This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

Achievable Rates of Secure Transmission in Gaussian MISO Channel with Imperfect Main Channel Estimation

Xinyu Zhou, Zouheir Rezki *Senior Member, IEEE*, Basel Alomair, *Member, IEEE*, Mohamed-Slim Alouini, *Fellow, IEEE*

***Abstract*—A Gaussian multiple-input single-output (MISO) fading channel is considered. We assume that the transmitter, in addition to the statistics of all channel gains, is aware instantaneously of a noisy version of the channel to the legitimate receiver. On the other hand, the legitimate receiver is aware instantaneously of its channel to the transmitter, whereas the eavesdropper instantaneously knows all channel gains. We evaluate an achievable rate using a Gaussian input without indexing an auxiliary random variable. A sufficient condition for beamforming to be optimal is provided. When the number of transmit antennas is large, beamforming also turns out to be optimal. In this case, the maximum achievable rate can be expressed in a simple closed form and scales with the logarithm of the number of transmit antennas. Furthermore, in the case when a noisy estimate of the eavesdropper's channel is also available at the transmitter, we introduce the SNR difference and the SNR ratio criterions and derive the related optimal transmission strategies and the corresponding achievable rates.**

I. INTRODUCTION

A. Motivation and Related Work

The notion of information theoretic security was first introduced by Shannon in [1]. According to the mathematical structure of secrecy systems set in this paper, the intended receiver and the eavesdropper (also called the "enemy") have direct access to the transmitted signal, and the system is said "perfectly secure" if the enemy's observation is independent of the secret message. Unfortunately, in this setting, "perfect secrecy" requires at least as many keys as secret messages [1]. Later, another setting for information theoretic security was proposed by Wyner in [2]. In the latter setting, the intended receiver and the the eavesdropper observe the transmitted signal, not directly as in [1], but via two discrete memoryless channels, with the channel to the eavesdropper being a degraded version of the one to the legitimate receiver. In particular, it was shown in [2] that "perfect secrecy" is achievable. Since then, there has been an intermittent gain of interest toward exploring information theoretical potentials as a new paradigm to secure or complement existing security mechanisms. For instance, [3] extended the information-theoretic secrecy to the Gaussian wiretap channel. Also, in [4] a setting that generalizes the wiretap channel is proposed. Specifically, a common message is transmitted to two users while a private message is transmitted to only one of them. Users should be able to recover the common messages without errors whereas the confidential message should be recovered strictly by the legitimate receiver. Differently from [2], the channel to the two receivers are not necessarily ordered. Again, a full characterization of the rates-equivocation region is provided.

Extensive efforts have been devoted to designing different schemes to achieve secrecy through wireless channel. In [5], artificial noise (AN), which can be canceled by the legitimate receiver, is injected through the channel in order to degrade the eavesdropper's channel. Secure communications can be guaranteed, regardless of the position of the eavesdropper, even if the eavesdropper has access to perfect channel state information (CSI). Similarly, another scheme with intended ambiguity injected is proposed in [6]. On the other hand, in [7], a helping interferer is used to assist the secrecy transmission. The interferer, which does not know the confidential message, can guarantee secrecy by sending independent interference signals.

Multi-antenna system has gained great popularity since it can provide both spatial multiplexing and diversity gains. Channel state information at the transmitter (CSIT) is particularly useful in improving the performance of multi-antenna systems. Although perfect CSIT is often hard to acquire because of the rapid time variations of the wireless channel, it has been shown that even partial CSIT may provide a substantial performance gain. For instance, in [8], the authors consider two types of feedback: the mean feedback and the covariance feedback. Optimal transmission strategies in both cases are derived in [8]. In [9], the authors derive a necessary and sufficient condition under which beamforming is the optimal transmission strategy to achieve the Shannon capacity. The result of [8] is extended to MIMO cases in [10] and [11], where the optimal covariance matrix of the Gaussian

This work has been done while Xinyu Zhou, an undergraduate student from Shanghai Jiaotong University, Shanghai, China, was doing his internship at King Abdullah University of Science and Technology (KAUST), (email: zhxyhhh@hotmail.com).

Zouheir Rezki and Mohamed-Slim Alouini are with the Electrical Engineering Program, Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Mekkah Province, Saudi Arabia, (email: {zouheir.rezki, slim.alouini}@kaust.edu.sa)

Basel Alomair is with The National Center for Cybersecurity Technology (C4C), King Abdulaziz City for Science and Technology (KACST), Riyadh, Saudi Arabia, (email: alomair@kacst.edu.sa).

This work has been supported by a grant from King Abdulaziz City of Science and Technology (KACST), Riyadh, Saudi Arabia.

Part of this work has been presented at the 2015 IEEE Global Communication Conference (Globecom'2015), Workshop on Trusted Communications with Physical Layer Security, San Diego, CA, USA, December 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

input vector is derived. Furthermore, the effect of channel uncertainty on the capacity is analyzed in [12], [13] and [14].

In the case of transmission with secrecy constraints, the effect of fading together with the knowledge of channel information at the transmitter has received a great deal of attention recently. In [15], when the channel input is restricted to be Gaussian, optimal transmission strategies are found for different channel fading assumptions. In particular, both in the case of a constant eavesdropper channel that is known to all parties and in the case of a fading eavesdropper channel that is unknown to the transmitter, the optimal transmission strategy turns out to be beamforming. In [16], secrecy capacity is characterized under the perfect CSI assumptions as well as the case that only the CSI of the legitimate receiver is known to the transmitter. The results show that coding over a large number of fading blocks, has a positive effect on the secrecy capacity and secure communication is possible even when the average SNR of the legitimate receiver is smaller than that of the eavesdropper. The secrecy outage with perfect main CSIT and partial eavesdropper's CSIT available at the transmitter is investigated in [17].

Further results focusing on imperfect main CSIT have been reported recently, e.g., [18]. In [19], the transmitter is provided a noisy version of the main CSI and only the statistics of the eavesdropper's CSI. The authors give a lower bound and an upper bound on the secrecy capacity under this setting. Furthermore, the result is extended to the MISO case in [20]. In [21], several results are derived regarding the optimal input covariance matrix when the eavesdropper and the legitimate channels have nontrivial covariance. The authors also proved that as long as perfect CSI of the legitimate channel is available at the transmitter, a rank-one input covariance matrix is optimal. Other fundamental results focusing more on quantized or delayed main CSI feedback at the transmitter have been reported in e.g., [22], [23] and [24].

B. Approach and Contributions

In this paper, we consider a fast fading MISO wiretap channel. The maximum achievable rate using a Gaussian input without an auxiliary random variable is considered. While the maximization problem is not convex and thus not straightforward to solve in general, we develop a sufficient condition for beamforming to be optimal in the case where the transmitter has only the eavesdropper's channel statistics. In the case of a large number of antennas at the transmitter, we also find that beamforming is optimal and the maximum achievable rate takes a simple closed form. Meanwhile, we also find that the maximum achievable rate scales with the logarithm of the number of transmit antennas.

In the general case, when the transmitter also has a noisy version of the eavesdropper's channel, we consider two important optimization problems related to the achievable rate: the SNR difference criterion and the SNR ratio criterion. The transmission strategies related to the maximization of the previous performance metrics are derived. We also argue that if the power constraint and the channel estimation errors satisfy certain conditions, then these two SNR criterions can provide achievable secrecy rates that are very close to the maximum achievable rates using Gaussian input.

We note that our problem somewhat resembles to the mean feedback problem treated in [8] and [9], but without secrecy constraint. Furthermore, our framework is also related to [25]. However, the result in [25] cannot be applied to our problem since perfect CSIT is not available. Our primary concern in this paper is on how to design secure transmission strategies leveraging CSIT.

We note that our work is quite different from [18]. Recall that [18] deals essentially with perfect main CSI, it devotes one section to highlight the effect of imperfect main CSI at both the transmitter and the legitimate receiver. However, the approach adopted therein is quite different from our framework. For instance, while [18] provides an achievable secrecy rate based on an AN scheme, our scheme does not rely on sending AN, but rather utilizes a wiretap code based scheme adapted to fast fading channel. Consequently, in its general form, our secrecy rate is different from the one reported in [18].

It is worth mentioning that another transmission strategy that may guarantee secrecy is the so-called AN aided beamforming (BF). Essentially, this strategy consists of transmitting both the information bearing signal and the AN along all directions and then maximize the rate by optimizing over both the directions space and the power profiles subject to a power constraint [26]–[28]. For a more comprehensive treatment of communications with secrecy constraint including transmission schemes, please refer to [29]. In our previous work [20], we have adopted a particular case of AN-BF transmission strategy and realized that it may not generally allow gaining insights on design guidelines. The fact that the transmitter has only noisy main CSI even worsen the situation. Hence why in this paper, our focus is on BF without AN, albeit we are aware that AN-BF performs better.

C. Outline of the Paper

The remainder of this paper is organized as follows: Section II contains the system model and the problem statement. Section III evaluates the achievable rate and gives the main results. Section IV addresses a more general case and introduces the SNR difference criterion and the SNR ratio criterion. Numerical results are included in Section V. Section VI concludes the paper.

Notation: $\mathcal{CN}(\boldsymbol{\mu}, K)$ denotes circularly symmetric complex distribution with mean vector $\boldsymbol{\mu}$ and covariance K. $\mathcal{N}(\boldsymbol{\mu}, K)$ represents real Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance K. All logarithms are natural logarithm. For two vector $\boldsymbol{m}$, $\boldsymbol{n}$, $\boldsymbol{m} \succeq \boldsymbol{n}$ means $\boldsymbol{m}$ majorizes $\boldsymbol{n}$. The function $f_x()$ and $F_x()$ denote the probability density function (pdf) and the cumulative distribution function (cdf) of the random variable x. More specifically, $\Phi(x)$ is the cumulative distribution function of standard Gaussian distribution with zero mean and unity variance, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$. $\mathrm{tr}(X)$ represents the trace of the matrix X. The Hermitian transpose of a matrix X is denoted as X^H. $\mathrm{diag}(X)$ represents the vector containing the diagonal elements of the matrix X. $O(.)$ represents the big O notation. The vectors are bolded

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

throughout the paper. For a random sequence X_n, $X_n \overset{a.s.}{\to} X$ means that the sequence X_n converges almost surely towards X, i.e., $P(\lim_{n\to\infty} X_n = X) = 1$. For a matrix U, $U[i]$ designates its ith column.

II. SYSTEM MODEL AND PROBLEM STATEMENT

We consider a fast fading multiple-input single-output (MISO) channel where a transmitter is communicating to a receiver in the presence of an eavesdropper. The transmitter is equipped with N_t antennas while each of the legitimate receiver and the eavesdropper is equipped with only one antenna. The channel gains to the legitimate receiver and the eavesdropper can be represented as two $N_t \times 1$ vectors $\boldsymbol{h}$ and $\boldsymbol{g}$. The received signals at the legitimate receiver and the eavesdropper can be written as,

$$\begin{cases} y_j = \boldsymbol{h}_j^H \boldsymbol{x}_j + v_j \\ z_j = \boldsymbol{g}_j^H \boldsymbol{x}_j + w_j, \end{cases} \tag{1}$$

where $\boldsymbol{x}_j$ is the transmitted signal at time j and v_j, w_j are unit-variance complex circularly symmetric Gaussian random variables. We assume that all random processes involved in (1) are stationary and thus the time index j will be omitted next for convenience. The channel gains $\boldsymbol{h}$ and $\boldsymbol{g}$ are assumed to be independent, but not identically distributed. Specifically, we assume without loss of generality that $\boldsymbol{g} \sim \mathcal{CN}(0, I)$ and $\boldsymbol{h} \sim \mathcal{CN}\left(0, \sigma_h^2 I\right)$. This will allow us to capture the cases where the main channel is better than the eavesdropper's channel on average $\sigma_h^2 > 1$, the main channel is worse than the eavesdropper's channel on average ($\sigma_h^2 < 1$) and the main channel is statistically similar to the eavesdropper's channel ($\sigma_h^2 = 1$). The transmitter is provided with a noisy version of $\boldsymbol{h}$ which can be written as:

$$\boldsymbol{h} = \sqrt{1-\alpha}\hat{\boldsymbol{h}} + \sqrt{\alpha}\tilde{\boldsymbol{h}}, \tag{2}$$

where $\hat{\boldsymbol{h}}$ is the estimate of the main channel, $\hat{\boldsymbol{h}}$ follows $\mathcal{CN}\left(0, \sigma_h^2 I\right)$. $\tilde{\boldsymbol{h}}$ represents the estimation error, $\tilde{\boldsymbol{h}} \sim \mathcal{CN}\left(0, \sigma_h^2 I\right)$ and α is the error variance, $\alpha \in [0, 1]$. For simplicity, it is assumed that $\hat{\boldsymbol{h}}$, $\tilde{\boldsymbol{h}}$ and $\boldsymbol{g}$ are independent of each other. We assume that the transmitter is only aware of an estimate $\hat{\boldsymbol{h}}$ of the main channel $\boldsymbol{h}$. On the other hand, the legitimate receiver is aware instantaneously of its channel to the transmitter, whereas the eavesdropper instantaneously knows all channel gains. It is also assumed that the transmitter would broadcast its channel estimate so that this estimate is known to all parties. Finally the input is subject to an average power constraint: $\mathbf{E}(\boldsymbol{x}\boldsymbol{x}^H) \leq P$. From [4], an ergodic achievable rate of the above channel is given by:

$$\begin{aligned} R_s &= \max_{\boldsymbol{u}\to\boldsymbol{x}\to\boldsymbol{g},y,\boldsymbol{h},z} I(\boldsymbol{u}; y, \boldsymbol{h}|\hat{\boldsymbol{h}}) - I(\boldsymbol{u}; z, \boldsymbol{g}|\hat{\boldsymbol{h}}) \quad (3) \\ &= \max_{\boldsymbol{u}\to\boldsymbol{x}\to\boldsymbol{g},y,\boldsymbol{h},z} I(\boldsymbol{u}; y|\boldsymbol{h}, \hat{\boldsymbol{h}}) - I(\boldsymbol{u}; z|\boldsymbol{g}, \hat{\boldsymbol{h}}), \quad (4) \end{aligned}$$

where $\boldsymbol{u}$, $\boldsymbol{x}$ and $(y, \boldsymbol{h}, z, \boldsymbol{g})$ form a Markov chain conditioned on $\hat{\boldsymbol{h}}$, and where (4) follows from (3) since given $\hat{\boldsymbol{h}}$, $\boldsymbol{u}$ and $\boldsymbol{h}$ and $\boldsymbol{u}$ and $\boldsymbol{g}$ are independent. We restrict ourselves to the potentially sub-optimal inputs without indexing, i.e., $\boldsymbol{x} = \boldsymbol{u}$, under which the achievable rate becomes:

$$R_s = \max_{p(x)} I(\boldsymbol{x}; y|\boldsymbol{h}, \hat{\boldsymbol{h}}) - I(\boldsymbol{x}; z|\boldsymbol{g}, \hat{\boldsymbol{h}}). \tag{5}$$

