# Non-Coherent Capacity of Secret-Key Agreement with Public Discussion

Anurag Agrawal, Zouheir Rezki, *Member, IEEE*, Ashish Khisti, *Member, IEEE,* and Mohamed-Slim Alouini, *Senior Member, IEEE*

*Abstract*—We study the Rayleigh fading non-coherent capacity of secret-key agreement with public discussion, where neither the sender nor the receivers have access to instantaneous channel state information (CSI) of any channel. We present two results. At high Signal-to-Noise Ratio (SNR), the secret-key capacity is bounded in SNR, regardless of the number of antennas at each terminal. Second, for a system with a single antenna at both the legitimate and the eavesdropper terminals and an arbitrary number of transmit antennas, the secret-key capacity-achieving input distribution is discrete, with a finite number of mass points. Numerically we observe that at low-SNR, the capacity achieving distribution has two mass points with one of them at the origin.

*Index Terms*—Non-coherent capacity, secret-key agreement, Rayleigh fading channels, information theoretic security, Karush-Kuhn-Tucker (KKT) condition, discrete input distribution.

## I. INTRODUCTION

Information theoretic secret-key agreement provides provably secure mechanisms for generating secret-keys between two or more legitimate terminals. In such protocols, the legitimate terminals need to have access to a source of correlated randomness e.g., communication channels or correlated sources [1], [2]. Furthermore a discussion channel of unlimited capacity is also available for communication, but is public to the wiretapper. The legitimate terminals distill a common secret-key that satisfies an equivocation constraint with respect to the eavesdropper.

The present paper studies capacity limits of secret-key agreement when the underlying channel from the sender to the receiver and the eavesdropper are modeled as independent identically distributed (i.i.d.) Rayleigh fading. We further assume the non-coherent model i.e., the instantaneous channel state information is not known to either of the terminals. The channel statistics are however globally known.

Note that for our proposed channel model the outputs at the legitimate receiver and the eavesdropper are conditionally independent given the channel input. A class of discrete memoryless channel models with this property was studied

Anurag Agrawal (anuragagrawal2006@gmail.com) is with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India. Zouheir Rezki and Mohamed-Slim Alouini are with King Abdullah University of Science and Technology (KAUST) {zouheir.rezki;slim.alouini}@kaust.edu.sa, Thuwal, KSA. Ashish Khisti (akhisti@comm.utoronto.ca) is with the Electrical and Computer Engineering Department, University of Toronto, Toronto, ON, Canada. This work was conducted when Anurag Agrawal and Zouheir Rezki were visiting University of Toronto. This work will be presented at the IEEE International Conference on Communications (ICC) 2011, Workshop on Physical Layer Security in Kyoto, Japan

in [1], [2] and a single-letter capacity expression was characterized. In particular a *source-emulation* strategy was shown to be optimal — the sender generates a discrete memoryless source, then transmits it over the channel to generate correlated sources at the two terminals and then the legitimate terminals distill a common key as in the source model. While their result can be extended using standard techniques to the (continuous-valued) Rayleigh fading channels studied in this work, finding the optimizing distribution is difficult in general. In the present work we show the following two properties (1) Unlike the case without secrecy constraint where the capacity scales as $\log \log (P)$ at high-SNR, the secret-key capacity is bounded in SNR, regardless of the number of antennas at each terminal and (2) The capacity achieving distribution is discrete with a finite number of mass points, for the MISO case.

In related works, [3] studies the secret-key agreement over Rayleigh fading channels for the case of receiver CSI and establishes that a Gaussian input distribution maximizes the secret-key capacity. [4], [5], [6], [7], [8], [9] and [10] also approach similar problems in different settings. [11] studies the problem of generating shared secret keys using channel reciprocity instead of public discussion. This approach is not considered in the present paper. Related work can also be found in [12], [13], [14], [15], [16], [17] and [18].

## II. THE CHANNEL MODEL

In this paper we consider the Channel-Type Model with Wiretapper [1, Section III]. The sender and receiver communicate over a discrete memoryless channel (DMC). In addition they can also access a public discussion channel of unlimited capacity. In our case of interest, the DMC consists of an i.i.d. multi-input-multi-output (MIMO) fading channel where the sender and receiver and eavesdropper have $n_T$, $n_R$ and $n_E$ antennas respectively. We have closely followed [19] and [20] while proposing our model. The outputs at both the legitimate destination and the eavesdropper, at time $i = 1, \ldots, L$, are expressed, respectively by:

$$\begin{cases} \mathbf{y}(i) = \mathbf{H}(i)\mathbf{x}(i) + \mathbf{v}(i) \\ \mathbf{z}(i) = \mathbf{G}(i)\mathbf{x}(i) + \mathbf{w}(i) \end{cases} \tag{1}$$

where $\mathbf{x}(i) \in \mathbb{C}^{n_t}$ is the transmitted signal, and $\mathbf{H}(i) \in \mathbb{C}^{n_R \times n_T}$, $\mathbf{G}(i) \in \mathbb{C}^{n_E \times n_T}$ represent the main channel and the eavesdropper channel gains, respectively; and $\mathbf{v}(i) \in \mathbb{C}^{n_R}$, $\mathbf{w}(i) \in \mathbb{C}^{n_E}$ are circularly symmetric white Gaussian noises with covariance matrices $\mathbb{E}[\mathbf{v}(i)\mathbf{v}(i)^{\dagger}] = \mathrm{I}_{n_R}$ and

$\mathbb{E}[\mathbf{w}(i)\mathbf{w}(i)^{\dagger}] = I_{n_E}$. We assume that $\mathbf{H}(i)$ and $\mathbf{G}(i)$ have independent and identically-distributed Gaussian entries with zero-mean and unit-variance. We assume that the Channel State Information (CSI) is not available at any terminal. That is, the transmitter, the legitimate receiver and the eavesdropper do not have access to the instantaneous channel realizations $\mathrm{H}(i)$ and $\mathrm{G}(i)$; but are aware of their statistic. The source is constrained according to a short-term average power constraint:

$$\mathbb{E}[\mathbf{x}(i)^{\dagger}\mathbf{x}(i)] = \operatorname{tr}(\mathrm{Q}_x) \leq P \qquad (2)$$

for all $i = 1,\ldots,L$. Given the normalization of the fading channels and the additive noise in our channel model, $P$ in (2) may be interpreted as a signal-to-noise ratio (SNR) per receive antenna. Since the channel defined in (1) is i.i.d. we may drop the time index $i$ in the sequel for convenience. The sender and receiver can interactively communicate over the discussion channel between each successive use of the channel. We refer to the reader to [1] for a formal description of the protocol. Furthermore we are interested in the secret-key capacity between the sender and the receiver. We again refer to the reader to [1] for a formal description and definition of key capacity.

## III. SECRET-KEY CAPACITY

In [1, Theorem 2], a single-letter formula of key capacity has been established and is given by:

$$C = \sup_{F(\mathbf{x})\in\mathcal{F}} I(\mathbf{x};\mathbf{y}|\mathbf{z}), \qquad (3)$$

where $\mathcal{F}$ is the set of all possible distribution functions $F(\mathbf{x})$ that satisfy the average power constraint. Note that in our model of interest $\mathbf{y}$ and $\mathbf{z}$ are conditionally independent given $\mathbf{x}$ and the distribution is given by [21]:

