Secret-Key Agreement over Non-Coherent Block-Fading Channels with Public Discussion

Ashish Khisti *Member, IEEE*

Abstract

Fundamental limits of secret-key generation over a two-way, reciprocal and block fading wireless channel are investigated. Only statistical channel state information (CSI) of the main channel is assumed to be available, whereas the eavesdropper has perfect CSI of its own channel. We establish upper and lower bounds on the secret-key generation capacity with public discussion. The upper bound can be expressed as a sum of two terms — one of the terms arises due to channel reciprocity, while the other term arises due to the communication. In the limit of long coherence period, the contribution from channel reciprocity vanishes to zero, whereas the other term prevails. The lower bound involves a separation based scheme consisting of channel training followed by source emulation in each coherence block. The resulting rate also consists of the contribution from each phase. For Rayleigh fading channels, in the high signal-to-noise ratio (SNR) regime, the gap between the upper and lower bounds decreases inversely with the coherence period. Numerical results indicate significant performance gains over training-only schemes even for moderate values of SNR and small coherence periods.

Index Terms

Information Theoretic Security, Secret-Key Agreement, Wireless Fading Channels, Two-Way Channels

I. Introduction

There has been a growing interest in using the physical layer for enhancing security in wireless systems in recent years. One area that has received significant interest (see e.g., [1]–[5]) is secret-key generation using channel reciprocity. When the legitimate terminals use the same carrier frequency for transmission, the phenomenon of channel reciprocity [6] applies. The channel measurements are strongly correlated and can be used to generate a common secret key. Channel reciprocity may not however exist in frequency division duplex (FDD) systems when the carrier frequencies used are different. In such systems secret-key generation techniques based on interactive communication between the legitimate terminals [7], [8] may be used. Such approaches rely on independence in channel fluctuations or noise between the legitimate receiver and eavesdropper channels, which may either exist naturally or can be engineered artificially.

A. Khisti is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: akhisti@comm.utoronto.ca).

A portion of the work was presented at the 50th Allerton Conference on Communications, Control and Computing, 2012, Montecillo, IL

In this paper we study secret-key generation over a two-way, block-fading channel with an external public discussion channel. We assume that in each coherence period the channel gains on the main channel are sampled from a known joint distribution, but the actual realization of the channel gains is not revealed to any of the terminals. In contrast the channel gains of the eavesdropper are perfectly known only to this terminal. We establish upper and lower bounds on the secret-key capacity. Our upper bound expression can be interpreted as a sum of two components. The first component is due to the correlation between the channel gains on the forward and reverse channel. The second component is due to interactive communication. The contribution of the first component vanishes inversely with the coherence period whereas the contribution of the second component remains bounded in this limit. We propose a two-phase coding scheme consisting of channel training followed by source-emulation in each coherence block. The achievable rate also admits an interpretation similar to the upper bound. Through suitable power allocation between the training and source emulation phases for Rayleigh fading channels, we show that the achievable rate approaches the upper bound in the limit of high SNR and long coherence blocks. Numerical evaluations further indicate significant improvements over training-only schemes.

In related works, information theoretic treatment of secret-key generation was first introduced in [9], [10] and has been extended in a number of recent works, see e.g., [11]–[15]. Motivated by applications to wireless channels references [16], [17] study secret-key generation over a state-controlled broadcast channel However the cost of acquiring CSI is not accounted for in these setups. Secret-key generation over a fast fading Rayleigh channel in the forward direction and a public discussion channel in the reverse direction have been studied in [18], [19] for the case of coherent channels and in [20] for the case of non-coherent channels. Note that these works only consider one-way fading channels and do not account for the role of channel reciprocity in secret-key generation, which has been the focus of practical approaches. The most closely related work to the present paper is reference [21], where the authors study a two-way setup with perfect reciprocity over the main channel. The authors propose a coding technique consisting of channel training and secret-message transmission. Unfortunately the secret-message rate is non-zero only when the eavesdropper's channel gain is, on average, weaker than the legitimate receiver's channel. In contrast the scheme presented in the present paper yields improvement over the training based scheme even when the eavesdropper's channel is stronger. Furthermore assuming *imperfect* reciprocity over the main channel the proposed paper also presents an upper bound, which is tight in certain cases. Secret-key generation is also an integral component in the delay-limited secrecy framework [22], [23]. The ergodic behaviour of the channel is exploited to maintain a steady buffer of the secret-key whereas each transmitted message is secured using a one-time pad. The role of channel sparsity in secret-key generation is studied in [24], [25].

The non-coherent capacity of fading channels has been extensively studied in many previous works, see, e.g., [26]–[28] and references therein. However to the best of our knowledge these models only consider one-way channels. As such, the use of interactive communication for channel-training and power control in reciprocal fading channel even in the context of classical communication has been considered in only a limited set of papers e.g., [29], [30] and remains a fertile area of research. One key assumption made in the present paper is that the channel gains in the forward and reverse channel gains are correlated, with the correlation coefficient assumed to be in the interval

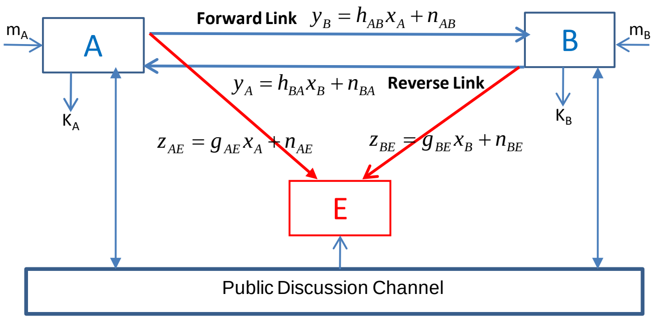

Forward Link $y_B = h_{AB}x_A + n_{AB}$
m_A
A
B
m_B
$y_A = h_{BA}x_B + n_{BA}$ Reverse Link
K_A
K_B
$z_{AE} = g_{AE}x_A + n_{AE}$
$z_{BE} = g_{BE}x_B + n_{BE}$
E
Public Discussion Channel

Fig. 1. Problem Setup. A two-way fading channel with two legitimate terminals and one eavesdropper terminal. The legitimate terminals A and B are required to agree on a common secret-key. The fading gains h_{AB} and h_{BA} are not revealed to any terminals, whereas the g_{AE} and g_{BE} are revealed to the eavesdropper. In addition to the wireless fading channel, terminals A and B can also exchange messages over an external public discussion channel, which are revealed to E.

$[0, 1)$. In general achieving perfect channel reciprocity in baseband is challenging because different terminals use different I/Q mixers, amplifiers and path lengths in the RF chains. While closed-loop calibration can be performed (see e.g. [31], [32]), such methods can become challenging if the calibration needs to be done in the open air. Hence we believe that our assumption of imperfect reciprocity could be more realistic. We note in advance that while our proposed coding technique can also be applied to the case of perfect reciprocity, the upper bound becomes degenerate in this special case.

II. System Model and Main Results

The problem setup is described in Fig. 1. Nodes, A and B are legitimate terminals, whereas node E is the eavesdropper. On the forward channel, at time $t \in \{1, 2, \ldots, N\}$, node A transmits a (complex valued) symbol $x_A(t) \in \mathbb{C}$, and nodes B and E observe $y_B(t)$ and $z_{AE}(t)$ as follows:

$$y_B(t) = h_{AB}(t)x_A(t) + n_B(t), \tag{1}$$

$$z_{AE}(t) = g_{AE}(t)x_A(t) + n_{AE}(t). \tag{2}$$

On the reverse channel, at time t, node B transmits a symbol $x_B(t) \in \mathbb{C}$, and nodes A and E observe $y_A(t)$ and $z_{BE}(t)$ as follows:

$$y_A(t) = h_{BA}(t)x_B(t) + n_A(t), \tag{3}$$

$$z_{BE}(t) = g_{BE}(t)x_B(t) + n_{BE}(t). \tag{4}$$

We assume that all the additive noise variables in (1)-(4) are mutually independent and sampled i.i.d. from

$\mathcal{CN}(0,1)$. The input symbols $x_A(t)$ and $x_B(t)$ satisfy an average power constraint i.e.,

$$E\left[\frac{1}{N}\sum_{t=1}^{N}|x_A(t)|^2\right]\leq P, \quad E\left[\frac{1}{N}\sum_{t=1}^{N}|x_B(t)|^2\right]\leq P. \tag{5}$$

We assume a block-fading channel model with a coherence period of T symbols and assume that the communication spans K coherence blocks i.e., $N=KT$. Thus for each $i\in\{0,1,\ldots,K-1\}$, the coherence block i spans the interval $[iT+1,(i+1)T]$. At time $t=iT+1$, the channel gains, $h_{AB}(t)$, $h_{BA}(t)$, $g_{AE}(t)$ and $g_{BE}(t)$, are sampled from a joint distribution

$$p_{h_{AB},h_{BA}}(h_{AB},h_{BA})\cdot p_{g_{AE},g_{BE}}(g_{AE},g_{BE}) \tag{6}$$

and remain constant in the interval $[iT+1,(i+1)T]$. Note that we assume that the channel gains of the eavesdropper are independent of the channel gains over the main channel. This assumption is only made to simplify the rate expressions. We believe that the results can be easily extended to the case where all the channel gains are drawn from some joint distribution, although the bounds will be more involved.

In the proposed setup, the eavesdropper observes the forward and reverse channel outputs over non-interfering channels (c.f. (2) and (4)). This is clearly a stronger model than the setup when only a superposition of $x_A(t)$ and $x_B(t)$ is observed by the eavesdropper. Note that nodes A and B are assumed to operate in a full-duplex mode i.e., they can transmit and receive simultaneously. Our results can also be easily generalized if the terminals operate in a half-duplex mode, provided that the transmission schedules are fixed (e.g., if the nodes alternate during transmission).

We will also assume that a public discussion channel is available for communication. After the transmission of $x_A(t)$, and before the transmission of $x_A(t+1)$ on the forward channel (and likewise after the transmission of $x_B(t)$ and before the transmission of $x_B(t+1)$) the terminals can exchange L rounds of messages over a separate *discussion channel*: $\boldsymbol{\psi}_A(t)\triangleq\{\psi_A(t,1),\ldots,\psi_A(t,L)\}$ and $\boldsymbol{\psi}_B(t)\triangleq\{\psi_B(t,1),\ldots,\psi_B(t,L)\}$, where

$$\psi_A(t,j)=\Psi_A(y_A^t,\boldsymbol{\psi}_B^{t-1},\psi_B(t,1),\ldots,\psi_B(t,j-1),m_A) \tag{7}$$

$$\psi_B(t,j)=\Psi_B(y_B^t,\boldsymbol{\psi}_A^{t-1},\psi_A(t,1),\ldots,\psi_A(t,j-1),m_B), \tag{8}$$

denote the messages transmitted by nodes A and B respectively. The messages transmitted over the discussion channel are also revealed to the eavesdropper. The transmitted symbols $x_A(t)$ and $x_B(t)$ can be expressed as:

$$x_A(t)=f_{A,t}(y_A^{t-1},m_A,\boldsymbol{\psi}_B^{t-1}) \tag{9}$$

$$x_B(t)=f_{B,t}(y_B^{t-1},m_B,\boldsymbol{\psi}_A^{t-1}). \tag{10}$$

The secret-key at the end of communication is generated as follows:

$$k_A=\mathcal{K}_A(m_A,y_A^N,\boldsymbol{\psi}_A^N,\boldsymbol{\psi}_B^N), \quad k_B=\mathcal{K}_B(m_B,y_B^N,\boldsymbol{\psi}_A^N,\boldsymbol{\psi}_B^N). \tag{11}$$

A secret-key rate $R=\frac{1}{N}H(k_A)$ is achievable if for every $\varepsilon>0$ there exists a N sufficiently large such that $\Pr(k_A\neq k_B)\leq\varepsilon$ and

$$\frac{1}{N}I(k_A;z_A^N,z_B^N,\boldsymbol{\psi}_A^N,\boldsymbol{\psi}_B^N,g_{AE}^N,g_{BE}^N)\leq\delta(\varepsilon), \tag{12}$$

for some function $\delta(\varepsilon)$ that goes to zero as $\varepsilon \to 0$.

The largest achievable rate is the secret-key capacity which is the quantity of interest. We next present upper and lower bounds on the secret-key capacity.

A. Upper Bound

We present the following upper bound on the secret-key capacity.

Theorem 1: An upper bound on the secret-key agreement capacity, with (or without) public discussion is given by the following:

$$R^{+} = \frac{1}{T} I(\mathsf{h}_{AB}; \mathsf{h}_{BA}) + R^{+}_{AB} + R^{+}_{BA} \tag{13}$$

where R^{+}_{AB} and R^{+}_{BA} are defined as follows:

$$R^{+}_{AB} = \max_{P(\cdot)\in\mathcal{P}_F} E\left[\log\left(1 + \frac{P(\mathsf{h}_{AB})|\mathsf{h}_{AB}|^2}{1 + P(\mathsf{h}_{AB})|\mathsf{g}_{AE}|^2}\right)\right] \tag{14}$$

$$R^{+}_{BA} = \max_{P(\cdot)\in\mathcal{P}_R} E\left[\log\left(1 + \frac{P(\mathsf{h}_{BA})|\mathsf{h}_{BA}|^2}{1 + P(\mathsf{h}_{BA})|\mathsf{g}_{BE}|^2}\right)\right], \tag{15}$$

where the maximization in (14) is over the set $\mathcal{P}_F$ of all power-allocation functions on the forward channel that satisfy the average power constraint $E[P(\mathsf{h}_{AB})] \leq P$, and the maximization in (15) is defined similarly.

□

An interpretation of the upper bound (13) is as follows. The term $I(\mathsf{h}_{AB}; \mathsf{h}_{BA})$ denotes the contribution from channel correlation. The scaling factor $\frac{1}{T}$ is due to the fact that the channel gains remain constant over the duration of one coherence block. The other two terms R^{+}_{AB} and R^{+}_{BA} denote the contribution to the secret-key from the forward and reverse channels using public discussion respectively [9], [10]. The upper bound expression (13) indicates that the total secret-key rate cannot be larger than the sum of these three terms.

B. Lower Bound

Our lower bound involves a separation based scheme. In each coherence block we reserve the first symbol for channel training and use the remainder of the block for source emulation [9], [10] .