It is well known that a Gaussian input maximizes the target in (5). Finally the problem of interest is:

$$\max_{Q(\hat{\boldsymbol{h}})} \left[\mathbf{E}_{\boldsymbol{h},\hat{\boldsymbol{h}}}[\log(1 + \boldsymbol{h}^H Q(\hat{\boldsymbol{h}})\boldsymbol{h})] - \mathbf{E}_{\boldsymbol{g},\hat{\boldsymbol{h}}}[\log(1 + \boldsymbol{g}^H Q(\hat{\boldsymbol{h}})\boldsymbol{g})]\right]^+, \tag{6}$$

where $Q(\hat{\boldsymbol{h}})$ is the covariance matrix of the channel input $\boldsymbol{x}$ and is subject to the constraint $\text{tr}(Q(\hat{\boldsymbol{h}})) \leq P$ and $[x]^+ = \max(0, x)$. We note that always using full power for each channel realization $\hat{\boldsymbol{h}}$ is suboptimal. However, since determining the structure of the optimal covariance matrix under a trace constraint inequality is very involved, we choose to consider a trace constraint equality instead, in order to gain an insight on signaling design. That is,

$$\mathbf{E}(\boldsymbol{x}^H\boldsymbol{x}) = \text{tr}(Q(\hat{\boldsymbol{h}})) = P. \tag{7}$$

As a consequence of this choice, it may happen that for some channel CSIT realizations $\hat{\boldsymbol{h}}$, the instantaneous rate $R_s(U, \Lambda, \hat{\boldsymbol{h}})$ is negative and for others, it is positive. The overall secrecy rate $\bar{R}_s$ cannot be negative since in the worst case scenario it would be equal to 0.

We note that if AN-BF transmission strategy is used, then it follows directly from (4) (by taking $\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{a}$ with $\boldsymbol{u} \sim \mathcal{CN}\left(\boldsymbol{0}, \phi_s(\hat{\boldsymbol{h}})\right)$ and $\boldsymbol{a} \sim \mathcal{CN}\left(\boldsymbol{0}, \phi_a(\hat{\boldsymbol{h}})\right)$) that the following secrecy rate is achievable:

$$\begin{aligned} \max_{\phi_s(\hat{\boldsymbol{h}}),\phi_a(\hat{\boldsymbol{h}})} \quad & \mathbf{E}_{\boldsymbol{h},\hat{\boldsymbol{h}},\boldsymbol{g}} \left[\log\left(1 + \frac{\boldsymbol{h}^H\phi_s(\hat{\boldsymbol{h}})\boldsymbol{h}}{1+\boldsymbol{h}^H\phi_a(\hat{\boldsymbol{h}})\boldsymbol{h}}\right)\right. \\ & \left. - \log\left(1 + \frac{\boldsymbol{g}^H\phi_s(\hat{\boldsymbol{h}})\boldsymbol{g}}{1+\boldsymbol{g}^H\phi_a(\hat{\boldsymbol{h}})\boldsymbol{g}}\right)\right] \end{aligned} \tag{8}$$

where $\phi_s(\hat{\boldsymbol{h}})$ and $\phi_a(\hat{\boldsymbol{h}})$ are the covariance matrix of the signal and AN respectively, and the maximization in (8) is subject to $\text{tr}(\phi_s(\hat{\boldsymbol{h}})) + \text{tr}(\phi_s(\hat{\boldsymbol{h}})) = P$.

Given our setting, it is clear that in case $\alpha = 1$ and $\sigma_h^2 \leq 1$, there is no advantage of the legitimate receiver over the eavesdropper that can be exploited by the transmitter. Hence, in this case, the secrecy capacity is equal to 0. Therefore, in the sequel, we assume that we either have $\alpha < 1$ or $\sigma_h^2 > 1$.

III. ACHIEVABLE RATE EVALUATION

For a given $\hat{\boldsymbol{h}}$, the target in (6) can be written as:

$$R_s(Q, \hat{\boldsymbol{h}}) \triangleq \mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}[\log(1 + \boldsymbol{h}^H Q\boldsymbol{h})|\hat{\boldsymbol{h}}] - \mathbf{E}_{\boldsymbol{g}}[\log(1 + \boldsymbol{g}^H Q\boldsymbol{g})]. \tag{9}$$

Note that $Q = Q(\hat{\boldsymbol{h}})$. However, to keep the notation convenient, we make the dependence on $\hat{\boldsymbol{h}}$ implicit and write it as Q instead of $Q(\hat{\boldsymbol{h}})$. We define the eigenvalue decomposition of the covariance matrix $Q = U\Lambda U^H$, where Λ is a diagonal

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

matrix and U is a unitary matrix with $U^HU = UU^H = I$. Then (9) can be rewritten as:

$$\begin{aligned} R_s(Q,\hat{\boldsymbol{h}}) &= R_s(U\Lambda U^H,\hat{\boldsymbol{h}}) && (10)\\ &\triangleq R_s(U,\Lambda,\hat{\boldsymbol{h}}) && (11)\\ &= \mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}[\log(1+\boldsymbol{h}^HU\Lambda U^H\boldsymbol{h})|\hat{\boldsymbol{h}}] \\ &\quad -\mathbf{E}_{\boldsymbol{g}}[\log(1+\boldsymbol{g}^HU\Lambda U^H\boldsymbol{g})] && (12)\\ &= \mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}[\log(1+\boldsymbol{h}^HU\Lambda U^H\boldsymbol{h})|\hat{\boldsymbol{h}}] \\ &\quad -\mathbf{E}_{\boldsymbol{g}}[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})]. && (13)\end{aligned}$$

Equation (13) follows from (12) since according to [30], when U is unitary and $\boldsymbol{g}$ is a random vector whose elements are independent and identically distributed (i.i.d.) zero-mean complex circularly symmetric Gaussian, then $U^H\boldsymbol{g}$ shares the same distribution as $\boldsymbol{g}$.

The expected form of the maximum achievable rate can be defined by:

$$\bar{R}_s \triangleq \mathbf{E}_{\hat{\boldsymbol{h}}}\left[\max_{U,\Lambda}\left\{\mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\left[\log(1+\boldsymbol{h}^HU\Lambda U^H\boldsymbol{h})|\hat{\boldsymbol{h}}\right] -\mathbf{E}_{\boldsymbol{g}}\left[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})\right]\right\}\right]. \quad (14)$$

A. Discussion

The problem is slightly related to the mean feedback case in [8], where the capacity without secrecy constraint is studied. In [8], the capacity is strictly concave in Q and thus the Frechet differential condition which is necessary for optimality is also sufficient. This was the key observation to solve the channel capacity with mean feedback problem. Unfortunately, this approach cannot be applied directly to our problem since the cost function is not concave in Q, and hence the Frechet differential only provides a necessary condition for optimality. While we do not solve the problem in its general form, the problem of interest can be solved in some non-trivial special cases. In *Theorem* 1, we give a sufficient condition for beamforming to be optimal. In *Theorem* 2, we claim that beamforming is also optimal in the case where the number of antennas at the transmitter approaches infinity. The maximum achievable rate takes a simple closed form and scales with the logarithm of the number of transmit antennas.

B. Sufficient Condition For Beamforming To Be Optimal

Theorem 1 provides a sufficient condition for beamforming to be the optimum transmission strategy. The outline of the proof is that departing from (13), beamforming can maximize the first term and minimize the second term simultaneously under this sufficient condition, and thus the difference of them is maximized.

Theorem 1: For a given noisy estimation of the main channel $\hat{\boldsymbol{h}}$, if the following condition is satisfied,

$$\mathbf{E}_{\omega_1}[\frac{1}{1+P\,\alpha\,\sigma_h^2\omega_1}] \le \frac{1}{1+P\,\alpha\,\sigma_h^2}, \quad (15)$$

where ω_1 has a noncentral chi-square distribution with pdf

$$f_{\omega_1}(\omega_1) = e^{-\frac{(1-\alpha)\|\hat{\boldsymbol{h}}\|^2}{\alpha\,\sigma_h^2}-\omega_1} I_0\left(2\sqrt{\frac{(1-\alpha)\,\|\hat{\boldsymbol{h}}\|^2\omega_1}{\alpha\sigma_h^2}}\right),$$

where $I_0(.)$ is the 0th-order modified Bessel function of the first kind, then the maximum achievable rate is attained by beamforming to the direction of $\hat{\boldsymbol{h}}$.

Proof: Here we rewrite the optimization problem:

$$\max_{U,\Lambda} R_s(U,\Lambda,\hat{\boldsymbol{h}}) = \max_{U,\Lambda}\left\{\mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\left[\log(1+\boldsymbol{h}^HU\Lambda U^H\boldsymbol{h})|\hat{\boldsymbol{h}}\right] - \mathbf{E}_{\boldsymbol{g}}\left[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})\right]\right\}. \quad (16)$$

We will first focus on the second term: $f(\Lambda) \triangleq \mathbf{E}_{\boldsymbol{g}}[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})]$. Define the diagonal elements of Λ as λ_i's, where $\Sigma\lambda_i = P$, and let $\boldsymbol{g} = (g_1,\ldots,g_{N_t})$. Note that $f(\Lambda)$ can also be written as $f(\Lambda) = \mathbf{E}_{\boldsymbol{g}}[\log(1+\sum_{i=1}^{N_t}\lambda_i|g_i|^2)]$ which is clearly symmetric in λ_i's. That is, we can exchange the order of λ_i and λ_j without modifying the value of $f(\Lambda)$. In addition, $f(\Lambda) = \mathbf{E}_{\boldsymbol{g}}[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})]$ is a continuous concave function with respect to Λ. Following *Proposition* 3.C.2 in [31], if a function is symmetric and concave, then it is Schur-concave. So the function $f(\Lambda)$ is Schur-concave with respect to Λ. According to the definition of a Schur-concave function, for Λ_1 and Λ_2, if $\mathrm{diag}(\Lambda_1) \succeq \mathrm{diag}(\Lambda_2)$, then $f(\Lambda_1) \le f(\Lambda_2)$. Since uni-rank Λ^o with $\mathrm{diag}(\Lambda^o) = [P,0,...0]$ majorizes any other Λ, the minimum of $f(\Lambda)$ is achieved by Λ^o.

Then we step back to the first term of (16). With the condition $\mathbf{E}_{\omega_1}[\frac{1}{1+P\,\alpha\,\sigma_h^2\,\omega_1}] \le \frac{1}{1+P\,\alpha\,\sigma_h^2}$, we can derive directly from [9] that the maximum of the first term can be achieved with uni-rank Λ^o with $\mathrm{diag}(\Lambda^o) = [P,0,...0]$ and the maximizing U^o is given by the first column $U^o[1] = \frac{\hat{\boldsymbol{h}}}{\|\hat{\boldsymbol{h}}\|}$, $U^o[2]...U^o[N_t]$ are arbitrarily chosen, except for the restriction that the columns of U^o are orthonormal.

Also note that the choice of U does not affect the value of $f(\Lambda)$ in (16). Combining the above two statements, we have that Λ^o and U^o both achieve the maximum of the first term and the minimum of the second term, thus achieve the maximum of the difference of them.

$$\begin{aligned} &\max_{U,\Lambda} R_s(U,\Lambda,\hat{\boldsymbol{h}}) \\ =&R_s(U^o,\Lambda^o) && (17)\\ =&\mathbf{E}_{v_1}\left[\log(1+P|v_1|^2)\right] - \mathbf{E}_{g_1}\left[\log(1+P|g_1|^2)\right], && (18)\end{aligned}$$

where $v_1 \sim \mathcal{CN}\left(\sqrt{1-\alpha}\|\hat{\boldsymbol{h}}\|, \alpha\,\sigma_h^2\right)$, and $g_1 \sim \mathcal{CN}(0,1)$. Hence the achievable rate is maximized by beamforming to the direction of $\hat{\boldsymbol{h}}$. ■

Therefore, if the beamforming condition in *Theorem* 1 holds true, then the secrecy rate simplifies to:

$$\bar{R}_s = \mathbf{E}_{\|\hat{\boldsymbol{h}}\|,v_1}\left[\log(1+P|v_1|^2)\right] - \mathbf{E}_{g_1}[\log(1+P|g_1|^2)], \quad (19)$$

where the averaging in (19) is over scalar r.v.'s. Although a closed-form expression of $\bar{R}_s$ is apparently hard to obtain, numerical evaluation of (19) is rather straightforward.

C. Large Scale Antennas Case

In this section, we consider the case where the number of antennas at the transmitter goes to infinity. We prove that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

beamforming is the optimal strategy and the maximum achievable rate can take a simple closed form. For convenience, the proof is streamlined into 3 lemmas.