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \frac{1}{\pi^{n_R}(1+\|\mathbf{x}\|^2)^{n_R}} \exp\left(\frac{-\|\mathbf{y}\|^2}{1+\|\mathbf{x}\|^2}\right) \qquad (4)$$

$$p_{\mathbf{z}|\mathbf{x}}(\mathbf{z}|\mathbf{x}) = \frac{1}{\pi^{n_E}(1+\|\mathbf{x}\|^2)^{n_E}} \exp\left(\frac{-\|\mathbf{z}\|^2}{1+\|\mathbf{x}\|^2}\right), \qquad (5)$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector. We note that $p_{\mathbf{y}|\mathbf{x}}$ and $p_{\mathbf{z}|\mathbf{x}}$ depend on $\mathbf{y}$, $\mathbf{z}$ and $\mathbf{x}$ only through their norms. We next define new random variables: $X = \|\mathbf{x}\|, Y = \|\mathbf{y}\|^2$ and $Z = \|\mathbf{z}\|^2$. Following [22, Appendix VI], we see that the pdf of the conditional distribution associated with the norms is given by:

$$p_{\|\mathbf{z}\| \,|\, \|\mathbf{x}\|}(\|\mathbf{z}\| \,|\, \|\mathbf{x}\|) = \phi'\, p_{\mathbf{z}|\mathbf{x}}(\mathbf{z}|\mathbf{x}) \qquad (6)$$

or equivalently

$$p_{Z|X}(z|x) = \phi\, p_{\mathbf{z}|\mathbf{x}}(\mathbf{z}|\mathbf{x}) \qquad (7)$$

, where

$$\phi' = \frac{\|\mathbf{z}\|^{n_E-1}\pi^{n_E}}{\Gamma(n_E)}, \phi = \frac{\|\mathbf{z}\|^{2(n_E-1)}\pi^{n_E}}{\Gamma(n_E)} \qquad (8)$$

are the Jacobian coordinate transformation factors applied in $n_E$ dimensions. Note that the slight difference in the expressions obtained in [22] occurs because the channel model in [22] considers the elements of the channel matrix $\mathbf{H}$ to

have unit variance in each dimension and also because of the difference in the definitions of the norms: $Y$ and $Z$. Then, from (4) and (5),

$$\begin{aligned} p_{Y|X}(y|x) &= \frac{y^{(n_R-1)}}{\Gamma(n_R)(1+x^2)^{n_R}} \exp\left(\frac{-y}{1+x^2}\right), \\ p_{Z|X}(z|x) &= \frac{z^{(n_E-1)}}{\Gamma(n_E)(1+x^2)^{n_E}} \exp\left(\frac{-z}{1+x^2}\right). \end{aligned} \qquad (9)$$

Also,

$$\begin{aligned} p_{\mathbf{z}}(\mathbf{z}) &= \mathbb{E}_{\mathbf{x}}\left[p_{\mathbf{z}|\mathbf{x}}(\mathbf{z}|\mathbf{x})\right] \\ &= \mathbb{E}_{\mathbf{x}}\left[\frac{p_{\|\mathbf{z}\|\,|\,\|\mathbf{x}\|}(\|\mathbf{z}\|\,|\,\|\mathbf{x}\|)}{\phi'}\right] \\ &= \mathbb{E}_{\|\mathbf{x}\|}\left[\frac{p_{\|\mathbf{z}\|\,|\,\|\mathbf{x}\|}(\|\mathbf{z}\|\,|\,\|\mathbf{x}\|)}{\phi'}\right] \\ &= \frac{p_{\|\mathbf{z}\|}(\|\mathbf{z}\|)}{\phi'} \end{aligned} \qquad (10)$$

Now, $I(\mathbf{x};\mathbf{y}|\mathbf{z})$ can be formulated as follows:

$$\begin{aligned} I(\mathbf{x};\mathbf{y}|\mathbf{z}) &= I(\mathbf{x};\mathbf{y},\mathbf{z}) - I(\mathbf{x};\mathbf{z}) \\ &= I(X;Y,Z) - I(X;Z) \\ &= I(X;Y|Z) \end{aligned} \qquad (11)$$

To establish (11), here we prove that $I(\mathbf{x};\mathbf{z}) = I(X;Z)$. The proof of $I(\mathbf{x};\mathbf{y},\mathbf{z}) = I(X;Y,Z)$ will follow on similar lines. The notion that $I(\mathbf{x};\mathbf{z}) = I(X;Z)$ has previously been used in [23], [19] and [20]. Here we provide an explicit proof to establish the claim.

$$\begin{aligned} &I(\mathbf{x};\mathbf{z}) \\ &= \iint p_{\mathbf{z},\mathbf{x}}(\mathbf{z},\mathbf{x}) \ln\left[\frac{p_{\mathbf{z}|\mathbf{x}}(\mathbf{z}|\mathbf{x})}{p_{\mathbf{z}}(\mathbf{z})}\right] d\mathbf{z}d\mathbf{x} \\ &= \iint p_{\mathbf{z},\mathbf{x}}(\mathbf{z},\mathbf{x}) \ln\left[\frac{p_{\|\mathbf{z}\|\,|\,\|\mathbf{x}\|}(\|\mathbf{z}\|\,|\,\|\mathbf{x}\|)/\phi'}{p_{\|\mathbf{z}\|}(\|\mathbf{z}\|)/\phi'}\right] d\mathbf{z}d\mathbf{x} \qquad (12) \\ &= \iint p_{\mathbf{z},\mathbf{x}}(\mathbf{z},\mathbf{x}) \ln\left[\frac{p_{\|\mathbf{z}\|\,|\,\|\mathbf{x}\|}(\|\mathbf{z}\|\,|\,\|\mathbf{x}\|)}{p_{\|\mathbf{z}\|}(\|\mathbf{z}\|)}\right] d\mathbf{z}d\mathbf{x} \\ &= \iint p_{\mathbf{z},\mathbf{x}}(\mathbf{z},\mathbf{x}) \ln\left[f(\|\mathbf{x}\|,\|\mathbf{z}\|)\right] d\mathbf{z}d\mathbf{x} \qquad (13) \\ &= \iint p_{\mathbf{z}|\mathbf{x}}(\mathbf{z}|\mathbf{x})p_{\mathbf{x}}(\mathbf{x}) \ln\left[f(\|\mathbf{x}\|,\|\mathbf{z}\|)\right] d\mathbf{z}d\mathbf{x} \\ &= \int \mathbb{E}_{\mathbf{x}}\left[p_{\mathbf{z}|\mathbf{x}}(\mathbf{z}|\mathbf{x}) \ln\left[f(\|\mathbf{x}\|,\|\mathbf{z}\|)\right]\right] d\mathbf{z} \\ &= \int \mathbb{E}_{\mathbf{x}}\left[h(\|\mathbf{x}\|,\|\mathbf{z}\|)\right] d\mathbf{z} \\ &= \int \mathbb{E}_{\|\mathbf{x}\|}\left[h(\|\mathbf{x}\|,\|\mathbf{z}\|)\right] d\mathbf{z} \\ &= \iint p_{\mathbf{z}|\mathbf{x}}(\mathbf{z}|\mathbf{x})p_{\|\mathbf{x}\|}(\|\mathbf{x}\|) \ln\left[f(\|\mathbf{x}\|,\|\mathbf{z}\|)\right] d\mathbf{z}d\|\mathbf{x}\| \\ &= \iint \frac{p_{\|\mathbf{z}\|\,|\,\|\mathbf{x}\|}(\|\mathbf{z}\|\,|\,\|\mathbf{x}\|)}{\phi'} p_{\|\mathbf{x}\|}(\|\mathbf{x}\|) \ln\left[f(\|\mathbf{x}\|,\|\mathbf{z}\|)\right] d\mathbf{z}d\|\mathbf{x}\| \\ &= \iint \frac{p_{\|\mathbf{z}\|,\|\mathbf{x}\|}(\|\mathbf{z}\|,\|\mathbf{x}\|)}{\phi'} \ln\left[f(\|\mathbf{x}\|,\|\mathbf{z}\|)\right] d\mathbf{z}d\|\mathbf{x}\| \\ &= \iint \frac{p_{\|\mathbf{x}\|\,|\,\|\mathbf{z}\|}(\|\mathbf{x}\|\,|\,\|\mathbf{z}\|)}{\phi'} p_{\|\mathbf{z}\|}(\|\mathbf{z}\|) \ln\left[f(\|\mathbf{x}\|,\|\mathbf{z}\|)\right] d\mathbf{z}d\|\mathbf{x}\| \\ &= \iint p_{\|\mathbf{x}\|\,|\,\|\mathbf{z}\|}(\|\mathbf{x}\|\,|\,\|\mathbf{z}\|)p_{\mathbf{z}}(\mathbf{z}) \ln\left[f(\|\mathbf{x}\|,\|\mathbf{z}\|)\right] d\mathbf{z}d\|\mathbf{x}\| \qquad (14) \end{aligned}$$

