MIMO Wiretap Channels with Unknown and Varying Eavesdropper Channel States

Xiang He, and Aylin Yener, *Senior Member, IEEE*

***Abstract*—In this work, a class of information theoretic secrecy problems is addressed where the eavesdropper channel state is *completely unknown* to the legitimate parties. In particular, a Gaussian MIMO wiretap channel is considered where the eavesdropper channel state can vary from one channel use to the next, and the overall channel state sequence is known only to the eavesdropper. When the eavesdropper has fewer antennas than the transmitter and its intended receiver, a positive secrecy rate in the sense of *strong secrecy* is proved to be achievable and shown to match with the converse in secure degrees of freedom. This yields the conclusion that secure communication is possible regardless of the location or the channel states of the eavesdropper. Additionally, it is observed that, the present setting renders the secrecy capacity problems for some multi-terminal wiretap-type channels more tractable as compared to the case with full or partial knowledge of eavesdropper channel states. To demonstrate this observation, secure degrees of freedom regions are derived for the Gaussian MIMO multiple access wiretap channel (MIMO MAC-WT) and the two-user Gaussian MIMO broadcast wiretap channel (MIMO BC-WT) where the transmitter(s) and the intended receiver(s) have the same number of antennas.**

***Index Terms*—Information theoretic secrecy, MIMO wiretap channel, MIMO MAC wiretap channel, MIMO BC wiretap channel, strong secrecy, eavesdroppers with unknown and varying channel gains.**

I. Introduction

Information theoretic secrecy dates back to the seminal work by Shannon [1], where it was shown that if the eavesdropper had perfect knowledge of the signals sent by the transmitter and had unbounded computational power, for perfect secrecy, the transmitter and the receiver would have to share a key whose rate equals that of the data.

Wyner, in [2], found that Shannon's result was overly pessimistic, and showed that for the wiretap channel, where the eavesdropper had a noisy observation of the signals sent by the transmitter, a positive rate could be supported for transmitting confidential messages without requiring the communicating parties to share a key. The model was generalized by Csiszár and Körner in [3].

The wiretap channel model in [2]–[4] has inspired considerable effort toward identifying secure communication limits of various channel models, e.g. [5]–[16]. In these works, it is assumed that the transmitter(s) has (have) perfect knowledge of the eavesdropper channel states, which may be difficult to obtain in a practical system, since the eavesdropper is by nature a passive entity. To resolve this issue, recent works attempt to relax this condition by assuming the transmitter only has *partial* knowledge about the channel states of the eavesdropper. Notably, this line of work includes the compound setting, where the eavesdropper channel can only be taken from a *finite* selection [17]–[20], and the fading channel, where the transmitter only knows the distribution of the eavesdropper channel [21]. These each call for different types of codebook design. For example, in [17]–[19], the coding scheme depends on the possible channel gains of the eavesdropper included in the finite set. For the fading setting [21], the duration of communication needs to be able to accommodate a sufficient number of channel uses to ensure that the ergodicity assumption is valid. In addition, the rate of the codebook depends on the fading parameter of the eavesdropper, e.g., the variance of the Rayleigh distribution, and thus needs to be acquired, which may be difficult to do with a passive but malicious entity. Given the absence of a robustness analysis toward understanding how sensitive the achievable secrecy rate is to errors in the aforementioned modeling parameters in [17]–[19], [21], it is difficult to ascertain how close these can model a realistic secure system design based on information theoretic guarantees.

The case where the eavesdropper's location is not perfectly known was also considered in the context of network coding [22], [23]. In [22], the eavesdropper is assumed to monitor no more than K edges in a network, while the locations of these edges can be arbitrary. The code designer uses the fact that there are more than K routes connecting the sender and the receiver of the confidential message, while the eavesdropper cannot monitor all routes[1]. The simple, yet powerful insight offered by references [22], [23] on the merit of utilizing the advantage enjoyed by the legitimate nodes via multiple routes, can be brought into the wireless setting by utilizing multiple antennas. Specifically, if the intended receiver has more antennas than the eavesdropper, then even though the eavesdropper can be anywhere, i.e., experience any channel state, it cannot monitor all antennas of the receiver.

Manuscript received July 20, 2010; revised May 31, 2013 and August 10, 2014; accepted September 4, 2014. This work was presented in part at the 2009 IEEE Global Telecommunications conference and the 2010 Allerton conference on Communication, Control and Computing. This work was supported in part by the National Science Foundation via Grants CNS-0716325, CCF-0964362, CCF-1319338 and the DARPA ITMANET Program via Grant W911NF-07-1-0028. Copyright ©2014 IEEE.

Xiang He was with the Department of Electrical Engineering at the Pennsylvania State University, University Park, PA 16802. He is now with Microsoft (email: xianghe@microsoft.com).

Aylin Yener is with the Department of Electrical Engineering at the Pennsylvania State University, University Park, PA 16802 (email: yener@ee.psu.edu).

[1] Results of similar spirit can be traced back to [24], where the eavesdropper has access to a K transmitted bits, but the bits it has access to are not specified.

Inspired by this observation, in this work, we study the MIMO wiretap channel, where the eavesdropper has fewer antennas than the transmitter and its intended receiver. The channel state of the eavesdropper can take any value at each channel use and vary from one channel use to the next. We assume the sequence constituted by these channel states are perfectly known by the eavesdropper but is completely unknown to the legitimate parties. Conditioned on any given channel state sequence, we assume that the eavesdropper channel is memoryless.

The main contribution of this work is to prove the existence of a *universal* coding scheme that secures the confidential message against any sequence of eavesdropper channel states for the MIMO wiretap setting described above. The universal nature of the coding scheme is what sets this work apart from the previous work that considered a time-varying eavesdropper channel [25]. Additionally, unlike [26] and [27] which considered the discrete arbitrarily varying wiretap channel, this work considers a Gaussian setting which does not lend itself to a direct extension from its discrete counterpart.

The achievable rates we prove in this work satisfy *strong* secrecy requirements [28], which may be better suited for practice [28]–[30] as compared to weak secrecy that is more frequently considered for secrecy capacity analysis in information theory [2]–[16], [25]. It is often argued that strong secrecy can be obtained from weak secrecy through privacy amplification, as shown in [28]. In the setting considered in this paper, however, how to use privacy amplification is still an open problem. Therefore, a direct proof of strong secrecy is provided[2].

