A Power Allocation Strategy for Multi-Antenna Amplify-and-Forward Fading Relay Channels

Nicola Varanese*, Osvaldo Simeone, Yeheskel Bar-Ness
*CWCSPR
New Jersey Institute of Technology,
Email: {nv35, osvaldo.simeone}@njit.edu,
barness@yegal.njit.edu

Umberto Spagnolini
*Dipartimento di Elettronica e Informazione
Politecnico di Milano
Email: spagnoli@elet.polimi.it

***Abstract*— In this paper[1], we consider a fading relay channel operating according to the amplify-and-forward protocol, where each node has full CSI and multiple-antennas are deployed at the relay node. Maximization of the achievable rate with respect to the linear processing at the relay and the power allocation at the source and relay is performed under an instantaneous sum-power constraint. In particular, it is assumed that the total power used by all the active nodes during each time slot is fixed.**

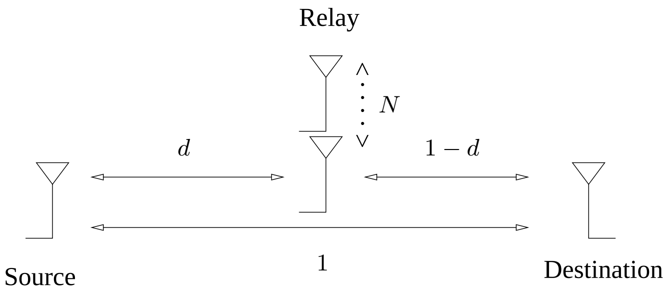

Fig. 1. The Relay channel with multiple-antennas at the relay node.

I. Introduction

Cooperative communications is envisaged to be a key technology for improving the performance of wireless ad-hoc and cellular networks. The simplest cooperative system is the relay channel, introduced by Cover in [1]. The capacity for this channel is still unknown, and the theory seems to suggest optimality of coding schemes that are far from being practical [1]. Simpler relaying schemes, namely Amplify-and-Forward (AF) and Decode-and-Forward (DF), have been proposed and proved to achieve full diversity in [2]. Modified versions of the original protocols, capable to achieve a higher rate, have been introduced in [3]. In all these works, fixed power allocation between source and relay is assumed.

Recently, power allocation strategies have been investigated for the fading relay channel by various authors. In [4] resource allocation strategies minimizing the energy-per-information-bit are presented for different protocols under an instantaneous total power constraint over the two time-slots. A rate-maximizing power allocation algorithm has been developed in [5] for an adaptive version of the DF protocol, with separate average power constraints for the source and the relay node. Coded protocols were also considered in [6], with an average total power constraint over the two time-slots. Finally, in [7] optimal power control for the minimization of the outage probability has been studied for decode- and estimate-and-forward protocols with an average sum-power constraint for the source and the relay. It should be emphasized that in these previous works all the nodes in the system were deployed with only one antenna.

In this paper, we consider the problem of resource allocation for a fading relay channel operating according to the AF protocol, where each node has full CSI and multiple-antennas are deployed at the relay node. The achievable rate is maximized over the linear processing at the relay and the power allocation between the source and the relay. An instantaneous sum-power constraint is imposed, such that in each time-slot the total power used by the active node(s) in the system must be equal. Notice that, in principle, a larger rate could be achieved by allowing a different total power for transmissions during the first and second time-slot [8]. However, this benefit would come at the expenses of an increased transmitted power dynamics and it will not be further investigated in this work.

[1]This work was partially supported by grant from Samsung Advanced Institute of Technology (SAIT), Korea.

II. System model and problem definition

The system under analysis, illustrated in fig. 1, consists of a source node, a destination node, and a multi-antenna relay node deployed with N antennas, which amplifies the received signal and forwards it towards the destination node. In order to comply with the practical half-duplex constraint, the relay transmission occurs according to a time division duplex (TDD) protocol. In the first time slot the source broadcasts the signal x_1 to the relay and the destination. In the second time slot, the relay re-transmits an amplified version of x_1, while the source sends a second signal x_2, chosen independently from x_1. The time durations of the two time-slots are equal. The destination then *jointly* decodes x_1 and x_2, given the symbols received in the two time-slots.