$$= \int \mathbb{E}_{\mathbf{z}} \left[ p_{\|\mathbf{x}\| \mid \|\mathbf{z}\|}(\|\mathbf{x}\| \mid \|\mathbf{z}\|) \ln \left[ f(\|\mathbf{x}\|, \|\mathbf{z}\|) \right] \right] d\|\mathbf{x}\|$$

$$= \int \mathbb{E}_{\|\mathbf{z}\|} \left[ p_{\|\mathbf{x}\| \mid \|\mathbf{z}\|}(\|\mathbf{x}\| \mid \|\mathbf{z}\|) \ln \left[ f(\|\mathbf{x}\|, \|\mathbf{z}\|) \right] \right] d\|\mathbf{x}\|$$

$$= \iint p_{\|\mathbf{x}\| \mid \|\mathbf{z}\|}(\|\mathbf{x}\| \mid \|\mathbf{z}\|) p_{\|\mathbf{z}\|}(\|\mathbf{z}\|) \ln \left[ f(\|\mathbf{x}\|, \|\mathbf{z}\|) \right] d\|\mathbf{z}\| d\|\mathbf{x}\|$$

$$= \iint p_{\|\mathbf{x}\|, \|\mathbf{z}\|}(\|\mathbf{x}\|, \|\mathbf{z}\|) \ln \left[ \frac{p_{\|\mathbf{z}\| \mid \|\mathbf{x}\|}(\|\mathbf{z}\| \mid \|\mathbf{x}\|)}{p_{\|\mathbf{z}\|}(\|\mathbf{z}\|)} \right] d\|\mathbf{z}\| d\|\mathbf{x}\|$$

$$= I(\|\mathbf{x}\|; \|\mathbf{z}\|)$$

$$= I(X; Z)$$

where (12) follows from (6) and (10). (13) says that $\frac{p_{\|\mathbf{z}\| \mid \|\mathbf{x}\|}(\|\mathbf{z}\| \mid \|\mathbf{x}\|)}{p_{\|\mathbf{z}\|}(\|\mathbf{z}\|)}$ is some function $f(\cdot)$ of the norms of the two vectors. (14) follows from (10). The result is summarized in the following lemma.

*Lemma 1:* The non-coherent secret-key capacity of the channel model (1) described above is given by:

$$C = \sup_{F(x) \in \mathcal{F}} I(X; Y | Z). \tag{15}$$

Lemma 1 states that secret-key communication is conveyed over the norms X, Y and Z. This property will be used in the sequel. As verified in Appendix A the capacity can be expressed as

$$C = \sup_{F(x) \in \mathcal{F}}$$
$$\iiint p_{Y|X}(y|x) p_{Z|X}(z|x) \ln \left[ \frac{p_{Y|X}(y|x)}{p_{Y|Z}(y|z; F)} \right] dy\, dz\, dF(x) \tag{16}$$

## IV. Capacity Results At A High-SNR Regime

In this section, we analyze the non-coherent secret key capacity asymptotically at high-SNR. By high-SNR we mean that this analysis is an asymptotic analysis, the purpose of which is to analyze the capacity as SNR$\to \infty$, a common formulation in information theory and communication theory. The reader is referred to [20] and [19] for futher details. Our result is rather negative as it establishes the non-efficiency of communication over this channel at high-SNR. Theorem 1 below formalizes this result.

*Theorem 1:* At high-SNR, the non-coherent secret-key capacity of the channel model (1) described above is given by:

$$C(P) = O(1) \tag{17}$$

*Proof:* Keeping in mind the result given by Lemma 1 and equations given by 9, we define two new random variables $W_1$ and $W_2$ such that

$$\begin{cases} Y = (1 + X^2) W_1 \\ Z = (1 + X^2) W_2 \end{cases} \tag{18}$$

where it follows via (9) that $W_1$ and $W_2$ are Gamma-distributed random variables, mutually independent and also independent of $X$, and with probability density functions (pdfs):

$$p_{W_1}(w_1) = \frac{w_1^{n_R - 1}}{\Gamma(n_R)} \exp(-w_1) \tag{19}$$

$$p_{W_2}(w_2) = \frac{w_2^{n_E - 1}}{\Gamma(n_E)} \exp(-w_2). \tag{20}$$

(18) can be viewed as a multiplicative channel model which is equivalent from a capacity perspective to our original model by Lemma 1.

Now, by letting $Y_1 = \ln(Y)$, $Z_1 = \ln(Z)$, $D_1 = \ln(1 + X^2)$, $N_1 = \ln(W_1)$ and $N_2 = \ln(W_2)$, and by applying the log function on both sides of (18), the following channel is obtained:

$$\begin{cases} Y_1 = D_1 + N_1 \\ Z_1 = D_1 + N_2 \end{cases} \tag{21}$$

Since the log function is a one-to-one transformation that does not entail any capacity loss, the secret-key capacity can be bounded as follows:

$$\begin{aligned} I(X; Y | Z) &= I(D_1; Y_1 | Z_1) \\ &= h(Y_1 | Z_1) - h(Y_1 | Z_1, D_1) \\ &= h(Y_1 - Z_1 | Z_1) - h(Y_1 | Z_1, D_1) \\ &\leq h(Y_1 - Z_1) - h(N_1) \tag{22} \\ &= h(N_1 - N_2) - h(N_1). \tag{23} \end{aligned}$$

Hence, it remains to show that the right hand side of (23) is bounded.

First note that using (19), the p.d.f. of $N_1$ can be easily derived as:

$$p_{N_1}(n_1) = \frac{\exp(n_R n_1 - e^{n_1})}{\Gamma(n_R)}, \tag{24}$$

thus, the second term on the RHS of (23) can be upper bounded as follows:

$$\begin{aligned} -h(N_1) &= \mathbb{E}[\ln p_{N_1}(n_1)] \\ &= n_R \mathbb{E}[N_1] - \mathbb{E}[W_1] - \ln(\Gamma(n_R)) \\ &\leq n_R \ln(\mathbb{E}[W_1]) - \mathbb{E}[W_1] - \ln(\Gamma(n_R)) \tag{25} \\ &= n_R \ln(n_R) - n_R - \ln(\Gamma(n_R)), \tag{26} \end{aligned}$$

where (25) follows from Jensen inequality and (26) holds because $\mathbb{E}[W_1] = n_R$.

To show that $h(N_1 - N_2)$ is upper bounded, it suffices to show that $Var(N_1 - N_2) = \mathbb{E}[N_1^2] + \mathbb{E}[N_2^2] - \mathbb{E}[N_1]^2 - \mathbb{E}[N_2]^2$ is bounded, since a Gaussian distribution with same variance upper bounds the differential entropy of any continuous distribution. In other words, it suffices to show that $\mathbb{E}[N_1^2] < \infty$ and $\mathbb{E}[N_2^2] < \infty$.

$$\begin{aligned} \mathbb{E}[N_1^2] &= \mathbb{E}\left[ (\ln(W_1))^2 \right] \\ &\leq \mathbb{E}\left[ (\ln(W_1))^2 \,|\, W_1 \leq 1 \right] \Pr(W_1 \leq 1) \\ &\quad + \mathbb{E}\left[ (\ln(W_1))^2 \,|\, W_1 > 1 \right] \Pr(W_1 > 1) \\ &\leq \mathbb{E}\left[ \frac{1}{W_1} \,|\, W_1 \leq 1 \right] \Pr(W_1 \leq 1) \\ &\quad + \mathbb{E}[W_1 | W_1 > 1] \Pr(W_1 > 1) \tag{27} \\ &\leq \mathbb{E}\left[ \frac{1}{W_1} \right] + \mathbb{E}[W_1] \\ &= \frac{1}{n_R - 1} + n_R \tag{28} \end{aligned}$$