The achieved rate derived in this work is shown to be *tight* in terms of secure degrees of freedom (s.d.o.f.), which is a high signal-to-noise[3] (SNR) characteristic of the secrecy capacity. For low and moderate SNR regimes, we provide the achievable rates, but the secrecy capacity characterization remains open.

We also extend our results to a MIMO MAC wiretap channel (MIMO MAC-WT) and a MIMO Broadcast wiretap channel (MIMO BC-WT), for the case with two users where legitimate transmitter(s) and receiver(s) have the same number of antennas, and identify their secure degrees of freedom region.

The remainder of the paper is organized as follows. The system models are introduced in Section II. In Section III, we state the main results, which are proved in Section IV. Section V presents a discussion on strong secrecy as well as a detailed comparison to related work. Section VI concludes the paper. Appendices A-G contain the various necessary proofs in support of the main achievability proof. Appendix H describes a simple upper bound on secrecy rate used for numerical comparison with the achievable rate. Appendix I provides the weak secrecy proof for completeness and consistency with the previous literature.

II. System Models

A. The (N_T, N_R, N_E) MIMO Wiretap Channel

The channel from the transmitter to the intended receiver, i.e., *the main channel*, is assumed to be static. Let $A(i)$ denote the value of the signal A during the ith channel use. The input and output of the main channel during the ith channel use are related as:

$$\mathbf{Y}_{N_R\times 1}(i) = \mathbf{H}_{N_R\times N_T}\mathbf{X}_{N_T\times 1}(i) + \mathbf{Z}_{N_R\times 1}(i) \tag{1}$$

where the subscripts denote the dimension of each term. $\mathbf{H}$ denotes the $N_R \times N_T$ channel matrix with complex entries[4]. It is assumed that $\mathbf{H}$ is perfectly known by the transmitter, the intended receiver and the eavesdropper, and $\mathbf{H}$ has full rank. $\mathbf{Z}$ is a $N_R \times 1$ vector representing the additive noise. $\mathbf{Z}$ is composed of independent rotationally invariant complex Gaussian random variables, each with zero mean and unit variance[5]. $\mathbf{X}$ and $\mathbf{Y}$ are the transmitted and received signals respectively.

The channel from the transmitter to the eavesdropper, i.e., *the eavesdropper channel*, varies from one channel use to the next. It can be expressed as:

$$\tilde{\mathbf{Y}}_{N_E\times 1}(i) = \tilde{\mathbf{H}}_{N_E\times N_T}(i)\mathbf{X}_{N_T\times 1}(i) \tag{2}$$

where $\tilde{\mathbf{Y}}(i)$ denotes the signals received by the eavesdropper during the ith channel use. $\tilde{\mathbf{H}}_{N_E\times N_T}(i)$ is the channel state matrix for the eavesdropper channel during the ith channel use. We use $\tilde{\mathbf{H}}^n$ to denote $\tilde{\mathbf{H}}(1), ..., \tilde{\mathbf{H}}(n)$. $\tilde{\mathbf{H}}^n$ is any arbitrary sequence. However, we stress that the channel states by the eavesdropper are not chosen in an adversarial manner adapting to the transmitted signals in previous channel uses. $\tilde{\mathbf{H}}^n$ is *not* known at the legitimate parties and is perfectly known by the eavesdropper.

Note that we assume the eavesdropper's channel is noiseless. This is obviously a worst case assumption, and if the eavesdropper's signals are corrupted by additive noise, they can always be considered as a degraded version of the signals received by the eavesdropper considered in this work.

Let W denote the confidential message transmitted to the intended receiver, over n channel uses using $\mathbf{X}^n$. In addition, we assume that there is a local random source F which is only known to the transmitter. $\mathbf{X}^n$ is computed by the transmitter from W, F and $\mathbf{H}$ using the following encoding function f_n:

$$\mathbf{X}^n = \mathrm{f}_n(W, F, \mathbf{H}). \tag{3}$$

Note that f_n does not depend on $\tilde{\mathbf{H}}^n$, since the transmitter does not know the channel state of the eavesdropper.

We represent $\mathbf{X}^n$ as a $N_T \times n$ matrix. The transmitter is constrained in terms of average transmission power:[6]

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{Tr}(\mathbf{X}^n(\mathbf{X}^n)^H) \le \bar{P}. \tag{4}$$

The decoder is defined as

$$\hat{W} = \psi_n(\mathbf{Y}^n, \mathbf{H}) \tag{5}$$

[2] As the proof techniques for weak and strong secrecy differ considerably, and for the sake of comprehensiveness and consistency with the recent information theoretic secrecy literature, we also provide the proof for weak secrecy in an Appendix.

[3] for the legitimate receiver and the eavesdropper

[4] Since we assume that the main channel is static, $\mathbf{H}$ remains fixed for all channel uses.

[5] The assumption on the equal noise variances is without loss of generality.

[6] Tr denotes the sum of the diagonal elements of a square matrix.

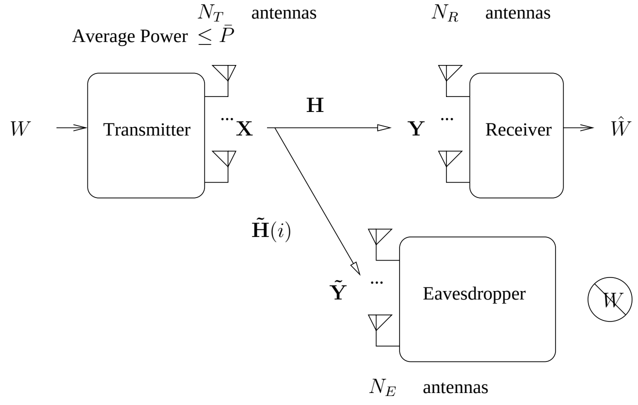

Fig. 1. The MIMO Wiretap Channel.

where $\hat{W}$ denotes the decoder output of the intended receiver from $\mathbf{Y}^n$. Then we require the average probability of decoding error to vanish:

$$\lim_{n\to\infty} \Pr(W \neq \hat{W}) = 0. \tag{6}$$

The message W must also be kept secret from the eavesdropper regardless of the channel state sequence it observes. This is represented by the following *strong secrecy* constraint:

$$\lim_{n\to\infty} \sup_{\tilde{\mathbf{h}}^n} I(W; \tilde{\mathbf{Y}}^n | \tilde{\mathbf{H}}^n = \tilde{\mathbf{h}}^n) = 0. \tag{7}$$

For the MIMO wiretap channel, the secrecy rate R_s is defined as

$$R_s = \lim_{n\to\infty} \frac{1}{n} H(W). \tag{8}$$

The rate R_s is said to be achievable if for each n there exists a *fixed* encoding function f_n and decoding function ψ_n as defined by (3) and (5), such that (4), (6), (7) and (8) are satisfied. The supremum of all possible values for R_s is called the *secrecy capacity* of this channel model.