Theorem 2: An achievable secret-key rate with a public discussion channel is given by:

$$R^{-} = \max_{P_1,P_2} \left\{ \frac{1}{T} I(\hat{\mathsf{h}}_{AB}; \hat{\mathsf{h}}_{BA}) + \frac{T-1}{T}\left(R^{-}_{AB} + R^{-}_{BA}\right)\right\} \tag{16}$$

where we have introduced,

$$\hat{\mathsf{h}}_{AB} \triangleq \mathsf{h}_{AB} + \frac{1}{\sqrt{P_1}} \hat{\mathsf{n}}_A \tag{17}$$

$$\hat{\mathsf{h}}_{BA} \triangleq \mathsf{h}_{BA} + \frac{1}{\sqrt{P_1}} \hat{\mathsf{n}}_B \tag{18}$$

to be noisy observations of the channel gains on the forward and reverse channels respectively, $\hat{n}_A \sim \mathcal{CN}(0,1)$ and $\hat{n}_B \sim \mathcal{CN}(0,1)$ are independent of all other random variables and

$$R^-_{AB} = E\left[\log\left(1+\frac{P_2|h_{AB}|^2}{1+P_2|g_{AE}|^2}\right)\right] - \log\left(1+\frac{P_2}{1+P_1}\right) \tag{19}$$

$$R^-_{BA} = E\left[\log\left(1+\frac{P_2|h_{BA}|^2}{1+P_2|g_{BE}|^2}\right)\right] - \log\left(1+\frac{P_2}{1+P_1}\right), \tag{20}$$

where P_1 and P_2 are non-negative constants that satisfy

$$P_1 + (T-1)P_2 \le TP. \tag{21}$$

□

The interpretation of the achievable rate is very similar to the upper bound. The first term in (16) is the contribution of channel reciprocity to the overall secret-key rate. The rates corresponding to R^-_{AB} and R^-_{BA} are the contributions to the secret-key rate by the forward and reverse channels when source emulation is used. The factor $\left(1-\frac{1}{T}\right)$ is due to the fact that the proposed scheme only uses $T-1$ symbols in each block for source emulation. The penalty terms in (19) and (20) arise because in our analysis of the source emulation phase, we will assume that the eavesdropper is revealed the true channel gains of the main channel, whereas the intended receivers only have access to the noisy channel estimates. Finally we note that the expressions in (19) and (20) do not involve power allocation over fading states as this is precluded in our coding scheme.

We note that the achievable rate (16) is structurally similar to the upper bound expression (13). The following result shows that the upper and lower bounds are indeed very close in the high signal to noise ratio (SNR) regime for Rayleigh fading channels.

Corollary 1: In the high SNR regime, assuming that h_{AB} and h_{BA} are jointly Gaussian, zero mean, unit variance random variables, the upper and lower bounds satisfy the following relation:

$$\lim_{P\to\infty} \left\{R^+(P) - R^-_{\mathrm{G}}(P)\right\} \le \frac{1}{T}\gamma \tag{22}$$

where R^+, and R^- are given by (13) and (16) respectively and

$$\gamma \triangleq E\left[\log\left(1+\frac{|h_{AB}|^2}{|g_{AE}|^2}\right)\right] + E\left[\log\left(1+\frac{|h_{BA}|^2}{|g_{BE}|^2}\right)\right]. \tag{23}$$

□

C. Numerical Comparisons

We present numerical comparisons between our upper and lower bounds in Fig. 2 and Fig. 3. Fig. 2 shows the bounds as a function of SNR when the coherence period $T = 10$. Fig. 3 shows the bounds as a function of the coherence period T when SNR = 30 dB. In both figures we assume that the channel gains are all drawn $\mathcal{CN}(0,1)$ and let ρ denote the cross-correlation. In Fig. 2 we fix $\rho = 0.95$, while in Fig. 3 it is fixed to $\rho = 0.97$.

The upper-most plot in Fig. 2 and Fig. 3, marked with squares, is the upper bound in Theorem 1. Note that for small values of T the contribution from channel-reciprocity term in (13) is dominant. However when T increases

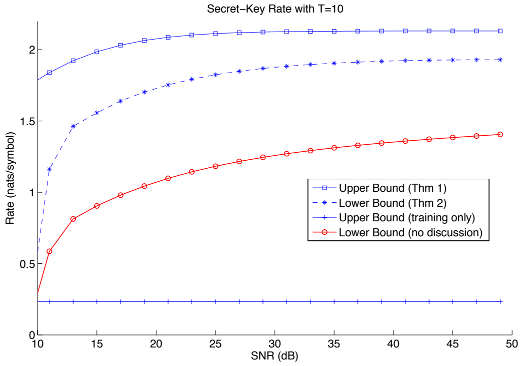


Fig. 2. Bounds on the SK capacity as a function of SNR for a coherence period of $T = 10$.

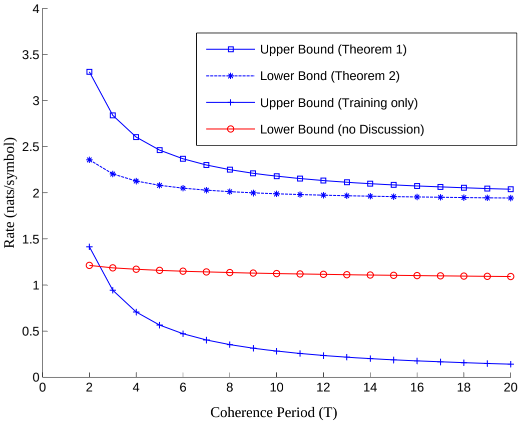


Fig. 3. Bounds on the SK capacity as a function of coherence period for SNR = 30dB.

the contribution of this term diminishes and the achievable rate saturates to the contribution from the remaining two terms in (13).

The plot below the upper bound, marked with asterisks, is the lower bound with public discussion in Theorem 2. As we increase T, we note that the gap between the upper and lower bound decreases inversely with the coherence period in Fig. 3.

We also have two additional plots for comparison in both Fig. 2 and Fig. 3. The plot marked with circles is an achievable rate using the separation based scheme when the public discussion channel is not available [33]. We do not develop this lower bound in the paper as it appears to exhibit a considerable gap from the upper bound. The lowermost plot is a simple upper bound on training based schemes $R^{+}_{\text{training}} = -\frac{1}{T}\log(1-\rho^2)$. This bound is attained by revealing the respective channel gains to the legitimate receivers. Note that the training only scheme is far from optimal even when correlation coefficient is close to 1.

III. Upper Bound

We provide a proof of Theorem 1 in this section. We assume that the communication spans over K coherence blocks and let $N = K \cdot T$. In our analysis below we use the notation that $\mathbf{g}^t = (g^t_{AE}, g^t_{BE})$ and $\mathbf{z}^t = (z^t_{AE}, z^t_{BE})$.

$$
\begin{aligned}
NR &= H(k_A) && (24)\\
&= I(k_A; k_B) + H(k_A|k_B) && (25)\\
&= I(k_A; k_B) - I(k_A; \mathbf{z}^N, \mathbf{g}^N, \psi^N_A, \psi^N_B) + I(k_A; \mathbf{z}^N, \mathbf{g}^N, \psi^N_A, \psi^N_B) + H(k_A|k_B) && (26)\\
&\le I(k_A; k_B) - I(k_A; \mathbf{z}^N, \mathbf{g}^N, \psi^N_A, \psi^N_B) + N\gamma(\varepsilon) && (27)\\
&\le I(k_A; k_B|\mathbf{z}^N, \mathbf{g}^N, \psi^N_A, \psi^N_B) + N\gamma(\varepsilon) && (28)
\end{aligned}
$$

$$\leq I(m_A, y_A^N; m_B, y_B^N | \mathbf{z}^N, \mathbf{g}^N, \boldsymbol{\psi}_A^N, \boldsymbol{\psi}_B^N) + N\gamma(\varepsilon) \tag{29}$$

$$\leq I(m_A, y_A^N, h_{BA}^N; m_B, y_B^N, h_{AB}^N | \mathbf{z}^N, \mathbf{g}^N, \boldsymbol{\psi}_A^N, \boldsymbol{\psi}_B^N) + N\gamma(\varepsilon) \tag{30}$$

where $\gamma(\varepsilon)$ is a function that approaches zero as $\varepsilon \to 0$ and we apply Fano's Inequality [34] and the secrecy constraint (12) in (27), and (29) uses the fact that the secret-keys are computed using $k_A = \mathcal{K}_A(m_A, y_A^N, \boldsymbol{\psi}_A^N, \boldsymbol{\psi}_B^N)$ and $k_B = \mathcal{K}_B(m_B, y_B^N, \boldsymbol{\psi}_A^N, \boldsymbol{\psi}_B^N)$ respectively. We next show the following lemma that successively reduces the N letter expression in (30).

We can successively reduce the above upper bound using the following two Lemmas. These lemmas are based on repeated application of the chain rule of mutual information. Their proofs are delegated to the Appendix.

Lemma 1: For each $t \in \{1, \dots, N\}$, we have that:

$$I(m_A, y_A^t, h_{BA}^N; m_B, y_B^t, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^t, \boldsymbol{\psi}_B^t) \leq I(m_A, y_A^t, h_{BA}^N; m_B, y_B^t, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}). \tag{31}$$

Eq. (31) states that removing the conditioning of the public messages $\boldsymbol{\psi}_A(t) \triangleq \{\psi_A(t,1), \dots, \psi_A(t,L)\}$ as well as $\boldsymbol{\psi}_B(t) \triangleq \{\psi_B(t,1), \dots, \psi_B(t,L)\}$ only increases the mutual information expression. The special case when $t = 0$ is separately stated below.

$$I(m_A, h_{BA}^N; m_B, h_{AB}^N | \mathbf{g}^N, \boldsymbol{\psi}_A(0), \boldsymbol{\psi}_B(0)) \leq I(m_A, h_{BA}^N; m_B, h_{AB}^N | \mathbf{g}^N) \tag{32}$$

□

Proof: See Appendix A

■

Lemma 2: For each $t \in \{1, 2, \dots, N\}$, we have that:

$$I(m_A, y_A^t, h_{BA}^N; m_B, y_B^t, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}) \leq I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1})$$
$$+ I(x_A(t); y_B(t) | z_{AE}(t), g_{AE}(t), h_{AB}(t)) + I(x_B(t); y_A(t) | z_{BE}(t), g_{BE}(t), h_{BA}(t)). \tag{33}$$

□

Proof: See Appendix B

■

We show how to reduce (30) by successively applying Lemma 1 and Lemma 2. Substituting (31) with $t = N$ into (30) we have:

$$NR \leq I(m_A, y_A^N, h_{BA}^N; m_B, y_B^N, h_{AB}^N | \mathbf{z}^N, \mathbf{g}^N, \boldsymbol{\psi}_A^N, \boldsymbol{\psi}_B^N) + N\gamma(\varepsilon) \tag{34}$$

$$\leq I(m_A, y_A^N, h_{BA}^N; m_B, y_B^N, h_{AB}^N | \mathbf{z}^N, \mathbf{g}^N, \boldsymbol{\psi}_A^{N-1}, \boldsymbol{\psi}_B^{N-1}) + N\gamma(\varepsilon) \tag{35}$$

Next substituting (33) in Lemma 2 with $t = N$, into (35)

$$NR \leq I(m_A, y_A^{N-1}, h_{BA}^N; m_B, y_B^{N-1}, h_{AB}^N | \mathbf{z}^{N-1}, \mathbf{g}^N, \boldsymbol{\psi}_A^{N-1}, \boldsymbol{\psi}_B^{N-1}) + N\gamma(\varepsilon)$$
$$+ I(x_A(N); y_B(N) | z_{AE}(N), g_{AE}(N), h_{AB}(N)) + I(x_B(N); y_A(N) | z_{BE}(N), g_{BE}(N), h_{BA}(N)) \tag{36}$$

By recursively applying (31) and (33) for $t \in \{N-1, N-2, \dots, 1\}$ and finally applying (32) we have that

$$NR \le I(\mathsf{m}_A, \mathsf{h}_{BA}^N; \mathsf{m}_B, \mathsf{h}_{AB}^N | \mathbf{g}^N) + \sum_{t=1}^{N} I(\mathsf{x}_A(t); \mathsf{y}_B(t) | \mathsf{z}_{AE}(t), \mathsf{g}_{AE}(t), \mathsf{h}_{AB}(t)) + \sum_{t=1}^{N} I(\mathsf{x}_B(t); \mathsf{y}_A(t) | \mathsf{z}_{BE}(t), \mathsf{g}_{BE}(t), \mathsf{h}_{BA}(t)) + N\gamma(\varepsilon). \tag{37}$$

Recall that the random variables m_A and m_B are mutually independent, and independent of $(\mathsf{h}_{AB}^N, \mathsf{h}_{BA}^N, \mathbf{g}^N)$. Furthermore since the channel is block fading channel model, each sequence of channel gains, e.g., h_{AB}^N is piecewise constant for a duration of T symbols and contains only $K = N/T$ independent random variables. Thus we have

$$I(\mathsf{m}_A, \mathsf{h}_{BA}^N; \mathsf{m}_B, \mathsf{h}_{AB}^N | \mathbf{g}^N) = I(\mathsf{h}_{BA}^N; \mathsf{h}_{AB}^N | \mathbf{g}^N) \tag{38}$$

$$= KI(\mathsf{h}_{BA}; \mathsf{h}_{AB} | \mathbf{g}) \tag{39}$$

$$= KI(\mathsf{h}_{BA}; \mathsf{h}_{AB}) \tag{40}$$

where (40) uses the fact that the channel gains $(\mathsf{h}_{AB}, \mathsf{h}_{BA})$ are sampled independently of $(\mathsf{g}_{AE}, \mathsf{g}_{BE})$ (see (6)).