In the above derivation, (27) follows from the fact that $x (\ln(x))^2 \leq 1$ for any $x \in [0, 1]$, whereas (28) holds when

$n_R > 1$ because $\frac{1}{W_1}$ follows inverse Gamma distribution with parameters $\beta = 1$ and $\alpha = n_R$. For $n_R = 1$ we can see that the random variable $N_1$, being the log of an exponential random variable, has density function $p_{N_1}(x) = e^x e^{-e^x}$ with finite mean and variance. By a similar argument, $\mathbb{E}\left[N_2^2\right]$ is finite, and thus $Var(N_1 - N_2)$ is also finite. Now, we have an upper bound on $I(X; Y|Z)$ that is itself bounded irrespective to the power $P$. We conclude that the secret-key capacity is asymptotically bounded at high-SNR. ∎

## V. THE KARUSH-KUHN-TUCKER (KKT) CONDITION

In this section, following [19], a necessary and sufficient condition for optimality in secret-key agreement settings is established. Our proof relies on two steps. First, we show that the supremum in (16) is achieved (Lemma 2). Next, we argue that $I(X; Y|Z)$ is also weak differentiable in $F$ over $\mathcal{F}$ (Lemma 3). The KKT follows then by the concavity of a modified objective function that includes the power constraint. In our proof, we are focussing on the MISOSE case, i.e., $n_R = n_E = 1$.

*Lemma 2:* The supremum in (16) is achievable by at least one $F$, say $F^*$, belonging to $\mathcal{F}^+$, where $\mathcal{F}^+$ is the set of all nonnegative input distributions that meet the power constraint.

*Proof:* A sufficient condition for the supremum in (16) to exist, is that the mutual information $I(X; Y|Z)$ be weak continuous in $F$ and the set $\mathcal{F}^+$ be weak compact. That $\mathcal{F}^+$ is weak compact follows from [19, Appendix 1.A]. The proof of weak continuity of $I(X; Y|Z)$ in $F$ is reported in Appendix B. ∎

*Lemma 3:* $I(X; Y|Z)$ is weak differentiable and concave in $F$ over $\mathcal{F}^+$.

*Proof:* The proof of weak differentiability is presented in Appendix C. The concavity of $I(X; Y|Z)$ in $F$, follows from [24, Fact 2]. ∎

Now, from Lemma 2 and Lemma 3, a necessary and sufficient condition for optimality, referred to as the KKT condition, can be obtained as stated below.

*Proposition 2:* For the channel given by (1) described above, where $n_R = n_E = 1$, an input random variable $X^*$ with distribution function $F^*$ achieves the secret-key capacity $C$ if and only if there exists a $\gamma \geq 0$ such that,

$$\gamma(x^2 - P) + C - \iint p_{Y|X}(y|x)p_{Z|X}(z|x)\ln\left[\frac{p_{Y|X}(y|x)}{p_{Y|Z}(y|z; F^*)}\right] dy\,dz \geq 0 \tag{29}$$

for all $x$, with equality if $x$ belongs to the support of $X^*$.

*Proof:* The proof is presented in Appendix D. ∎

Using (19), (20) and letting $s = \frac{1}{1+x^2}$ with $s \in [0, 1]$, (29) can be expressed as:

$$\gamma\left(\frac{1}{s} - 1 - P\right) + C - \ln(s) + 1 + \iint s^2 e^{-s(y+z)} \ln\left[p_{Y|Z}(y|z; F^*)\right] dy\,dz \geq 0. \tag{30}$$

## VI. CHARACTERIZATION OF $X^*$

Here we follow [19] and [25] to use the Kuhn-Tucker condition (29) to prove that $X^*$ is discrete. Although our framework parallels these previous works, several modifications are necessary to account for the conditional mutual information. The existence of a secret-key capacity achieving input implies that $X^*$ should satisfy one of the following properties:

1) Its support contains an interval;
2) It is discrete, with an infinite number of mass points in some bounded interval;
3) It is discrete and infinite, but with only a finite number of mass points on any bounded interval;
4) It is discrete with a finite number of mass points.

Now, let us assume that 1) or 2) holds and define the function $H(\cdot)$ by:

$$H(z_c) = \gamma\left(\frac{1}{z_c} - 1 - P\right) + C - \ln(z_c) + 1 + \iint z_c^2 e^{-z_c(y+z)} \ln\left[p_{Y|Z}(y|z; F^*)\right] dy\,dz \tag{31}$$

for all $z_c$ belonging to the set of complex numbers, where $\ln(\cdot)$ is the principal branch of the logarithm. We note that $H(z_c)$ is analytic over the domain $D$ defined by $\Re(z_c) > 0$. We now make the following observations:

- From our assumption it is evident that there exists an $A$ such that the support of $X^*$ contains infinitely many points in $[0, A]$ or equivalently the support of $S^*$ contains an infinite set of distinct points in $S_s \subseteq [1/(1 + A^2), 1]$,
- The interval $[1/(1 + A^2), 1]$ is compact, hence by Bolzano-Weierstrass theorem $S_s$ has an accumulation point in $[1/(1 + A^2), 1] \subset [0, 1]$,
- From the Kuhn-Tucker condition (30), $H(z_c) = 0$ on the support of $S^*$ and thus on $S_s$.

Hence we have an analytic function over $D$ that vanishes on a set having an accumulation point of in $D$. From the identity theorem [26], we conclude that $H(z_c) = 0$ over the whole $D$ and in particular, over $z_c \in [0, \infty)$. Consequently, (31) can be written as:

$$\iint e^{-z_c(y+z)} \ln\left[p_{Y|Z}(y|z; F^*)\right] dy\,dz = \frac{-1}{z_c^2}\left[\gamma(\frac{1}{z_c} - 1 - P) + C - \ln(z_c) + 1\right], \tag{32}$$

where $z_c \in D$. We next show that there cannot be a valid conditional probability density function $p_{Y|Z}(\cdot|\cdot)$ that satisfies (32). This is done by multiplying both sides of (32) by $z_c^2$ and by taking the limit $z_c \to \infty$ we show that while the right hand side diverges, the left hand side remains bounded. Towards this end we show the following.

*Lemma 4:* For each $z_c$ with $\Re(z_c) > 0$ we have that

$$\iint e^{-z_c(y+z)} \ln\left[p_{Y|Z}(y|z; F^*)\right] dy\,dz = I_1(z_c) - I_2(z_c), \tag{33}$$

where $I_1(\cdot)$ and $I_2(\cdot)$ denote the Laplace transforms of the following

$$I_1(\cdot) = \mathcal{L}(w \ln f(w)) \tag{34}$$

$$I_2(\cdot) = \mathcal{L}\left(\int_0^w \ln g(\tau)d\tau\right) \tag{35}$$

where we define

$$f(t) = \int_0^1 s^2 e^{-st} dF^*(s) \qquad (36)$$

$$g(t) = \int_0^1 s e^{-st} dF^*(s). \qquad (37)$$

Furthermore $I_1(z_c)$ and $I_2(z_c)$ are well defined for all $z_c$ with $\Re(z_c) > 0$.

