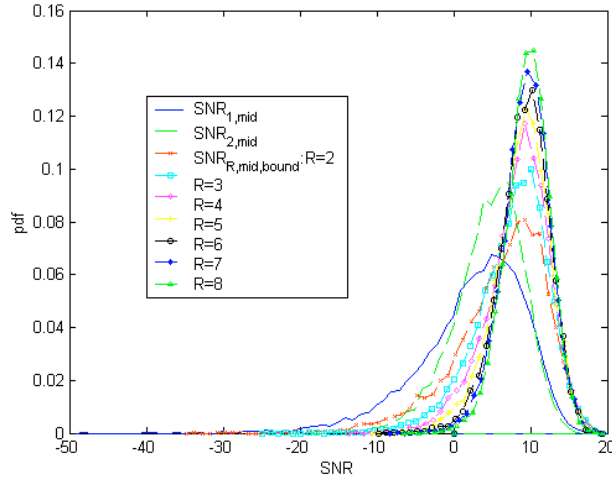


Figure 4: pdf of $SNR_{1,mid}$, $SNR_{2,mid}$ and $SNR_{R,mid,bound}$ (all in dB) for $R=2,3,4,5,6,7,8$ with $SNR^* = 10dB$.

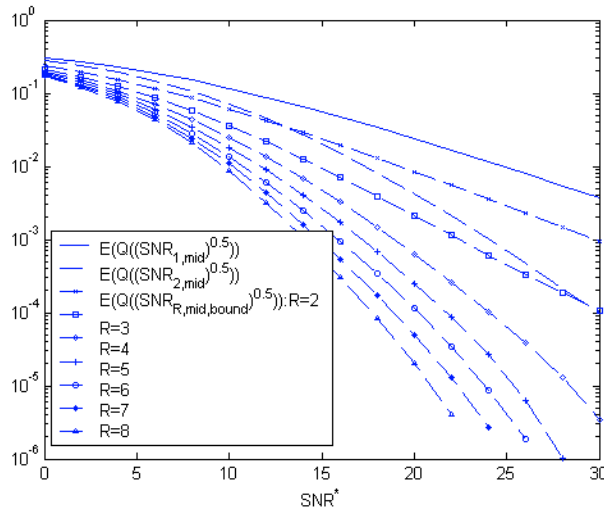


Figure 5: Average BER, i.e., $E(Q(\sqrt{SNR_{1,mid}}))$, $E(Q(\sqrt{SNR_{2,mid}}))$ and $E(Q(\sqrt{SNR_{R,mid,bound}}))$ for $R=2,3,4,5,6,7,8$ versus SNR^* in dB.

3.3 Other Important Issues

3.3.1 Channel Estimation

Channel estimation is an essential part in any situation of wireless communications. With wireless antennas, channel estimation must take place at both the relays and the destination. At the relays, channel estimation can follow the conventional approach. Typically, the source sends out one or more pilot symbols within each packet of data. The receiver (i.e., each relay in our case) then uses the pilot symbols and the received symbols to estimate the channel fading factor (which combines both the large scale fading and the small scale fading). Therefore, the fading factors $\sqrt{G\gamma}h_i$ can be relatively easily obtained by the i th relay. The large scale fading factor η can be obtained by averaging consecutive estimates of $\sqrt{G\gamma}h_i$. The source needs not to know h_i (which may be too fast for the source to keep track of, anyway). But to some degree, $\sqrt{G\gamma}\eta$ should and can be made known to the source.

For wireless relays, h_i and η need not be further passed down to the destination. With pilot symbols embedded in \mathbf{s} , the destination should be able to estimate the final effective channel matrix \mathbf{H} such as in [35a]. For the two relays case, we only need a minimum of two pilot symbols in \mathbf{s} . Assume that the two pilot symbols are adjacent to each other and belong to one block of 2×2 space-time code, then we have from (9) that

$$\begin{bmatrix} y(1) \\ y^*(2) \end{bmatrix} = \begin{bmatrix} a & -b^* \\ b & a^* \end{bmatrix} \cdot \begin{bmatrix} s(1) \\ s^*(2) \end{bmatrix} + \begin{bmatrix} n(1) \\ n^*(2) \end{bmatrix} \quad (56)$$