Lemma 1: When $N_t \to \infty$, define $\|\hat{\boldsymbol{h}}\|^2 = |\hat{h}_1|^2 + |\hat{h}_2|^2 ... + |\hat{h}_{N_t}|^2$, where $\hat{h}_i$'s are i.i.d. random variables and $\hat{h}_i \sim \mathcal{CN}(0,1)$. Then for any real number u,

$$\lim_{N_t\to\infty} \Pr(\|\hat{\boldsymbol{h}}\|^2 > u) = 1. \tag{20}$$

That is, the value of $\|\hat{\boldsymbol{h}}\|^2$ converges to infinity almost surely, $\lim_{N_t\to\infty} \|\hat{\boldsymbol{h}}\|^2 \overset{a.s.}{\to} \infty$.

Proof: For each $\hat{h}_i$, $|\hat{h}_i|^2$ can be regarded as the sum of square of two i.i.d random variables which follow $\mathcal{N}(0, \frac{\sigma_h^2}{2})$. The random variable $\|\hat{\boldsymbol{h}}\|^2 = \sum_{i=1}^{N_t} |\hat{h}_i|^2$, by definition, has a chi-square distribution $\chi^2(2N_t)$ with $2N_t$ degrees of freedom. By the central limit theorem, because the chi-squared distribution is the sum of $2N_t$ independent random variable with finite mean and variance, it converges to a normal distribution for large N_t. That is,

$$\lim_{N_t\to\infty} \frac{\|\hat{\boldsymbol{h}}\|^2/\sigma_h^2 - N_t}{\sqrt{N_t}} \sim \mathcal{N}(0,1). \tag{21}$$

For any real number u:

$$\lim_{N_t\to\infty} \Pr(\|\hat{\boldsymbol{h}}\|^2 > u) = \lim_{N_t\to\infty} \Pr(\|\hat{\boldsymbol{h}}\|^2/\sigma_h^2 > u/\sigma_h^2) \tag{22}$$

$$= \lim_{N_t\to\infty} 1 - \Phi(\frac{u/\sigma_h^2 - N_t}{\sqrt{N_t}}) = 1. \tag{23}$$

■

The following lemma is also useful to establish our result.

Lemma 2: Let v be a circularly-symmetric complex Gaussian random variable with real mean m and variance σ^2. Then

$$\lim_{m\to\infty} \{avr[\log(1+|v|^2)] - \log(\mathbf{E}[1+|v|^2])\} = 0. \tag{24}$$

Proof: For convenience, the proof is provided in Appendix A. ■

Finally, we claim the following result.

Lemma 3: Given any $\hat{\boldsymbol{h}}$. If $\|\hat{\boldsymbol{h}}\|^2 < \tau$, where τ is a positive real number, then $R_s(Q, \hat{\boldsymbol{h}})$ defined in (9) is bounded.

Proof: $R_s(Q, \hat{\boldsymbol{h}})$ can be written as:

$$R_s(Q,\hat{\boldsymbol{h}}) = \mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\left[\log(1+\boldsymbol{h}^H Q \boldsymbol{h})|\hat{\boldsymbol{h}}\right] - \mathbf{E}_{\boldsymbol{g}}\left[\log(1+\boldsymbol{g}^H Q \boldsymbol{g})\right] \tag{25}$$

$$= \mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\left[\log(1+\boldsymbol{h}^H U\Lambda U^H \boldsymbol{h})|\hat{\boldsymbol{h}}\right] - \mathbf{E}_{\boldsymbol{g}}\left[\log(1+\boldsymbol{g}^H \Lambda \boldsymbol{g})\right], \tag{26}$$

where $U\Lambda U^H$ is the eigenvalue decomposition of Q. Define $\boldsymbol{v} = U^H\boldsymbol{h}$ and $\boldsymbol{m}$ as the mean vector of $\boldsymbol{v}$ with $\boldsymbol{m}^H\boldsymbol{m} = (1-\alpha)\hat{\boldsymbol{h}}^H\hat{\boldsymbol{h}}$, we have:

$$|R_s(U,\Lambda,\hat{\boldsymbol{h}})| = \left|\mathbf{E}_{\boldsymbol{v}}\left[\log(1+\boldsymbol{v}^H\Lambda\boldsymbol{v})\right] - \mathbf{E}_{\boldsymbol{g}}[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})]\right| \tag{27}$$

$$\leq \left|\mathbf{E}_{\boldsymbol{v}}\left[\log(1+\boldsymbol{v}^H\Lambda\boldsymbol{v})\right]\right| + \left|\mathbf{E}_{\boldsymbol{g}}\left[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})\right]\right| \tag{28}$$

$$\leq \log\left(1+\mathbf{E}\left[\Sigma_i\lambda_i|v_i|^2\right]\right) + \log\left(1+\mathbf{E}\left[\Sigma_i\lambda_i|g_i|^2\right]\right) \tag{29}$$

$$\leq \log(1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\sigma_h^2)) + \log(1+P), \tag{30}$$

where (29) follows from (28) using Jensen's inequality. Equation (30) follows from (29) since if we suppose that $t = \text{argmax}_i \mathbf{E}\left[|v_i|^2\right]$, then

$$\mathbf{E}\left[\Sigma_i\lambda_i|v_i|^2\right] \leq (\sum_i \lambda_i)\mathbf{E}\left[|v_t|^2\right] = P\left(|m_t|^2+\alpha\sigma_h^2\right) \leq P\left((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\sigma_h^2\right). \tag{31}$$

Since $\|\hat{\boldsymbol{h}}\|^2 < \tau$, then it follows immediately from (30) that $R_s(Q,\hat{\boldsymbol{h}})$ is bounded. ■

Now we are ready to give the result.

Theorem 2: As $N_t \to \infty$, the achievable rate can be maximized by beamforming to the direction of $\hat{\boldsymbol{h}}$. The expectation of the maximum achievable rate $\bar{R}_s$, defined in (14), can be expressed as:

$$\lim_{N_t\to\infty}\left\{\bar{R}_s - \left[\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log(1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\sigma_h^2))\right] - \mathbf{E}_{g_1}\left[\log(1+P|g_1|^2)\right]\right]\right\} = 0. \tag{32}$$

Proof: For convenience, the proof is also provided in the Appendix. ■

There is also a heuristic explanation for beamforming to be optimal when N_t goes to infinity. By Lemma 1, $\|\hat{\boldsymbol{h}}\|$ converges to infinity almost surely. When $\|\hat{\boldsymbol{h}}\|$ goes to infinity, it is clear that the left hand side of Equation (15) converges to 0 and the sufficient condition for beamforming to be optimal is always satisfied. From *Theorem* 2, if the number of antennas at the transmitter is large enough, beamforming to the direction of $\hat{\boldsymbol{h}}$ is optimal. The maximum can be also expressed as the upper bound yielded by Jensen's equality. Moreover, the following proposition shows that the maximum achievable rate scales with $\log(N_t)$.

Proposition 1: As $N_t \to \infty$, the expectation of the maximum achievable rate $\bar{R}_s$, defined in (14), can be expressed as:

$$\lim_{N_t\to\infty}\left\{\bar{R}_s - \log(N_t) - \left[\log(P(1-\alpha)\,\sigma_h^2) - \mathbf{E}_{g_1}\left[\log(1+P|g_1|^2)\right]\right]\right\} = 0. \tag{33}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

Proof: According to *Theorem* 2, we have:

$$\lim_{N_t\to\infty}\left\{\bar{R}_s-\left[\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log(1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\,\sigma_h^2))\right]-\mathbf{E}_{g_1}\left[\log\left(1+P|g_1|^2\right)\right]\right]\right\}=0 \tag{34}$$

$$\Longrightarrow \lim_{N_t\to\infty}\left\{\bar{R}_s-\log(N_t)+\mathbf{E}_{g_1}\left[\log\left(1+P|g_1|^2\right)\right]-\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log\left(1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\,\sigma_h^2)\right)-\log(N_t)\right]\right\}=0 \tag{35}$$

$$\Longrightarrow \lim_{N_t\to\infty}\left\{\bar{R}_s-\log(N_t)+\mathbf{E}_{g_1}\left[\log\left(1+P|g_1|^2\right)\right]-\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log\frac{1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\,\sigma_h^2)}{N_t}\right]\right\}=0 \tag{36}$$

$$\Longrightarrow \lim_{N_t\to\infty}\left\{\bar{R}_s-\log(N_t)+\mathbf{E}_{g_1}\left[\log\left(1+P|g_1|^2\right)\right]\right\}=\lim_{N_t\to\infty}\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log\frac{1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\,\sigma_h^2)}{N_t}\right]. \tag{37}$$

We will show that the limit on the RHS of (37) exists and is equal to $\log(P(1-\alpha))+\log\left(\sigma_h^2\right)$. Firstly, we have:

$$\lim_{N_t\to\infty}\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log\frac{1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\,\sigma_h^2)}{N_t}\right] \geq \lim_{N_t\to\infty}\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log\frac{P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2)}{N_t}\right] \tag{38}$$

$$=\log(P(1-\alpha))+\lim_{N_t\to\infty}\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log(\frac{\|\hat{\boldsymbol{h}}\|^2}{N_t})\right]. \tag{39}$$

Since $\frac{\|\hat{\boldsymbol{h}}\|^2}{\sigma_h^2}$ has a chi-square distribution with $2N_t$ degrees of freedom, then it can be verified that:

$$\mathbf{E}_{\hat{\boldsymbol{h}}}[\log(\frac{\|\hat{\boldsymbol{h}}\|^2}{\sigma_h^2\,N_t})]=\log(\frac{1}{N_t})+\psi(N_t), \tag{40}$$

where $\psi(z)$ denotes the digamma function $\psi(z)=\frac{\Gamma'(z)}{\Gamma(z)}$. According to [32]:

$$\lim_{N_t\to\infty}\{\log(\frac{1}{N_t})+\psi(N_t)\}=0. \tag{41}$$

Combining (39) and (41), we have:

$$\lim_{N_t\to\infty}\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log\frac{1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\sigma_h^2)}{N_t}\right] \geq\log(P(1-\alpha))+\log\left(\sigma_h^2\right). \tag{42}$$

On the other hand, by Jensen's inequality, we have:

$$\lim_{N_t\to\infty}\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log\frac{1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\,\sigma_h^2)}{N_t}\right] \leq \lim_{N_t\to\infty}\left\{\log\frac{1+P((1-\alpha)\mathbf{E}_{\hat{\boldsymbol{h}}}[\|\hat{\boldsymbol{h}}\|^2]+\alpha\,\sigma_h^2)}{N_t}\right\} \tag{43}$$

$$=\lim_{N_t\to\infty}\log\left(\frac{1+P((1-\alpha)N_t\sigma_h^2+\alpha\sigma_h^2)}{N_t}\right) \tag{44}$$

$$=\log\left(P(1-\alpha)\,\sigma_h^2\right). \tag{45}$$

Combining again (42) and (45), we have:

$$\lim_{N_t\to\infty}\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log\frac{1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\,\sigma_h^2)}{N_t}\right] =\log\left(P(1-\alpha)\,\sigma_h^2\right). \tag{46}$$

■

Proposition 1 also highlights the impact of channel estimation errors through α. Should α be equal to 0, one obtains directly from (33) that $\lim_{N_t\to\infty}\{\bar{R}_s-\log(N_t)-\log(P(1-\alpha)\sigma_h^2)+\mathbf{E}_{g_1}[\log(1+P|g_1|^2)]\}=0$. Moreover, if we treat $\bar{R}_s$ as a function of α, i.e., $\bar{R}_s(\alpha)$, we can also derive from (33) that,

$$\lim_{N_t\to\infty}\{\bar{R}_s(\alpha)-\bar{R}_s(0)\}=\log(1-\alpha). \tag{47}$$

Note that (47) characterizes the penalty due to CSI error. For example, if $\alpha=0.5$, there should be a $|\log(0.5)|\approx 0.7$ npcu gap between the achievable rates with $\alpha=0.5$ and $\alpha=0$ as the number of transmit antennas goes to infinity. This result is also demonstrated numerically in Figure 7 of Section V.

IV. ACHIEVABLE RATE BASED ON SNR CRITERIONS

In this section, apart from the estimation of main channel, the transmitter is also provided with an estimation of the eavesdropper's channel. Thus the channel to the eavesdropper can be expressed as:

$$\boldsymbol{g}=\sqrt{1-\beta}\hat{\boldsymbol{g}}+\sqrt{\beta}\tilde{\boldsymbol{g}}, \tag{48}$$

where $\hat{\boldsymbol{g}}$ is the estimate of the eavesdropper's channel which follows $\mathcal{CN}(0,I)$. $\tilde{\boldsymbol{g}}$ represents the estimation error and $\tilde{\boldsymbol{g}}\sim\mathcal{CN}(0,I)$. β is the error variance. Similar to the main channel, we assume that $\hat{\boldsymbol{g}}$ and $\tilde{\boldsymbol{g}}$ are independent of each other. Furthermore, the transmitter would also broadcast its estimate of the eavesdropper's channel and this estimate is known to all parties. In our analysis, we mainly focus on the SNR as the measure of performance. Although there are differences between SNR and mutual information criterion in a sense that SNR characterizes the performance of uncoded systems and mutual information measures the maximum rate achieved by coded systems, these two metrics are still highly related. Moreover, in some cases, the optimization involving mutual information is far less tractable than that with respect to SNR.