*Proof:* The proof of Lemma 4 is provided in Appendix E. ∎

Note that $f(t)$ and $g(t)$ in (34) and (35) are infinitely differentiable decreasing functions over the positive real axis. Furthermore,

$$\lim_{t \to 0} g(t) = g(0) = \mathbb{E}\left[\frac{1}{(1+X^2)}\right] \quad \in \quad [0,1] \qquad (38)$$

Similarly, we have:

$$\lim_{t \to 0} f(t) = \mathbb{E}\left[\frac{1}{(1+X^2)^2}\right] \quad \in \quad [0,1] \qquad (39)$$

$$\lim_{t \to 0} t f'(t) = 0. \qquad (40)$$

Finally, multiplying both sides of (32) by $z_c^2$ and taking the limit as $z_c \to \infty$, we show that the RHS goes to infinity whereas LHS is finite. To see this, we first recall that:

$$\lim_{z_c \to \infty} z_c^2 LHS = \lim_{z_c \to \infty} \left[ z_c^2 \left( I_1(z_c) - I_2(z_c) \right) \right]. \qquad (41)$$

Then, the limit of the second term on the RHS of (41) is equal to:

$$\lim_{z_c \to \infty} z_c^2 I_2(z_c) = \lim_{z_c \to \infty} \left[ z_c \int_0^\infty e^{-z_c w} \ln[g(w)] dw \right]$$
$$= \lim_{w \to 0} [\ln(g(w))] \qquad (42)$$
$$= \ln\left( \mathbb{E}\left[ 1/\left( 1+X^2 \right) \right] \right), \qquad (43)$$

where (42) follows by the Initial Value Theorem [27] and (43) is obtained from (38). Next, we note that the second derivative $(w \ln(f(w)))''$ exists for all positive $w$, that $\lim_{w \to 0} [w \ln(f(w))] = 0$ and that $\lim_{w \to 0} \left[ (w \ln(f(w)))' \right] = \ln\left( \mathbb{E}\left[ 1/\left( 1+X^2 \right)^2 \right] \right)$. Hence, applying the identity (from [27]),

$$\int_0^\infty e^{-z_c t} p''(t) dt = z_c^2 \int_0^\infty e^{-z_c t} p(t) dt - z_c p(0) - p'(0) \qquad (44)$$

to the function $w \ln(f(w))$ and taking the limit as $z_c \to \infty$ on both sides of (44) yield,

$$0 = \lim_{z_c \to \infty} \left[ z_c^2 (I_1(z_c)) \right] - \ln\left( \mathbb{E}\left[ 1/\left( 1+X^2 \right)^2 \right] \right), \qquad (45)$$

which confirms that $\lim_{z_c \to \infty} \left[ z_c^2 (I_1(z_c)) \right]$ is finite too. Therefore, the limit in (41) exists and is finite. This implies that (32) does not hold for all $z_c \in D$. But, this contradicts our initial assumption that either 1) or 2) holds. Consequently, neither 1) nor 2) can happen. We are then left with 3) and 4) as the only possibilities.

Let us assume that 3) holds. In this case we argue that the Lagrange multiplier in the KKT condition (30) is zero and

in turn obtain a contradiction. Since $X^*$ has infinitely many mass points and only finitely many in any bounded interval, $S^*$ has an accumulation point only at zero and its support can thus be written as a sequence $\{s_i\}$ converging to zero. Let $\Pr\{S^* = s_i\} = p_i$. Then we have:

$$p_{Y|Z}(y|z; F^*) = \frac{\sum_{i=1}^\infty p_i p_{Y|S^*}(y|s_i) p_{Z|S^*}(z|s_i)}{\sum_{i=1}^\infty p_i p_{Z|S^*}(z|s_i)} \qquad (46)$$

$$= \frac{\sum_{i=1}^\infty p_i s_i^2 e^{-s_i(y+z)}}{\sum_{i=1}^\infty p_i s_i e^{-s_i z}}. \qquad (47)$$

Note that the denominator of (47) is smaller than $E[S]$ which is less than 1 (c.f. (38)). This implies that $p_{Y|Z}(y|z; F^*) > p_i s_i^2 e^{-s_i(y+z)}$ for all $y, z \geq 0$ and all $i = 1, 2, \ldots$. Consequently, we have:

$$\iint s^2 e^{-s(y+z)} \ln\left[ p_{Y|Z}(y|z; F^*) \right] dy dz$$
$$> \iint s^2 e^{-s(y+z)} \ln\left[ p_i s_i^2 e^{-s_i(y+z)} \right] dy dz \qquad (48)$$
$$= \ln\left( p_i s_i^2 \right) - \frac{2 s_i}{s} \qquad (49)$$

Now, using (49), bound the LHS of (30) as follows:

$$\text{LHS} >$$
$$\gamma(\frac{1}{s} - 1 - P) + C - \ln(s) + 1 + \ln(p_i s_i^2) - \frac{2s_i}{s} \,(50)$$
$$= \frac{\gamma - 2s_i}{s} + o(\frac{1}{s}) \qquad (51)$$

Where the $o(\cdot)$ term applies when $s \to 0$ for a fixed $i$. Now, if $\gamma > 2s_i$ then (51) goes to infinity as $s \to 0$, but the LHS of (30) should be zero on the support of $S^*$ which by our initial assumption, contains a point of accumulation at 0. Hence, $\gamma \leq 2s_i$. As this is true for all $s_i$, and $s_i \to 0$, we see that $\gamma \leq 0$. As the Lagrange multiplier is non-negative, we conclude that $\gamma = 0$. Then from (30) we get,

$$C - \ln(s_i) + 1 + \iint s_i^2 e^{-s_i(y+z)} \ln[p(y|z; F^*)] dy dz = 0 \,(52)$$

for all $s_i$. Now taking the limit at $s_i \to 0$ on both sides of (52), we see that the integrand on the LHS tends toward 0, which implies that

$$C = \lim_{s_i \to 0} \ln(s_i) - 1 \qquad (53)$$

and consequently the capacity $C$ goes to $-\infty$, which contradicts $C \geq 0$. Hence, the assumption 3) is ruled out as well. Therefore, the optimum input distribution must be discrete with a finite number of mass points which is what we wanted to prove.

We conclude the section by enumerating the key differences in the results obtained in [19] with those obtained in this paper. These differences primarily arise because we are dealing with the conditional version of the mutual information compared to [19]:

1) To eliminate the conditions 1) and 2) in [19, Section IV], the authors obtain a unique distribution function $p(y)$ that satisfies the Kuhn-Tucker condition and claim that this function fails to satisfy the properties of a probability distribution function. On the contrary, (32) of this paper may have infinite solutions for $p_{Y|Z}(y|z)$ but
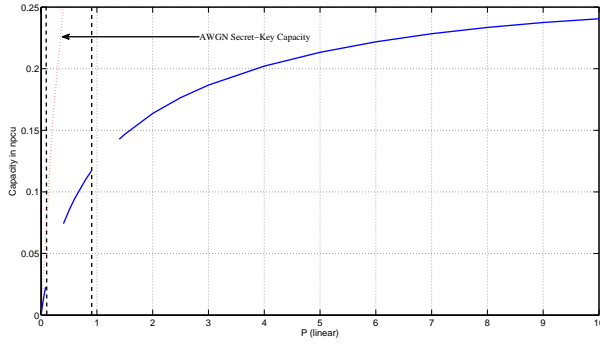
Fig. 1.  The SISOSE secret-key capacity versus the SNR value $P$.

we prove that none of them is a valid solution because (32) does not hold good as $z_c \to \infty$;

2) To disprove 3) in [19, Section IV], the authors establish the impossibility of $\gamma = 0$ by proposing a family of distributions with strictly monotonically increasing mutual information, whereas to disprove (52), we exploit the fact that the capacity is non-negative;

3) In [19], the authors establish that the mutual information is strictly concave in $F$. Our result on concavity relies on [24] which does not establish strict concavity. Consequently we do not pursue the problem of establishing uniqueness of the input distribution in this work.

## VII. NUMERICAL RESULTS

We employed the Gauss-Laguerre quadrature method to evaluate all the concerned integrals in obtaining capacity-achieving input distributions. We obtain some useful insights related to the variation of the number of mass points and their respective probabilities with the SNR. Furthermore we exploit the variation of KKT, a necessary and sufficient condition for optimality, with $x$ to exactly predict the location of a new mass point and to validate the optimal input distribution for a given SNR value. It has been observed that there exists a mass point at the origin for all SNR values. Fig. 1 represents the non-coherent secret-key capacity in nats per channel use (npcu) in function of SNR (average power constraint in Joules per second). As shown in Fig. 1, the secret-key capacity shown in solid blue is monotonically increasing in SNR. In this figure, discontinuities of the capacity plot can be observed. Indeed, these regions represent zones where a new mass point is about to appear and where numerical optimization becomes very unstable to an extent that the results obtained in these zones do not fulfill the KKT condition. Furthermore, it may be seen in Fig. 1, that the capacity is bounded at high-SNR in full agreement with Theorem 1. Also shown in Fig. 1 as a benchmark, is the Additive White Gaussian Noise (AWGN) channel secret-key capacity (in dashed red). While the gap between the two plots is marginal at very-low SNR (below 0.5 J/s), the AWGN secret-key capacity prevails remarkably as SNR increases.