where $a = \sqrt{G\gamma}\eta g_1 h_1$, $b = \sqrt{G\gamma}\eta g_2^* h_2$, and the noise term can be found from (9). In order to obtain the sufficient statistics (10), we only need to know a and b (not individual factors of these two numbers). With the knowledge of a and b over several packets, the large scale fading factor $\sqrt{G\gamma}\eta$ can be easily found and made known to the source for the power control purpose. If the two symbols in (56) are pilots, then the destination should consider the dual form of (56):

$$\begin{bmatrix} y(1) \\ y(2) \end{bmatrix} = \begin{bmatrix} s(1) & -s(2)^* \\ s(2) & s(1)^* \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} n(1) \\ n(2) \end{bmatrix} \quad (56a)$$

where the noise vector is white and Gaussian, and the coefficient matrix is orthogonal. Then, the maximum likelihood estimates of a and b are given by

$$\begin{bmatrix} a \\ b \end{bmatrix}_{MLE} = \begin{bmatrix} s(1) & -s(2)^* \\ s(2) & s(1)^* \end{bmatrix}^H \begin{bmatrix} y(1) \\ y(2) \end{bmatrix} \quad (57)$$

which is a very simple task to perform. The estimates of a and b can be used for symbol detection for other 2×1 blocks within the same packet.

For more than two relays, the story is a bit more complicated. Let us now consider an equivalent form of (47):

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2^* \end{bmatrix} = \begin{pmatrix} \sum_{i=1}^R a_i \mathbf{A}_i \\ \sum_{i=1}^R a_i^* \mathbf{A}_i \end{pmatrix} \mathbf{s} + \begin{pmatrix} \mathbf{n}_1 \\ \mathbf{n}_2^* \end{pmatrix} \quad (58)$$

where $a_i = \sqrt{G\gamma}\eta g_i |h_i|^2$ for $i = 1, 2, \dots, R$, and the noise vector is white and (circular) Gaussian. We only need to know a_i , $i = 1, 2, \dots, R$, (not their individual factors). With the knowledge of a_i , $i = 1, 2, \dots, R$, over a single or multiple packets, we can find the large scale fading factor $\sqrt{G\gamma}\eta$

relatively easily. Once again, it is only the large scale fading factor that the source needs to know. The dual form of (58) is

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{pmatrix} \mathbf{A}_1 \mathbf{s} & \mathbf{A}_2 \mathbf{s} & \cdots & \mathbf{A}_R \mathbf{s} \\ \mathbf{A}_1 \mathbf{s}^* & \mathbf{A}_2 \mathbf{s}^* & \cdots & \mathbf{A}_R \mathbf{s}^* \end{pmatrix} \mathbf{a} + \begin{pmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \end{pmatrix} \quad (58a)$$

or simply written as

$$\mathbf{y} = \mathbf{S} \mathbf{a} + \mathbf{n} \quad (58b)$$

where $\mathbf{a} = [a_1 \ a_2 \ \cdots \ a_R]^T$, the coefficient matrix \mathbf{S} is complex orthogonal and the noise vector \mathbf{n} is white and (circular) Gaussian. Note that the covariance matrix of the noise vector in (58a) is diagonal, but the covariance matrix of the noise vector in (58) is not. This seemingly surprising result is simply because \mathbf{w}_i is white complex circular Gaussian¹³, i.e., $E(\mathbf{w}_i \mathbf{w}_j^H) = \delta_{ij} \sigma_w^2 \mathbf{I}_L$ and $E(\mathbf{w}_i \mathbf{w}_j^T) = 0$. Therefore, if the coefficient matrix \mathbf{S} of (58a) is given, the maximum likelihood estimate of \mathbf{a} is:

$$\mathbf{a}_{MLE} = \frac{1}{2\|\mathbf{s}\|^2} \mathbf{S}^H \mathbf{y}. \quad (59)$$

Based on the structure of (58), the number of pilot symbols required is generally equal to the number R of relays, and hence the rest of the L symbols in the packet can be used for data. If R is 2, 4 or 8, then it is possible to rearrange (59) such that all data symbols are separated from the computation for estimation of \mathbf{a} . Otherwise, the data symbols can not be completely separated from the estimation of \mathbf{a} unless all L symbols are used as pilot symbols. When R is larger than 8, L increases much faster than R , i.e., for very large R , $R \approx 2 \log_2 L$.