In the transmission with secrecy constraint, the signal can be received by two parties: the legitimate receiver and the eavesdropper. An appealing SNR criterion is the one indicating the input covariance matrix that can enlarge the expected received SNR of the legitimate receiver while putting the eavesdropper at a disadvantage. Recall that the rate in (6) is achievable for any positive definite Q satisfying $\mathrm{tr}(Q)=P$, we are free to choose Q such that it either maximizes the difference of the expected SNR at the legitimate receiver and the one at the eavesdropper, or their ratio. Two SNR criterions

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

are introduced here, the first one is the SNR difference maximization criterion:

$$\begin{aligned} Q_d = & \quad \text{argmax}_Q\{\mathbf{E}[SNR_r] - \mathbf{E}[SNR_e]\} && (49)\\ = & \quad \text{argmax}_Q\Big\{\mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}\left[\boldsymbol{h}^H Q\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}\right] \\ & -\mathbf{E}_{\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}\left[\boldsymbol{g}^H Q\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}\right]\Big\}. && (50)\end{aligned}$$

The second one is the SNR ratio maximization criterion:

$$\begin{aligned} Q_r &= \text{argmax}_Q \frac{\mathbf{E}[SNR_r]+1}{\mathbf{E}[SNR_e]+1} && (51)\\ &= \text{argmax}_Q \frac{\mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}[\boldsymbol{h}^H Q\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}]+1}{\mathbf{E}_{\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}[\boldsymbol{g}^H Q\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}]+1}, && (52)\end{aligned}$$

where the SNR_r and SNR_e represent the received SNR of the legitimate receiver and the eavesdropper respectively. In (51), we add 1 in the nominator and denominator of the SNR ratio criterion in order to ensure that our result is consistent with the result in [33] in the perfect feedback case, i.e., when both α and β are equal to 0.

A. The SNR Difference Criterion

Here, we want to obtain the covariance matrix Q that maximizes the received SNR difference between the legitimate receiver and the eavesdropper. The optimization function can be written as:

$$\begin{cases} \max_Q S(Q,\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}) & \triangleq \mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}\left[\boldsymbol{h}^H Q\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}\right] \\ & \quad -\mathbf{E}_{\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}\left[\boldsymbol{g}^H Q\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}\right] \\ \text{s.t.} \quad \text{tr}(Q) & = P. \end{cases} \quad (53)$$

It turns out that beamforming is once again optimal as formalized in *Proposition* 2.

Proposition 2: The optimal solution to (53) is given by Q^* with $Q^* = P\boldsymbol{w}_0\boldsymbol{w}_0^H$, where $\boldsymbol{w}_0$ is the eigenvector corresponding to the largest eigenvalue of $[(1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H - (1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H]$, that is, the optimal transmission strategy based on the SNR difference criterion is beamforming to the direction of $\boldsymbol{w}_0$.

Proof: Let $U\Lambda U^H$ denote the spectral decomposition of Q, we rewrite the cost function as:

$$\begin{aligned} S(U,\Lambda,\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}) = & \quad \mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}[\boldsymbol{h}^H U\Lambda U^H\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}] \\ & -\mathbf{E}_{\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}[\boldsymbol{g}^H U\Lambda U^H\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}]. \quad (54)\end{aligned}$$

We firstly show that uni-rank Λ is optimal, that is, beamforming is the optimal transmission strategy based on the SNR difference criterion. Define $\boldsymbol{v} = U^H\boldsymbol{h}$, $\boldsymbol{w} = U^H\boldsymbol{g}$ and $\boldsymbol{m}$, $\boldsymbol{n}$ as the mean vectors of $\boldsymbol{v}$ and $\boldsymbol{w}$ respectively. For any given estimation of CSI of the main channel and the eavesdropper channel, $\boldsymbol{h}|\hat{\boldsymbol{h}} \sim \mathcal{CN}\left(\sqrt{1-\alpha}\hat{\boldsymbol{h}}, \alpha\,\sigma_h^2\, I\right)$ and $\boldsymbol{g}|\hat{\boldsymbol{g}} \sim \mathcal{CN}\left(\sqrt{1-\beta}\hat{\boldsymbol{g}}, \beta\, I\right)$. Because U is unitary matrix and $\text{Cov}(\boldsymbol{h}|\hat{\boldsymbol{h}}) = \alpha\,\sigma_h^2\, I$, $\text{Cov}(\boldsymbol{v}|\hat{\boldsymbol{h}}) = U^H\text{Cov}(\boldsymbol{h}|\hat{\boldsymbol{h}})U = U^H \times (\alpha\sigma_h^2 I) \times U = \alpha\sigma_h^2 U^H U = \alpha\sigma_h^2 I$, which indicates that the covariance of $\boldsymbol{v}$ is also white. Similarly, $\text{Cov}(\boldsymbol{w}|\hat{\boldsymbol{g}}) = \beta I$.

$$\begin{aligned} \boldsymbol{m} &= \sqrt{1-\alpha}U^H\hat{\boldsymbol{h}} && (55)\\ \boldsymbol{m}^H\boldsymbol{m} &= (1-\alpha)\hat{\boldsymbol{h}}^H UU^H\hat{\boldsymbol{h}} = (1-\alpha)\hat{\boldsymbol{h}}^H\hat{\boldsymbol{h}}. && (56)\end{aligned}$$

With the same reasoning, $\boldsymbol{n}^H\boldsymbol{n} = (1-\beta)\hat{\boldsymbol{g}}^H\hat{\boldsymbol{g}}$. For any given U, the optimization function can be expanded as follows:

$$\begin{aligned} & S(U,\Lambda,\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}) \\ =& \mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}\left[\boldsymbol{h}^H U\Lambda U^H\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}\right] - \mathbf{E}_{\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}\left[\boldsymbol{g}^H U\Lambda U^H\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}\right] && (57)\\ =& \mathbf{E}_{\boldsymbol{v}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}\left[\sum_i \lambda_i|v_i|^2\right] - \mathbf{E}_{\boldsymbol{w}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}\left[\sum_i \lambda_i|w_i|^2\right] && (58)\\ =& \sum_i \lambda_i\left(|m_i|^2 + \alpha\sigma_h^2\right) - \sum_i \lambda_i\left(|n_i|^2 + \beta\right) && (59)\\ =& \sum_i \lambda_i\left(|m_i|^2 - |n_i|^2\right) + P\left(\alpha\sigma_h^2 - \beta\right), && (60)\end{aligned}$$

where λ_i's are the diagonal elements of Λ, $\text{tr}(\Lambda) = \sum_i \lambda_i = \text{tr}(Q) = P$. Without loss of generality, we assume that there exists at least one i that $|m_i|^2 - |n_i|^2 > 0$. Thus, we can find a particular t that satisfies: $t = \text{argmax}_i\left(t|m_i|^2 - |n_i|^2\right)$. According to the above assumption, we have $t|m_t|^2 - |n_t|^2 > 0$. Thus, the optimization function can be maximized as follows:

$$\begin{aligned} & S(U,\Lambda,\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}) \\ \le& (\lambda_1 + \ldots + \lambda_{N_t})(|m_t|^2 - |n_t|^2) + P(\alpha\sigma_h^2 - \beta) && (61)\\ =& P(|m_t|^2 - |n_t|^2) + P(\alpha\sigma_h^2 - \beta). && (62)\end{aligned}$$

The equality can be achieved when $\lambda_t = P$ and $\lambda_i = 0$, $i \neq t$. This result shows that the best transmission strategy under the SNR difference criterion is beamforming. It only remains to find the optimal beamforming direction $\boldsymbol{w_0}$.

$$\begin{aligned} \max_{\boldsymbol{w}} \quad & (1-\alpha)|\hat{\boldsymbol{h}}^H\boldsymbol{w}|^2 - (1-\beta)|\hat{\boldsymbol{g}}^H\boldsymbol{w}|^2 \\ \text{s.t.} \quad & \boldsymbol{w}^H\boldsymbol{w} = 1. \end{aligned} \quad (63)$$

We apply Lagrange multiplier method to get the solution. Let $f(\boldsymbol{w},\lambda) = (1-\alpha)|\hat{\boldsymbol{h}}^H\boldsymbol{w}|^2 - (1-\beta)|\hat{\boldsymbol{g}}^H\boldsymbol{w}|^2 - \lambda(|\boldsymbol{w}|^2 - 1)$ and set partial derivatives with respect to $\boldsymbol{w}$ and λ to zero:

$$\begin{cases} \langle\boldsymbol{w},\boldsymbol{w}\rangle = 1 \\ 2(1-\alpha)\left\langle\hat{\boldsymbol{h}},\boldsymbol{w}\right\rangle\hat{\boldsymbol{h}} - 2(1-\beta)\langle\hat{\boldsymbol{g}},\boldsymbol{w}\rangle\hat{\boldsymbol{g}} - 2\lambda\boldsymbol{w} = 0. \end{cases} \quad (64)$$

The second equation is equivalent to:

$$((1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H - (1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H)\boldsymbol{w} = \lambda\boldsymbol{w}. \quad (65)$$

Hence, $\boldsymbol{w}$ which maximizes (63) should be an eigenvector of $[(1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H - (1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H]$ with respect to the eigenvalue λ. Subsequently, (63) can be expressed as:

$$\begin{aligned} & (1-\alpha)|\hat{\boldsymbol{h}}^H\boldsymbol{w}|^2 - (1-\beta)|\hat{\boldsymbol{g}}^H\boldsymbol{w}|^2 \\ =& \boldsymbol{w}^H((1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H - (1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H)\boldsymbol{w} && (66)\\ =& \boldsymbol{w}^H\lambda\boldsymbol{w} = \lambda. && (67)\end{aligned}$$

Thus, in order to maximize expression (63), $\boldsymbol{w}$ is the eigenvector with respect to the largest eigenvalue of $[(1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H - (1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H]$. In summary, under the SNR difference maximization criterion, the optimal transmission strategy is to beamform to the direction of eigenvector with respect to the largest eigenvalue of $[(1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H - (1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H]$. ■

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

B. The SNR Ratio Criterion

In this section, our main objective is to find the optimal covariance matrix Q maximizing the ratio of received SNR of the legitimate receiver and the eavesdropper. The optimization problem can be expressed as follows:

$$\begin{array}{rcl} \max\limits_{Q} \quad R(Q,\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}) & \triangleq & \frac{\mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}(\boldsymbol{h}^H Q \boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}})+1}{\mathbf{E}_{\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}(\boldsymbol{g}^H Q \boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}})+1} \\ \text{s.t.} \quad \text{tr}(Q) & = & P. \end{array} \tag{68}$$

Once again, beamforming is the right choice.

Proposition 3: The optimal solution to (68) is given by $Q^* = P\boldsymbol{w}_0\boldsymbol{w}_0^H$, where $\boldsymbol{w}_0$ is the generalized eigenvector corresponding to the largest generalized eigenvalue of

$$(P((1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H+\alpha\sigma_h^2 I)+I, P((1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H+\beta I)+I). \tag{69}$$

The optimal transmission strategy based on the SNR ratio criterion is beamforming to the direction of $\boldsymbol{w}_0$.

Proof: Let $Q = U\Lambda U^H$ and rewrite the optimization function:

$$R(U,\Lambda,\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}) = \frac{\mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}(\boldsymbol{h}^H U\Lambda U^H \boldsymbol{h}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}})+1}{\mathbf{E}_{\boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}}(\boldsymbol{g}^H U\Lambda U^H \boldsymbol{g}|\hat{\boldsymbol{h}},\hat{\boldsymbol{g}})+1} \tag{70}$$

$$= \frac{\sum_i \lambda_i \mathbf{E}[|v_i|^2]+1}{\sum_i \lambda_i \mathbf{E}[|w_i|^2]+1} \tag{71}$$

$$= \frac{\sum_i \lambda_i(|m_i|^2+\alpha\sigma_h^2)+\frac{1}{P}(\sum_i\lambda_i)}{\sum_i \lambda_i(|n_i|^2+\beta)+\frac{1}{P}(\sum_i\lambda_i)} \tag{72}$$

$$= \frac{\sum_i \lambda_i(|m_i|^2+\alpha\sigma_h^2+\frac{1}{P})}{\sum_i \lambda_i(|n_i|^2+\beta+\frac{1}{P})}. \tag{73}$$

Here the v, w, m, n, have the same meaning as the symbols in subsection 4.1. We note that for 4 positive real number, a, b, c, d, if $\frac{a}{b} \geq \frac{c}{d}$, then $\frac{c}{d} \leq \frac{a+c}{b+d} \leq \frac{a}{b}$. The proof is trivial and only needs several steps of basic transformation. Now, define $t = \text{argmax}_i \frac{|m_i|^2+\alpha\sigma_h^2+\frac{1}{P}}{|n_i|^2+\beta+\frac{1}{P}}$, in addition to the result mentioned above, we have:

$$R(U,\Lambda,\hat{\boldsymbol{h}},\hat{\boldsymbol{g}}) = \frac{\sum_i \lambda_i(|m_i|^2+\alpha\sigma_h^2+\frac{1}{P})}{\sum_i \lambda_i(|n_i|^2+\beta+\frac{1}{P})} \tag{74}$$

$$\leq \frac{P(|m_t|^2+\alpha\sigma_h^2+\frac{1}{P})}{P(|n_t|^2+\beta+\frac{1}{P})} \tag{75}$$

$$= \frac{P(|m_t|^2+\alpha\sigma_h^2)+1}{P(|n_t|^2+\beta)+1}. \tag{76}$$

The equality is achieved if $\lambda_t = P$ and $\lambda_i = 0$ for $i \neq t$. So the beamforming is also the best transmission strategy under the SNR ratio maximization criterion. Again it remains to figure out the optimal beamforming direction $\boldsymbol{w_0}$. We note that $m_t = \sqrt{1-\alpha}\hat{\boldsymbol{h}}^{\boldsymbol{H}}\boldsymbol{w}$ and $n_t = \sqrt{1-\beta}\hat{\boldsymbol{g}}^{\boldsymbol{H}}\boldsymbol{w}$, $\boldsymbol{w}$ is a column of unitary matrix U with $\boldsymbol{w}^H\boldsymbol{w} = 1$. The original optimization problem is simplified to:

$$\begin{array}{rc} \max\limits_{\boldsymbol{w}} & \frac{P(|\sqrt{1-\alpha}\hat{\boldsymbol{h}}^H\boldsymbol{w}|^2+\alpha\sigma_h^2)+1}{P(|\sqrt{1-\beta}\hat{\boldsymbol{g}}^H\boldsymbol{w}|^2+\beta)+1} \\ \text{s.t.} & \boldsymbol{w}^H\boldsymbol{w} = 1. \end{array} \tag{77}$$