The high SNR behavior of the secrecy rate is characterized by the secure degrees of freedom defined as:

$$\text{s.d.o.f.} = \limsup_{\bar{P}\to\infty} \frac{R_s(\bar{P})}{\log_2\left(\bar{P}\right)} \tag{9}$$

where we write R_s as $R_s(\bar{P})$ to emphasize its dependence on $\bar{P}$.

We use the term *the secure degrees of freedom* of a channel to represent the largest possible value of (9).

Remark 1: While (7) represents the strong secrecy constraint, the following *weak secrecy constraint* has been more frequently used for secrecy capacity analysis, e.g., [2]–[16], [25].

$$\lim_{n\to\infty} \sup_{\tilde{\mathbf{h}}^n} \frac{1}{n} I(W; \tilde{\mathbf{Y}}^n | \tilde{\mathbf{H}}^n = \tilde{\mathbf{h}}^n) = 0. \tag{10}$$

The achievability proof in our setting for this weaker secrecy notion is relatively simpler and is provided in Appendix I for completeness and for reference for the case where the eavesdropper channel does not change over time. □

Remark 2: Another way to model the eavesdropper channel is to define the distribution of $\tilde{\mathbf{H}}^n$ for each n, see [21], [30] for example. The secrecy constraint in (7) implies that the message is secure for this setting if the distribution for $\tilde{\mathbf{H}}^n$ is defined to be independent from the channel inputs $\mathbf{X}^n$. To show this, first note that the secrecy constraint in this case is given by:

$$\lim_{n\to\infty} I(W; \tilde{\mathbf{Y}}^n, \tilde{\mathbf{H}}^n) = 0. \tag{11}$$

Since $\tilde{\mathbf{H}}^n$ is independent from the channel inputs $\mathbf{X}^n$, (11) can be written as:

$$\lim_{n\to\infty} I(W; \tilde{\mathbf{Y}}^n | \tilde{\mathbf{H}}^n) = 0 \tag{12}$$

which is implied by (7).

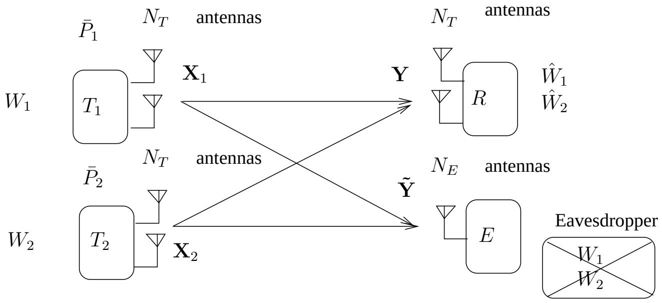

Fig. 2. The (N_T, N_T, N_T, N_E) MIMO MAC wiretap channel where legitimate nodes have $N_T = 2$ antennas each, and the eavesdropper has $N_E = 1$ antenna.

B. The two-user MIMO MAC Wiretap Channel and MIMO Broadcast Wiretap Channel

The single user MIMO Wiretap channel defined earlier can be extended to the two-user MAC channel and broadcast channel. In this section, we briefly discuss these two extensions when each legitimate transmitter and each intended receiver has N_T antennas, and the eavesdropper has N_E antennas.

These two channel models are shown in Figure 2 and Figure 3 respectively, where $(N_T,\ N_T,\ N_T,\ N_E)$ represents the antenna number configuration. During the ith channel use, the (N_T, N_T, N_T, N_E) MAC channel is defined as:

$$\mathbf{Y}(i) = \mathbf{H}_1\mathbf{X}_1(i) + \mathbf{H}_2\mathbf{X}_2(i) + \mathbf{Z}(i), \tag{13}$$

$$\tilde{\mathbf{Y}}(i) = \tilde{\mathbf{H}}_1(i)\mathbf{X}_1(i) + \tilde{\mathbf{H}}_2(i)\mathbf{X}_2(i) \tag{14}$$

and the (N_T, N_T, N_T, N_E) broadcast channel is given by

$$\mathbf{Y}_k(i) = \mathbf{H}_k\mathbf{X}(i) + \mathbf{Z}_k(i), \quad k = 1, 2, \tag{15}$$

$$\tilde{\mathbf{Y}}(i) = \tilde{\mathbf{H}}(i)\mathbf{X}(i) \tag{16}$$

where $\mathbf{H}_k, k = 1, 2$, $\tilde{\mathbf{H}}_k(i), k = 1, 2$ and $\tilde{\mathbf{H}}(i)$ are the channel matrices. $\mathbf{Z}$ and $\mathbf{Z}_k, k = 1, 2$ are the additive Gaussian noise observed by the intended receivers, which has the same distribution as $\mathbf{Z}$ in (1).

We assume the main channels $\mathbf{H}_k, k = 1, 2$ are known by both the legitimate parties and the eavesdropper. The eavesdropper channel state sequence, $\tilde{\mathbf{H}}_k^n, k = 1, 2$ for the MAC channel and $\tilde{\mathbf{H}}^n$ for the broadcast channel, is unknown to the legitimate parties but known by the eavesdropper.

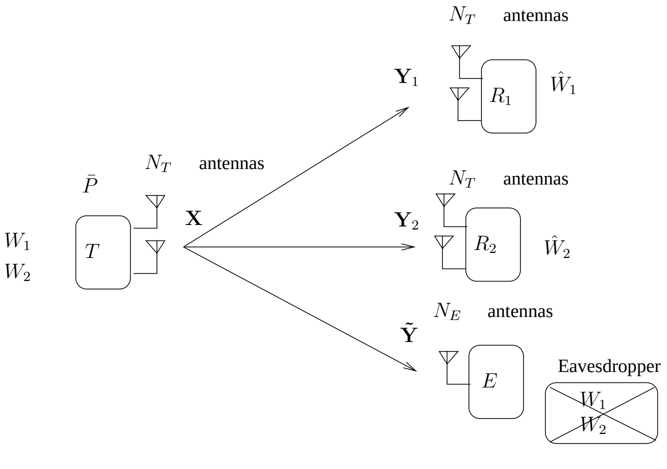

Fig. 3. The two-user (N_T, N_T, N_T, N_E) MIMO BC Wiretap Channel where legitimate nodes have $N_T = 2$ antennas each, and the eavesdropper has $N_E = 1$ antenna.