All the fading channels between pair of antennas are assumed to be affected by independent Rayleigh flat fading processes and additive white gaussian noise (AWGN). We also assume that every node in the system knows the state of the channels $\mathcal{H} = \{h_0, \mathbf{h}_1, \mathbf{h}_2\}$, where h_0 is the scalar source-destination channel, $\mathbf{h}_1$ is the $N \times 1$ vector source-relay channel and $\mathbf{h}_2$ is the $N \times 1$ vector relay-destination channel. The state $\mathcal{H}$ is a stationary ergodic process, and it is constant over the two time-slots. The signal received by the destination in the first time-slot can be written as

$$y_1 = h_0 x_1 + n_1, \tag{1}$$

where x_1 is the symbol transmitted by the source in the first time-slot and n_1 is the noise sample at the destination, assumed to have distribution $\mathcal{CN}(0, N_0)$. At the same time, the relay receives a signal

$$\mathbf{y}_R = \mathbf{h}_1 x_1 + \mathbf{n}_R, \tag{2}$$

where $\mathbf{n}_R$ is the noise vector at the relay, with distribution $\mathcal{CN}(\mathbf{0}, N_0 \mathbf{I}_N)$. In the second time-slot, both the source and the relay are active. The relay processes the vector signal previously received by multiplication with a $N \times N$ matrix $\mathbf{G}$. Thus, the symbol transmitted by the relay can be expressed as

$$\mathbf{x}_R = \mathbf{G}\mathbf{y}_R = \mathbf{G}\mathbf{h}_1 x_1 + \mathbf{G}\mathbf{n}_R. \tag{3}$$

The overall signal received by the destination in the second time-slot is

$$y_2 = \mathbf{h}_2^H \mathbf{G}\mathbf{h}_1 x_1 + h_0 x_2 + \mathbf{h}_2^H \mathbf{G}\mathbf{n}_R + n_2, \tag{4}$$

where x_2 is the signal transmitted by the source and n_2 is the noise sample at the destination, assumed to have distribution $\mathcal{CN}(0, N_0)$. Finally, we can more compactly express the vector input/output relation for this channel as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_0 & 0 \\ \mathbf{h}_2^H \mathbf{G}\mathbf{h}_1 & h_0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 & \mathbf{0}^H & 0 \\ 0 & \mathbf{h}_2^H \mathbf{G} & 1 \end{bmatrix} \begin{bmatrix} n_1 \\ \mathbf{n}_R \\ n_2 \end{bmatrix}. \tag{5}$$

Note that, since x_1 and x_2 are assumed to be independent, the model (5) corresponds to a two-user multiple access channel with two antennas at the receiver, and thus successive decoding is a capacity-achieving decoding strategy [9].

Based on the channel state information $\mathcal{H}$, the source node allocates the power $P_{s_1}(\mathcal{H})$ and $P_{s_2}(\mathcal{H})$ for transmission during the first and second time-slot respectively, while the relay node employs the power allocation policy $P_r(\mathcal{H})$. This later assumption affects the design of the matrix $\mathbf{G}(\mathcal{H})$ since the power transmitted by the relay is

$$P_r(\mathcal{H}) = tr(\mathbf{x}_R \mathbf{x}_R^H) = tr(\mathbf{G}(\mathcal{H})\mathbf{h}_1 P_{s_1}(\mathcal{H})\mathbf{h}_1^H \mathbf{G}^H(\mathcal{H}) + \mathbf{G}(\mathcal{H}) N_0 \mathbf{G}^H(\mathcal{H})). \tag{6}$$

Power allocation and relay processing are jointly optimized so as to maximize the *instantaneous* achievable rate of the protocol

$$R = \frac{1}{2} I(x_1, x_2; y_1, y_2), \tag{7}$$

where $I(x_1, x_2; y_1, y_2)$ is the mutual information between the source input (x_1, x_2) and the output at the destination (y_1, y_2), and the factor $1/2$ accounts for the time-division operation. An instantaneous power constraint on each time-slot is enforced so that the problem can be stated as

$$\max_{\mathrm{G}(\mathcal{H}),\theta(\mathcal{H})} R \tag{8}$$
$$\text{s.t.} \begin{cases} P_{s_1}(\mathcal{H}) &= 1 \\ P_{s_2}(\mathcal{H}) &= \theta(\mathcal{H}) \\ P_r(\mathcal{H}) &= 1 - \theta(\mathcal{H}) \end{cases}$$

$\forall \mathcal{H}$, where $\theta(\mathcal{H}) \in [0, 1]$.