When R is not 2, 4 or 8, an efficient method must pack both pilot symbols and data symbols (or part of the data symbols) into the vector \mathbf{s} in (59). In this case, joint estimation of channel parameters and data symbols should be carried out based on (58). A simple iterative method is to estimate channel parameters and data symbols alternately until convergence. This is an example of blind estimation methods, e.g., see [HUA(e)].

3.3.2 Symbol Synchronization

Symbol synchronization is critical for the wireless relays to achieve the desired results. A simple method of symbol synchronization can be based on the symbol sequence transmitted from the source. As long as the circuits in each relay can keep track of the timing of its received symbols, the relays are in general well synchronized. Note that if the relays are at the half way between the

source and the destination, they are within l m diameter of each other, and $\left(\frac{l}{D/2}\right)^2 \ll 1$ where D

is the distance between the source and the destination, then the maximum time difference of the signals arriving at the destination can be shown to be $\Delta T_{\max} = \frac{2l^2}{Dc}$ where c is the velocity of light. For example, if $D = 2000$ m and $l = 100$ m , then $\Delta T_{\max} = 33ns$ which is much smaller than the symbol duration of any existing (or possible) wireless standards.

If the relays are near the source or the destination, then the corresponding distance between the relays should be decreased to maintain symbol synchronization. If the maximum difference of the path distances¹⁴ is less than 20 m , then the corresponding time difference is 0.067 μs , which is still much smaller than the typical symbol duration (around 1 μs) used in the wireless standards.

¹³ The real part and imaginary part of a complex circular Gaussian random variable are independent and each of the same variance.

¹⁴ Each path distance is the distance between the source and a relay plus the distance between the relay and the destination.

It seems that only in some extreme and unpredictable environment, special circuits may be needed for symbol synchronization.

3.3.3 Coordination of relays

The wireless relays must be coordinated for many reasons. The coordination of the relays should be handled by a processor at the destination. Each relay must know its identification from the destination so that the correct code is used at the relay. Each relay should also inform the destination of its average signal strength so that the destination can decide whether this relay should be activated or de-activated. The relays should also inform the source of their average signal strength so that the source can adjust its transmission power and choose a proper coding scheme.

The roles of the source and the destination can be reversed with respect to the same relays (especially when the relays are at the half way between the source and the destination). But this does not cause any additional fundamental problem although separate channels (in time or frequency) should be allocated.

The communication for relay coordination only requires a small amount of information (or low bit rate). This information can be embedded in each packet. Special medium access control (MAC) protocols may be designed for the wireless relays to maximize the overall efficiency.

3.3.4 Signal processing hardware

The key operation required for the wireless antennas is the space-time encoding at the relays and the space-time decoding at the destination. Such an operation must be implemented on DSP chips. There are a few space-time codes that can be applied to the wireless antennas. But the Hurwitz-Radon space-time code introduced in this paper has the simplicity that no other code can compete with. We also believe that it is not a very difficult task to have Hurwitz-Radon code implemented on off-the-shelves DSP devices.

4. Final Remarks

This paper has presented a new framework of wireless communications with emphasis on signal processing concepts. These concepts were derived from the basics of radio physics, communication theory, and the needs of a fully mobile wireless network. We have established a good reason to believe that wireless antennas or wireless relays will significantly enhance the capacity of mobile wireless networks.

It took nearly 50 years for researchers to develop and implement efficient coding techniques to achieve less than 9 dB power saving (from the baseline of uncoded symbols) for spectrum limited wireline communications [FOR]. Our study has shown that more than 10 dB power saving, from the baseline of a single (regenerative) relay system, can be achieved with eight parallel wireless mobile (non-regenerative) relays. Further research investment into this area is surely justified. We believe that this is only a beginning of an exciting field for research and development.