In Fig. 2, the LHS of the KKT condition (29) is plotted versus $x$ for an SNR= 0.09 and where numerical optimization

was set to two mass point input distributions ($N = 2$). From Fig. 2, it can be seen that although the KKT is null in two points, the results obtained is not optimal since the plot goes below zero for a certain value of $x$, suggesting that a new mass point is more likely to appear. In order to confirm our claim, we set $N = 3$ in our optimization problem and increase the power constraint away from this unstable zone, to find that three mass point is in fact optimal and that a new mass point shows up around $x = 3$ at approximately $P = 0.4$. This explains the discontinuities in our capacity plot as depicted by Fig. 1. We conjecture that a new mass point shows up first at $x \to \infty$ and then decreases as SNR increases. Note that this peculiar behavior of the optimal input distribution has also been observed previously in non-coherent fading channels without secrecy [19].

Fig. 3 depicts optimal non zero mass point locations versus SNR. Likewise in Rayleigh fading channel without secrecy, two mass point is optimal at low SNR (below $P = 0.08$). As SNR increases, the number of mass point increases gradually to be $N = 3$ for $P$ below 0.9 and then $N = 4$ for $P$ below 10. At high-SNR, we observe that the optimal distribution has 3 non zero mass points and more interestingly as SNR increases, only the biggest of the three tends to increase whereas the two others of lower values tend to attain constant values of approximately 1.9 and 3.5, respectively.

Finally, the non zero mass point probability versus $P$ is shown in Fig. 4 at low-SNR, where it can be seen that the non zero mass point probability seems to increase almost linearly with SNR. On the other hand, the non zero mass point location versus $P$ is also displayed in Fig. 5 at low-SNR, where it can be observed that as SNR increases the non zero mass point decreases in magnitude.

## VIII. CONCLUSION

The secret-key capacity under an average power constraint of a Rayleigh fading channel, where the instantaneous CSI is not available at any terminal, has been studied. When the legitimate receiver and the eavesdropper have one antenna each, i.e., the MISOSE ($n_R = n_E = 1$) setting, it has been shown that the capacity-achieving input distribution is discrete with a finite number of mass points. Although in this case we have focused on the MISOSE case, our proof technique can be extended in a straightforward manner to encompass a channel where the number of receive antennas at both the legitimate receiver and the eavesdropper are arbitrary, using techniques similar to the ones discussed in [19, Appendix III]. At high-SNR, it is established that the secret-key capacity is bounded irrespective of SNR, regardless of the number of antennas at each terminal. At low-SNR, it has been observed through numerical results that an input distribution with two mass points, one of them at the origin, is optimal.

## APPENDIX A
## PROOF OF (16)
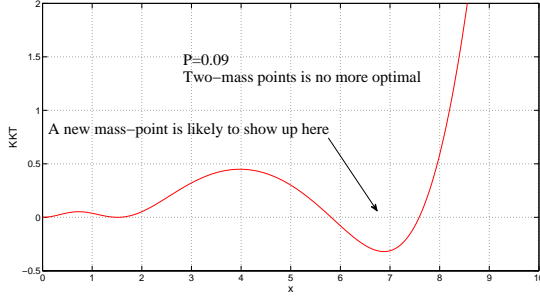
$$I(X; Y|Z) = I(X; Y, Z) - I(X; Z)$$

Fig. 2. The value of the LHS of the KKT condition (29) versus $x$ at SNR=0.09.



Fig. 3. The optimal non zero mass point locations versus the SNR value $P$.
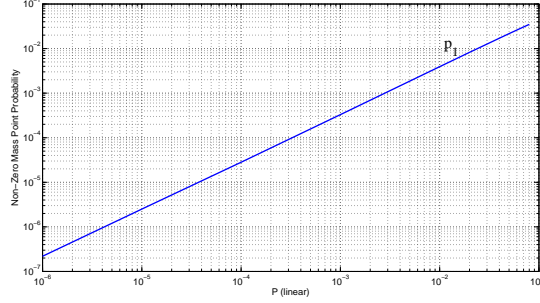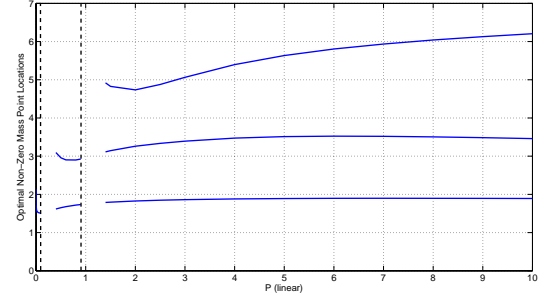


Fig. 4. The non zero mass point probability versus the SNR value $P$ at very low SNR region.



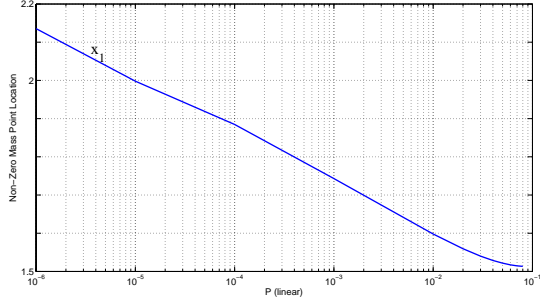Fig. 5. The non zero mass point location versus the SNR value $P$ at very low SNR region .

$$=h(Y,Z) - h(Y,Z|X) - h(Z) + h(Z|X)$$
$$=h(Y|Z) - h(Y|X) \qquad (54)$$
$$=\iint p_{Y,X}(y,x) \ln\big(p_{Y|X}(y|x)\big) dy dx$$
$$\quad - \iint p_{Y,Z}(y,z) \ln\big(p_{Y|Z}(y|z)\big) dy dz$$
$$=\iiint p_{Y,X,Z}(y,x,z) \ln\big(p_{Y|X}(y|x)\big) dy dx dz$$
$$\quad - \iiint p_{Y,Z|X}(y,z|x) \ln\big(p_{Y|Z}(y|z;F)\big) dy dz dF(x)$$
$$=\iiint p_{Y,Z|X}(y,z|x) \ln\big(p_{Y|X}(y|x)\big) dy dF(x) dz$$
$$\quad -\iiint p_{Y|X}(y|x)p_{Z|X}(z|x) \ln\big(p_{Y|Z}(y|z)\big) dy dz dF(x) \quad (55)$$
$$=\iiint p_{Y|X}(y|x)p_{Z|X}(z|x) \ln\big(p_{Y|X}(y|x)\big) dy dF(x) dz$$
$$\quad -\iiint p_{Y|X}(y|x)p_{Z|X}(z|x) \ln\big(p_{Y|Z}(y|z)\big) dy dz dF(x) \quad (56)$$

where to obtain (54), (55) and (56), we used the fact that given $X$, $Y$ and $Z$ are independent. Here $p_{Y|Z}(y|z;F)$ signifies the conditional distribution of Y given Z induced by F.