The power constraints in (8) allow the transmission scheme to encompass and generalize direct transmission and the AF cooperative protocols in [2] and [3]. In fact, if $\theta(\mathcal{H}) = 1$ the relay is not used at all, neither during the first nor in the second time-slot, and the communication scheme boils down to direct transmission from the source to the relay node; if $\theta(\mathcal{H}) = 0$ only the relay is active during the second time-slot and the communication protocol is the same as in [2]; finally, if $\theta(\mathcal{H}) = 1/2$ the communication protocol employs both the relay and the source in the second time slot, and thus resembles the scheme introduced in [3]. Notice that, under the considered ergodicity assumption, we could have also enforced a long-term power constraint in (8): this modification would complicate the analysis and is not further considered in this work.

In order to tackle the optimization problem (8), is convenient to expand the instantaneous achievable rate (7) by using the chain rule for the mutual information [9]

$$R = \underbrace{\frac{1}{2} I(x_1; y_1)}_{I_1} + \underbrace{\frac{1}{2} I(x_1, x_2; y_2 | y_1)}_{I_2}, \tag{9}$$

where we have used the fact that $I(x_2; y_1 | x_1) = 0$. In (9) I_1 accounts for the source transmission during the first time-slot, while I_2 measures the contribution of the relay and source transmissions during the second time-slot. Given (5) and (9), for any power allocation policies $P_{s_1}(\mathcal{H})$, $P_{s_2}(\mathcal{H})$, $P_r(\mathcal{H})$ and linear processing at the relay $\mathbf{G}(\mathcal{H})$, the achievable rate components read

$$I_1 = \frac{1}{2} C\left(\frac{|h_0|^2 P_{s_1}(\mathcal{H})}{N_0}\right) \tag{10}$$

$$I_2 = \frac{1}{2} C\left(\frac{|h_0|^2 P_{s_2}(\mathcal{H}) + \frac{\mathbf{h}_2^H \mathbf{G}(\mathcal{H})\mathbf{h}_1 P_{s_1}(\mathcal{H})\mathbf{h}_1^H \mathbf{G}^H(\mathcal{H})\mathbf{h}_2}{1 + \frac{|h_0|^2 P_{s_1}(\mathcal{H})}{N_0}}}{N_0(1 + \mathbf{h}_2^H \mathbf{G}(\mathcal{H})\mathbf{G}^H(\mathcal{H})\mathbf{h}_2)}\right), \tag{11}$$

where we have defined $C(x) := \log(1 + x)$. From (8), (9), (10), (11) it is clear that the optimal power allocation requires $P_{s_1}(\mathcal{H}) = 1, \forall \mathcal{H}$. On the other hand, the optimization of powers $P_{s_2}(\mathcal{H})$ and $P_r(\mathcal{H})$, and linear processing $\mathbf{G}(\mathcal{H})$ boils down to the maximization of the term I_2 (11). In the next Sections, we first discuss the optimization of the relay linear processing $\mathbf{G}(\mathcal{H})$ (Sec. III), and the of power policies $P_{s_2}(\mathcal{H})$ and $P_r(\mathcal{H})$ (Sec. IV).

III. Optimization of the relay linear processing

In this Section, the expression for the optimal linear processing matrix at the relay $\mathbf{G}(\mathcal{H})$ is derived. From (11) it is easy to prove that the optimal $\mathbf{G}(\mathcal{H})$ can be written as the outer product of the beamformers for the channels $\mathbf{h}_1$ and $\mathbf{h}_2$

$$\mathbf{G} = \frac{g\mathbf{h}_2\mathbf{h}_1^H}{\|\mathbf{h}_2\|\|\mathbf{h}_1\|}, \tag{12}$$

where the scalar normalization factor g in (12) is determined by the relay power constraint (6). This can be restated, using the constraints in (8) and (12) as

$$P_r(\mathcal{H}) = \|\mathbf{h}_1\|^2 |g|^2 + N_0 |g|^2 = 1 - \theta(\mathcal{H}), \tag{13}$$