Next we show that there exists a power allocation function $P(\cdot)$ satisfying $E[P(\mathsf{h}_{AB})] \le P$, such that:

$$\frac{1}{N} \sum_{t=1}^{N} I(\mathsf{x}_A(t); \mathsf{y}_B(t) | \mathsf{z}_{AE}(t), \mathsf{g}_{AE}(t), \mathsf{h}_{AB}(t)) \le E\left[\log\left(1 + \frac{P(\mathsf{h}_{AB})|\mathsf{h}_{AB}|^2}{1 + P(\mathsf{h}_{AB})|\mathsf{g}_{AE}|^2}\right)\right] \tag{41}$$

Let $P_t(h_{AB})$ denote the average power of symbol $\mathsf{x}_{AB}(t)$ when the gain on forward-channel equals h_{AB} i.e., we have $E[|\mathsf{x}_{AB}(t)|^2 | \mathsf{h}_{AB} = h_{AB}] = P_t(h_{AB})$. Using the fact that a Gaussian input distribution maximizes the conditional mutual information (see e.g., [35]) it follows that:

$$I(\mathsf{x}_A(t); \mathsf{y}_B(t) | \mathsf{z}_{AE}(t), \mathsf{g}_{AE}(t), \mathsf{h}_{AB}(t)) \le E\left[\log\left(1 + \frac{P_t(\mathsf{h}_{AB})|\mathsf{h}_{AB}|^2}{1 + P_t(\mathsf{h}_{AB})|\mathsf{g}_{AE}|^2}\right)\right]. \tag{42}$$

Thus we have that

$$\sum_{t=1}^{N} I(\mathsf{x}_A(t); \mathsf{y}_B(t) | \mathsf{z}_{AE}(t), \mathsf{g}_{AE}(t), \mathsf{h}_{AB}(t)) \le \sum_{t=1}^{N} E\left[\log\left(1 + \frac{P_t(\mathsf{h}_{AB})|\mathsf{h}_{AB}|^2}{1 + P_t(\mathsf{h}_{AB})|\mathsf{g}_{AE}|^2}\right)\right] \tag{43}$$

$$= E\left[\sum_{t=1}^{N} \log\left(1 + \frac{P_t(\mathsf{h}_{AB})|\mathsf{h}_{AB}|^2}{1 + P_t(\mathsf{h}_{AB})|\mathsf{g}_{AE}|^2}\right)\right] \tag{44}$$

$$\le NE\left[\log\left(1 + \frac{\frac{1}{N}\sum_{t=1}^{N} P_t(\mathsf{h}_{AB})|\mathsf{h}_{AB}|^2}{1 + \frac{1}{N}\sum_{t=1}^{N} P_t(\mathsf{h}_{AB})|\mathsf{g}_{AE}|^2}\right)\right] \tag{45}$$

$$= NE\left[\log\left(1 + \frac{P(\mathsf{h}_{AB})|\mathsf{h}_{AB}|^2}{1 + P(\mathsf{h}_{AB})|\mathsf{g}_{AE}|^2}\right)\right], \tag{46}$$

where $P(\mathsf{h}_{AB}) \triangleq \frac{1}{N}\sum_{t=1}^{N} P_t(\mathsf{h}_{AB})$ is the average power allocated when the fading state equals h_{AB} and (45) uses the fact that the function $f(x) = \log\left(1 + \frac{ax}{1+bx}\right)$ is a concave function in x and hence Jensen's inequality [34] applies. Also note that

$$E[P(\mathsf{h}_{AB})] = E\left[\frac{1}{N}\sum_{t=1}^{N} P_t(\mathsf{h}_{AB})\right] \tag{47}$$

$$= \frac{1}{N}\sum_{t=1}^{N} E[P_t(\mathsf{h}_{AB})] \tag{48}$$

$$\leq \frac{1}{N}\sum_{t=1}^{N} P_t \leq P. \tag{49}$$

This establishes (41). In a similar fashion we can show that there exists a power allocation function $P(\cdot)$ satisfying $E[P(h_{BA})] \leq P$ such that

$$\frac{1}{N}\sum_{t=1}^{N} I(x_B(t); y_A(t)|z_{BE}(t), g_{BE}(t), h_{BA}(t)) \leq E\left[\log\left(1 + \frac{P(h_{BA})|h_{BA}|^2}{1+P(h_{BA})|g_{BE}|^2}\right)\right]. \tag{50}$$

$$\tag{51}$$

Substituting (40), (46) and (50) into (37) and using $N = KT$, we have that for a certain power allocation functions the secret-key rate satisfies:

$$R \leq \frac{1}{T} I(h_{AB}; h_{BA}) + E\left[\log\left(1+\frac{P(h_{AB})|h_{AB}|^2}{1+P(h_{AB})|g_{AE}|^2}\right)\right] + E\left[\log\left(1+\frac{P(h_{BA})|h_{BA}|^2}{1+P(h_{BA})|g_{BE}|^2}\right)\right] + \gamma(\varepsilon). \tag{52}$$

Since (52) must hold for every $\varepsilon > 0$, and $\gamma(\varepsilon) \to 0$ as $\varepsilon \to 0$, this completes the proof of Theorem 1.

IV. Lower Bound

Our coding scheme in Theorem 2 is a two-phase scheme consisting of training followed by source-emulation. In each coherence block, the first symbol transmitted on each of the forward and reverse channels is a training symbol i.e., $x_A(iT+1) = x_B(iT+1) = \sqrt{P_1}$ for $i \in \{0, 2, \ldots, K-1\}$.

For each $j \in \{2, \ldots, T\}$ we sample $x_A(iT+j) \sim p_{x_A}(\cdot)$ and $x_B(iT+j) \sim p_{x_B}(\cdot)$ independently of all other symbols where $p_{x_A}(\cdot)$ and $p_{x_B}(\cdot)$ are fixed distributions such that both $E[|x_A|^2] \leq P_2$ and $E[|x_B|^2] \leq P_2$. Note that P_1 and P_2 denote the power in the training phase and the transmission phase. For convenience we denote $\bar{\mathbf{x}}_A(i) \in \mathbb{C}^{T-1}$ to be the vector of $T-1$ symbols transmitted in the i-th coherence block on the forward channel and similarly let $\bar{\mathbf{x}}_B(i) \in \mathbb{C}^{T-1}$ denote the vector of $T-1$ transmitted symbols on the reverse channel. Using (1)-(4), the corresponding output in the i-th coherence block is expressed as[1]:

$$\text{(Forward Channel):} \qquad \bar{\mathbf{y}}_B(i) = h_{AB}(i)\cdot\bar{\mathbf{x}}_A(i) + \bar{\mathbf{n}}_B(i), \qquad \bar{\mathbf{z}}_{AE}(i) = g_{AE}(i)\cdot\bar{\mathbf{x}}_A(i) + \bar{\mathbf{n}}_{AE}(i), \tag{53}$$

$$\text{(Reverse Channel):} \qquad \bar{\mathbf{y}}_A(i) = h_{BA}(i)\cdot\bar{\mathbf{x}}_B(i) + \bar{\mathbf{n}}_A(i), \qquad \bar{\mathbf{z}}_{BE}(i) = g_{BE}(i)\cdot\bar{\mathbf{x}}_B(i) + \bar{\mathbf{n}}_{BE}(i), \tag{54}$$

where $\bar{\mathbf{y}}_A(i), \bar{\mathbf{y}}_B(i) \in \mathbb{C}^{T-1}$ denote the output symbols in coherence block i over the forward and reverse channels respectively and $\bar{\mathbf{z}}_{AE}(i)$ and $\bar{\mathbf{z}}_{BE}(i)$ denote the associated outputs at the eavesdropper. All the additive noise vectors have i.i.d. entries sampled from $\mathcal{CN}(0,1)$.

Let $\hat{h}_{AB}(i)$ and $\hat{h}_{BA}(i)$ denote the channel estimates obtained by terminals B and A over the forward and reverse channels respectively in coherence block i as defined in (17) and (18). Thereafter terminal A observes $\bar{\mathbf{y}}_A(i)$ whereas terminal B observes $\bar{\mathbf{y}}_B(i)$. At the end of K such coherence blocks, as indicated in Table I, terminal A has access to

[1] In this section, in the analysis of our proposed coding scheme, it is convenient to let $h_{AB}(i)$ and $h_{BA}(i)$ denote the fading gains in coherence block i. Thus in Table I, the sequence $\hat{h}_{AB}^K$ denotes a length K sequence of channel gains corresponding to the K coherence blocks. The other channel sequences have a similar interpretation. The reader is alerted that notation is different from the channel model (e.g. (1)) where the channel sequences must be of length N.

$(\hat{h}_{BA}^K, \bar{\mathbf{x}}_A^K, \bar{\mathbf{y}}_A^K)$ whereas terminal B has access to $(\hat{h}_{AB}^K, \bar{\mathbf{x}}_B^K, \bar{\mathbf{y}}_B^K)$. The eavesdropper observes $(g_{AE}^K, g_{BE}^K, \bar{\mathbf{z}}_{AE}^K, \bar{\mathbf{z}}_{BE}^K)$. The sequences generated in this fashion are then used to extract a common secret-key as discussed below.

TABLE I

SIDE INFORMATION GENERATED AT THE TERMINALS IN THE TWO-PHASE SCHEME

Sequence/Terminal	A	B	E
Channel Sequence	$\hat{h}_{BA}^K$	$\hat{h}_{AB}^K$	(g_{AE}^K, g_{BE}^K)
Source Sequence - Forward Channel	$\bar{\mathbf{x}}_A^K$	$\bar{\mathbf{y}}_B^K$	$\bar{\mathbf{z}}_{AE}^K$
Source Sequence - Reverse Channel	$\bar{\mathbf{y}}_A^K$	$\bar{\mathbf{x}}_B^K$	$\bar{\mathbf{z}}_{BE}^K$

A. Discretization of Continuous Valued Random Variables

In the analysis of the coding theorem, we need to consider discrete valued random variables at the legitimate receivers. We adapt following the technique outlined in [36, pp. 50-51].

- The channel estimates $\hat{h}_{BA}$ and $\hat{h}_{AB}$ are discretized as follows. Let us define:

$$\mathcal{I}_1 = \{-j\Delta_1, -(j-1)\Delta_1, \ldots, (j-1)\Delta_1, j\Delta_1\}, \tag{55}$$

  where j is an integer and we select $\Delta_1 = \frac{1}{\sqrt{j}}$. We find elements in $\mathcal{I}_1$ closest to the real and imaginary parts of $\hat{h}_{AB}$ and denote $\left[\hat{h}_{AB}\right]_j$ as the resulting quantization. We define $\left[\hat{h}_{BA}\right]_j$ in a similar fashion.
- We discretize $x_A(t)$ and $x_B(t)$ to $[x_A(t)]_k$ and $[x_B(t)]_k$ respectively, whose real and imaginary parts take values in the set

$$\mathcal{I}_2 = \{-k\Delta_2, -(k-1)\Delta_2, \ldots, (k-1)\Delta_2, k\Delta_2\}, \tag{56}$$

  where k is an integer and $\Delta_2 = \frac{1}{\sqrt{k}}$. We select $[x_A(t)]_k$ to be the closest such point to $x_A(t)$ with $|[x_A(t)]_k| \leq |x_A(t)|$. Thus we have that $E\left[|[x_A(t)]_k|^2\right] \leq E[|x_A(t)|^2] \leq P_2$. We define $[x_B(t)]_k$ in an analogous manner.
- We also discretize the channel output at each legitimate receiver. Note that from (1), we have $y_{B,k}(t) = h_{AB}(t)[x_A(t)]_k + n_{AB}(t)$. Note that $y_{B,k}(t)$ is continuous valued even through $[x_A(t)]_k$ is discrete. We discretize $y_{B,k}(t)$ to $[y_{B,k}(t)]_l$ whose real and imaginary parts take values over the interval

$$\mathcal{I}_3 = \{-l\Delta_3, -(l-1)\Delta_3, \ldots, (l-1)\Delta_3, l\Delta_3\} \tag{57}$$

  where l is an integer and $\Delta_3 = \frac{1}{\sqrt{l}}$.

Note that since we use scalar quantization and the sequences $\hat{h}_{AB}^K$ and $\hat{h}_{BA}^K$ are sampled i.i.d. the sequences $\left[\hat{h}_{AB}^K\right]_j$ and $\left[\hat{h}_{BA}^K\right]_j$ are also sampled i.i.d. In a similar way the sequences $\left[\bar{\mathbf{x}}_A^K\right]_k$ and $\left[\bar{\mathbf{x}}_B^K\right]_k$ are also sampled i.i.d. Further defining $z_{AE,k}(t) = g_{AE}(t)[x_A(t)]_k + n_{AE}(t)$ and $z_{BE,k}(t) = g_{BE}(t)[x_B(t)]_k + n_{BE}(t)$ and using $\bar{\mathbf{z}}_k^K \triangleq (\bar{\mathbf{z}}_{AE,k}^K, \bar{\mathbf{z}}_{BE,k}^K)$ and $\mathbf{g}^K \triangleq (g_{AE}^K, g_{BE}^K)$, we have that:

$$\left[\bar{\mathbf{y}}_{A,k}^K\right]_l \leftrightarrow \left(h_{BA}^K, \left[\bar{\mathbf{x}}_B^K\right]_k\right) \leftrightarrow \left(\bar{\mathbf{z}}_k^K, \mathbf{g}^K\right) \tag{58}$$

$$\left[\bar{\mathbf{y}}_{B,k}^{K}\right]_l \leftrightarrow \left(h_{AB}^{K}, \left[\bar{\mathbf{x}}_A^{K}\right]_k\right) \leftrightarrow \left(\bar{\mathbf{z}}_k^{K}, \mathbf{g}^{K}\right) \tag{59}$$

are satisfied. Note that for each $\hat{h}_{AB}$ we have that $\left|\hat{h}_{AB} - \left[\hat{h}_{AB}\right]_j\right| \to 0$, as $j \to \infty$, and similarly $\left|\hat{h}_{BA} - \left[\hat{h}_{BA}\right]_j\right| \to 0$, as $j \to \infty$ in a point wise fashion. In a similar manner, as $k \to \infty$, we have that $|x_A(t) - [x_A(t)]_k| \to 0$, $|x_B(t) - [x_B(t)]_k| \to 0$. Furthermore for each fixed k, as $l \to \infty$, we have that $\left|y_{A,k}(t) - [y_{A,k}(t)]_l\right| \to 0$ and $|y_{B,k}(t) - [y_{B,k}(t)]_l| \to 0$.

In the subsequent sections, we present our proposed coding scheme and the corresponding achievable rate for the discretized variables above. Thereafter we argue that in the limit of vanishingly small quantization errors, the rate converges as expected.

B. Achievable Rate using Discretized Variables

For the discretized set of random variables we have the following achievable rate.