## APPENDIX B
## WEAK CONTINUITY OF THE CONDITIONAL MUTUAL INFORMATION

From [28, Section 5.10], we first note that if a function $f$ is weak* continuous on a weak* compact set $\mathcal{F}^+$, then $f$ achieves its maximum on $\mathcal{F}^+$. Since it has been established that $\mathcal{F}^+$ is convex and compact in [19, Appendix 1.A], we only need to show that the mutual information is weak continuous on $\mathcal{F}^+$ to prove the existence of a capacity achieving input

distribution. Recall that from (11), we have:
$$I(X;Y|Z) = -h_{Y|X}(F) + h_{Y|Z}(F)$$
$$= -h_{Y|X}(F) + h_{Y,Z}(F) - h_Z(F).$$

That $h_{Y|X}(F)$ and $h_Z(F)$ are weak* continuous in $F$ follows along similar lines as the proof in [19, Appendix 1.B]. We now prove that $h_{Y,Z}(F)$ and $h_Z(F)$ are weak* continuous in $F$. Suppose the exists a sequence $\{F_n\}$ that converges to $F$, which will be referred to as $F_n \xrightarrow{w^*} F$. First, recall that

$$p_{Y,Z}(y,z;F) = \int \frac{1}{(1+x^2)^2} e^{-(y+z)/(1+x^2)} dF(x). \quad (57)$$

Since the integrand in (57) is a bounded continuous function of $x$, then by the definition of weak* topology, $p_{Y,Z}$ is a continuous function of $F$ for all $y, z \geq 0$. Moreover, $x \ln(x)$ is continuous and thus $p_{Y,Z}(y,z;F) \ln[p_{Y,Z}(y,z;F)]$ is also continuous in $F$. Hence by the definition of weak continuity, we have:

$$\lim_{n\to\infty} p_{Y,Z}(y,z;F_n) \ln[p_{Y|Z}(y|z;F_n)]$$
$$= p_{Y,Z}(y,z;F) \ln[p_{Y,Z}(y,z;F)] \quad (58)$$

We now claim that:

$$\lim_{n\to\infty} h_{Y,Z}(F_n)$$
$$= -\lim_{n\to\infty} \iint p_{Y,Z}(y,z;F_n) \ln[p_{Y,Z}(y,z;F_n)] dy dz$$
$$= -\iint \lim_{n\to\infty} p_{Y,Z}(y,z;F_n) \ln[p_{Y,Z}(y,z;F_n)] dy dz \quad (59)$$
$$= -\iint p_{Y,Z}(y,z;F) \ln[p_{Y,Z}(y,z;F)] dy dz \quad (60)$$
$$= h_{Y,Z}(F)$$

where (60) follows from (58). We next justify the interchange of the limit and the integral in (59). To do this, we note that by Lebesgue dominated convergence theorem, it suffices to find an integrable function $g$ such that

$$p_{Y,Z}(y,z;F_n)\ln\left[p_{Y,Z}(y,z;F_n)\right] \leq g(y,z), \qquad (61)$$

for all $F_n$. Let $g(y) = \min\left(1, 1/y^2\right)$. First, note that both $p_{Y|X}(y|x) \leq g^2(y)$ and $p_{Z|X}(z|x) \leq g^2(z)$. Thus, by (57), we also have $p_{Y,Z}(y,z) \leq g^2(y)g^2(z)$ for all $x$, $y$, $z$ and $F_n$. Using the fact that $|x\ln(x)|^2 \leq x$ for all $x \in [0,1]$, the following derivation holds:

$$|p_{Y,Z}(y,z;F_n)\ln\left[p_{Y,Z}(y,z;F_n)\right]|^2 \leq p_{Y,Z}(y,z;F_n)$$
$$= \int p_{Y,Z|X}(y,z|x)dF_n(x)$$
$$\leq \int g^2(y)g^2(z)dF_n(x)$$
$$= g^2(y)g^2(z)$$

Since $g$ is integrable over $[0,\infty)$, (61) is true. Consequently, $h_{Y,Z}(F)$ is weak* continuous. By a similar argument, $h_Z(F)$ is also weak* continuous. Hence being a sum of weak* continuous functions, $I(X;Y|Z)$ is weak* continuous in $F$.

## APPENDIX C
### WEAK DIFFERENTIABILITY OF THE CONDITIONAL MUTUAL INFORMATION

The idea of weak differentiability has been previously used in [19] in channels without secrecy constraint. Here, we extend this idea to show that the conditional mutual information is weak differentiable. Let us first define the mapping

$$g(F) = \int x^2 dF(x) - P$$

from $\mathcal{F}^+$ to $\mathbb{R}$, the set of real numbers. Note that $g(F)$ is linear and hence convex in $F$. Since $I(F)$ is also convex, there exists a $\gamma \geq 0$ in $\mathbb{R}$ such that

$$C = \sup_{F\in\mathcal{F}^+} \left[I(F) - \gamma g(F)\right]. \qquad (62)$$

Moreover, since this capacity is achieved by some $F_0$, then we necessarily have:

$$\gamma g(F_0) = 0 \qquad (63)$$

We next prove that the functions $I$ and $g$ are weak differentiable in $F$. Let us define:

$$F_\theta = (1-\theta)F_0 + \theta F$$
$$i(x;F) = \iint p_{Y|X}(y|x)p_{Z|X}(z|x)\ln\left[\frac{p_{Y|X}(y|x)}{p_{Y|Z}(y|z;F)}\right]dydz$$

Computing $(I(F_\theta) - I(F_0))$, we get:

$$I(F_\theta) - I(F_0) =$$
$$\iiint p_{Y|X}(y|x)p_{Z|X}(z|x)\ln\left[\frac{p_{Y|Z}(y|z;F_0)}{p_{Y|Z}(y|z;F_\theta)}\right]dydzdF_0$$
$$+ \theta \iint i(x;F_\theta)dF(x)$$
$$- \theta \iint i(x;F_\theta)dF_0(x).$$

Next we see that

$$p_{Y,Z}(y,z;F_\theta) = (1-\theta)p_{Y,Z}(y,z;F_0) + \theta p_{Y,Z}(y,z;F)$$
$$p_Z(z;F_\theta) = (1-\theta)p_Z(z;F_0) + \theta p_Z(z;F).$$

Then from the definition of weak differentiability [19, Appendix II-B], we have:

$$I'_{F_0}(F) = \lim_{\theta\downarrow 0}\frac{I(F_\theta) - I(F_0)}{\theta}$$
$$= \int i(x;F_0)dF(x) - I(F_0), \qquad (64)$$

which holds because both terms are finite due to the power constraint. Similarly for $g$ we obtain:

$$g'_{F_0}(F) = g(F) - g(F_0) \qquad (65)$$

Since (64) and (65) are valid for any $F_0$ and $F$ in $\mathcal{F}^+$, we conclude that $I$ and $g$ are weak differentiable and so is $I-\gamma g$.

## APPENDIX D
### THE KKT CONDITION

By Lemma 3, $I$ is concave in $F$. Furthermore, $g$ is linear in $F$ and hence $I - \gamma g$ is concave in $F$. Now from [19, Theorem 3], a necessary and sufficient condition for $F_0$ to achieve the supremum in (62) is

$$I'_{F_0}(F) - \gamma g'_{F_0}(F) \leq 0 \qquad \forall F \in \mathcal{F}^+, \qquad (66)$$

or equivalently,

$$\int \left[i(x;F_0) - \gamma x^2\right]dF(x) \leq I(F_0) - \gamma \int x^2 dF_0(x)$$
$$= C - \gamma\left(g\left(F_0\right) + P\right)$$
$$= C - \gamma P, \qquad (67)$$

where (67) follows from or (63). Now we conclude the section by noting that from [19, Theorem 4], (67) is also equivalent to:

$$\begin{cases} i(x;F_0) \leq C + \gamma(x^2 - P) & \forall x \\ i(x;F_0) = C + \gamma(x^2 - P), & \forall x \in \mathbb{E}_0, \end{cases} \qquad (68)$$

where $\mathbb{E}_0$ is the set of points of increase of $F_0$. From (68), the KKT condition (29) follows immediately.

## APPENDIX E
### PROOF OF LEMMA 4

The probability $p_{Y|Z}(y|z)$ can be written as

$$p_{Y|Z}(y|z;F^*) = \frac{p_{Y,Z}(y,z;F^*)}{p_Z(z;F^*)}$$
$$= \frac{f(y+z)}{g(z)}, \qquad (69)$$

where it follows via (9) that when $n_R = n_E = 1$, the functions $f(\cdot)$ and $g(\cdot)$ in (69) are defined by (36) and (37) respectively.