Then, after transforming expression (77) to the form of Rayleigh quotient, we obtain:

$$\max_{\boldsymbol{w}} \frac{\boldsymbol{w}^H[P((1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H+\alpha\sigma_h^2 I)+I]\boldsymbol{w}}{\boldsymbol{w}^H[P((1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H+\beta I)+I]\boldsymbol{w}}. \tag{78}$$

According to definition 1 and fact 1 in [33], for a Hermitian matrix $A \in \mathbb{C}^{n\times n}$ and positive definite matrix $B \in \mathbb{C}^{n\times n}$. The generalized eigenvector of (A,B) are the stationary point solution to a particular Rayleigh quotient. Specially, the largest generalized eigenvalue is the maximum of the Rayleigh quotient

$$\lambda_{max}(A,B) = \max_{\boldsymbol{\psi}\in\mathbb{C}^n} \frac{\boldsymbol{\psi}^H A\boldsymbol{\psi}}{\boldsymbol{\psi}^H B\boldsymbol{\psi}}, \tag{79}$$

and the optimum is attained by the eigenvector with respect to the largest eigenvalue. It is easy to verify that the matrix $[P((1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H+\alpha\sigma_h^2 I)+I]$ is a Hermitian matrix. For any non-zero vector $\boldsymbol{x} \in \mathbb{C}^n$:

$$\boldsymbol{x}^H[P((1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H+\beta I)+I]\boldsymbol{x}$$

$$= P(1-\beta)|\hat{\boldsymbol{g}}^H\boldsymbol{x}|^2 + (P\beta+1)|x|^2 \tag{80}$$

$$> 0. \tag{81}$$

Hence, the matrix $[P((1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H+\beta I)+I]$ is positive definite. With the result of [33], expression (78) is optimized if $\boldsymbol{w}$ is the generalized eigenvector corresponding to the largest generalized eigenvalue of

$$(P((1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H+\alpha\sigma_h^2 I)+I, P((1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H+\beta I)+I). \tag{82}$$

Thus the optimal transmission strategy under the SNR ratio criterion is beamforming to the direction of the eigenvector corresponding to the largest generalized eigenvalue of $(P((1-\alpha)\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H+\alpha\sigma_h^2 I)+I, P((1-\beta)\hat{\boldsymbol{g}}\hat{\boldsymbol{g}}^H+\beta I)+I)$. ■

Remark: if the values of α and β are set to 0, the optimal transmission strategy becomes beamforming to the direction of the generalized eigenvector corresponding to the largest generalized eigenvalue of $(P\boldsymbol{h}\boldsymbol{h}^H+I, P\boldsymbol{g}\boldsymbol{g}^H+I)$, which is consistent with the result in [33].

C. Special Case When $\beta = 1$

Here we discuss the special case when $\beta = 1$, which indicates that there is only an estimate of the main channel at the transmitter. We will show that both the SNR difference and the SNR ratio criterions imply that the optimal transmission strategy is beamforming to the direction of $\hat{\boldsymbol{h}}$. With β equal to 1, we have:

$$\mathbf{E}_{\boldsymbol{g}|\hat{\boldsymbol{h}}}[\boldsymbol{g}^H Q\boldsymbol{g}|\hat{\boldsymbol{h}}] = \mathbf{E}_{\boldsymbol{g}|\hat{\boldsymbol{h}}}[\boldsymbol{g}^H U\Lambda U^H\boldsymbol{g}|\hat{\boldsymbol{h}}] \tag{83}$$

$$= \mathbf{E}_{\boldsymbol{g}|\hat{\boldsymbol{h}}}[\boldsymbol{g}^H\Lambda\boldsymbol{g}|\hat{\boldsymbol{h}}] = \mathbf{E}[\Sigma\lambda_i|g_i|^2] \tag{84}$$

$$= \Sigma\lambda_i\mathbf{E}|g_i|^2 = P. \tag{85}$$

From equation (85), it is clear that the expected SNR of the eavesdropper remains constant regardless of the choice of Q. Thus, we only need to maximize the expected SNR of the legitimate channel in order to maximize the SNR difference or the ratio between the legitimate receiver and the eavesdropper.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

As a result, both the SNR difference and the ratio criterions can be reduced to the following form:

$$\begin{array}{ll} \max\limits_{Q} & \mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}[\boldsymbol{h}^H Q \boldsymbol{h}|\hat{\boldsymbol{h}}] \\ \text{s.t.} & \text{tr}(Q) = P. \end{array} \tag{86}$$

Define $\boldsymbol{v} = U^H \boldsymbol{h}$. Then, we have:

$$\mathbf{E}_{\boldsymbol{h}|\hat{\boldsymbol{h}}}\mathbf{E}[\boldsymbol{h}^H Q \boldsymbol{h}|\hat{\boldsymbol{h}}] = \Sigma_i \lambda_i |m_i|^2 + P\alpha\sigma_h^2. \tag{87}$$

Suppose $t = \max_i |m_i|^2$, then:

$$\Sigma_i \lambda_i |m_i|^2 \leq (\Sigma_i \lambda_i)|m_t|^2 \leq P(1-\alpha)\|\hat{\boldsymbol{h}}\|^2. \tag{88}$$

The equality is achieved with rank-one Λ^o with $\text{diag}(\Lambda^o) = [P, 0..., 0]$, and U^o whose first column is $\frac{\hat{\boldsymbol{h}}}{\|\hat{\boldsymbol{h}}\|}$. Therefore, when no estimate of the eavesdropper's channel is available at the transmitter, the optimal transmission strategy based on the SNR difference and the ratio criterions is beamforming to the direction of $\hat{\boldsymbol{h}}$.

V. NUMERICAL RESULTS

In this section, we present selected numerical results. All the achievable rates have been obtained using Monte Carlo methods by averaging over 100 of channel realizations $\hat{\boldsymbol{h}}$. First, we compare the AN-BF secrecy rate in (8) and the BF-only secrecy rate in (6) numerically. In all our curves, the unit of the transmission rate is in bits per channel use (bpcu). In order to simplify the optimization problem related to (8), we choose the first column of $\phi_s(\hat{\boldsymbol{h}})$ and $\phi_a(\hat{\boldsymbol{h}})$ to be $\frac{\hat{\boldsymbol{h}}}{\|\hat{\boldsymbol{h}}\|}$ and the other columns orthogonal to the first one. In Fig. 1, both the optimal secrecy rates with and without artificial noise were found by an exhaustive search. As shown in Fig. 1, AN-BF expectedly outperforms BF-only at the cost of more complexity in finding the structure of the optimal covariance matrices. However, as argued above, the focus of this paper is on a BF-only scheme as it is simpler, provides more insights into the problem via finding analytically optimal transmission strategies in various important configurations.

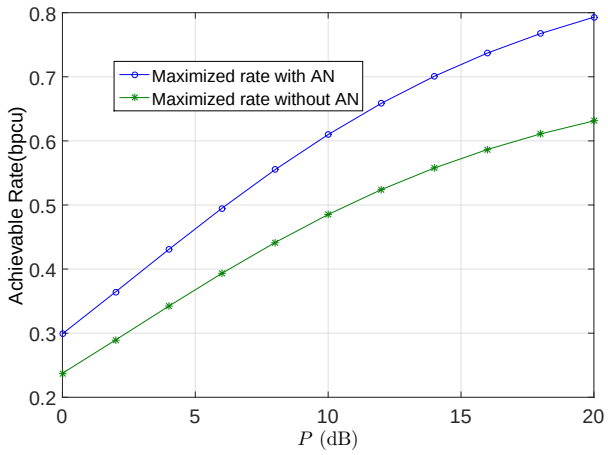


Figure 1: Secrecy rates with and without AN with $N_t = 2$, $\sigma_h^2 = 1$ and $\alpha = 0.5$.

Fig. 2 displays the sufficient condition for optimality of beamforming given in *Theorem* 1. The optimality of beamforming depends only on the main channel SNR ($P\,\alpha\,\sigma_h^2$) and the accuracy of channel estimate ($\frac{1-\alpha}{\alpha\,\sigma_h^2}\|\hat{\boldsymbol{h}}\|^2$). As can be seen in Fig. 2, for perfect main CSI at the transmitter ($\frac{1-\alpha}{\alpha\,\sigma_h^2}\|\hat{\boldsymbol{h}}\|^2 \to \infty$), the optimal strategy is beamforming. From Fig. 2, we note that for a given channel SNR, as the quality of estimate improves, beamforming becomes optimal.

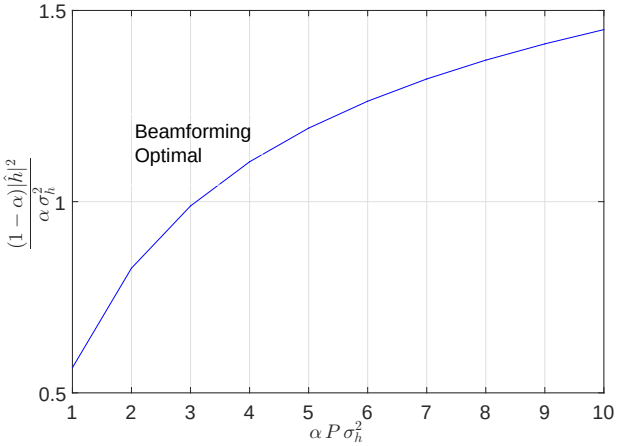


Figure 2: Sufficient condition for optimality of beamforming.

One may be curious to know how often the beamforming condition in *Theorem* 1 is satisfied. For this purpose, we have depicted in Fig. 3 below the probability that the beamforming condition holds true versus the power P in dB, for different α values. As can be seen from Fig. 3, the probability for the sufficient condition to hold is relatively high for any P value when the channel estimation quality is relatively good (Figs. 3.a, 3.b). However, the probability for the sufficient condition to hold decreases drastically with P when the channel estimation quality is relatively poor (Fig. 3.d).

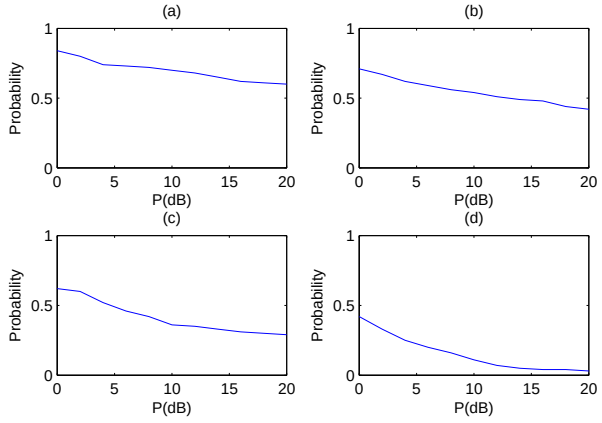


Figure 3: Probability for the beamforming condition in *Theorem* 1 to be satisfied versus P, for $\sigma_h^2 = 1$ and with different values of α: (a) $\alpha = 0.2$, (b) $\alpha = 0.3$, (c) $\alpha = 0.4$ and (d) $\alpha = 0.6$.

Fig. 4 depicts the maximum achievable rate when beamforming is optimal. In our simulation, the number of antennas N_t is equal to 2. $\|\hat{\boldsymbol{h}}\|$ is set to 1.5. The pairs $\left(\alpha, \sigma_h^2\right)$ in this

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

figure are chosen such that the sufficient condition in *Theorem* 1 is satisfied and beamforming is optimal. Furthermore, the maximum instantaneous rate can be computed using (18). As can be seen in Fig. 4, it may happen that the instantaneous rate is not increasing with P and it may be even negative (blue curve).

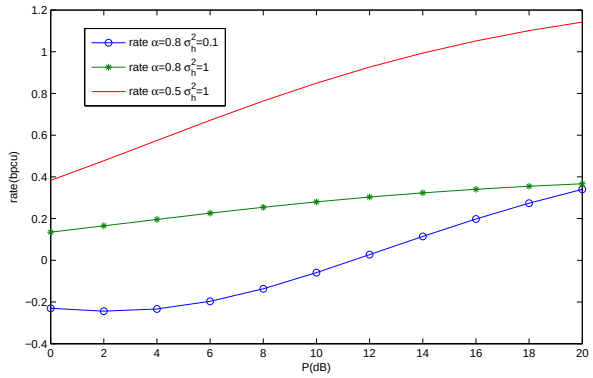


Figure 4: Instantaneous rate when beamforming is optimal, for different $\left(\alpha, \sigma_h^2\right)$ pair values, and a given $\hat{\boldsymbol{h}}$ with $\|\hat{\boldsymbol{h}}\| = 1.5$.

Fig. 5 displays the secrecy rate versus P for different values of the main channel variance σ_h^2. As shown in Fig. 5, the secrecy rate improves substantially with the main channel variance. For instance, when σ_h^2 increases from 0.5 to 1 and from 1 to 2, the secrecy rates roughly triples and doubles, respectively, when P is large enough.

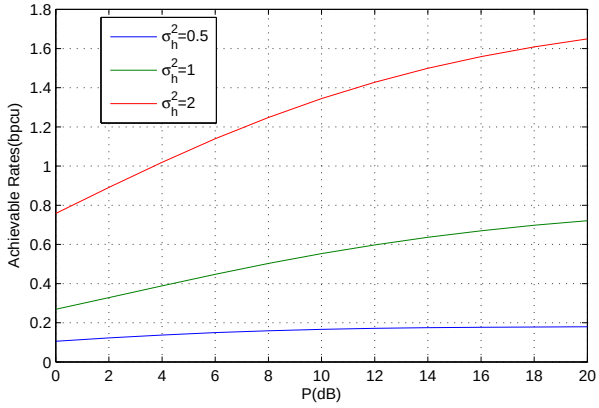


Figure 5: Achievable rates for different σ_h with $N_t = 2$, $\alpha = 0.5$ and $\beta = 1$

In order to assess the effect of adding more antennas on the secrecy rate, we present in Fig. 6 the performance for $N_t = 3$. Again, the secrecy rates improves substantially with σ_h^2.