which implies the condition

$$g = \sqrt{\frac{1-\theta(\mathcal{H})}{N_0 + \|\mathbf{h}_1\|^2}}. \tag{14}$$

Substituting the optimal expressions (12) and (14) into the expression of I_2 (11), we get

$$I_2 = \frac{1}{2}C\left(\frac{|h_0|^2\theta(\mathcal{H}) + \frac{\|\mathbf{h}_2\|^2\|\mathbf{h}_1\|^2}{1+\frac{|h_0|^2}{N_0}}\frac{1-\theta(\mathcal{H})}{N_0+\|\mathbf{h}_1\|^2}}{N_0(1+\|\mathbf{h}_2\|^2\frac{1-\theta(\mathcal{H})}{N_0+\|\mathbf{h}_1\|^2})}\right), \tag{15}$$

which depends only on the power allocation policy $\theta(\mathcal{H})$.

IV. Optimization of the power allocation

In this section, the optimal power allocation policy is derived. As discussed in the previous Section, this problem boils down to the maximization of I_2 in (15) over the fraction of power allocated to the source in the second time-slot $\theta(\mathcal{H})$. Therefore the problem can be stated as

$$\max_{\theta\in[0,1]} f_0(\theta) \tag{16}$$

with

$$f_0(\theta) = C\left(\frac{|h_0|^2\theta + \frac{\|\mathbf{h}_2\|^2\|\mathbf{h}_1\|^2}{1+\frac{|h_0|^2}{N_0}}\frac{1-\theta}{N_0+\|\mathbf{h}_1\|^2}}{N_0(1+\|\mathbf{h}_2\|^2\frac{1-\theta}{N_0+\|\mathbf{h}_1\|^2})}\right).$$

After tedious calculations, it can be verified that

$$\frac{df_0(\theta)}{d\theta} = \frac{\alpha}{\beta(\theta)}, \tag{17}$$

where

$$\alpha = -(N_0 + \|\mathbf{h}_1\|^2)\left[\|\mathbf{h}_1\|^2\|\mathbf{h}_2\|^2 N_0 - |h_0|^2(N_0+|h_0|^2)(N_0+\|\mathbf{h}_1\|^2+\|\mathbf{h}_2\|^2)\right], \tag{18}$$

and $\beta(\theta)$ has a cumbersome expression which is positive for every value of $\theta \in [0,1]$. We can then conclude that $f_0(\theta)$ is monotone in the interval $[0,1]$, and that the maximum is achieved in $\theta = \{0,1\}$, according to the sign of α. We can summarize this conclusion as

$$\begin{cases} \theta = 0, & \text{if} \quad \alpha < 0 \\ \theta = 1, & \text{if} \quad \alpha > 0 \end{cases}. \tag{19}$$

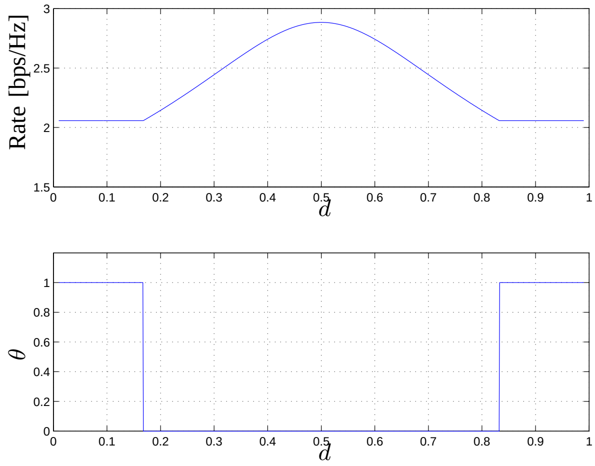

Fig. 2. Achievable rate and optimal power allocation for the AWGN system ($N = 2$, $\gamma = 4$, $\frac{1}{N_0} = 5dB$). $\theta = 1$ implies that the relay node is silent during the second time-slot and direct transmission is employed, while $\theta = 0$ implies that the source node is silent during the second time-slot and the AF protocol in [2] is employed.