The confidential messages are denoted by $W_k, k = 1, 2$. For the MAC channel, the secrecy constraint for these messages is given by:

$$\lim_{n\to\infty} \sup_{\tilde{\mathbf{h}}_k^n, k=1,2} I\left(W_1, W_2; \tilde{\mathbf{Y}}^n | \tilde{\mathbf{H}}_k^n = \tilde{\mathbf{h}}_k^n, k = 1, 2\right) = 0. \tag{17}$$

For the broadcast channel, the secrecy constraint is

$$\lim_{n\to\infty} \sup_{\tilde{\mathbf{h}}_k^n, k=1,2} I\left(W_1, W_2; \tilde{\mathbf{Y}}^n | \tilde{\mathbf{H}}_k^n = \tilde{\mathbf{h}}_k^n, k = 1, 2\right) = 0. \tag{18}$$

For the MAC channel, the average power constraints of the two transmitters are given by

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{Tr}(\mathbf{X}_k^n(\mathbf{X}_k^n)^H) \le \bar{P}_k, k = 1, 2. \tag{19}$$

For the broadcast channel, the average power constraint of the transmitter is given by

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{Tr}(\mathbf{X}^n(\mathbf{X}^n)^H) \le \bar{P}. \tag{20}$$

The secrecy rate for W_k, $R_{s,k}$, is defined as

$$R_{s,k} = \lim_{n\to\infty} \frac{1}{n} H(W_k), k = 1, 2 \tag{21}$$

such that the average power constraint and the secrecy constraint(s) are satisfied and the average probability of decoding error at the intended receiver(s) $\to 0$ when $n \to \infty$.

The secure degrees of freedom region for these models is defined as[7]:

$$\left\{(d_1, d_2) : d_k = \limsup_{\bar{P}\to\infty} \frac{R_{s,k}}{\log_2 \bar{P}}, k = 1, 2\right\}. \tag{22}$$

III. Main Results

Let $[x]^+$ equal to x if $x \ge 0$ and 0 if $x < 0$. Define $C(x)$ as $\log_2(1 + x)$. Define $N_{T,R}$ as

$$N_{T,R} = \min\{N_T, N_R\}. \tag{23}$$

Let s_i, $1 \le i \le N_{T,R}$, be the $N_{T,R}$ singular values of $\mathbf{H}$. For a positive constant σ^2, define P as

$$P = \max\{\bar{P} - N_{T,R}\sigma^2, 0\}. \tag{24}$$

As we shall show later, the variable $N_{T,R}\sigma^2$ represents the total power of the artificial noise [25] injected by the transmitter. The variable P represents the remaining power available to the transmitter.

Equipped with the notation defined above, the first result of this work is given by the following proposition.

Proposition 1: Any secrecy rate R_s that satisfies

$$0 \le R_s < \begin{cases} \sup_{\sigma^2>0} \left[\begin{array}{c} \sum\limits_{i=1}^{N_{T,R}} C\left(\frac{s_i^2 P}{(s_i^2\sigma^2+1)N_{T,R}}\right) \\ -N_E C\left(\frac{P}{N_{T,R}\sigma^2}\right) \end{array} \right]^+, & N_{T,R} > N_E \\ 0, & N_{T,R} \le N_E \end{cases} \tag{25}$$

is achievable for the MIMO-wiretap channel described in Section II-A.

Proposition 1 is proved in Section IV.

In Section IV-F, we show that the achievable secrecy rate given by Proposition 1 matches the converse in terms of secure degrees of freedom. Hence we have the following theorem.

Theorem 1: If $\mathbf{H}$ has rank $N_{T,R}$, then the s.d.o.f. of the MIMO-wiretap channel described in Section II-A is

$$\max\{N_{T,R} - N_E, 0\}. \tag{26}$$

The achievable secrecy rate given by Proposition 1 can be easily extended to the MIMO (N_T, N_T, N_T, N_E) MAC wiretap channel and the MIMO (N_T, N_T, N_T, N_E) BC wiretap channel using time sharing. Let α be the time sharing parameter, and $\bar{\alpha} = 1 - \alpha$. For a positive constant σ_k, Define $P_{k,\alpha}, k = 1, 2$ as:

$$P_{k,\alpha} = \left[\frac{\bar{P}_k}{\alpha} - N_T\sigma_k^2\right]^+, \quad k = 1, 2. \tag{27}$$

Let "co" denote the convex hull operation, and $s_{k,i}, k = 1, 2$ denote the N_T singular values of $\mathbf{H}_k$. Then, for the MIMO (N_T, N_T, N_T, N_E) MAC wiretap channel, the achievable secrecy rate region is given by:

$$\mathrm{co} \bigcup_{\alpha, \sigma_k^2, k=1,2} \left\{ \begin{array}{l} (R_{s,1}, R_{s,2}) : \\ 0 \le R_{s,1} < \\ \alpha \left[\sum\limits_{i=1}^{N_T} C\left(\frac{s_{1,i}^2 P_{1,\alpha}}{(s_{1,i}^2\sigma_1^2+1)N_T}\right) - N_E C\left(\frac{P_{1,\alpha}}{N_T\sigma_1^2}\right)\right]^+, \\ 0 \le R_{s,2} < \\ \bar{\alpha} \left[\sum\limits_{i=1}^{N_T} C\left(\frac{s_{2,i}^2 P_{2,\bar{\alpha}}}{(s_{2,i}^2\sigma_2^2+1)N_T}\right) - N_E C\left(\frac{P_{2,\bar{\alpha}}}{N_T\sigma_2^2}\right)\right]^+ \end{array} \right\}. \tag{28}$$

[7] For the MAC channel, we assume $\bar{P}_k = \bar{P}, k = 1, 2$.