Fig. 7 describes the convergence of the lower bound and the upper bound on the maximum achievable rate in (133). The maximum achievable rate given by (14) when $\alpha = 0$ is also included in Fig. 7 as a benchmark. In our simulation, the power P is set to 17 dB. The expectation over $\hat{h}$ is computed using Monte Carlo methods. As indicated in Fig. 7, the upper bound and the lower bound converge as the number of antennas converges to infinity. As a result, the maximum

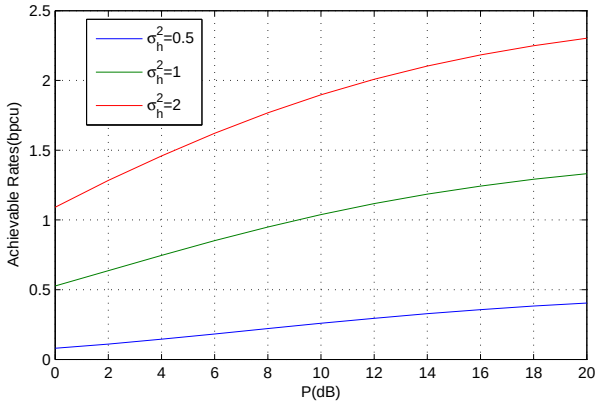


Figure 6: Achievable Rates for different σ_h with $N_t = 3$, $\alpha = 0.5$ and $\beta = 1$

achievable rate is fully characterized in this regime. Moreover, the gap between the maximum achievable rates with $\alpha = 0$ and $\alpha = 0.5$ is about 0.7 npcu, which is consistent with our analytical result after *Proposition* 1.

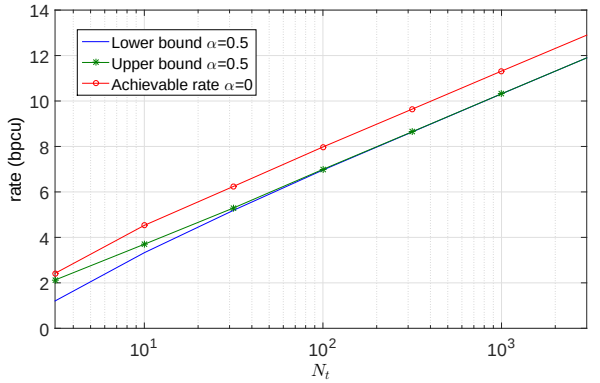


Figure 7: The upper bound and lower bound of the maximum achievable rate, with $\sigma_h^2 = 1$.

In order to depict how the secrecy rate varies with the estimation error variances of the main channel and the eavesdropper's channel, Fig. 8 and Fig. 9 display the secrecy rate given by (14) versus α and β, respectively. In both these figures, the number of transmit antenna is set to $N_t = 2$. Both these figures confirm that the achievable secrecy rate decreases as α and β increase.

To assess the impact of the main channel strength in comparison with the eavesdropper's channel strength, Fig. 10 displays the secrecy rate for large antenna systems for different σ_h^2 values. As can be seen in Fig. 10, even when the main channel is worst that the eavesdropper's channel on average ($\sigma_h^2 = \frac{1}{2} < \sigma_g^2 = 1$, where σ_g^2 is the eavesdropper's channel variance), a positive secrecy rate is achievable. Notice that the three curves in Fig. 10 are parallel as predicted by *Proposition* 1.

Figs. 11 to 15 display the performances of the SNR ratio and SNR difference criterions for different power constraints and error variances. For a particular power constraint and a particular error variance, we derive the optimal input covari-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

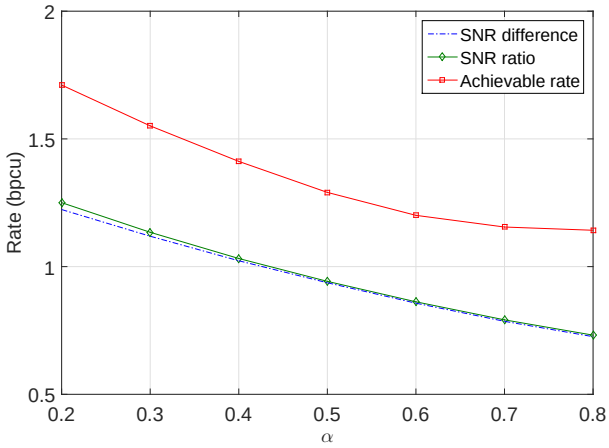


Figure 8: Achievable secrecy rate versus α, for $P = 10$ dB, $\sigma_h^2 = 1$ and $\beta = 0.5$.

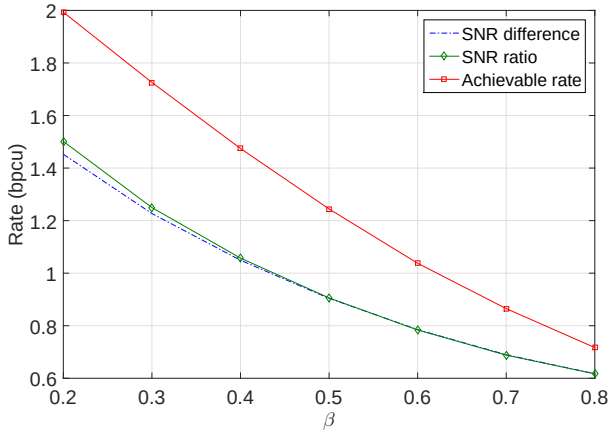


Figure 9: Achievable secrecy rate versus β, for $P = 10$ dB, $\sigma_h^2 = 1$ and $\alpha = 0.5$.

ance matrix under the SNR ratio and difference criterions respectively according to the results in Section IV, then apply the derived covariance matrix to equation (6) to get the transmission rate. The maximum achievable rates are derived by solving (14) through an exhaustive numerical search. The number of antennas N_t in all these figures is equal to 2.

Fig. 11 depicts the performance of the SNR difference and the SNR ratio criterions in the low SNR regime, when $\alpha = 0.3$ and $\beta = 0.2$. We can see from Fig. 11 that the transmission rates achieved by both the SNR ratio and difference criterions are close to the achievable rate given by (14). Another interesting feature shown in Fig. 11 is that the SNR difference and the SNR ratio criterions have almost the same performance in low SNR regime. This could be explained by the following:

$$
\begin{aligned}
Q_o = \ & \text{argmax}_Q \left\{ (\mathbf{E}[SNR_r] + 1)(\mathbf{E}[SNR_e] + 1)^{-1} \right\} && (89)\\
\approx \ & \text{argmax}_Q \left\{ (\mathbf{E}[SNR_r] + 1)(1 - \mathbf{E}[SNR_e]) \right\} && (90)\\
= \ & \text{argmax}_Q \Big\{ (1 + \mathbf{E}[SNR_r] - \mathbf{E}[SNR_e] \\
& \qquad + \mathbf{E}[SNR_r]\mathbf{E}[SNR_e] \Big\} && (91)\\
\approx \ & \text{argmax}_Q \left\{ \mathbf{E}[SNR_r] - \mathbf{E}[SNR_e] \right\}. && (92)
\end{aligned}
$$

From equation (92), the optimal covariance derived by the SNR ratio and the SNR difference criterions turn out to be the same in low SNR regime.

Fig. 12 targets medium to high power regimes. Compared with Fig. 11, the gaps between the maximum achievable rate given by (14) and the transmission rates based on SNR criterions are further enlarged.

Fig. 13 shows the performance in case of low error variance of both the legitimate channel and the eavesdropper channel. Both α and β are set to 0.02. The result shows that the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

transmission rate determined by the SNR ratio criterion can approach quite closely the maximum achievable rate given by (14). However, as the power increases, even the low error variance cannot guarantee satisfactory performance for the SNR ratio criterion, and there exists a 0.8 npcu gap between the rate based on the SNR ratio criterion and the maximum achievable rate as the power reaches 20 dB.

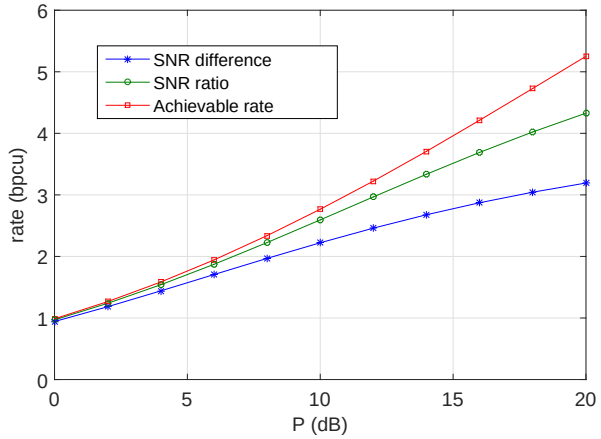


Figure 13: The transmission rates based on the SNR difference and the SNR ratio criterions, in case of low error variances, $\alpha = 0.02$ and $\beta = 0.02$. Here, $\sigma_h^2 = 1$.

Fig. 14 depicts the performance of the SNR ratio and difference criterions at high SNR. From Fig. 14, it can be inferred that a power increase may enlarge the gaps between the maximum achievable rate given by (14) and the transmission rates based on the two SNR criterions. As indicated in Fig. 14, the transmission rates achieved by the two criterions cannot match the maximum achievable rate any more. This can be explained by the fact that for both the SNR difference and SNR ratio maximization criterions, beamforming is the optimal strategy. However, our exhaustive search in evaluating the maximum achievable rate revealed that beamforming is actually suboptimal and that exploiting all directions might be optimal.

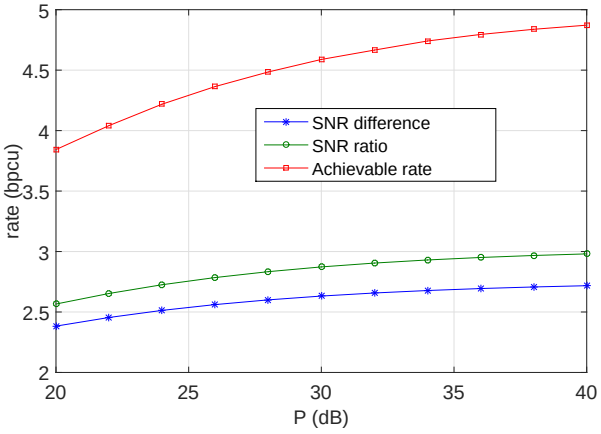


Figure 14: The transmission rates based on the SNR difference and the SNR ratio criterions at high SNR, with $\alpha = 0.2$ and $\beta = 0.2$. Here, $\sigma_h^2 = 1$.

Fig. 15 depicts the performance of the SNR ratio and SNR difference criterions when $\beta = 1$, i.e., no knowledge of the eavesdropper's channel is available at the transmitter. Furthermore, the maximum achievable rate given by (14) when $\alpha = 0$ is also presented in Fig. 15. According to Section 4.3, beamforming to the direction of $\hat{\boldsymbol{h}}$ is optimal based on both the SNR difference and the SNR ratio criterions. From the simulation result, we can see that beamforming is also optimal to achieve the maximum achievable rate, so that the curve representing the maximum achievable rate overlaps with the curves representing the rates based on SNR criterions.

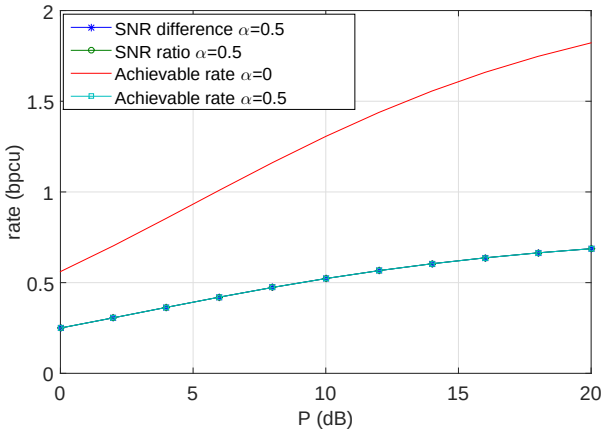


Figure 15: The transmission rates based on SNR difference and SNR ratio criterions, with $\beta = 1$, for $\alpha = 0.5$. Here, $\sigma_h^2 = 1$.

VI. CONCLUSION

We have considered secret transmission in fading MISO channel with imperfect channel state information at the transmitter. A sufficient condition is provided for beamforming to be optimal. Moreover, when the number of transmit antennas is very large, beamforming is also optimal, the achievable rate takes a simple closed form and scales with the logarithm of the number of transmit antennas.

Furthermore, when the transmitter is also provided with an estimate of the eavesdropper's channel, the SNR difference and the SNR ratio criterions are proposed in order to provide suboptimal transmission strategies. We have found numerically that the performance of these two criterions is satisfactory in the low SNR regime in the sense that the corresponding achievable rates are asymptotically (in the low SNR regime) optimal. However, the performance of the SNR difference and SNR ratio criterions degrades as the transmit power increases. Furthermore, the performance of the SNR ratio criterion also depends on the channel estimate error variance; smaller is this variance, better is its performance.