Proposition 1: An achievable secret-key rate using public discussion is given by:

$$R_{\Delta}^{-} = \frac{1}{T} I\left(\left[\hat{h}_{AB}\right]_j ; \left[\hat{h}_{BA}\right]_j\right) + \frac{T-1}{T}\left(R_{AB,\Delta}^{-} + R_{BA,\Delta}^{-}\right) \tag{60}$$

where $\hat{h}_{AB}$ and $\hat{h}_{BA}$ are the MMSE estimates of channel gains h_{AB} and h_{BA} respectively (c.f. (17), (18)) and

$$R_{AB,\Delta}^{-} = I\left([y_{B,k}]_l ; [x_A]_k, [u]_j\right) - I\left([y_{B,k}]_l ; z_{AE,k}, g_{AE}, h_{AB}\right) \tag{61}$$

$$R_{BA,\Delta}^{-} = I\left([y_{A,k}]_l ; [x_B]_k, [u]_j\right) - I\left([y_{A,k}]_l ; z_{BE,k}, g_{BE}, h_{BA}\right) \tag{62}$$

where we have introduced $[u]_j \triangleq \left(\left[\hat{h}_{AB}\right]_j, \left[\hat{h}_{BA}\right]_j\right)$. The variables $[x_A]_k$ and $[x_B]_k$ are obtained by discretizing x_A and x_B respectively as in section IV-A. Furthermore $x_A \sim p_{x_A}(\cdot)$ and $x_B \sim p_{x_B}(\cdot)$ are sampled independently and satisfy $E[|x_A|^2] \leq P_2$ and $E[|x_B|^2] \leq P_2$. The powers P_1 (c.f. (17), (18)) and P_2 must satisfy $P_1 + (T-1)P_2 \leq TP$. □

The coding theorem associated with Prop. 1 is presented in Appendix C.

C. Achievable Rate: Extension to Continuous Valued Inputs

In this section, we further analyze the achievable rate in the limit that the quantization error due to discretization in section IV-A approaches zero. In this limit, the rate required over the public discussion channel will also increase to infinity. However we note that there is no rate constraint imposed over the discussion channel and thus the proposed rate can be approached arbitrarily closely. In our analysis we make use of the following fact:

Fact 1: [36, pg. 23] Let u and v be two arbitrary random variables (discrete or continuous) with a joint probability measure $\mu(u, v)$ and marginal measures $\mu(u)$ and $\mu(v)$. The mutual information between u and v, defined as,

$$I(u; v) = \int \log \frac{d\mu(u, v)}{d(\mu(u) \times \mu(v))} d\mu(u, v) \tag{63}$$

where $d\mu(u, v)/d(\mu(u) \times \mu(v))$ denotes the Radon-Nikodym derivative of the joint probability measure with respect to the product measure, can be equivalently expressed as:

$$I(u; v) = \lim_{j,k \to \infty} I\left([u]_j ; [v]_k\right) \tag{64}$$

where $[u]_j$ and $[v]_k$ are any sequence of fine quantizations of u and v respectively such that $\left|u-[u]_j\right|$ and $|v-[v]_k|$ tend to zero as pointwise as $j \to \infty$, and $k \to \infty$, respectively

Upon examining (61),(62) and taking $j, l \to \infty$, with k fixed, we have that with $u = (\hat{h}_{AB}, \hat{h}_{BA})$:

$$\lim_{l,j\to\infty} I\left([y_{B,k}]_l \ ; \ [x_A]_k, [u]_j\right) = I(y_{B,k}; [x_A]_k, u) \tag{65}$$

$$\lim_{l\to\infty} I\left([y_{B,k}]_l \ ; \ z_{AE,k}, g_{AE}, h_{AB}\right) = I(y_{B,k}; z_{AE,k}, g_{AE}, h_{AB}) \tag{66}$$

$$\lim_{l,j\to\infty} I\left([y_{A,k}]_l \ ; \ [x_B]_k, [u]_j\right) = I(y_{A,k}; [x_B]_k, u) \tag{67}$$

$$\lim_{l\to\infty} I\left([y_{A,k}]_l \ ; \ z_{BE,k}, g_{BE}, h_{BA}\right) = I(y_{A,k}; z_{BE,k}, g_{BE}, h_{BA}) \tag{68}$$

Similarly the first term in (60) converges as follows

$$\lim_{j\to\infty} I\left(\left[\hat{h}_{AB}\right]_j \ ; \ \left[\hat{h}_{BA}\right]_j\right) = I(\hat{h}_{AB}; \hat{h}_{BA}). \tag{69}$$

We now consider the term

$$I(y_{B,k}; [x_A]_k, u) - I(y_{B,k}; z_{AE,k}, g_{AE}, h_{AB}) = \hbar(y_{B,k}|z_{AE,k}, g_{AE}, h_{AB}) - \hbar(y_{B,k}|[x_A]_k, u) \tag{70}$$

We will show that when $x_A \sim \mathcal{CN}(0, P_2)$ we have that

$$\liminf_{k\to\infty} \hbar(y_{B,k}|z_{AE,k}, g_{AE}, h_{AB}) \geq E\left[\log 2\pi e\left(1 + \frac{P_2|h_{AB}|^2}{1+P_2|g_{AE}|^2}\right)\right] \tag{71}$$

$$\limsup_{k\to\infty} \hbar(y_{B,k}|[x_A]_k, u) \leq \log 2\pi e\left(1 + \frac{P_2}{1+P_1}\right) \tag{72}$$

Together (71) and (72) imply that:

$$\liminf_{k\to\infty} \left\{I(y_{B,k}; [x_A]_k, u) - I(y_{B,k}; z_{AE,k}, g_{AE}, h_{AB})\right\} \geq E\left[\log\left(1 + \frac{P_2|h_{AB}|^2}{1+P_2|g_{AE}|^2}\right)\right] - \log\left(1 + \frac{P_2}{1+P_1}\right) \tag{73}$$

We first establish (71). Note that

$$\hbar(y_{B,k}|z_{AE,k}, g_{AE}, h_{AB}) = \hbar(y_{B,k}, z_{AE,k}|g_{AE}, h_{AB}) - \hbar(z_{AE,k}|g_{AE}, h_{AB}) \tag{74}$$

$$= \hbar(y_{B,k}, z_{AE,k}|g_{AE}, h_{AB}) - \hbar(z_{AE,k}|g_{AE}) \tag{75}$$

Furthermore since $E\left[|[x_A]_k|^2\right] \leq P_2$, and $[x_A]_k$ is independent of g_{AE}, we have, using (2),

$$E[|z_{AE,k}|^2 | g_{AE} = g_{AE}] \leq |g_{AE}|^2 P_2 + 1. \tag{76}$$

Thus using the fact that the differential entropy under a variance constraint is maximized by a Gaussian distribution we have that

$$\hbar(z_{AE,k}|g_{AE}) \leq E[\log\left(2\pi e(\,|g_{AE}|^2 P_2 + 1)\right)]. \tag{77}$$

Next we show

$$\liminf_{k\to\infty} \hbar(y_{B,k}, z_{AE,k}|h_{AB}, g_{AE}) \geq E\left[(\log 2\pi e)^2\left(1 + P_2|h_{AB}|^2 + P_2|g_{AE}|^2\right)\right], \tag{78}$$

using an argument analogous to [36, pp. 77]. Eq. (77) and (78) together complete the proof of (71). Note that for each $(\mathsf{h}_{AB} = h_{AB}, \mathsf{g}_{AE} = g_{AE})$ we have that

$$f_{\mathsf{y}_{B,k},\mathsf{z}_{AE,k}}(y,z|\mathsf{h}_{AB} = h_{AB}, \mathsf{z}_{AE} = g_{AE}) = \int f_{\mathsf{n}_B}\left(y - h_{AB}[\mathsf{x}_A]_k\right) \cdot f_{\mathsf{n}_{AE}}\left(z - g_{AE}[\mathsf{x}_A]_k\right) dF_{[\mathsf{x}_A]_k}(x) \tag{79}$$

$$= E_{[\mathsf{x}_A]_k}\left[f_{\mathsf{n}_B}\left(y - h_{AB}[\mathsf{x}_A]_k\right) \cdot f_{\mathsf{n}_{AE}}\left(z - g_{AE}[\mathsf{x}_A]_k\right)\right] \tag{80}$$

where n_B and n_{AE} are the additive noise variables in (1) and (2). Since $f_{\mathsf{n}_B}(\cdot)$ and $f_{\mathsf{n}_{AE}}(\cdot)$ are bounded and continuous functions $f_{\mathsf{y}_{B,k},\mathsf{z}_{AE,k}}(y,z|\mathsf{h}_{AB} = h_{AB}, \mathsf{g}_{AE} = g_{AE})$ converges to $f_{\mathsf{y}_B,\mathsf{z}_{AE}}(y,z|\mathsf{h}_{AB} = h_{AB}, \mathsf{g}_{AE} = g_{AE})$ as $k \to \infty$. Furthermore we have that

$$f_{\mathsf{y}_{B,k},\mathsf{z}_{AE,k}}(y,z|\mathsf{h}_{AB} = h_{AB}, \mathsf{z}_{AE} = g_{AE}) \leq \max_{n_B,n_{AE}} f_{\mathsf{n}_B}(n_B) f_{\mathsf{n}_{AE}}(n_{AE}) = \frac{1}{2\pi}. \tag{81}$$

Therefore, for each $a > 0$ we can express:

$$\hbar(\mathsf{y}_{B,k}, \mathsf{z}_{AE,k}|\mathsf{h}_{AB}, \mathsf{g}_{AE})$$

$$= \oint_{h_{AB}} \oint_{g_{AE}} \oint_{y} \oint_{z} f_{\mathsf{h}_{AB}}(h_{AB}) f_{\mathsf{g}_{AE}}(g_{AE}) f_{\mathsf{y}_{B,k},\mathsf{z}_{AE,k}}(y,z|h_{AB},g_{AE}) \log \frac{1}{f_{\mathsf{y}_{B,k},\mathsf{z}_{AE,k}}(y,z|h_{AB},g_{AE})} dydzdg_{AE}dh_{AB} \tag{82}$$

$$\geq \oint_{|h_{AB}|\leq a} \oint_{|g_{AE}|\leq a} \oint_{|y|\leq a} \oint_{|z|\leq a} f(h_{AB}) f(g_{AE}) f_{\mathsf{y}_{B,k},\mathsf{z}_{AE,k}}(y,z|h_{AB},g_{AE}) \log \frac{1}{f_{\mathsf{y}_{B,k},\mathsf{z}_{AE,k}}(y,z|h_{AB},g_{AE})} dydzdg_{AE}dh_{AB} \tag{83}$$

where (83) follows from the fact that from (81) the integrand in (82) is always positive and therefore (83) is integrating over a smaller set. By the dominated convergence theorem [36, Appendix B], upon taking $k \to \infty$ for each fixed a, the right hand side converges to

$$\oint_{|h_{AB}|\leq a} \oint_{|g_{AE}|\leq a} \oint_{|y|\leq a} \oint_{|z|\leq a} f(h_{AB}) f(g_{AE}) f_{\mathsf{y}_B,\mathsf{z}_{AE}}(y,z|h_{AB},g_{AE}) \log \frac{1}{f_{\mathsf{y}_B,\mathsf{z}_{AE}}(y,z|h_{AB},g_{AE})} dydzdg_{AE}dh_{AB}. \tag{84}$$

Upon taking $a \to \infty$, the above expression in turn converges to $\hbar(\mathsf{y}_B, \mathsf{z}_{AE}|\mathsf{h}_{AB}, \mathsf{g}_{AE})$. Since $\mathsf{x}_A \sim \mathcal{CN}(0, P_2)$ the relation (78) now follows.

To establish (72) we let $\tilde{\mathsf{h}}_{AB}$ be the MMSE estimation error of h_{AB} given the observation $\sqrt{P_1}\mathsf{h}_{AB} + \mathsf{n}_B$. We can express $\mathsf{h}_{AB} = \tilde{\mathsf{h}}_{AB} + \mathsf{e}_{AB}$, where the estimation error e_{AB} satisfies $E[|\mathsf{e}_{AB}|^2] \leq \frac{1}{1+P_1}$. Using (1), we can express

$$\hbar\left(\mathsf{y}_{B,k}|[\mathsf{x}_A]_k, \mathsf{u}\right) \leq \hbar\left(\mathsf{y}_{B,k} \mid [\mathsf{x}_A]_k, \tilde{\mathsf{h}}_{AB}\right) \tag{85}$$

$$= \hbar\left(\mathsf{h}_{AB}[\mathsf{x}_A]_k + \mathsf{n}_B \mid [\mathsf{x}_A]_k, \tilde{\mathsf{h}}_{AB}\right) \tag{86}$$

$$\leq \hbar\left(\mathsf{e}_{AB}[\mathsf{x}_A]_k + \mathsf{n}_B \mid [\mathsf{x}_A]_k\right) \tag{87}$$

$$\leq \hbar\left(\mathsf{e}_{AB}[\mathsf{x}_A]_k + \mathsf{n}_B\right) \tag{88}$$

$$\leq \log 2\pi e\left(1 + \frac{P_2}{1+P_1}\right) \tag{89}$$

where (89) uses the fact that since the input $[x_A]_k$ is generated independently of h_{AB} and satisfies $E[|[x_A]_k|^2] \leq P_2$ we have

$$E\left[|e_{AB}[x_A]_k|^2\right] = E[|e_{AB}|^2]E[|[x_A]_k|^2] \tag{90}$$

$$\leq \frac{1}{1+P_1}P_2, \tag{91}$$

and we use the fact that differential entropy is maximized by a Gaussian distribution under a variance constraint. Since (89) holds for every k, the inequality in (72) now follows. The proof of (73) is now complete.

In a similar fashion we can show that

$$\liminf_{k\to\infty}\{I(y_{A,k};[x_B]_k,u) - I(y_{A,k};z_{B,k},g_{BE},h_{BA})\} \geq E\left[\log\left(1+\frac{P_2|h_{BA}|^2}{1+P_2|g_{BE}|^2}\right)\right] - \log\left(1+\frac{P_2}{1+P_1}\right) \tag{92}$$

The proof of (16) follows by substituting (69), (73) and (92) into (60).