After some algebraic manipulations the optimal strategy for power sharing can be expressed as

$$\begin{cases} \theta(\mathcal{H}) = 0 & \text{if} \quad K_D(\mathcal{H}) < K_L(\mathcal{H}) \\ \theta(\mathcal{H}) = 1 & \text{if} \quad K_D(\mathcal{H}) > K_L(\mathcal{H}) \end{cases} \tag{20}$$

where $K_D(\mathcal{H}) = \frac{|h_0|^2}{N_0}$ is the argument of $C(x)$ in the second time-slot term in (15) if $\theta = 1$ (i.e., it corresponds to direct transmission without the use of the relay), whereas $K_L(\mathcal{H}) = \frac{\|\mathbf{h}_1\|^2\|\mathbf{h}_2\|^2}{(1+\frac{|h_0|^2}{N_0})N_0(N_0+\|\mathbf{h}_1\|^2+\|\mathbf{h}_2\|^2)}$ is the argument of $C(x)$ in the second time-slot term in (15) if $\theta = 0$, and corresponds to the AF protocol presented in [2]. In other words, for any realization of the channels it is optimal to use either direct transmission or the protocol in [2], depending upon which scheme achieves the higher rate. Under the constraint of a fixed total power for each time-slot, the scheme presented in [3] is never optimal.

As a final remark we note that if we let $N_0 \to 0$ (i.e., asymptotically with respect to the signal-to-noise ratio), it is easily seen that $K_D(\mathcal{H}) \to \infty$, while

$$\lim_{N_0\to 0} K_L(\mathcal{H}) = \frac{\|\mathbf{h}_1\|^2\|\mathbf{h}_2\|^2}{|h_0|^2(\|\mathbf{h}_1\|^2+\|\mathbf{h}_2\|^2)}. \tag{21}$$

Thus for high SNR $K_D > K_L$ is always verified and the asymptotic optimal allocation is $\theta = 1$, i.e., direct transmission is optimal and the AF protocol is not advantageous.

V. Simulations results

In order to get insight into our results, we assume that the relay is located on a line between the source and the destination, such that the average power of the channels gains depends on the source-relay distance d and the path-loss exponent γ (see fig. 1). Let us first consider the unfaded case (AWGN channel), where $|h_0|^2 = 1$, $\|\mathbf{h}_1\|^2 = \frac{1}{d^\gamma}N$, $\|\mathbf{h}_2\|^2 = \frac{1}{(1-d)^\gamma}N$. In this case, fig. 2 shows the achievable rate and the optimal

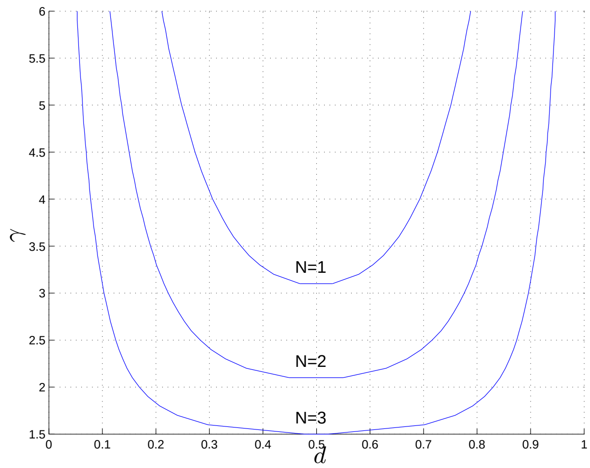

Fig. 3. Optimality regions for different values of N ($\frac{1}{N_0} = 5dB$). In the region above each curve the AF scheme in [2] is optimal, while in the region below each curve direct transmission is optimal.