For the MIMO (N_T, N_T, N_T, N_E) BC wiretap channel, for $P = [\bar{P} - N_T\sigma^2]^+$, the achievable secrecy rate region is given by:

$$\text{co}\left\{\begin{array}{l} (R_{s,1}, R_{s,2}) : \qquad (0,0)\,, \\ \left(\sup_{\sigma^2>0}\left[\begin{array}{c}\sum_{i=1}^{N_T} C\left(\frac{s_{1,i}^2 P}{(s_{1,i}^2\sigma^2+1)N_T}\right) \\ -N_E C\left(\frac{P}{N_T\sigma^2}\right)\end{array}\right]^+, 0\right), \\ \left(0, \sup_{\sigma^2>0}\left[\begin{array}{c}\sum_{i=1}^{N_T} C\left(\frac{s_{2,i}^2 P}{(s_{2,i}^2\sigma^2+1)N_T}\right) \\ -N_E C\left(\frac{P}{N_T\sigma^2}\right)\end{array}\right]^+\right)\end{array}\right\}. \tag{29}$$

We shall show next that these achievable regions match their converse in terms of secure degrees of freedom region. Hence we have the following theorem.

Theorem 2: If $\mathbf{H}_k, k = 1, 2$ has full rank, the secure degrees of freedom region of the two-user MIMO MAC wiretap channel in Figure 2 and the two-user MIMO BC wiretap channel in Figure 3 are both given by

$$d_1 + d_2 \leq \max\{N_T - N_E, 0\} \tag{30}$$

$$d_i \geq 0, \quad i = 1, 2. \tag{31}$$

Proof: For the MIMO MAC-WT, the secrecy rate region in (28) is achieved by time sharing, i.e., letting user 1 transmit during α fraction of the channel uses alone, and user 2 for the remaining channel uses. The region then readily follows from Proposition 1. The achievability of the s.d.o.f. region follows from the achievable secrecy rate region (28), by letting $\bar{P}_1 = \bar{P}_2 = \bar{P} \to \infty$. The achievable secrecy rate region for the MIMO BC-WT is also obtained by time sharing. The secrecy rates achieved by receivers 1 and 2 are found to be (29) as a consequence of Proposition 1. The achieved s.d.o.f. region then follows from the achievable secrecy rate region (29), by letting $\bar{P}_1 = \bar{P}_2 = \bar{P} \to \infty$.

For the converse, we simply combine the two transmitters for the MIMO MAC-WT and combine the two receivers for the MIMO broadcast channel. The channel then becomes a single-user MIMO wiretap channel in each case. The inequality $d_1 + d_2 \leq \max\{N_T - N_E, 0\}$ then follows from the converse of Theorem 1. ■

Remark 3: When the eavesdropper's channel state is fixed and known by the transmitters, the s.d.o.f. region for the MIMO MAC wiretap channel is still an open problem [19], [31], [32]. When the eavesdropper's channel state can take more than one possible value from a finite set and the set is known by the transmitters, the s.d.o.f. region for the MIMO BC wiretap channel is also open. On the other hand if the eavesdropper's channel state sequence is arbitrary and all legitimate nodes have the same number of antennas, the s.d.o.f. capacity region of both problems are found in this paper. □

IV. The MIMO Wiretap Channel

In this section, we present our detailed results on the MIMO wiretap channel with strong secrecy. We first present the notation. Section IV-B provides the channel transformation and signaling scheme. Section IV-C presents the codebook construction. Section IV-D presents the achievability proof for the static eavesdropper channel, which is extended to the varying eavesdropper channel in Section IV-E. A diagram summarizing the steps leading to the achievability proof is provided in Figure 4. Finally, in Section IV-F, we present the converse to establish the secure degrees of freedom result in Theorem 1.

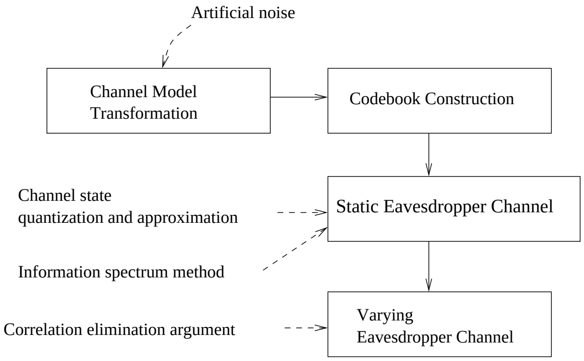
Artificial noise
Channel Model Transformation
Codebook Construction
Channel state quantization and approximation
Static Eavesdropper Channel
Information spectrum method
Correlation elimination argument
Varying Eavesdropper Channel

Fig. 4. Organization of the proof for Proposition 1.

A. Notation

We use $p_W(w)$ to denote the probability mass function (p.m.f.) of a random variable W evaluated at w. $f_{\gamma,A}(a)$ denotes the probability density function (p.d.f.) of a random variable A at value a with parameter γ. $f_{\gamma,A|B}(a|b)$ denotes the conditional p.d.f. of a random variable A conditioned on a random variable B when $A = a, B = b$ with parameter γ. For a vector x^n, we let $\|x^n\|$ denote its L_2-norm. For a matrix $\mathbf{A}$, we let $\|\mathbf{A}\|^2$ denote the sum of the L_2-norm squared of all the row vectors of $\mathbf{A}$. $\mathrm{E}_B[A]$ denotes the expectation of A averaged over B. We define

$$\mathbf{A}^n - \mathbf{B}^n\mathbf{C}^n \tag{32}$$

as the row concatenation of the matrices $\{\mathbf{A}(i) - \mathbf{B}(i)\mathbf{C}(i), 1 \leq i \leq n\}$ and

$$\mathbf{B}^n\mathbf{C}^n \tag{33}$$

as the row concatenation of the matrices $\{\mathbf{B}(i)\mathbf{C}(i), 1 \leq i \leq n\}$.

B. Channel Model Transformation

We first observe that we only need to consider the case where $N_T = N_R$ and that it is sufficient to consider the main channel matrix $\mathbf{H}$ to be diagonal without loss of generality. For a general channel matrix $\mathbf{H}$, we can always transform it into this form by (1) performing singular value decomposition (SVD) on it, (2) canceling the right and left unitary matrices of its SVD decomposition and (3) discarding channel inputs that cannot reach the receiver and channel outputs that only contain channel noise when designing the coding scheme.