APPENDIX
PROOF OF LEMMA 2

Let $Y = \log(1 + |v|^2)$. Note that $Y \geq 0$ by definition, by Markov's inequality, for any $a > 0$:

$$\Pr(Y > a) \leq \frac{\mathbf{E}[Y]}{a} \implies \mathbf{E}[Y] \geq a \Pr(Y > a). \qquad (93)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

On the other hand, by Jensen's inequality, we have:

$$\mathbf{E}[Y] \leq \log\left(1+\mathbf{E}[|v|^2]\right) = \log(1+m^2+\sigma^2). \quad (94)$$

Combining these inequalities, we obtain that for any $a>0$:

$$a\ \Pr(Y>a) \leq \mathbf{E}[Y] \leq \log(1+m^2+\sigma^2). \quad (95)$$

First, note that

$$\Pr(Y>a) = \Pr(|v|^2 > e^a-1) = \Pr(|v| > \sqrt{e^a-1}), \quad (96)$$

which can be low bounded as follows:

$$\begin{aligned}\Pr(Y>a) =& \Pr(|v|>\sqrt{e^a-1}) && (97)\\ \geq& \Pr(|\Re(v)|>\sqrt{e^a-1}) && (98)\\ =& 1-\Phi(\frac{\sqrt{2}(\sqrt{e^a-1}-m)}{\sigma}) \\ &+\Phi(-\frac{\sqrt{2}(\sqrt{e^a-1}+m)}{\sigma}). && (99)\end{aligned}$$

Thus, for any $a>0$, we have:

$$\begin{aligned}&a\left(1-\Phi(\frac{\sqrt{2}(\sqrt{e^a-1}-m)}{\sigma})+\Phi(-\frac{\sqrt{2}(\sqrt{e^a-1}+m)}{\sigma})\right)\\ \leq&\mathbf{E}[Y] && (100)\\ \leq&\log(1+m^2+\sigma^2). && (101)\end{aligned}$$

Now, let $a=k\log(1+m^2+\sigma^2)$, where $k=1-\frac{1}{\log^2 m}$. Then it follows from (99) that

$$\begin{aligned}&a\ \Pr(Y>a)\\ \geq& k\ \log(1+m^2+\sigma^2)\\ &\cdot\left[1-\Phi(\frac{\sqrt{2}((m^2+\sigma^2)^{\frac{k}{2}}-m)}{\sigma})\right.\\ &\left.+\Phi(-\frac{\sqrt{2}((m^2+\sigma^2)^{\frac{k}{2}}+m)}{\sigma})\right]. \quad (102)\end{aligned}$$

Define $D \triangleq \lim_{m\to\infty} k\log(1+m^2+\sigma^2)\left[\Phi(\frac{\sqrt{2}((m^2+\sigma^2)^{\frac{k}{2}}-m)}{\sigma})-\Phi(-\frac{\sqrt{2}((m^2+\sigma^2)^{\frac{k}{2}}+m)}{\sigma})\right]$, and $M \triangleq \frac{\sqrt{2}((m^2+\sigma^2)^{\frac{k}{2}}-m)}{\sigma}$. Note that as $m\to\infty$, $M\to-\infty$, and

$$\begin{aligned}O(M) =& O((m^2+\sigma^2)^{\frac{k}{2}}-m) && (103)\\ =& O(m^k-m). && (104)\end{aligned}$$

Furthermore, it holds also that:

$$\begin{aligned}m-m^k =& m-m^{1-\frac{1}{\log^2 m}} && (105)\\ =& m(1-e^{-\frac{\log m}{\log^2 m}}) && (106)\\ =& m(1-e^{-\frac{1}{\log m}}) && (107)\\ =& m(1-1+\frac{1}{\log m}+O(\frac{1}{\log^2 m})) && (108)\\ =& \frac{m}{\log m}+O(\frac{m}{\log^2 m}) && (109)\end{aligned}$$

First, because $\Phi(x)$ is monotonically increasing, and $\frac{\sqrt{2}((m^2+\sigma^2)^{\frac{k}{2}}-m)}{\sigma} > -\frac{\sqrt{2}((m^2+\sigma^2)^{\frac{k}{2}}+m)}{\sigma}$, then $D\geq 0$. Second, we have:

$$\begin{aligned}D \leq& \lim_{m\to\infty} k\log(1+m^2+\sigma^2)\Phi(\frac{\sqrt{2}((m^2+\sigma^2)^{\frac{k}{2}}-m)}{\sigma}) && (110)\\ =& \frac{1}{\sqrt{2\pi}}\lim_{m\to\infty} k\log(1+m^2+\sigma^2)\int_{-\infty}^{M} e^{\frac{-x^2}{2}}\,dx && (111)\\ \leq& \frac{1}{\sqrt{2\pi}}\lim_{m\to\infty} k\log(1+m^2+\sigma^2)\int_{-M}^{\infty} e^{-x}\,dx && (112)\\ =& \frac{1}{\sqrt{2\pi}}\lim_{m\to\infty} k\log(1+m^2+\sigma^2)e^{M} && (113)\\ =& 0, && (114)\end{aligned}$$

where (112) follows from (111) since for large X, $e^{-\frac{X^2}{2}} \leq e^{-X}$. Equation (114) follows from (113) because $M = \frac{\sqrt{2}((m^2+\sigma^2)^{\frac{k}{2}}-m)}{\sigma} = O(\frac{m}{\log(m)})$, and the order of $e^{\frac{m}{\log(m)}}$ is much higher than that of $\log(m^2)$. Since $D\leq 0\leq D$, we have $D=0$. Therefore, when m approaches infinity, $a\ \Pr(Y>a) \geq k\log(1+m^2+\sigma^2)$ (in the limit sense obviously). Using the latter property along with (101), we obtain:

$$\begin{aligned}&\lim_{m\to\infty}\{k\log(1+m^2+\sigma^2)-\mathbf{E}[Y]\}\\ \leq&0\\ \leq&\lim_{m\to\infty}\{\log(1+m^2+\sigma^2)-\mathbf{E}[Y]\}. && (115)\end{aligned}$$

Then we substitute k with $1-\frac{1}{\log^2 m}$ in Equation (115), to get

$$\begin{aligned}&\lim_{m\to\infty} k\log(1+m^2+\sigma^2)-\mathbf{E}[Y]\\ =&\lim_{m\to\infty}\left(1-\frac{1}{\log^2 m}\right)\log\left(1+m^2+\sigma^2\right)-\mathbf{E}[Y] && (116)\\ =&\lim_{m\to\infty}\left(\log(1+m^2+\sigma^2)-\mathbf{E}[Y]\right)\\ &-\lim_{m\to\infty}\frac{\log(1+m^2+\sigma^2)}{\log^2 m} && (117)\\ =&\lim_{m\to\infty}\left(\log(1+m^2+\sigma^2)-\mathbf{E}[Y]\right) && (118)\end{aligned}$$

Combining (115) and (118), we obtain:

$$\begin{aligned}&\lim_{m\to\infty}\{\log(1+m^2+\sigma^2)-\mathbf{E}[Y]\}\\ \leq&0\\ \leq&\lim_{m\to\infty}\{\log(1+m^2+\sigma^2)\mathbf{E}[Y]\}. && (119)\end{aligned}$$

By the Squeeze theorem, we have then:

$$\lim_{m\to\infty}\{\log(1+m^2+\sigma^2)-\mathbf{E}[Y]\}=0. \quad (120)$$

Substituting $\mathbf{E}[Y]$ with $\mathbf{E}[\log(1+|v|^2)]$ and $\log(1+m^2+\sigma^2)$ with $\log(\mathbf{E}[1+|v|^2])$ yields:

$$\lim_{m\to\infty}\{\mathbf{E}[\log(1+|v|^2)]-\log(\mathbf{E}[1+|v|^2])\}=0, \quad (121)$$

which is the desired result.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

PROOF OF THEOREM 2

For any given $\hat{\boldsymbol{h}}$, define a new random variable vector $\boldsymbol{v} = U^H \boldsymbol{h}$, and $\boldsymbol{m}$ as the mean vector of $\boldsymbol{v}$. By Jensen's inequality

$$R_s(U,\Lambda) = \mathbf{E}_{\boldsymbol{v}}[\log(1+\boldsymbol{v}^H\Lambda\boldsymbol{v})] - \mathbf{E}_{\boldsymbol{g}}[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})] \quad (122)$$

$$\leq \log(1+\mathbf{E}[\boldsymbol{v}^H\Lambda\boldsymbol{v}]) - \mathbf{E}_{\boldsymbol{g}}[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})]. \quad (123)$$

Note that

$$\mathbf{E}[\boldsymbol{v}^H\Lambda\boldsymbol{v}] = \lambda_1\mathbf{E}|v_1|^2 + \ldots + \lambda_{N_t}\mathbf{E}|v_{N_t}|^2, \quad (124)$$

where λ_i's are the diagonal elements of Λ and v_i's are the elements of the random vector $\boldsymbol{v}$. Suppose $t = \text{argmax}_i \mathbf{E}\left[|v_i|^2\right]$, then

$$\mathbf{E}(\boldsymbol{v}^H\Lambda\boldsymbol{v}) \leq \left(\sum_i \lambda_i\right)\mathbf{E}\left[|v_t|^2\right] \quad (125)$$

$$= P\,\mathbf{E}\left[|v_t|^2\right] \quad (126)$$

$$= P\left(|m_t|^2 + \alpha\,\sigma_h^2\right) \quad (127)$$

$$\leq P\left((1-\alpha)\|\hat{\boldsymbol{h}}\|^2 + \alpha\,\sigma_h^2\right). \quad (128)$$

The equality is achieved when $\lambda_1 = P$, $\lambda_i = 0$ for $i \neq 1$, $m_1 = \sqrt{1-\alpha}\|\hat{\boldsymbol{h}}\|$ and $m_i = 0$ for $i \neq 1$. Furthermore, the optimal U^o which results in $\boldsymbol{m} = [\sqrt{1-\alpha}\|\hat{\boldsymbol{h}}\|, 0...0]$ is given by $U^o[1] = \frac{\hat{\boldsymbol{h}}}{\|\hat{\boldsymbol{h}}\|}$, $U^o[2]...U^o[N_t]$ are arbitrarily chosen, except for the restriction that the columns of U^o are orthonormal. The function $f(\Lambda) = \mathbf{E}_g[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})]$ is Schur-concave, so the minimum of $f(\Lambda)$ can be achieved by the rank-one Λ^o which majorizes any other Λ. So rank-one Λ^o with $\text{diag}(\Lambda^o) = [P,0,...,0]$, achieves the maximum of the first term in equation (123) and minimum of the second term, thus maximizes the difference of them. Note that the choice of U has no impact on the second term, the U^o that maximizes the first term also maximizes the right hand side (RHS) of (123). Combining the above facts gives :

$$\max_{U,\Lambda} R_s(U,\Lambda|\hat{\boldsymbol{h}})$$

$$= \max_{U,\Lambda}\{\mathbf{E}_v[\log(1+\boldsymbol{v}^H\Lambda\boldsymbol{v})] - \mathbf{E}_{\boldsymbol{g}}[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})]\} \quad (129)$$

$$\leq \max_{U,\Lambda}\{\log(1+\mathbf{E}[\boldsymbol{v}^H\Lambda\boldsymbol{v}]) - \mathbf{E}_{\boldsymbol{g}}[\log(1+\boldsymbol{g}^H\Lambda\boldsymbol{g})]\} \quad (130)$$

$$= \log(1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2 + \alpha\,\sigma_h^2)) - \mathbf{E}_{g_1}[\log(1+P|g_1|^2)]. \quad (131)$$

Then we choose Λ^* to be rank-one with $\text{diag}(\Lambda^*) = [P,0,...,0]$, and $U^* = \frac{\hat{\boldsymbol{h}}\hat{\boldsymbol{h}}^H}{\|\hat{\boldsymbol{h}}\|^2}$ to obtain:

$$R_s(U^*,\lambda^*) = \mathbf{E}_{v_1}[\log(1+P|v_1|^2)] - \mathbf{E}_{g_1}[\log(1+P|g_1|^2)], \quad (132)$$

where $v_1 \sim \mathcal{CN}(\sqrt{1-\alpha}\|\hat{\boldsymbol{h}}\|, \alpha\,\sigma_h^2)$, note that $\max_{U,\Lambda} R_s(U,\Lambda|\hat{\boldsymbol{h}})$ must be larger than or equal to $R_s(U^*,\Lambda^*)$. From equation (131) and (132), we have that:

$$LB(\hat{\boldsymbol{h}}) \leq \max_{U,\Lambda} R(U,\Lambda|\hat{\boldsymbol{h}}) \leq UB(\hat{\boldsymbol{h}}), \quad (133)$$

where the upper bound in equation (133) $UB(\hat{\boldsymbol{h}}) = \log(1 + P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2 + \alpha\,\sigma_h^2)) - \mathbf{E}_{g_1}[\log(1+P|g_1|^2)]$, and the lower bound $LB(\hat{\boldsymbol{h}}) = \mathbf{E}_{v_1}[\log(1+P|v_1|^2)] - \mathbf{E}_{g_1}[\log(1+P|g_1|^2)]$. By Lemma 2, the upper bound and lower bound converge as $\|\hat{\boldsymbol{h}}\| \to \infty$, that is,

$$\lim_{\|\hat{\boldsymbol{h}}\|\to\infty}\left\{UB(\hat{\boldsymbol{h}}) - LB(\hat{\boldsymbol{h}})\right\}$$

$$= \lim_{\|\hat{\boldsymbol{h}}\|\to\infty}\Big\{\log(1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2 + \alpha\,\sigma_h^2)) - \mathbf{E}_{v_1}\left[\log(1+P|v_1|^2)\right]\Big\} \quad (134)$$

$$= \lim_{\|\hat{\boldsymbol{h}}\|\to\infty}\{\log\mathbf{E}_{v_1}[1+P|v_1|^2] - \mathbf{E}_{v_1}[\log(1+P|v_1|^2)]\} \quad (135)$$

$$= 0. \quad (136)$$

Hence, according to the Squeeze Theorem, we deduce that:

$$\lim_{\|\hat{\boldsymbol{h}}\|\to\infty}\left\{\max_{U,\Lambda} R(U,\Lambda|\hat{\boldsymbol{h}}) - UB(\hat{\boldsymbol{h}})\right\} = 0. \quad (137)$$

This maximum is achieved by choosing Λ to be rank-one and $U[1] = \frac{\hat{\boldsymbol{h}}}{\|\hat{\boldsymbol{h}}\|}$. The optimal transmission strategy is to beamform to the direction of $\hat{\boldsymbol{h}}$.