V. Proof of Corollary 1

In this section we analyze the high SNR behaviour of the upper and lower bounds. In particular we will show that in the limit $P \to \infty$, the upper and lower bounds in (13) and (16) reduce to the following:

$$\lim_{P\to\infty} R^+ \leq -\frac{1}{T}\log(1-\rho^2) + \gamma \tag{93}$$

$$\lim_{P\to\infty} R^- = -\frac{1}{T}\log(1-\rho^2) + \frac{T-1}{T}\gamma. \tag{94}$$

where γ is defined in (23). The claim in Corollary 1 immediately follows from (93) and (94). Note that the first term in (13) is independent of P and furthermore since h_{AB} and h_{BA} are jointly Gaussian, zero mean, unit variance and with a cross-correlation of ρ it can be readily shown that

$$I(h_{AB};h_{BA}) = -\log(1-\rho^2). \tag{95}$$

The remaining two terms are upper bounded as follows. For any power allocation $P(h_{AB})$ note that

$$E\left[\log\left(1+\frac{P(h_{AB})|h_{AB}|^2}{1+P(h_{AB})|g_{AE}|^2}\right)\right] \leq E\left[\log\left(1+\frac{|h_{AB}|^2}{|g_{AE}|^2}\right)\right] \tag{96}$$

which follows since the function $f(x) = \frac{ax}{1+bx}$ is increasing in x for any $a, b > 0$. Similarly we have

$$E\left[\log\left(1+\frac{P(h_{BA})|h_{BA}|^2}{1+P(h_{BA})|g_{BE}|^2}\right)\right] \leq E\left[\left(1+\frac{|h_{BA}|^2}{|g_{BE}|^2}\right)\right]. \tag{97}$$

The upper bound (93) follows by substituting (95), (96) and (97) into (13).

For establishing (94), we consider the secret-key rate expression (16) in Thm 2. We select $P_2 = \frac{\sqrt{P}}{T-1}$ and $P_1 = P - \sqrt{P}$. Note that as $P \to \infty$ we have that $P_1, P_2 \to \infty$ and $\frac{P_2}{P_1} \to 0$. Since the function $f(x) = \frac{ax}{1+bx}$ is bounded for all $a, b > 0$ we can apply the Dominated Convergence Theorem to interchange the limit and expectation,

$$\lim_{P_2\to\infty} E\left[\log\left(1+\frac{P_2|h_{AB}|^2}{1+P_2|g_{AE}|^2}\right)\right] = E\left[\log\left(1+\frac{|h_{AB}|^2}{|g_{AE}|^2}\right)\right] \tag{98}$$

$$\lim_{P_2\to\infty} E\left[\log\left(1+\frac{P_2|h_{BA}|^2}{1+P_2|g_{BE}|^2}\right)\right] = E\left[\log\left(1+\frac{|h_{BA}|^2}{|g_{BE}|^2}\right)\right] \tag{99}$$

Furthermore since $\lim_{P\to\infty}\log\left(1+\frac{P_2}{1+P_1}\right)=0$

clearly holds, using (19) and (20) it follows that

$$\lim_{P\to\infty} R_{AB}^{-}(P)+R_{BA}^{-}(P)=\gamma \tag{100}$$

We next consider the first term in (16). When h_{AB} and h_{BA} are zero mean, unit variance, jointly Gaussian random variables with a cross-correlation coefficient of ρ, using (17) and (18) it can be easily shown that

$$I(\hat{h}_{AB};\hat{h}_{BA})=\log(1-\alpha^2\rho^2) \tag{101}$$

where $\alpha=\frac{P_1}{P_1+1}$. It immediately follows that

$$\lim_{P_1\to\infty} -\log(1-\alpha^2\rho^2)=-\log(1-\rho^2), \tag{102}$$

Eq. (94) follows by using (100) and (102) in (16).

The proof of Corollary 1 is now completed.

VI. Conclusions

In this paper we study the secret-key generation capacity over a two-way reciprocal block fading channel model with public discussion. There are two principle factors that contribute to the secret-key: (i) the reciprocity between channel gain of the main channel and (ii) interactive communication. We propose a two-phase separation scheme consisting of channel training and source emulation in each coherence block. Through a suitable allocation of power between the two phases, we show that the proposed scheme is near optimal in the high SNR regime for Rayleigh fading channels. More generally our results show that, even for moderately long coherence periods, the contribution of channel reciprocity towards the secret-key generation capacity is generally far less significant than source emulation. Thus in designing practical secret-key generation schemes over fading channels, one should aim to exploit the interactive nature of the system as opposed to passive training.

In future work it will be interesting to study the problem setup when an external public discussion channel is not available. The overhead for transmitting public messages needs to be explicitly considered in such systems and corresponding upper bounds may also have to be developed. In the low SNR regime it appears that the associated scheme may be very different than the proposed two-phase scheme [37]. Similarly extending the secret-key capacity results to the case of multiple antennas also remains an interesting topic for future research.

Appendix A
Proof of Lemma 1

Recall that we can express

$$\boldsymbol{\psi}_A^t=\left(\boldsymbol{\psi}_A^{t-1},\psi_A(t,1),\ldots,\psi_A(t,L)\right) \tag{103}$$

$$\boldsymbol{\psi}_B^t=\left(\boldsymbol{\psi}_B^{t-1},\psi_B(t,1),\ldots,\psi_B(t,L)\right) \tag{104}$$

where $\psi_A(t,j)$ denotes the j-th public message transmitted by node A after $x_A(t)$. It can be expressed as

$$\psi_A(t,j) = \Psi_A(y_A^t, \boldsymbol{\psi}_B^{t-1}, \psi_B(t,1), \dots, \psi_B(t,j-1), m_A) \tag{105}$$

The message $\psi_B(t,j)$ is defined in a similar fashion. Now we use the following inequalities:

$$
\begin{aligned}
&I(m_A, y_A^t, h_{BA}^N; m_B, y_B^t, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^t, \boldsymbol{\psi}_B^t) && (106)\\
&= I(m_A, y_A^t, h_{BA}^N; m_B, y_B^t, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}, \{\psi_A(t,j)\}_{1\le j\le L}, \{\psi_B(t,j)\}_{1\le j\le L}) && (107)\\
&\le I(m_A, y_A^t, h_{BA}^N, \psi_A(t,L); m_B, y_B^t, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}, \{\psi_A(t,j)\}_{1\le j\le L-1}, \{\psi_B(t,j)\}_{1\le j\le L}) && (108)\\
&= I(m_A, y_A^t, h_{BA}^N; m_B, y_B^t, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}, \{\psi_A(t,j)\}_{1\le j\le L-1}, \{\psi_B(t,j)\}_{1\le j\le L}) && (109)\\
&\le I(m_A, y_A^t, h_{BA}^N; m_B, y_B^t, h_{AB}^N, \psi_B(t,L) | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}, \{\psi_A(t,j)\}_{1\le j\le L-1}, \{\psi_B(t,j)\}_{1\le j\le L-1}) && (110)\\
&= I(m_A, y_A^t, h_{BA}^N; m_B, y_B^t, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}, \{\psi_A(t,j)\}_{1\le j\le L-1}, \{\psi_B(t,j)\}_{1\le j\le L-1}) && (111)
\end{aligned}
$$

where the equality in (109) follows from the fact that $\psi_A(t,L)$ is a deterministic function of

$$\{m_A, y_A^t, \boldsymbol{\psi}_B^{t-1}, \psi_B(t,1), \dots, \psi_B(t,L-1)\}$$

as indicated in (105). In a similar manner (111) follows from the fact that $\psi_B(t,L)$ is a deterministic function of

$$\{m_B, y_B^t, \boldsymbol{\psi}_A^{t-1}, \psi_A(t,1), \dots, \psi_A(t,L-1)\}.$$

Note that in following (106)-(111) we have eliminated the messages $(\psi_A(t,L), \psi_B(t,L))$. By recursively applying the same sequence of steps we successively eliminate the entire sequence of messages $\{\psi_A(t,j), \psi_B(t,j)\}$ and the claim in (31) follows.

The proof of (32) uses the sequence of steps analogous to (106)-(111) along with the fact that $\psi_A(0,j)$ is a deterministic function of $(m_A, \{\psi_B(0,k)\}_{1\le k\le j-1})$ and likewise $\psi_B(0,j)$ is a function of $(m_B, \{\psi_A(0,k)\}_{1\le k\le j-1})$. The details are completely analogous and will be omitted.

Appendix B
Proof of Lemma 2

Using the chain rule of mutual information we have

$$
\begin{aligned}
&I(m_A, y_A^t, h_{BA}^N; m_B, y_B^t, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1})\\
&= I(m_A, y_A^t, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1})\\
&\quad + I(m_A, y_A^t, h_{BA}^N; y_B(t) | \mathbf{z}^t, \mathbf{g}^N, y_B^{t-1}, h_{AB}^N, m_B, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}) && (112)\\
&= I(m_A, y_A^t, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1})\\
&\quad + I(m_A, y_A^{t-1}, h_{BA}^N; y_B(t) | \mathbf{z}^t, \mathbf{g}^N, y_B^{t-1}, h_{AB}^N, m_B, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1})\\
&\qquad + I(y_A(t); y_B(t) | \mathbf{z}^t, \mathbf{g}^N, y_B^{t-1}, h_{AB}^N, m_B, y_A^{t-1}, h_{BA}^N, m_A, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}) && (113)\\
&= I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1})
\end{aligned}
$$

$$+ I(y_A(t); m_B, y_B^{t-1}, h_{AB}^N | m_A, y_A^{t-1}, h_{BA}^N, \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1})$$
$$+ I(m_A, y_A^{t-1}, h_{BA}^N; y_B(t) | \mathbf{z}^t, \mathbf{g}^N, y_B^{t-1}, h_{AB}^N, m_B, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1})$$
$$+ I(y_A(t); y_B(t) | \mathbf{z}^t, \mathbf{g}^N, y_B^{t-1}, h_{AB}^N, m_B, y_A^{t-1}, h_{BA}^N, m_A, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}) \tag{114}$$

where we use the chain rule of mutual information in (112), (113) and (114). Further note that in the proposed coding scheme $x_A(t) = f_{A,t}(m_A, y_A^{t-1}, \boldsymbol{\psi}_B^{t-1})$ and $x_B(t) = f_{B,t}(m_B, y_B^{t-1}, \boldsymbol{\psi}_A^{t-1})$. Therefore we can express:

$$I(y_A(t); y_B(t) | \mathbf{z}^t, \mathbf{g}^N, y_B^{t-1}, h_{AB}^N, m_B, y_A^{t-1}, h_{BA}^N, m_A, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1})$$
$$= I(y_A(t); y_B(t) | \mathbf{z}^t, \mathbf{g}^N, y_B^{t-1}, h_{AB}^N, m_B, y_A^{t-1}, h_{BA}^N, m_A, x_A(t), x_B(t), \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}) = 0 \tag{115}$$

where we use the fact that the forward and reverse channels are memoryless and the additive noise variables $n_A(t)$ and $n_B(t)$ are independent (c.f. (1), (3)) and therefore:

$$y_B(t) \leftrightarrow (x_A(t), h_{AB}(t), \boldsymbol{\Omega}, x_B(t), h_{BA}(t)) \leftrightarrow y_A(t) \tag{116}$$

where

$$\boldsymbol{\Omega} \triangleq \left\{\mathbf{z}^t, \mathbf{g}^N, y_A^{t-1}, y_B^{t-1}, h_{AB}^{t-1}, h_{AB,t+1}^N, h_{BA}^{t-1}, h_{BA,t+1}^N, m_A, m_B, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}\right\} \tag{117}$$

denotes the remainder of the terms in (115) which are all independent of $\{n_A(t), n_B(t)\}$.

We next consider the second and third term in (114) and show that:

$$I(y_A(t); m_B, y_B^{t-1}, h_{AB}^N | m_A, y_A^{t-1}, h_{BA}^N, \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}) \le I(y_A(t); x_B(t) | z_{BE}(t), g_{BE}(t), h_{BA}(t)) \tag{118}$$
$$I(m_A, y_A^{t-1}, h_{BA}^N; y_B(t) | \mathbf{z}^t, \mathbf{g}^N, y_B^{t-1}, h_{AB}^N, m_B, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}) \le I(x_A(t); y_B(t) | z_{AE}(t), g_{AE}(t), h_{AB}(t)) \tag{119}$$

We establish (118) below. Eq. (119) can be established in an analogous fashion and its proof will be omitted.

$$I(y_A(t); m_B, y_B^{t-1}, h_{AB}^N | m_A, y_A^{t-1}, h_{BA}^N, \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1})$$
$$= I(y_A(t); m_B, y_B^{t-1}, h_{AB}^N, x_B(t) | m_A, y_A^{t-1}, h_{BA}^N, \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}, x_A(t)) \tag{120}$$
$$= \hbar(y_A(t) | m_A, y_A^{t-1}, h_{BA}^N, \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}, x_A(t))$$
$$- \hbar(y_A(t) | m_B, y_B^{t-1}, h_{AB}^N, x_B(t), m_A, y_A^{t-1}, h_{BA}^N, \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}, x_A(t)) \tag{121}$$
$$= \hbar(y_A(t) | m_A, y_A^{t-1}, h_{BA}^N, \mathbf{z}^t, \mathbf{g}^N, \boldsymbol{\psi}_A^{t-1}, \boldsymbol{\psi}_B^{t-1}, x_A(t)) - \hbar(y_A(t) | x_B(t), h_{BA}(t)) \tag{122}$$
$$\le \hbar(y_A(t) | h_{BA}(t), z_{BE}(t), g_{BE}(t)) - \hbar(y_A(t) | x_B(t), h_{BA}(t)) \tag{123}$$
$$= I(y_A(t); x_B(t) | h_{BA}(t), z_{BE}(t), g_{BE}(t)) \tag{124}$$

where (120) uses the fact that $x_A(t)$ is a function of $(m_A, y_A^{t-1}, \boldsymbol{\psi}_B^{t-1})$ and similarly $x_B(t)$ is a function of $(m_B, y_B^{t-1}, \boldsymbol{\psi}_A^{t-1})$. Eq. (122) uses the fact that with $\boldsymbol{\Omega}$ defined in (117), the Markov relation

$$y_A(t) \leftrightarrow (h_{BA}(t), x_B(t)) \leftrightarrow (\boldsymbol{\Omega}, x_A(t), h_{AB}(t)) \tag{125}$$

holds. Eq. (123) follows from the fact that conditioning reduces entropy. This completes the proof of (118).