We only consider $N_E < N_T$, since the achievable secrecy rate in Proposition 1 is zero otherwise. For this case, $\tilde{\mathbf{H}}(i)$ has the following form of SVD decomposition:

$$\tilde{\mathbf{H}}(i) = [\mathbf{I}_{N_E \times N_E}, \mathbf{0}_{N_E \times (N_T - N_E)}]\mathbf{U}(i) \tag{34}$$

where $\mathbf{U}(i)$ is a $N_T \times N_T$ unitary matrix. $\mathbf{I}$ is an identity matrix. This can be achieved by canceling the left unitary matrix of the SVD decomposition of $\tilde{\mathbf{H}}(i)$ and normalizing the singular values at the eavesdropper. Note that the transmitter does not know $\mathbf{U}(i)$.

Remark 4: Note that if $\tilde{\mathbf{H}}(i)(\mathbf{U}(i))^{-1}$ has all zero rows, we can alter appropriate entries to be 1s so that the resulting channel matrix has the form in (34). The signals received by the original eavesdropper is always degraded compared to the signals received by the eavesdropper after this modification. Hence, it is sufficient to consider the eavesdroppers with $\tilde{\mathbf{H}}(i)$ in the form given by (34). □

Since the eavesdropper channel is assumed to be noiseless, we have to introduce artificial noise [25] at the transmitter to limit the receiving capability of the eavesdropper. We express $\mathbf{X}(i)$ as

$$\mathbf{X}(i) = \bar{\mathbf{X}}(i) + \mathbf{N}(i) \tag{35}$$

where $\mathbf{N}$ is the $N_T \times 1$ artificial noise vector consisting of independent rotationally invariant complex Gaussian random variables with zero mean and variance σ^2. Coding is over $\bar{\mathbf{X}}$.

Define $\bar{\mathbf{N}}$ and $\tilde{\mathbf{N}}(i)$ as

$$\bar{\mathbf{N}}(i) = \mathbf{H}\mathbf{N}(i), \tag{36}$$

$$\tilde{\mathbf{N}}(i) = \tilde{\mathbf{H}}(i)\mathbf{N}(i). \tag{37}$$

Viewing $\bar{\mathbf{X}}$ as the input to the channel, the channel model can be expressed as:

$$\mathbf{Y}(i) = \mathbf{H}\bar{\mathbf{X}}(i) + \bar{\mathbf{N}}(i) + \mathbf{Z}(i), \tag{38}$$

$$\tilde{\mathbf{Y}}(i) = \tilde{\mathbf{H}}(i)\bar{\mathbf{X}}(i) + \tilde{\mathbf{N}}(i). \tag{39}$$

From (34) and (37), we observe that $\tilde{\mathbf{N}}$ has zero mean and is Gaussian distributed. The covariance matrix of $\tilde{\mathbf{N}}$ is

$$\begin{aligned} &\mathrm{E}[\tilde{\mathbf{H}}(i)\mathbf{N}(i)(\mathbf{N}(i))^H(\tilde{\mathbf{H}}(i))^H] \\ =&\tilde{\mathbf{H}}(i)\mathrm{E}[\mathbf{N}(i)(\mathbf{N}(i))^H](\tilde{\mathbf{H}}(i))^H & (40) \\ =&\sigma^2\tilde{\mathbf{H}}(i)(\tilde{\mathbf{H}}(i))^H & (41) \\ =&\sigma^2\mathbf{I}_{N_E \times N_E}. & (42) \end{aligned}$$

C. Codebook Construction

The *codebook ensemble* is denoted by $\{\mathcal{C}\}$. Each codebook $\mathcal{C}$ in the ensemble is constructed as follows:

Recall that P was defined in (24). Let $\bar{\mathbf{X}}_G$ denote a rotationally invariant zero mean complex Gaussian random variable with covariance matrix $(\frac{P(1-\bar{\epsilon})}{N_T})\mathbf{I}_{N_T \times N_T}$, where $\bar{\epsilon}$ is a constant such that $0 < \bar{\epsilon} < 1$. Let $Q_{\bar{\mathbf{X}}_G}(x)$ denote the probability density function of $\bar{\mathbf{X}}_G$.

Define the n-letter truncated Gaussian distribution $Q_{\bar{\mathbf{X}}_T^n}(x^n)$ as follows: Let x_i denote the ith component of x^n. $Q_{\bar{\mathbf{X}}_T^n}(x^n)$ is given by:

$$Q_{\bar{\mathbf{X}}_T^n}(x^n) = \mu_{n,\bar{\epsilon}}^{-1}\varphi(x^n)\prod_{i=1}^{n} Q_{\bar{\mathbf{X}}_G}(x_i) \tag{43}$$

where

$$\varphi(x^n) = \begin{cases} 1, & \text{if } \frac{1}{n}\|x^n\|^2 \le P \\ 0, & \text{otherwise} \end{cases} \tag{44}$$

$$\mu_{n,\bar{\epsilon}} = \int \varphi(x^n)\prod_{i=1}^{n} Q_{\bar{\mathbf{X}}_G}(x_i)dx^n. \tag{45}$$

Note that $0 < \mu_{n,\bar{\epsilon}} < 1$, and for a given $\bar{\epsilon}$, we have there exists an $\alpha(\bar{\epsilon}) > 0$, such that [33, (B2)]

$$1 - \mu_{n,\bar{\epsilon}} \le e^{-n\alpha(\bar{\epsilon})}, \tag{46}$$

$$\lim_{\bar{\epsilon}\to 0} \alpha(\bar{\epsilon}) = 0. \tag{47}$$

Let $\bar{\mathbf{X}}_G^n$ denote the length-n sequence sampled in an independent and identically distributed (i.i.d.) fashion from the input distribution $Q_{\bar{\mathbf{X}}_G}(x)$. Let $\bar{\mathbf{X}}_T^n$ denote the length-n sequence sampled in an i.i.d. fashion from the n-letter truncated Gaussian input distribution $Q_{\bar{\mathbf{X}}_T^n}$.[8]

The codebook $\mathcal{C}$ contains 2^{nR} sequences sampled from the distribution $Q_{\bar{\mathbf{X}}_T^n}$ in an i.i.d. fashion. R is chosen as

$$R = I(\bar{\mathbf{X}}_G; \mathbf{Y}_G) - \delta', \tag{48}$$

where $\mathbf{Y}_G$ denotes the outputs observed by the intended receiver when $\bar{\mathbf{X}}_G$ is used as inputs in (38). The variable δ' is a positive constant that can be arbitrarily small.