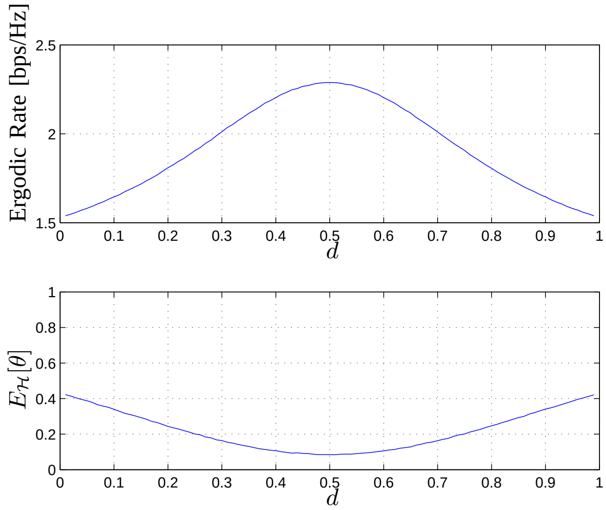

Fig. 4. Ergodic achievable rate $E_{\mathcal{H}}[R]$ and average optimal power allocation $E_{\mathcal{H}}[\theta]$ ($N = 2$, $\gamma = 4$, $\frac{1}{N_0} = 5dB$).

power allocation for $N = 2$, $\gamma = 4$, $1/N_0 = 5dB$. It is seen that the scheme in [2] is optimal when the relay is located in an interval around halfway between the source and the destination. Similar results can be obtained for different values of γ and N, showing that, increasing either the number of antennas or the path-loss exponent, the interval of d in which the relay is used grows. The regions of the $\gamma - d$ plane where either technique is optimal are plotted in fig. 3 for different values of N, with $1/N_0 = 5dB$. In the region above each curve $\theta = 0$ and the AF scheme in [2] is optimal, while in the region below each curve $\theta = 1$ and direct transmission is optimal. As expected, for larger number of antennas N the region in which the scheme in [2] is advantageous becomes larger.

Introducing uncorrelated Rayleigh fading, fig. 4 shows the ergodic achievable rate $E_{\mathcal{H}}[R]$ and the average power sharing $E_{\mathcal{H}}[\theta]$ ($E_{\mathcal{H}}[.]$ denotes the expectation with respect to fading states) versus different values of the source-relay distance d for $N = 2$, $\gamma = 4$, $1/N_0 = 5dB$. It is seen that $E_{\mathcal{H}}[\theta(\mathcal{H})]$ is symmetric around $d = 0.5$ and that the relay is more frequently used when it is located halfway between the source and the destination.

Acknowledgment

The authors would like to thank Dongho Kim and Cheolwoo You for their valuable comments.

VI. Conclusions

In this paper, optimal power allocation and linear processing have been investigated for an amplify-and-forward fading relay channel with multiple antennas at the relay node and under an instantaneous sum-power constraint. The optimal linear processing at the relay node is the outer product of the beamformers for the source-relay and relay-destination channels. Moreover, the optimal transmission scheme is either direct transmission (the relay remains silent in the second time slot), or the scheme proposed in [2] (the source remains silent in the second time-slot).

References

[1] T. M. Cover and A. A. El Gamal , "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 572-584, Sep. 1979.

[2] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behaviour," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3062-3080, Dec. 2004.

[3] R. U. Nabar, H. Bölcskei, and F. W. Kneubuhler, "Fading relay channels: Performance limits and space-time signal design," *IEEE Jour. on Selected Areas in Communications*, vol. 22, Issue 6, pp. 1099-1109, June 2004.

[4] Y. Yingwei, C. Xiaodong and G. B. Giannakis, "On energy efficiency and optimum resource allocation of relay transmissions in the low-power regime", *IEEE Trans. on Wireless Communications*, vol. 4, Issue 6, pp. 2917-2927, Nov. 2005.

[5] L. Yingbin and V. V. Veeravalli, "Resource allocation for wireless relay channels", *38th Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1902-1906, 7-10 Nov. 2004 .

[6] A. Host-Madsen and Z. Junshan, "Capacity bounds and power allocation for wireless relay channels", *IEEE Trans. Inf. Theory*, vol. 51, Issue 6, pp. 2020-2040, June 2005.

[7] N. Ahmed, M. A. Khojastepour and B. Aazhang, "Outage minimization and optimal power control for the fading relay channel", *IEEE Information Theory Workshop*, pp. 458-462, 24-29 Oct. 2004. vol. 51, Issue 6, pp. 2020-2040, June 2005.

[8] M. A. Khojastepour, A. Sabharwal and B. Aazhang, "Lower bounds on the capacity of Gaussian relay channel", *38th Annual Conf. on Info. Sci. and Sys. (CISS 2004)*, Princeton, NJ.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Interscience, 1991.