Now, we prove the asymptotic behavior of the secrecy rate as $N_t \to \infty$. Define a new random variable Z :

$$Z = \begin{cases} 0 & \text{if for any } u>0\ \|\hat{\boldsymbol{h}}\|^2 > u \\ 1 & \text{if there exists } N>0 \text{ such that } \|\hat{\boldsymbol{h}}\|^2 \text{ is bounded by } N. \end{cases} \quad (138)$$

From Lemma 1, we have $\lim_{N_t\to\infty}\Pr(Z=0) = 1$ and $\lim_{N_t\to\infty}\Pr(Z=1) = 0$. For convenience, define $M(\hat{\boldsymbol{h}}) \triangleq \max_{U,\Lambda} R_s(U,\Lambda|\hat{\boldsymbol{h}}) - UB(\hat{\boldsymbol{h}})$, then, we have:

$$\lim_{N_t\to\infty}\mathbf{E}_{\hat{\boldsymbol{h}}}[M(\hat{\boldsymbol{h}})] = \lim_{N_t\to\infty}\mathbf{E}_Z[\mathbf{E}_{\hat{\boldsymbol{h}}}[M(\hat{\boldsymbol{h}})|Z]] \quad (139)$$

$$= \lim_{N_t\to\infty}\Pr(Z=0)\mathbf{E}_{\hat{\boldsymbol{h}}}[M(\hat{\boldsymbol{h}})|Z=0] + \Pr(Z=1)\mathbf{E}_{\hat{\boldsymbol{h}}}[M(\hat{\boldsymbol{h}})|Z=1] \quad (140)$$

$$= \lim_{N_t\to\infty}\Pr(Z=0)\mathbf{E}_{\hat{\boldsymbol{h}}}[M(\hat{\boldsymbol{h}})|Z=0] \quad (141)$$

$$= \lim_{N_t\to\infty}\mathbf{E}_{\hat{\boldsymbol{h}}}[M(\hat{\boldsymbol{h}})|Z=0] \quad (142)$$

$$= 0, \quad (143)$$

where (142) follows from (141) since $\lim_{N_t\to\infty}\Pr(Z=1) = 0$ and according to *Lemma* 2, if $\|\hat{\boldsymbol{h}}\|^2$ is bounded by N then $\mathbf{E}[M(\hat{\boldsymbol{h}})]$ is also bounded. Equation (143) follows from (142) since according to (137) when $\|\hat{\boldsymbol{h}}\|^2$ approaches infinity, $\mathbf{E}[M(\hat{\boldsymbol{h}})]$ is equal to 0. Averaging over $\hat{\boldsymbol{h}}$, the secrecy rate $\bar{R}_s$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

may be expressed by:

$$\lim_{N_t\to\infty}\left\{\bar{R}_s-\left[\mathbf{E}_{\hat{\boldsymbol{h}}}\left[\log(1+P((1-\alpha)\|\hat{\boldsymbol{h}}\|^2+\alpha\,\sigma_h^2))\right]-\mathbf{E}_{g_1}\left[\log(1+P|g_1|^2)\right]\right]\right\} \tag{144}$$

$$=\lim_{N_t\to\infty}\mathbf{E}_{\hat{\boldsymbol{h}}}[M(\hat{\boldsymbol{h}})] \tag{145}$$

$$=0, \tag{146}$$

which completes the proof.

REFERENCES

[1] C. E. Shannon, "Communication theory of secrecy systems," *Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[2] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal, The*, vol. 54, no. 8, pp. 1355–1387, 1975.

[3] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inform. Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978.

[4] I. Csiszár and J. Korner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[5] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2180–2189, Jun. 2008.

[6] X. Li, M. Chen, and E. P. Ratazzi, "Space-time transmissions for wireless secret-key agreement with information-theoretic secrecy," in *the 6th IEEE Workshop on Signal Processing Advances in Wireless Communications*, 2005, pp. 811–815.

[7] X. Tang, R. Liu, P. Spasojevic, and H. V. Poor, "Interference assisted secret communication," *IEEE Trans. Inform. Theory*, vol. 57, no. 5, pp. 3153–3167, 2011.

[8] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2632–2639, 2001.

[9] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, July 2004.

[10] D. H?sli and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," in *Proceedings of the 5th International ITG Conference on Source and Channel Coding (SCC), Jan, 2004*, pp. 381–385.

[11] S. Venkatesan, S. H. Simon, and R. A. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," in *IEEE Vehicular Technology Conference (VTC'2003), Oct. 2003*, vol. 3, pp. 1767–1771.

[12] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933–946, 2000.

[13] A. Lapidoth, P. Narayan *et al.*, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2148–2177, 1998.

[14] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1423–1436, 1998.

[15] S. Shafiee, N. Liu, and S. Ulukus, "Towards the secrecy capacity of the gaussian MIMO wire-tap channel: The 2-2-1 channel," *IEEE Trans. Inform. Theory*, vol. 55, no. 9, pp. 4033–4039, 2009.

[16] P. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4687–4698, Oct. 2008.

[17] S. Gerbracht, C. Scheunert, and E. Jorswieck, "Secrecy outage in MISO systems with partial channel information," *IEEE Trans. Inform. Forensics and Security*, vol. 7, no. 2, pp. 704–716, 2012.

[18] X. Zhou and M. McKay, "Secure transmission with artificial noise over fading channels: Achievable rate and optimal power allocation," *IEEE Trans. Veh. Technol.*, vol. 59, no. 8, pp. 3831–3842, Oct. 2010.

[19] Z. Rezki, A. Khisti, and M.-S. Alouini, "On the secrecy capacity of the wiretap channel with imperfect main channel estimation," *IEEE Trans. Commun.*, vol. 62, no. 10, pp. 3652–3664, Oct 2014.

[20] Z. Rezki, B. Alomair, and M.-S. Alouini, "On the secrecy capacity of the MISO wiretap channel under imperfect channel estimation," in *Proceeding of IEEE Global Communications Conference (GLOBECOM'2014), Austin, TX, USA, Dec. 2014*, pp. 1602–1607.

[21] J. Li and A. Petropulu, "On ergodic secrecy rate for Gaussian MISO wiretap channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1176–1187, April 2011.

[22] M. Bloch and J. Laneman, "Exploiting partial channel state information for secrecy over wireless channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1840–1849, September 2013.

[23] Z. Rezki, A. Khisti, and M. Alouini, "Ergodic secret message capacity of the wiretap channel with finite-rate feedback," *IEEE Trans. on Wireless Communications*, vol. 13, no. 6, pp. 3364–3379, Jun. 2014.

[24] Y. Yang, W. Wang, H. Zhao, and L. Zhao, "Transmitter beamforming and artificial noise with delayed feedback: Secrecy rate and power allocation," *Journal of Communications and Networks*, vol. 14, no. 4, pp. 374–384, Aug. 2012.

[25] S. Shafiee and S. Ulukus, "Achievable rates in Gaussian MISO channels with secrecy constraints," in *IEEE International Symposium on Information Theory (ISIT'2007), Nice, France, June, 2007*, pp. 2466–2470.

[26] P.-H. Lin, S.-H. Lai, S.-C. Lin, and H.-J. Su, "On secrecy rate of the generalized artificial-noise assisted secure beamforming for wiretap channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1728–1740, September 2013.

[27] B. Wang, P. Mu, and Z. Li, "Secrecy rate maximization with artificial-noise-aided beamforming for MISO wiretap channels under secrecy outage constraint," *IEEE Comm. Letters*, vol. 19, no. 1, pp. 18–21, Jan 2015.

[28] E. Jorswieck, S. Tomasin, and A. Sezgin, "Broadcasting into the uncertainty: Authentication and confidentiality by physical-layer processing," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1702–1724, Oct 2015.

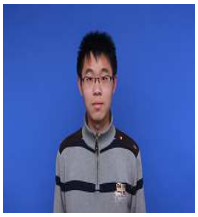
[29] A. Mukherjee, S. Fakoorian, J. Huang, and A. Swindlehurst, "Principles of physical layer security in multiuser wireless networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1550–1573, Third 2014.

[30] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European transactions on telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[31] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of majorization and its applications*. Springer Science & Business Media, 2010.

[32] I. Muqattash and M. Yahdi, "Infinite family of approximations of the digamma function," *Mathematical and computer modelling*, vol. 43, no. 11, pp. 1329–1336, 2006.

[33] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas I: The MISOME wiretap channel," *IEEE Trans. Inform. Theory*, vol. 56, no. 7, pp. 3088–3104, 2010.



Xinyu Zhou was born in Yancheng, China. Currently, he is a senior student in electrical engineering departmant of Shanghai Jiaotong University, Shanghai, China. From September 2014 to March 2015, he was a research intern in King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. His research interests include physical layer security, information theory and digital signal processing.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

Basel Alomair ((M11) received his Bachelors, Masters, and Ph.D. degrees from King Saud University, University of Wisconsin-Madison, and University of Washington-Seattle, respectively. He is an Assistant Professor and Founding Director of the National Center for Cybersecurity Technology (C4C) in King Abdulaziz City for Science and Technology (KACST), an Affiliate Professor and co-director of the Network Security Lab (NSL) at the University of Washington-Seattle, an Affiliate Professor at King Saud University (KSU), and a cryptology consultant at various agencies. He was recognized by the IEEE Technical Committee on Fault-Tolerant Computing (TC-FTC) and the IFIP Working Group on Dependable Computing and Fault Tolerance (WG 10.4) with the 2010 IEEE/IFIP William Carter Award for his significant contributions in the area of dependable computing. His research in information security was recognized with the 2011 Outstanding Research Award from the University of Washington. He was also the recipient of the 2012 Distinguished Dissertation Award from the Center for Information Assurance and Cybersecurity at the University of Washington (UW CIAC). He was awarded the 2015 Early Career Award in Cybersecurity by the NSA/DHS Center of Academic Excellence in Information Assurance Research for his contributions to Modern Cryptographic Systems and Visionary Leadership. He authored/co-authored multiple best paper awards and he is a member of the IEEE since 2010.

Mohamed-Slim Alouini (S'94, M'98, SM'03, F'09) was born in Tunis, Tunisia. He received the Ph.D. degree in Electrical Engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998. He served as a faculty member in the University of Minnesota, Minneapolis, MN, USA, then in the Texas A&M University at Qatar, Education City, Doha, Qatar before joining King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia as a Professor of Electrical Engineering in 2009. His current research interests include the modeling, design, and performance analysis of wireless communication systems.

Zouheir Rezki (S'01, M'08, SM'13) was born in Casablanca, Morocco. He received the Diplôme d'Ingénieur degree from the École Nationale de l'Industrie Minérale (ENIM), Rabat, Morocco, in 1994, the M.Eng. degree from École de Technologie Supérieure, Montreal, Québec, Canada, in 2003, and the Ph.D. degree from École Polytechnique, Montreal, Québec, in 2008, all in electrical engineering. From October 2008 to September 2009, he was a postdoctoral research fellow with Data Communications Group, Department of Electrical and Computer Engineering, University of British Columbia. He is now a Senior Research Scientist at King Abdullah University of Science and Technology (KAUST), Thuwal, Mekkah Province, Saudi Arabia. His research interests include: security of data networks, performance limits of communication systems, cognitive and sensor networks, optical communications and low-complexity detection algorithms.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TWC.2016.2542128, IEEE Transactions on Wireless Communications

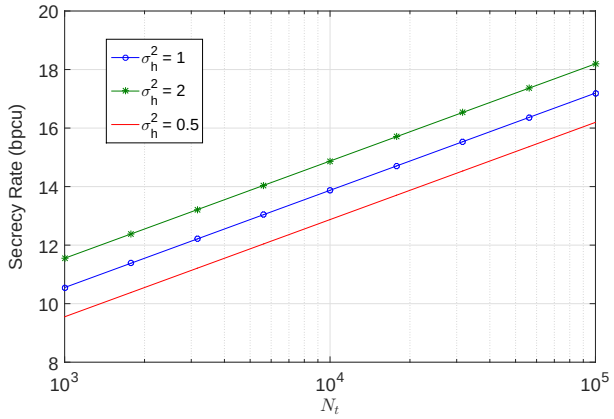


Figure 10: Secrecy rate versus N_t for large antenna systems, for channel estimation error $\alpha = 0.5$, transmit power $P = 10$ dB and different σ_h^2 values.

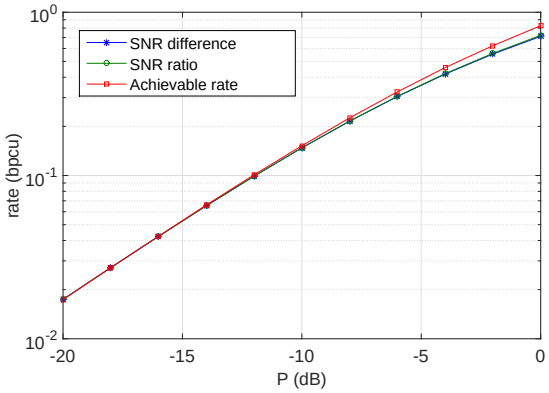


Figure 11: The transmission rates based on SNR difference and ratio criterions in low SNR regime with $\alpha = 0.3$ and $\beta = 0.2$. Here, $\sigma_h^2 = 1$.

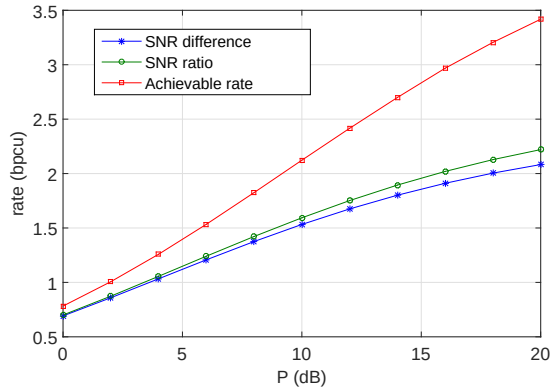


Figure 12: The transmission rates based on SNR difference and ratio criterions in medium to high SNR regimes with $\alpha = 0.3$ and $\beta = 0.2$. Here, $\sigma_h^2 = 1$.