We finally show that the first term in (114) can be upper bounded as

$$I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \psi_A^{t-1}, \psi_B^{t-1}) \le I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, \psi_A^{t-1}, \psi_B^{t-1}). \tag{126}$$

This is also established using the chain rule of mutual information as follows.

$$\begin{aligned}
&I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \psi_A^{t-1}, \psi_B^{t-1}) \\
&= I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, z_{AE}(t), z_{BE}(t), \psi_A^{t-1}, \psi_B^{t-1}) && (127)\\
&\le I(m_A, y_A^{t-1}, h_{BA}^N, z_{AE}(t); m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, z_{BE}(t), \psi_A^{t-1}, \psi_B^{t-1}) && (128)\\
&= I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, z_{BE}(t), \psi_A^{t-1}, \psi_B^{t-1}) \\
&\qquad + I(z_{AE}(t); m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, z_{BE}(t), m_A, y_A^{t-1}, h_{BA}^N, \psi_A^{t-1}, \psi_B^{t-1}) && (129)
\end{aligned}$$

Next recall that since $x_A(t) = f_{A,t}(m_A, y_A^{t-1}, \psi_B^{t-1})$ we have:

$$\begin{aligned}
&I(z_{AE}(t); m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, z_{BE}(t), m_A, y_A^{t-1}, h_{BA}^N, \psi_A^{t-1}, \psi_B^{t-1}) \\
&= I(z_{AE}(t); m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, z_{BE}(t), m_A, y_A^{t-1}, h_{BA}^N, x_A(t), \psi_A^{t-1}, \psi_B^{t-1}) \\
&= \hbar(z_{AE}(t) | \mathbf{z}^{t-1}, \mathbf{g}^N, z_{BE}(t), m_A, y_A^{t-1}, h_{BA}^N, x_A(t), \psi_A^{t-1}, \psi_B^{t-1}) \\
&\qquad - \hbar(z_{AE}(t) | \mathbf{z}^{t-1}, \mathbf{g}^N, z_{BE}(t), m_A, y_A^{t-1}, h_{BA}^N, x_A(t), m_B, y_B^{t-1}, h_{AB}^N, \psi_A^{t-1}, \psi_B^{t-1}) \\
&= \hbar(z_{AE}(t) | x_A(t), g_{AE}(t)) - \hbar(z_{AE}(t) | x_A(t), g_{AE}(t)) = 0 && (130)
\end{aligned}$$

where (130) uses the fact that the forward channel (c.f. (2)) satisfies

$$z_{AE}(t) \leftrightarrow (x_A(t), g_{AE}(t)) \leftrightarrow (\mathbf{z}^{t-1}, \mathbf{g}^{t-1}, \mathbf{g}_{t+1}^N, g_{BE}(t), z_{BE}(t), m_A, y_A^{t-1}, h_{BA}^N, m_B, y_B^{t-1}, h_{AB}^N, \psi_A^{t-1}, \psi_B^{t-1}). \tag{131}$$

Thus substituting (130) into (129) we have that:

$$\begin{aligned}
&I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^t, \mathbf{g}^N, \psi_A^{t-1}, \psi_B^{t-1}) \\
&\le I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, z_{BE}(t), \psi_A^{t-1}, \psi_B^{t-1}) && (132)\\
&\le I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N, z_{BE}(t) | \mathbf{z}^{t-1}, \mathbf{g}^N, \psi_A^{t-1}, \psi_B^{t-1}) && (133)\\
&\le I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, \psi_A^{t-1}, \psi_B^{t-1}) \\
&\qquad + I(m_A, y_A^{t-1}, h_{BA}^N; z_{BE}(t) | m_B, y_B^{t-1}, h_{AB}^N, \mathbf{z}^{t-1}, \mathbf{g}^N, \psi_A^{t-1}, \psi_B^{t-1}) && (134)\\
&\le I(m_A, y_A^{t-1}, h_{BA}^N; m_B, y_B^{t-1}, h_{AB}^N | \mathbf{z}^{t-1}, \mathbf{g}^N, \psi_A^{t-1}, \psi_B^{t-1}) && (135)
\end{aligned}$$

The justification that the second term in (134) equals zero can be established in an identical fashion as (130), and will be omitted. This completes the proof of (126).

The claim in Lemma 2 follows upon substituting (115), (118), (119) and (126) into (114).

APPENDIX C
PROOF OF PROP. 1

We provide the coding theorem associated with Prop. 1 below. As stated in section IV, the two-phase transmission scheme generates correlated sequences in Table I, which are then discretized at the legitimate terminals, as explained in section IV-A. For convenience table II shows the sequences available at each terminal after discretization below.

TABLE II
CORRELATED SOURCE AND CHANNEL SEQUENCES AFTER DISCRETIZATION

Sequence/Terminal	A	B	E
Channel Sequence	$\left[\hat{h}^K_{BA}\right]_j$	$\left[\hat{h}^K_{AB}\right]_j$	(g^K_{AE}, g^K_{BE})
Source Sequence - Forward Channel	$[\bar{\mathbf{x}}^K_A]_k$	$\left[\bar{\mathbf{y}}^K_{B,k}\right]_l$	$\bar{\mathbf{z}}^K_{AE,k}$
Source Sequence - Reverse Channel	$\left[\bar{\mathbf{y}}^K_{A,k}\right]_l$	$[\bar{\mathbf{x}}^K_B]_k$	$\bar{\mathbf{z}}^K_{BE,k}$

We establish the existence of a codebook $\tilde{\mathcal{C}}^{\otimes}$ that satisfies the following for sufficiently large K:

1) Both the legitimate terminals can decode sequences $\left(\left[\hat{h}^K_{AB}\right]_j, \left[\hat{h}^K_{BA}\right]_j, \left[\bar{\mathbf{y}}^K_{A,k}\right]_l, \left[\bar{\mathbf{y}}^K_{B,k}\right]_l\right)$ with an error probability no greater than $\delta_K(\varepsilon)$.
2) The equivocation at the eavesdropper satisfies:
$$\frac{1}{N} H\left(\left[\hat{h}^K_{AB}\right]_j, \left[\hat{h}^K_{BA}\right]_j, [\bar{\mathbf{y}}^K_{A,k}]_l, [\bar{\mathbf{y}}^K_{B,k}]_l \mid \Theta, \bar{\mathbf{z}}^K_{AE,k}, \bar{\mathbf{z}}^K_{BE,k}, g^K_{AE}, g^K_{BE}\right) \geq R^-_{\Delta} - \delta^1_K(\varepsilon) \tag{136}$$
where the quantity R^-_{Δ} is defined in (60), Θ denotes the collection of messages that need to be exchanged over the discussion channel, and $\delta_K(\varepsilon), \delta^1_K(\varepsilon)$ are function that can be made sufficiently small by selecting K sufficiently large.

An overview of the coding scheme is illustrated in Fig. 4. Note that the proposed code construction $\tilde{\mathcal{C}}^{\otimes}$ will only generate a common tuple of sequences at the legitimate receivers satisfying (136). The secret-key generation step will be discussed subsequently.

A. Codebook Construction and Error Analysis

We define:

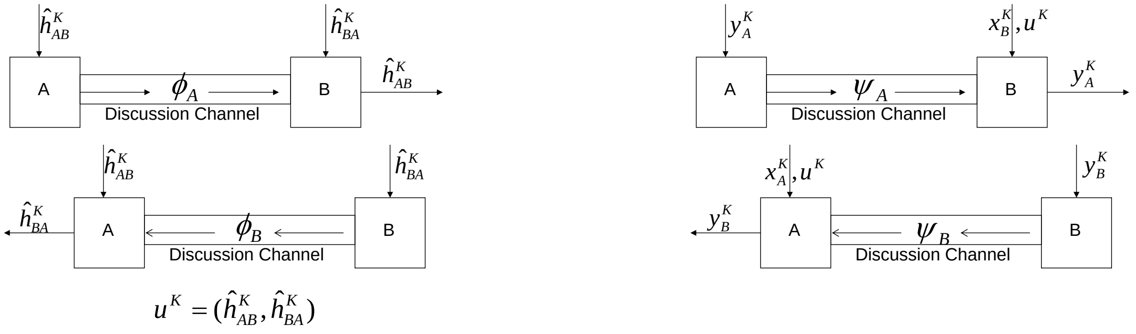
$$R^-_{B,c} = H\left(\left[\hat{h}_{AB}\right]_j \middle| \left[\hat{h}_{BA}\right]_j\right) + 2\varepsilon \tag{137}$$

$$R^-_{A,c} = H\left(\left[\hat{h}_{BA}\right]_j \middle| \left[\hat{h}_{AB}\right]_j\right) + 2\varepsilon \tag{138}$$

$$R^-_{B,s} = H\left([\bar{\mathbf{y}}_{B,k}]_l \mid [\bar{\mathbf{x}}_A]_k, [u]_j\right) + 4\varepsilon \tag{139}$$

$$R^-_{A,s} = H\left([\bar{\mathbf{y}}_{A,k}]_l \mid [\bar{\mathbf{x}}_B]_k, [u]_j\right) + 4\varepsilon \tag{140}$$

where $[u]_j \triangleq \left(\left[\hat{h}_{AB}\right]_j, \left[\hat{h}_{BA}\right]_j\right)$ and ε is a sufficiently small positive constant that will be specified in the sequel.



Phase I: Reconciliation of channel-estimates.

Phase II: Reconciliation of source sequences.

Fig. 4. Reconciliation of source and channel sequences in the proposed coding scheme with public discussion. The public messages are denoted by $\Theta = (\Phi_A, \Phi_B, \Psi_A, \Psi_B)$. In the analysis of the coding scheme, we use discretized version of the source and channel sequences.

We next randomly sample four codebooks $\mathcal{C}_A$, $\mathcal{C}_B$, $\mathcal{S}_A$ and $\mathcal{S}_B$ as discussed below.

- The codebook $\mathcal{C}_B$ is sampled by randomly and uniformly partitioning the set of all typical sequences $\left[\hat{h}_{AB}^K\right]_j \in T_\varepsilon^K\left(\left[\hat{h}_{AB}\right]_j\right)$ into $2^{K\bar{R}_{B,c}}$ bins, such that there are $2^{K\left(I\left(\left[\hat{h}_{AB}\right]_j \,;\, \left[\hat{h}_{BA}\right]_j\right)-\varepsilon\right)}$ sequences per bin.
- The codebook $\mathcal{C}_A$ is sampled by randomly and uniformly partitioning the set of all typical sequences $\left[\hat{h}_{BA}^K\right]_j \in T_\varepsilon^K\left(\left[\hat{h}_{BA}\right]_j\right)$ into $2^{K\bar{R}_{A,c}}$ bins, such that there are $2^{K\left(I\left(\left[\hat{h}_{AB}\right]_j \,;\, \left[\hat{h}_{BA}\right]_j\right)-\varepsilon\right)}$ sequences per bin.
- The codebook $\mathcal{S}_B$ is sampled by randomly and uniformly partitioning the set of all typical sequences $\left[\bar{\mathbf{y}}_{B,k}^K\right]_l \in T_\varepsilon^K\left([\bar{\mathbf{y}}_{B,k}]_l\right)$ into $2^{K\bar{R}_{B,s}}$ bins such that there are $2^{K\left(I\left([\mathbf{y}_{B,k}]_l;[\bar{\mathbf{x}}_A]_k,[u]_j\right)-3\varepsilon\right)}$ sequences per bin.
- The codebook $\mathcal{S}_A$ is sampled by randomly and uniformly partitioning the set of all typical sequences $\left[\bar{\mathbf{y}}_{A,k}^K\right]_l \in T_\varepsilon^K\left([\bar{\mathbf{y}}_{A,k}]_l\right)$ into $2^{K\bar{R}_{A,s}}$ bins such that there are $2^{K\left(I\left([\bar{\mathbf{y}}_{A,k}]_l \,;\, [\bar{\mathbf{x}}_B]_k,[u]_j\right)-3\varepsilon\right)}$ sequences in each bin.

The encoding steps are as follows.

- Given the sequence $\left[\hat{h}_{AB}^K\right]_j$, terminal B finds bin in $\mathcal{C}_B$ to which it belongs and transmits the bin-index ϕ_B over the discussion channel.
- Given the sequence $\left[\hat{h}_{BA}^K\right]_j$, terminal A finds the bin in $\mathcal{C}_A$ to which it belongs and transmits the bin index ϕ_A over the discussion channel.
- Given the sequence $\left[\bar{\mathbf{y}}_{A,k}^K\right]_l$, terminal A finds the bin in $\mathcal{S}_A$ to which it belongs and transmits the bin index ψ_A over the discussion channel.
- Given the sequence $\left[\bar{\mathbf{y}}_{B,k}^K\right]_l$, terminal B finds the bin in $\mathcal{S}_B$ to which it belongs and transmits the bin index ψ_B over the discussion channel.

We next sketch the decoding at legitimate terminals. The decoding proceeds in two phases. In the first phase the terminals attempt to reconstruct channel sequences and in the second phase they attempt to reconstruct the source sequences.

Upon receiving ϕ_B, terminal A searches for all sequences in the bin associated with ϕ_B that are jointly typical with $\left[\hat{h}_{BA}^K\right]_j$. If there is a unique sequence $\left[\tilde{h}_{AB}^K\right]_j$ that satisfies this, then it is selected as the reconstruction sequence. Terminal A declares $\left[\hat{u}_A^K\right]_j = \left(\left[\tilde{h}_{AB}^K\right]_j, \left[\hat{h}_{BA}^K\right]_j\right)$ to be the estimate of the joint sequence pair. In a similar fashion, upon receiving ϕ_A, terminal B generates $\left[\hat{u}_B^K\right]_j = \left(\left[\hat{h}_{AB}^K\right]_j, \left[\tilde{h}_{BA}^K\right]_j\right)$ to be the joint sequence pair. We declare an error if either $\left\{\left[\hat{u}_A^K\right]_j \neq \left[u^K\right]_j\right\}$ or $\left\{\left[\hat{u}_B^K\right]_j \neq \left[u^K\right]_j\right\}$. Through standard arguments [36, Chapter 10], it can be shown that the error probability approaches zero as $K \to \infty$ given (137) and (138).