The codewords are then divided into N_B bins, each containing N_C codewords. This is done by labeling each sampled codeword with label (i, j), with $i \in \{1, ..., N_B\}$ and $j \in \{1, ..., N_C\}$, where i is the bin this codeword belongs to, and j is the index of the codeword in the bin. Let $\tilde{\mathbf{Y}}_G$ denote the output signal from the eavesdropper channel when its input is $\bar{\mathbf{X}}_G$. Then N_B and N_C are given by:

$$N_B = 2^{n(R - I(\bar{\mathbf{X}}_G; \tilde{\mathbf{Y}}_G) - \delta)}, \tag{49}$$

$$N_C = 2^{n(I(\bar{\mathbf{X}}_G; \tilde{\mathbf{Y}}_G) + \delta)} \tag{50}$$

where the variable δ is a positive constant that can be made arbitrarily small. Note that the value of the mutual information $I(\bar{\mathbf{X}}_G; \tilde{\mathbf{Y}}_G)$ does not depend on the value of $\tilde{\mathbf{H}}^n$ due to the fact that the eavesdropper channel state matrix can be transformed to the form given by (34) thanks to the eavesdropper channel being noiseless.

Let $x_{i,j}^n$ denote the codeword in the codebook $\mathcal{C}$ that is labeled with (i, j).

As in [34], for a given codebook $\mathcal{C}$, the intended receiver uses a maximum likelihood decoder: Upon receiving $\mathbf{Y}^n = y^n$, the decoder $\psi_{\mathcal{C}}(y^n)$ is given by

$$\psi_{\mathcal{C}}(y^n) = \arg \min_{i,j: x_{i,j}^n \in \mathcal{C}} \|y^n - \mathbf{H}^n x_{i,j}^n\|. \tag{51}$$

The probability of decoding error for each codeword, and the average probability of decoding error for each codebook and the codebook ensembles are defined as:

$$\lambda_{\mathcal{C},i,j} = \Pr\left(\psi_{\mathcal{C}}(\mathbf{Y}^n) \ne (i,j) | \bar{\mathbf{X}}^n = x_{i,j}^n\right), \tag{52}$$

$$\lambda_{\mathcal{C}} = \frac{1}{N_B N_C}\sum_{i,j} \lambda_{\mathcal{C},i,j}, \tag{53}$$

[8] This input distribution was also used in [34, Section 7.3].

$$\lambda = \mathrm{E}_{\mathcal{C}}\left[\lambda_{\mathcal{C}}\right]. \tag{54}$$

We next present the achievability proof for strong secrecy rate when the eavesdropper's channel is static, i.e., $\tilde{\mathbf{H}}(i) = \tilde{\mathbf{H}}, \quad 1 \leq i \leq n$.

D. Unknown Static Eavesdropper Channel

For a given codebook $\mathcal{C}$, the encoder $\mathrm{f}_{n,\mathcal{C}}$ used by the transmitter is described as follows: Let $\{i\}$ denote the set of possible values of the confidential message. We assume the confidential message W is uniformly distributed over $\{i\}$. Given $W = i$, $\mathrm{f}_{n,\mathcal{C}}$ selects a codeword from all the codewords with label i in codebook $\mathcal{C}$ according to a uniform distribution. With this encoder, we observe that (i,j) has a uniform distribution.

Let $\tilde{\mathbf{Y}}_G^n, \tilde{\mathbf{Y}}_T^n, \tilde{\mathbf{Y}}_{\mathcal{C}}^n$ denote the outputs of the eavesdropper channel when its input $\bar{\mathbf{X}}^n$ is $\bar{\mathbf{X}}_G^n$, $\bar{\mathbf{X}}_T^n$ or uniformly distributed over the codebook $\mathcal{C}$ respectively.

Let $\tilde{\mathbf{h}}^n$ denote the eavesdropper channel state sequence over n channel uses. For the static eavesdropper channel, $\tilde{\mathbf{h}}^n$ is composed of n copies of an $N_E \times N_T$ matrix denoted by $\tilde{\mathbf{h}}$. We use $d_{\tilde{\mathbf{h}}^n,\mathcal{C}}$ to denote the variational distance between two distribution $p_W f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n}$ and $p_W f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}$, which is defined as:

$$d_{\tilde{\mathbf{h}}^n,\mathcal{C}} = d\left(p_W f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n}, p_W f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}\right) = \sum_w \int \left| p_W(w) f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n}(y^n) - p_W(w) f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|w)\right| dy^n \tag{55}$$

$$= \sum_w p_W(w) \int \left| f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|w)\right| dy^n. \tag{56}$$

The proof of achievability we present can be outlined in four steps:

1) As in [30], [35], we first prove for any eavesdropper channel state sequence $\tilde{\mathbf{h}}^n$, $d_{\tilde{\mathbf{h}}^n,\mathcal{C}}$ averaged over an ensemble of wiretap codebooks decreases uniformly and exponentially fast with respect to the code length n. As in [30], [36], the proof here uses the information spectrum method from [37].
2) We then quantize the channel gains and construct a finite subset of values of the eavesdropper channel state. We show that for this subset, there must exist a good codebook that retains the property of the codebook ensemble that $d_{\tilde{\mathbf{h}}^n,\mathcal{C}}$ is small.
3) We show that when the eavesdropper channel state is not in the finite subset, the resulting variational distance can be approximated by the variational distance when eavesdropper channel state sequence is in the finite set and hence is also small. This is the approximation argument from [38].
4) Building on 3), we then use [35, Lemma 1] to prove that the secrecy constraint (7) is satisfied, and hence the codebook secures the message for all possible values of eavesdropper channel states.

We start the proof with the following lemma.

Lemma 1: [30, Appendix II, Section D] For a fixed codebook in the ensemble, we have:

$$d_{\tilde{\mathbf{h}}^n,\mathcal{C}} \leq 2\sum_w p_W(w) \int \left| f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_T^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|w)\right| dy^n. \tag{57}$$

For each integral in the sum in (57), using the triangle inequality, we can write

$$\begin{aligned}
&\int \left| f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_T^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|w)\right| dy^n \\
&\leq \int \left| f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|w)\right| dy^n + \\
&\int \left| f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_T^n}(y^n)\right| dy^n.
\end{aligned} \tag{58}$$

The second term in (58) can be readily upper bounded with the following lemma.