Acknowledgement

The authors thank Brian Sadler and Ananthram Swami of ARL for their input, encouragement and support for this work.

References:

- [ALA] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas Commun.*, Vol. 16, pp. 1451-1458, Oct 1998.
- [ANG] P. A. Anghel, G. Leus, and M. Kaveh, "Multi-user space-time coding in a cooperative networks", *Proc of IEEE ICASSP'2003*, Hong Kong, May 2003.
- [DHA] N. Al-Dhahir, G. B. Giannakis, B. Hochwald, B. L. Hughes, and T. L. Marzetta, editors, *Special Issue on Signal Processing for Communications, IEEE Transactions on Signal Processing*, Vol. 50, No. 10, Oct 2002.
- [DOH] M. Dohler, and H. Aghvami, "A step towards MIMO: virtual antenna arrays," *Center for Telecommunication Research, King's College*, Longdon, UK, Jan 2003.
- [FOR] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *Information Theory – 50 Years of Discovery*, edited by S. Verdu, IEEE Press, 2000.
- [GAN] G. Ganesan, and P. Stoica, "Space-time block codes: a maximum SNR approach," *IEEE Trans. Information Theory*, Vol. 47, No. 4, pp. 1650-1656, May 2001.
- [GER] A. V. Geramita and J. Seberry, *Orthogonal Designs, Quadratic Forms and Hadamard Matrices*, Lecture Notes in Pure and Applied Mathematics, Vol. 43, New York and Basel: Marcel Dekker, 1979.
- [GIA] G. Giannakis, Y. Hua, P. Stoica and L. Tong, editors, *Signal Processing Advances for Wireless and Mobile Communications*, Volumes I-II, Prentice-Hall, 2001.
- [HUA(a)] Y. Hua, A. Gershman, and Q. Cheng, editors, *High-Resolution and Robust Signal Processing – Recent Advances in Sensor Array Processing*, Marcel Dekker, 2003.
- [HUA(b)] Y. Hua, "Ad hoc arrays", First Report, *ARL CTA Review Meeting*, John Hopkins Univ., Baltimore, MD, March 10-11, 2003.
- [HUA(c)] Y. Hua, Y. Mei, Y. Chang, "Parallel wireless mobile relays with space-time modulation," *IEEE Workshop on Statistical Signal Processing*, St. Louis, Missouri, USA, Sept. 28 - Oct 1, 2003.
- [HUA(d)] Y. Hua, Y. Mei, Y. Chang, "Smart wireless antenna arrays," *IEEE AP-S Topical Conference on Wireless Communications Technology*, Honolulu, Hawaii, USA, Oct 15-16, 2003.
- [HUA(e)] Y. Hua, S. An, Y. Xiang, "Blind identification and equalization of FIR MIMO channels driven by colored signals," *IEEE Transactions on Signal Processing*, pp. 1143-1155, No. 5, Vol. 51, May 2003.
- [JAF] H. Jafarkhani, and N. Seshadri, "Super-orthogonal space-time Trellis codes," *IEEE Transactions on Information Theory*, to appear.
- [LAN] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, submitted in Jan 2002.
- [PRO] J. G. Proakis, *Digital Communications*, 4th edition, McGraw-Hill, 2001.
- [RAP] T. S. Rappaport, *Wireless Communications – Principles and Practice*, 2nd edition, Prentice-Hall, 2002.
- [SCA] A. Scaglione, and Y.-W. Hong, "Opportunistic large arrays: cooperative transmission in wireless multi-hop ad hoc networks to reach far distance," *IEEE Trans. Signal Processing*, to appear.
- [TAR] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, Vol. 45, No. 5, July 1999.
- [TEL] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommunications*, Vol. 10, pp. 585-595, Nov. 1999.
- [VAN] H. L. Van Trees, *Optimal Array Processing*, Wiley, 2002.
- [VIS] P. Viswanath, D. Tse, R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, Vol. 48, No. 6, pp. 1277-1294, June, 2002.
- [ZHE] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple antenna channels," *IEEE Trans. Info. Theory*, to appear.