Thus we can split the left hand side (LHS) of (26) into two parts as stated in (33)

$$\iint e^{-z_c(y+z)}\ln\left[p_{Y|Z}(y|z;F^*)\right]dydz = I_1\left(z_c\right) - I_2\left(z_c\right),$$

where $I_1(z_c) = \iint e^{-z_c(y+z)} \ln [f(y+z)] dy dz$ and $I_2(z_c) = \iint e^{-z_c(y+z)} \ln [g(z)] dy dz$. Now, transforming the coordinate system from Y-Z to W-Z where W = Y + Z, we find that the Jacobian of the transformation is 1 and hence $I_1$ can be written as:

$$I_1(z_c) = \int_0^\infty \int_z^\infty e^{-z_c w} \ln [f(w)] dw dz \qquad (70)$$

$$= \int_0^\infty \int_0^w e^{-z_c w} \ln [f(w)] dz dw \qquad (71)$$

$$= \int_0^\infty w e^{-z_c w} \ln [f(w)] \, dw \qquad (72)$$

On simplification of $I_2(z_c)$, we also get:

$$I_2(z_c) = \frac{1}{z_c} \int_0^\infty e^{-z_c w} \ln [g(w)] dw \qquad (73)$$

Note that (72) and (73) represent the Laplace transforms of $w \ln [f(w)]$ and $\int_0^w \ln [g(\tau)] d\tau$, respectively. Since the integrands in (72) and (73) are integrable over the interval $[0, a]$ for all $a > 0$, then to prove that $I_1(z_c)$ and $I_2(z_c)$ are well-defined for all $\Re(z_c) > 0$, it suffices to show that $|we^{-z_c w} \ln [f(w)]|$ and $|e^{-z_c w} \ln [g(w)]|$ are bounded by some integrable functions on $[0, \infty]$. This can be established as follows:

$$|we^{-z_c w} \ln f(w)|$$
$$= -we^{-\Re(z_c)w} \ln (f(w)) \qquad (74)$$
$$= -we^{-\Re(z_c)w} \ln \left( \mathbb{E}[S^2 e^{-Sw}] \right)$$
$$\leq -we^{-\Re(z_c)w} \mathbb{E}[\ln (S^2 e^{-Sw})] \qquad (75)$$
$$= -we^{-\Re(z_c)w} \mathbb{E}[\ln (S^2) - Sw]$$
$$= we^{-\Re(z_c)w} \left( w\mathbb{E}\left[\frac{1}{1+X^2}\right] + 2\mathbb{E}[\ln (1 + X^2)] \right)$$
$$\leq we^{-\Re(z_c)w} (w + 2\ln (1 + P)) \qquad (76)$$

(74) holds because $f(w)$ is a decreasing function over $w \geq 0$ and $f(0) < 1$ (c.f. (39)); (75) follows from Jensen inequality; (76) is true because $\frac{1}{1+X^2} \leq 1$ and applying Jensen inequality again. Since $\int_0^\infty e^{-\Re(z_c)w} (w + 2\ln (1 + P)) dw$ exists for all $\Re(z_c) > 0$, then so does $I_1(z_c)$. By a similar technique, the following upper bound may be obtained:

$$|e^{-z_c w} \ln g(w)| \leq e^{-\Re(z_c)w} (w + \ln (1 + P)) \qquad (77)$$

to justify the convergence of the integral in (73). We conclude that $I_2(z_c)$ is also well-defined for all $\Re(z_c) > 0$.

## REFERENCES

[1] R. Ahlswede and I. Csiszar, "Common Randomness in Information Theory and Cryptography- Part I: Secret Sharing," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1121 –1132, jul 1993.

[2] U. Maurer, "Secret Key Agreement by Public Discussion from Common Information ," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 733 –742, may 1993.

[3] T. F. Wong, M. Bloch, and J. M. Shea, "Secret Sharing over Fast-Fading MIMO Wiretap Channels," *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, pp. 1–1, 2009.

[4] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. McLaughlin, "Wireless Information-Theoretic Security," *IEEE Transactions on Information Theory, Special Issue on Information-Theoretic Security*, vol. 54, no. 6, pp. 2515–2534, June 2008.

[5] V. Prabhakaran, K. Eswaran, and K. Ramchandran, "Secrecy via sources and channels - A Secret Key - Secret Message Rate Trade-off Region," *IEEE Trans. Information Theory, [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04595139*.

[6] L. Xiao, L. Greenstein, N. Mandayam, and W. Trapp, "Fingerprints in the ether: Using the physical layer for wireless authentication," *in IEEE Int.. Conf. on Comm*, pp. 4646–4651, 2007.

[7] A. A. Gohari and V. Anantharam, "Information-theoretic key agreement of multiple terminals - part I: Source model," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3997–4010, 2010.

[8] A. E. Gamal and Y. H. Kim, " Lecture Notes on Network Information Theory," *Available: http://arxiv.org/abs/1001.3404*.

[9] T. Aono, K. Higuchi, T. Ohira, B. Komiyama, and H. Sasaoka, "Wireless secret key generation exploiting reactance-domain scalar response of multipath fading channel," *IEEE Transactions on Antennas and Propagation*, pp. 3776–3784, 2005.

[10] C. Ye, A. Reznik, and Y. Shah, "Extracting secrecy from jointly Gaussian random variables," *in Proc. Int. Symp. Inform. Theory*, June 2006.

[11] R. Wilson, D. Tse, and R. Scholtz, "Channel Identification: Secret Sharing Using Reciprocity in Ultrawideband Channels," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 364 –375, sept. 2007.

[12] Y. Liang, H. V. Poor, and S. S. (Shitz), "Secrecy Capacity Region of Fading Broadcast Channels," *in Proc. IEEE Int. Symp. Information Theory, Nice, France*, pp. 1291–1295, June 2007.

[13] A. O. Hero, "Secure space-time communication," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3235–3249, Dec. 2008.

[14] A. Khisti and G. W. Wornel, "Secure transmission with multiple antennas I: The MISOME wiretap channel," *IEEE Trans. Information Theory*, vol. 56, 2009.

[15] M. H. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, May 1983.

[16] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, pp. 289–293, Oct 1958.

[17] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, "Radiotelepathy: extracting a secret key from an unauthenticated wireless channel," *in 14th ACM international conference on Mobile computing and networking*, pp. 128–139, 2008.

[18] Y. Chia and A. E. Gamal, "Wiretap channel with causal state information," *in Proc. Int. Symp. Inform. Theory*, June 2010.

[19] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The Capacity of Discrete-Time Memoryless Rayleigh-Fading Channels," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[20] R. Perera, T. Pollock, and T. Abhayapala, "Upper bound on non-coherent MIMO channel capacity in Rayleigh fading," *2005 Asia Pacific Conference on Communications*, pp. 72–76, 2005.

[21] T. Marzetta and B. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139 –157, jan 1999.

[22] R. Perera, T. Pollock, and T. Abhayapala, "Performance of Gaussian Distributed Input in Non-Coherent Rayleigh Fading MIMO Channels," *2005 Fifth International Conference on Information, Communications and Signal Processing*, pp. 791 –795, 2005.

[23] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *Electronics Letters- IEE*, vol. 33, no. 16, pp. 1368–1369, July 1997.

[24] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure Broadcasting Over Fading Channels," *Information Theory, IEEE Transactions on*, vol. 54, no. 6, pp. 2453–2469, 2008. [Online]. Available: http://dx.doi.org/10.1109/TIT.2008.921861

[25] S. Shamai and I. Bar-David, "The Capacity of Average and Peak-Power-Limited Quadrature Gaussian Channels," *IEEE Transactions on Information Theory*, vol. 41, no. 4, pp. 1060 –1071, jul 1995.

[26] H. Silverman, Complex Variables. Boston, MA: Houghton Mifflin, 1975.

[27] Davies Brian , Integral Transforms and their Applications, 3rd ed. New York: Springer, 2002.

[28] D. G. Luenberger, Optimization by Vector Space Methods. New York: Wiley, 1969.