In the second phase of decoding, the terminals attempt to decode the source sequences. Upon receiving ψ_B, and given $\left(\left[\bar{\mathbf{x}}_A^K\right]_k, \left[\hat{u}_A^K\right]_j\right)$, terminal A searches for all sequences in the bin associated ψ_B that satisfy $\left(\left[\mathbf{y}_{B,k}^K\right]_l, \left[\bar{\mathbf{x}}_A^K\right]_k, \left[\hat{u}_A^K\right]_j\right) \in T_{2\varepsilon}^K\left([\bar{\mathbf{y}}_{B,k}]_l, [\bar{\mathbf{x}}_A]_k, u\right)$. If there is a unique sequence $\left[\hat{\mathbf{y}}_{B,k}^K\right]_l$ that satisfies this, then it is selected as the reconstruction sequence. We declare $\Gamma_A = \left(\left[\hat{u}_A^K\right]_j, \left[\mathbf{y}_{A,k}^K\right]_l, \left[\hat{\mathbf{y}}_{B,k}^K\right]_l\right)$ to be the common reconstruction sequences at terminal A. Likewise terminal B reconstructs $\left[\hat{\mathbf{y}}_{A,k}^K\right]_l$ upon receiving ψ_A by searching for all sequences in the associated bin of ψ_A that satisfy $\left(\left[\mathbf{y}_{A,k}^K\right]_l, \left[\bar{\mathbf{x}}_B^K\right]_k, \left[\hat{u}_B^K\right]_j\right) \in T_{2\varepsilon}^K\left([\mathbf{y}_{A,k}]_l, [\bar{\mathbf{x}}_B]_k, [u]_j\right)$. We declare $\Gamma_B = \left(\left[\hat{u}_B^K\right]_j, \left[\hat{\mathbf{y}}_{A,k}^K\right]_l, \left[\mathbf{y}_{B,k}^K\right]_l\right)$ to be the common reconstruction sequences at terminal B. We declare an error if $\{\Gamma_A \neq \Gamma_B\}$. Through standard analysis, it can be shown that

$$E[\Pr(\Gamma_A \neq \Gamma \bigcup \Gamma_B \neq \Gamma) |\ \mathcal{C}^\otimes] \leq \delta_K(\varepsilon), \tag{141}$$

where $\delta_K(\varepsilon) \to 0$ as $\varepsilon \to 0$ and $K \to \infty$ where $\Gamma \triangleq \left(\left[u^K\right]_j, \left[\mathbf{y}_{A,k}^K\right]_l, \left[\mathbf{y}_{B,k}^K\right]_l\right)$ and $\mathcal{C}^\otimes \triangleq (\mathcal{C}_A, \mathcal{C}_B, \mathcal{S}_A, \mathcal{S}_B)$ denotes the collection of all the four codebooks.

B. Equivocation Analysis

We show that our proposed random code ensemble also satisfies:

$$\frac{1}{N} H\left(\left[\hat{h}_{AB}^K\right]_j, \left[\hat{h}_{BA}^K\right]_j, [\bar{\mathbf{y}}_{A,k}^K]_l, [\bar{\mathbf{y}}_{B,k}^K]_l \mid \Theta, \bar{\mathbf{z}}_{AE}^K, \bar{\mathbf{z}}_{BE}^K, g_{AE}^K, g_{BE}^K, \mathcal{C}^\otimes\right) \geq R_\Delta^- - \delta_K^1(\varepsilon) \tag{142}$$

with $\Theta = (\psi_A, \psi_B, \phi_A, \phi_B)$ and $\mathcal{C}^\otimes \triangleq (\mathcal{C}_A, \mathcal{C}_B, \mathcal{S}_A, \mathcal{S}_B)$. In the following analysis we introduce $\bar{\mathbf{z}}_k^K = \left(\bar{\mathbf{z}}_{AE,k}^K, \bar{\mathbf{z}}_{BE,k}^K\right)$ and $\mathbf{g}^K = \left(g_{AE}^K, g_{BE}^K\right)$. First consider the following:

$$\begin{aligned} & H\left(\left[\hat{h}_{AB}^K\right]_j, \left[\hat{h}_{BA}^K\right]_j, [\bar{\mathbf{y}}_{A,k}^K]_l, [\bar{\mathbf{y}}_{B,k}^K]_l \mid \psi_A, \psi_B, \phi_A, \phi_B, \bar{\mathbf{z}}_k^K, \mathbf{g}^K, \mathcal{C}^\otimes\right) \\ & \quad \geq H\left(\left[\hat{h}_{AB}^K\right]_j, \left[\hat{h}_{BA}^K\right]_j, [\bar{\mathbf{y}}_{A,k}^K]_l, [\bar{\mathbf{y}}_{B,k}^K]_l \mid \bar{\mathbf{z}}_k^K, \mathbf{g}^K, \mathcal{C}^\otimes\right) - H(\phi_A|\mathcal{C}^\otimes) - H(\phi_B|\mathcal{C}^\otimes) - H(\psi_A|\mathcal{C}^\otimes) - H(\psi_B|\mathcal{C}^\otimes) \end{aligned} \tag{143}$$

$$= H\left(\left[\hat{h}_{AB}^K\right]_j, \left[\hat{h}_{BA}^K\right]_j, [\bar{\mathbf{y}}_{A,k}^K]_l, [\bar{\mathbf{y}}_{B,k}^K]_l \mid \bar{\mathbf{z}}_k^K, \mathbf{g}^K\right) - H(\phi_A|\mathcal{C}^\otimes) - H(\phi_B|\mathcal{C}^\otimes) - H(\psi_A|\mathcal{C}^\otimes) - H(\psi_B|\mathcal{C}^\otimes) \tag{144}$$

where the last step follows from the fact that all the source and channel sequences in Table II are sampled independently of the codebooks $(\mathcal{C}_A, \mathcal{C}_B, \mathcal{S}_A, \mathcal{S}_B)$ in our code construction. The first term in (144) can be lower bounded as follows:

$$
\begin{aligned}
&H\left(\left[\hat{h}^K_{AB}\right]_j, \left[\hat{h}^K_{BA}\right]_j, [\bar{\mathbf{y}}^K_{A,k}]_l, [\bar{\mathbf{y}}^K_{B,k}]_l \mid \bar{\mathbf{z}}^K_k, \mathbf{g}^K\right)\\
&\geq H\left(\left[\hat{h}^K_{AB}\right]_j, \left[\hat{h}^K_{BA}\right]_j\right) + H\left([\bar{\mathbf{y}}^K_{A,k}]_l, [\bar{\mathbf{y}}^K_{B,k}]_l | \bar{\mathbf{z}}^K_k, \mathbf{g}^K, \left[\hat{h}^K_{AB}\right]_j, \left[\hat{h}^K_{BA}\right]_j\right) && (145)\\
&\geq H\left(\left[\hat{h}^K_{AB}\right]_j, \left[\hat{h}^K_{BA}\right]_j\right) + H([\bar{\mathbf{y}}^K_{A,k}]_l, [\bar{\mathbf{y}}^K_{B,k}]_l | \bar{\mathbf{z}}^K_k, \mathbf{g}^K, h^K_{AB}, h^K_{BA}) && (146)\\
&= H\left(\left[\hat{h}^K_{AB}\right]_j, \left[\hat{h}^K_{BA}\right]_j\right) + H([\bar{\mathbf{y}}^K_{A,k}]_l | \bar{\mathbf{z}}^K_{BE,k}, g^K_{BE}, h^K_{BA}) + H([\bar{\mathbf{y}}^K_{B,k}]_l | \bar{\mathbf{z}}^K_{AE,k}, g^K_{AE}, h^K_{AB}) && (147)\\
&= KH\left(\left[\hat{h}_{AB}\right]_j, \left[\hat{h}_{BA}\right]_j\right) + KH\left([\bar{\mathbf{y}}_{A,k}]_l | \bar{\mathbf{z}}_{BE,k}, g_{BE}, h_{BA}\right) + KH([\bar{\mathbf{y}}_{B,k}]_l | \bar{\mathbf{z}}_{AE,k}, g_{AE}, h_{AB}) && (148)
\end{aligned}
$$

where we use the chain rule of entropy and the fact that $(\hat{h}^K_{AB}, \hat{h}^K_{BA})$ is independent of $(\bar{\mathbf{z}}^K, \mathbf{g}^K)$ in (145); in (146) we use the fact that conditioning reduces entropy as well as $\hat{h}^K_{AB}$ and $\hat{h}^K_{BA}$ are independent of the remaining variables given h^K_{AB} and h^K_{BA} respectively; in (147) we use the fact that $\bar{\mathbf{x}}^K_A$ and $\bar{\mathbf{x}}^K_B$ are sampled independently and hence the following Markov conditions hold:

$$
\begin{aligned}
\bar{\mathbf{y}}^K_A \leftrightarrow (\bar{\mathbf{z}}^K_{BE}, g^K_{BE}, h^K_{BA}) \leftrightarrow (\bar{\mathbf{y}}^K_B, g^K_{AE}, h^K_{AB}), && (149)\\
\bar{\mathbf{y}}^K_B \leftrightarrow (\bar{\mathbf{z}}^K_{AE}, g^K_{AE}, h^K_{AB}) \leftrightarrow (\bar{\mathbf{y}}^K_A, g^K_{BE}, h^K_{BA}). && (150)
\end{aligned}
$$

which are established in Appendix D. The discretized random variables are also inherit the same properties. Eq. (148) follows from the fact that the sequence pair (h^K_{AB}, h^K_{BA}) is sampled i.i.d. and furthermore $\bar{\mathbf{x}}^K_A$ and $\bar{\mathbf{x}}^K_B$ are also sampled i.i.d.

The remaining terms in (144) can be upper bounded using (137)-(140).

$$
\begin{aligned}
H(\phi_A \mid \mathcal{C}^{\otimes}) &\leq K\left\{H\left(\left[\hat{h}_{BA}\right]_j \middle| \left[\hat{h}_{AB}\right]_j\right) + 2\varepsilon\right\}, && (151)\\
H(\phi_B \mid \mathcal{C}^{\otimes}) &\leq K\left\{H\left(\left[\hat{h}_{AB}\right]_j \middle| \left[\hat{h}_{BA}\right]_j\right) + 2\varepsilon\right\}, && (152)\\
H(\psi_A \mid \mathcal{C}^{\otimes}) &\leq K\left\{H\left([\bar{\mathbf{y}}_{A,k}]_l \mid [\bar{\mathbf{x}}_B]_k, [u]_j\right) + 4\varepsilon\right\}, && (153)\\
H(\psi_B \mid \mathcal{C}^{\otimes}) &\leq K\left\{H\left([\bar{\mathbf{y}}_{B,k}]_l \mid [\bar{\mathbf{x}}_A]_k, [u]_j\right) + 4\varepsilon\right\}. && (154)
\end{aligned}
$$

Using (151) and (152), we have the following:

$$
\begin{aligned}
&KH\left(\left[\hat{h}_{AB}\right]_j, \left[\hat{h}_{BA}\right]_j\right) - H(\phi_A | \mathcal{C}^{\otimes}) - H(\phi_B | \mathcal{C}^{\otimes})\\
&\geq KH\left(\left[\hat{h}_{AB}\right]_j, \left[\hat{h}_{BA}\right]_j\right) - KH\left(\left[\hat{h}_{AB}\right]_j \middle| \left[\hat{h}_{BA}\right]_j\right) - KH\left(\left[\hat{h}_{BA}\right]_j \middle| \left[\hat{h}_{AB}\right]_j\right) - 4K\varepsilon && (155)\\
&= K \cdot I\left(\left[\hat{h}_{AB}\right]_j ; \left[\hat{h}_{BA}\right]_j\right) - 4K\varepsilon. && (156)
\end{aligned}
$$

Using the fact that each element in $[\bar{\mathbf{x}}_B]_k = ([x_B(1)]_k, \ldots, [x_B(T-1)]_k)$ is sampled i.i.d. and using (153) and letting $\varepsilon' = \frac{\varepsilon}{T-1}$ we have

$$
KH([\bar{\mathbf{y}}_{A,k}]_l | \bar{\mathbf{z}}_{BE,k}, g_{BE}, h_{BA}) - H(\psi_A | \mathcal{C}^{\otimes})
$$

$$\geq KH([\bar{\mathbf{y}}_{A,k}]_l|\bar{\mathbf{z}}_{BE},\mathsf{g}_{BE},\mathsf{h}_{BA})-KH([\bar{\mathbf{y}}_{A,k}]_l|[\bar{\mathbf{x}}_B]_k,[\mathsf{u}]_j)-4K\varepsilon \tag{157}$$

$$\geq K(T-1)\left\{H([\mathsf{y}_{A,k}]_l|\mathsf{z}_{BE},\mathsf{g}_{BE},\mathsf{h}_{BA})-H([\mathsf{y}_{A,k}]_l|[\mathsf{x}_B]_k,[\mathsf{u}]_j)-4\varepsilon'\right\} \tag{158}$$

$$=K(T-1)\Big\{I([\mathsf{y}_{A,k}]_l;[\mathsf{x}_B]_k,[\mathsf{u}]_j)-I([\mathsf{y}_{A,k}]_l;\mathsf{z}_{BE,k},\mathsf{g}_{BE},\mathsf{h}_{BA})-4\varepsilon'\Big\}. \tag{159}$$

The justification of (158) is as follows. From the channel model (2) and (4) the sequence $[\bar{\mathbf{y}}_{A,k}]_l = ([\mathsf{y}_{A,k}(1)]_l,\ldots,[\mathsf{y}_{A,k}(T-1)]_l)$ of $T-1$ random variables, when conditioned on h_{BA} is an i.i.d. sequence and similarly the sequence $\bar{\mathbf{z}}_{BE,k}=(\mathsf{z}_{BE,k}(1),\ldots,\mathsf{z}_{BE,k}(T-1))$ when conditioned on g_{BE} is an i.i.d. sequence. Therefore the first term in (157) can be expressed as

$$H([\bar{\mathbf{y}}_{A,k}]_l|\bar{\mathbf{z}}_{BE,k},\mathsf{g}_{BE},\mathsf{h}_{BA})=(T-1)H([\mathsf{y}_{A,k}]_l|\mathsf{z}_{BE,k},\mathsf{g}_{BE},\mathsf{h}_{BA}). \tag{160}$$

Furthermore using the fact that conditioning reduces entropy, the second term in (157) can be upper bounded as

$$H\left([\bar{\mathbf{y}}_{A,k}]_l|[\bar{\mathbf{x}}_B]_k,[\mathsf{u}]_j\right)\leq (T-1)H\left([\mathsf{y}_{A,k}]_l|[\mathsf{x}_B]_k,[\mathsf{u}]_j\right). \tag{161}$$

Note that (158) follows by substituting (160) and (161) into (157). In a similar fashion we can show that

$$\begin{aligned}&KH([\bar{\mathbf{y}}_{B,k}]_l|\bar{\mathbf{z}}_{AE,k},\mathsf{g}_{AE},\mathsf{h}_{AB})-H(\psi_B|\mathcal{C}^{\otimes})\\&\geq K(T-1)\left\{I([\mathsf{y}_{B,k}]_l;[\mathsf{x}_A]_k,[\mathsf{u}]_j)-I([\mathsf{y}_{B,k}]_l;\mathsf{z}_{AE,k},\mathsf{g}_{AE},\mathsf{h}_{AB})-4\varepsilon'\right\}\end{aligned} \tag{162}$$

Combining (156), (159) and (162) we obtain that

$$\begin{aligned}&H([\bar{\mathbf{y}}_{A,k}^K]_l,[\bar{\mathbf{y}}_{B,k}^K]_l,[\mathsf{u}^K]_j|\bar{\mathbf{z}}_k^K,\mathbf{g}^K,\phi_A,\phi_B,\psi_A,\psi_B,\mathcal{C}^{\otimes})\\&\quad\geq KI\left(\left[\hat{\mathsf{h}}_{AB}\right]_j;\left[\hat{\mathsf{h}}_{BA}\right]_j\right)-4K\varepsilon-8K\varepsilon'\\&\quad+K(T-1)\left\{I([\mathsf{y}_{A,k}]_l;[\mathsf{x}_B]_k,[\mathsf{u}]_j)-I([\mathsf{y}_{A,k}]_l;\mathsf{z}_{BE,k},\mathsf{g}_{BE},\mathsf{h}_{BA})-2\varepsilon'\right\}\\&\quad+K(T-1)\left\{I([\mathsf{y}_{B,k}]_l;[\mathsf{x}_A]_k,[\mathsf{u}]_j)-I([\mathsf{y}_{B,k}]_l;\mathsf{z}_{AE,k},\mathsf{g}_{AE},\mathsf{h}_{AB})–2\varepsilon'\right\}\end{aligned} \tag{163}$$

Dividing both sides by $N=K\cdot T$ the relation (142) now follows. By examining (141) and (142) it follows that there exists a deterministic codebook $\tilde{\mathcal{C}}^{\otimes}$ in the ensemble, that simultaneously satisfies the equivocation constraint (136) and

$$\Pr(\Gamma_A\neq\Gamma\bigcup\Gamma_B\neq\Gamma)\leq\delta(\varepsilon). \tag{164}$$

C. Secret-Key Generation

Our analysis thus far has established the existence of a codebook $\tilde{\mathcal{C}}^{\otimes}$ that generates a common tuple of sequences $\left([\mathsf{u}^K],\left[\bar{\mathbf{y}}_{A,k}^K\right]_l,\left[\bar{\mathbf{y}}_{B,k}^K\right]_l\right)$, satisfies the equivocation constraint (136) and the reliability constraint (164). We next discuss how one can use the codebook $\tilde{\mathcal{C}}^{\otimes}$ to generate a common secret-key at the two terminals.

We consider transmission over a total of M macro-blocks. Each macro-block i spans K coherence blocks the terminals sample sequences $\bar{\mathbf{x}}_A^K(i)$ and $\bar{\mathbf{x}}_B^K(i)$ in an i.i.d. fashion, independently from the previous blocks. Thereafter we execute the steps discussed in the previous sections, which results in the following observations at the legitimate receivers and the eavesdropper:

$$\Gamma_A(i) = \left(\left[\hat{u}_A^K(i)\right]_j, \left[\mathbf{y}_{A,k}^K(i)\right]_l, \left[\hat{\mathbf{y}}_{B,k}^K(i)\right]_l\right) \tag{165}$$

$$\Gamma_B(i) = \left(\left[\hat{u}_B^K(i)\right]_j, \left[\hat{\mathbf{y}}_{A,k}^K(i)\right]_l, \left[\mathbf{y}_{B,k}^K(i)\right]_l\right) \tag{166}$$

$$\Omega(i) = \left(\bar{\mathbf{z}}_k^K(i), \mathbf{g}^K(i), \psi_A(i), \psi_B(i), \phi_A(i), \phi_B(i)\right). \tag{167}$$

Note that $\{\Gamma_A(i), \Gamma_B(i), \Omega(i)\}$ are sampled i.i.d. across the macro-blocks. Furthermore for each $i \in \{1, 2, \ldots, M\}$, they satisfy (164) and (136). Thus an achievable *one-way* secret-key generation rate for the sequence $\{\Gamma_A(i), \Gamma_B(i), \Omega(i)\}$ is (see e.g., [15]):

$$\begin{aligned} R^- &= \frac{1}{N}\left[I(\Gamma_A;\Gamma_B) - I(\Gamma_A;\Omega)\right] && (168)\\ &= \frac{1}{N}\left[H(\Gamma_A|\Omega) - H(\Gamma_A|\Gamma_B)\right] && (169)\\ &\geq \frac{1}{N}\left[H(\Gamma|\Omega) - N\gamma_K(\varepsilon)\right] && (170)\\ &\geq R_\Delta^- - \delta_K(\varepsilon) - \gamma_K(\varepsilon) && (171) \end{aligned}$$

where (170) follows Fano's inequality and (171) follows from (136). Since we can make ε arbitrarily small and K sufficiently large, the achievability of the proposed rate follows.

APPENDIX D
PROOF OF (149)

To establish (149) consider

$$p(\bar{\mathbf{y}}_A^K \mid \bar{\mathbf{z}}_{BE}^K, \mathsf{g}_{BE}^K, \mathsf{h}_{BA}^K, \bar{\mathbf{y}}_B^K, \mathsf{g}_{AE}^K, \mathsf{h}_{AB}^K) = \sum_{\bar{\mathbf{x}}_B^K} p(\bar{\mathbf{y}}_A^K, \bar{\mathbf{x}}_B^K | \bar{\mathbf{z}}_{BE}^K, \mathsf{g}_{BE}^K, \mathsf{h}_{BA}^K, \bar{\mathbf{y}}_B^K, \mathsf{g}_{AE}^K, \mathsf{h}_{AB}^K) \tag{172}$$

$$= \sum_{\bar{\mathbf{x}}_B^K} p(\bar{\mathbf{y}}_A^K | \bar{\mathbf{z}}_{BE}^K, \mathsf{g}_{BE}^K, \mathsf{h}_{BA}^K, \bar{\mathbf{y}}_B^K, \mathsf{g}_{AE}^K, \mathsf{h}_{AB}^K, \bar{\mathbf{x}}_B^K) p(\bar{\mathbf{x}}_B^K | \bar{\mathbf{z}}_{BE}^K, \mathsf{g}_{BE}^K, \mathsf{h}_{BA}^K, \bar{\mathbf{y}}_B^K, \mathsf{g}_{AE}^K, \mathsf{h}_{AB}^K) \tag{173}$$

$$= \sum_{\bar{\mathbf{x}}_B^K} p(\bar{\mathbf{y}}_A^K | \mathsf{h}_{BA}^K, \bar{\mathbf{x}}_B^K) p(\bar{\mathbf{x}}_B^K | \bar{\mathbf{z}}_{BE}^K, \mathsf{g}_{BE}^K) \tag{174}$$

$$= \sum_{\bar{\mathbf{x}}_B^K} p(\bar{\mathbf{y}}_A^K | \mathsf{h}_{BA}^K, \bar{\mathbf{x}}_B^K, \ \bar{\mathbf{z}}_{BE}^K, \mathsf{g}_{BE}^K) p(\bar{\mathbf{x}}_B^K | \bar{\mathbf{z}}_{BE}^K, \mathsf{g}_{BE}^K, \mathsf{h}_{BA}^K) \tag{175}$$

$$= \sum_{\bar{\mathbf{x}}_B^K} p(\bar{\mathbf{y}}_A^K, \bar{\mathbf{x}}_B^K | \mathsf{h}_{BA}^K, \ \bar{\mathbf{z}}_{BE}^K, \mathsf{g}_{BE}^K) = p(\bar{\mathbf{y}}_A^K | \mathsf{h}_{BA}^K, \ \bar{\mathbf{z}}_{BE}^K, \mathsf{g}_{BE}^K). \tag{176}$$

where (174) follows from the fact that $\bar{\mathbf{x}}_A^K$ is sampled independently of $\bar{\mathbf{x}}_B^K$ and the channel estimates, and using the structure of the channel (3) we also have that

$$\bar{\mathbf{y}}_A^K \leftrightarrow (\bar{\mathbf{x}}_B^K, \mathsf{h}_{BA}^K) \leftrightarrow (\ \bar{\mathbf{z}}_{BE}^K, \mathsf{g}_{BE}^K, \bar{\mathbf{y}}_B^K, \mathsf{g}_{AE}^K, \mathsf{h}_{AB}^K) \tag{177}$$

This establishes (149). Eq (150) can be established in an analogous fashion and its proof will be omitted.

REFERENCES

[1] R. Wilson, D. Tse, and R. A. Scholtz, "Channel identification: Secret sharing using reciprocity in ultrawideband channels," *IEEE Trans. Inform. Forensics and Security*, vol. 3, no. 2, pp. 364–375, 2007.

[2] S. Mathur, A. Reznik, C. Ye, R. Mukherjee, A. Rahman, Y. Shah, W. Trappe, and N. B. Mandayam, "Exploiting the physical layer for enhanced security," *IEEE Wireless Commun.*, vol. 17, no. 5, pp. 63–70, 2010.

[3] J. Croft, "Shared secret key establishment using wireless channel measurements," Ph.D. dissertation, University of Utah, 2011.

[4] Y. Wei, K. Zeng, and P. Mohapatra, "Adaptive wireless channel probing for shared key generation," in *INFOCOM*, 2011, pp. 2165–2173.

[5] M. Zafer, D. Agrawal, and M. Srivatsa, "Limitations of generating a secret key using wireless fading under active adversary," *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, pp. 1440–1451, 2012.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[7] I. Safaka, C. Fragouli, K. J. Argyraki, and S. N. Diggavi, "Creating shared secrets out of thin air," in *HotNets*, 2012, pp. 73–78.

[8] L. Czap, V. Prabhakaran, C. Fragouli, and S. Diggavi, "On interactive message secrecy over erasure networks," in *International Symposium on Communications, Control, and Signal Processing (ISCCSP)*, 2012.

[9] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inform. Theory*, vol. 39, pp. 733–742, Mar. 1993.

[10] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography – Part I: Secret sharing," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1121–1132, July 1993.

[11] I. Csiszar and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, 2004.

[12] A. Khisti, S. N. Diggavi, and G. W. Wornell, "Secret-key generation using correlated sources and channels," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 652–670, 2012.

[13] V. M. Prabhakaran, K. Eswaran, and K. Ramchandran, "Secrecy via sources and channels," *IEEE Trans. Inform. Theory*, vol. 58, no. 11, pp. 6747–6765, 2012.

[14] S. Nitinawarat and P. Narayan, "Secret Key Generation for Correlated Gaussian Sources," *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3373–3391, 2012.

[15] S. Watanabe and Y. Oohama, "Secret Key Agreement from Correlated Gaussian Sources by Rate Limited Public Communication," *IEICE Transactions*, vol. 93-A, no. 11, pp. 1976–1983, 2010.

[16] A. Khisti, S. N. Diggavi, and G. W. Wornell, "Secret-key agreement with channel state information at the transmitter," *IEEE Trans. Inform. Forensics and Security*, vol. 6, no. 3, 2011.

[17] M. El-Halabi, T. Liu, C. N. Georghiades, and S. Shamai, "Secret writing on dirty paper: A deterministic view," *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3419–3429, 2012.

[18] T. F. Wong, M. Bloch, and J. M. Shea, "Secret sharing over fast-fading MIMO wiretap channels," *Special Issue on Wireless Physical Layer Security, EURASIP J. Wireless Commun. Netw.*, 2009.

[19] A. Khisti and S. N. Diggavi, "A remark on secret-key generation over correlated fading channels," in *Globecom, Workshop on Physical Layer Security*, Houston, TX, 2011.

[20] A. Agrawal, Z. Rezki, A. Khisti, and M. S. Alouini, "Noncoherent secret-key agreement with public discussion," *IEEE Trans. Inform. Forensics and Security*, vol. 6, no. 3, pp. 565–574, 2011.

[21] L. Lai, Y. Liang, and H. V. Poor, "A unified framework for key agreement over wireless fading channels," *IEEE Trans. Inform. Forensics and Security*, vol. 7, no. 2, pp. 480–490, Apr. 2012.

[22] O. Güngör, J. Tan, C. E. Koksal, H. E. Gamal, and N. B. Shroff, "Joint power and secret key queue management for delay limited secure communication," in *INFOCOM*, 2010, pp. 677–685.

[23] ——, "Secrecy outage capacity of fading channels," in *CISS*, 2012, pp. 1–6.

[24] Y. Liu, S. C. Draper, and A. M. Sayeed, "Exploiting channel diversity in secret key generation from multipath fading randomness," *IEEE Trans. Inform. Forensics and Security*, vol. 7, no. 5, pp. 1484–1497, 2012.

[25] T.-H. Chou, S. C. Draper, and A. M. Sayeed, "Key generation using external source excitation: Capacity, reliability, and secrecy exponent," *IEEE Trans. Inform. Theory*, vol. 58, no. 4, pp. 2455–2474, 2012.

[26] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, 2003.

[27] L. Zheng and D. N. C. Tse, “Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel,” *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[28] S. Borade and L. Zheng, “Wideband fading channels with feedback,” *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6058–6065, 2010.

[29] F. Teng, D. Guo, and M. L. Honig, “Bidirectional channel estimation using adaptive pilots,” in *ISIT*, 2012, pp. 2381–2385.

[30] U. Salim, D. Gesbert, and D. T. M. Slock, “Combining training and quantized feedback in multiantenna reciprocal channels,” *IEEE Trans. Signal Processing*, vol. 60, no. 3, pp. 1383–1396, 2012.

[31] S. Haile, “Investigation of channel reciprocity for OFDM TDD systems,” Master’s thesis, U. Waterloo, 2009.

[32] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, “Relative channel reciprocity calibration in MIMO/TDD systems,” in *Future Netw. Mobile Summit*, 2010.

[33] A. Khisti, “Interactive secret-key agreement over fading channels,” in *Proc. Allerton Conf. Commun., Contr., Computing*, Monticello, IL, Oct. 2012.

[34] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.

[35] A. Khisti and G. W. Wornell, “Secure transmission with multiple antennas I: The MISOME wiretap channel,” *IEEE Trans. Inform. Theory*, vol. 56, 2009.

[36] A. E. Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, UK: Cambridge University Press, 2012.

[37] M. Andersson, A. Khisti, and M. Skoglund, “Secure key agreement over reciprocal fading channels in the low snr regime,” in *SPAWC*, Darmstaad, Germany, June 2013.