Lemma 2: For sufficiently large n, such that $1/2 > e^{-n\alpha(\bar{\epsilon})}$, we have:

$$\int \left| f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_T^n}(y^n)\right| dy^n < 4e^{-n\alpha(\bar{\epsilon})} \tag{59}$$

where $\alpha(\bar{\epsilon})$ is the positive exponent defined in (46).

Proof: This lemma is a natural consequence of the Data Processing Inequality for variational distance stated in Lemma 11. A proof is given in Appendix A for completeness. ■

Bounding the first term in (58) takes a few more steps.

Let $f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}|\bar{\mathbf{X}}}$ denote the conditional p.d.f. implied by the channel matrix $\tilde{\mathbf{H}}^n = \tilde{\mathbf{h}}^n$. Define *information density* [37], $i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}^n, \tilde{\mathbf{Y}}^n\right)$, as :

$$i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}^n, \tilde{\mathbf{Y}}^n\right) = \log_2 \frac{\prod_{i=1}^n f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}|\bar{\mathbf{X}}}\left(\tilde{\mathbf{Y}}_i|\bar{\mathbf{X}}_i\right)}{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}\left(\tilde{\mathbf{Y}}^n\right)}. \tag{60}$$

Then we have the following lemma.

Lemma 3: For a given $\varepsilon > 0$, there exists a constant $\alpha'(\varepsilon) > 0$, such that

$$\Pr\left[\frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_G^n, \tilde{\mathbf{Y}}_G^n\right) > I\left(\bar{\mathbf{X}}_G; \tilde{\mathbf{Y}}_G\right) + \varepsilon\right] \leq e^{-n\alpha'(\varepsilon)}. \tag{61}$$

Proof: The proof utilizes the fact that the probability for the L_2 norm of a length-n sequence sampled from a Gaussian distribution divided by n to be larger than the variance of the Gaussian distribution is negligible [33]. The details are provided in Appendix B. ■

Remark 5: Note that the subscript of $i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}^n, \tilde{\mathbf{Y}}^n\right)$ simply indicates the p.d.f.s we use to compute the information spectrum, which are $f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}|\bar{\mathbf{X}}}$ and $f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}$ in this case. The arguments of $i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}^n, \tilde{\mathbf{Y}}^n\right)$, $\bar{\mathbf{X}}^n$ and $\tilde{\mathbf{Y}}^n$, can have a different p.d.f. than the one indicated by the subscript of $i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}^n, \tilde{\mathbf{Y}}^n\right)$. □

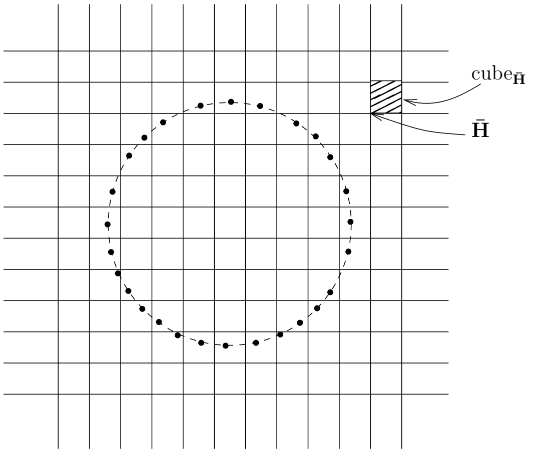

Fig. 5. Conceptual illustration of the construction of the finite set S_M: The dashed circle indicates the set of all $\tilde{\mathbf{H}}$ such that $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H = \mathbf{I}_{N_E \times N_E}$. Each square represents a hyper-cube $\text{cube}_{\bar{\mathbf{H}}}$ for different $\bar{\mathbf{H}}$ which is one of its vertex. The set S_M is composed of all the black dots on the dashed circle. Each square contains exactly one black dot if it intersects with the dashed circle.

Remark 6: The random variables $\bar{\mathbf{X}}_G^n, \tilde{\mathbf{Y}}_G^n$ satisfying Lemma 3 are called to be exponentially information stable in [35, Section 2]. □

We next use Lemma 3 to bound the first term of (58), and use Lemma 2 to bound the second term, which leads to the following lemma.

Lemma 4: If δ in (49)-(50) is positive, then there exists a constant c' such that for sufficiently large n, we have:

$$\begin{aligned}&\mathrm{E}_{\mathcal{C}}\left[2\sum_w p_W(w)\int_{y^n}\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|w)\right|dy^n\right]\\&\le \exp(-c'n) \qquad (62)\end{aligned}$$

and

$$\begin{aligned}&\mathrm{E}_{\mathcal{C}}\left[d_{\tilde{\mathbf{h}}^n,\mathcal{C}}\right] \le\\&\mathrm{E}_{\mathcal{C}}\left[2\sum_w p_W(w)\int_{y^n}\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_T^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|w)\right|dy^n\right]\\&\le \exp(-c'n). \qquad (63)\end{aligned}$$

The value of c' depends only on δ and $\bar{\epsilon}$. The minimum n for (62) and (63) to hold depends only on $\bar{\epsilon}$.

Proof: Two proofs utilizing different techniques are provided in Appendix C and Appendix D respectively[9]. The proof in Appendix C utilizes results from [37, Proof of Theorem 4] and the proof in Appendix D suggested by a reviewer uses results from [39]. ■

Note that (63) is the result mentioned in the first step in the proof outline.

We next construct the finite set S_M of quantized eavesdropper channel state values as mentioned in the second step of the proof outline. As illustrated in Figure 5, S_M is defined as follows:

Note that from (34),

$$\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H = \mathbf{I}_{N_E \times N_E}. \qquad (64)$$

Hence the absolute value of the real and imaginary parts of each element in $\tilde{\mathbf{H}}$ cannot exceed 1. Define $\bar{\mathbf{H}}$ as any matrix such that $M\bar{\mathbf{H}}$ is composed of elements with integral real and imaginary parts taking values in the set $\{-M, -M+1, ..., M-1\}$. For such a $\bar{\mathbf{H}}$, define a hyper-cube over $N_T \times N_E$ matrices, denoted by $\text{cube}_{\bar{\mathbf{H}}}$, as

$$\text{cube}_{\bar{\mathbf{H}}} = \left\{\begin{array}{ll}\mathbf{H}: & 0 \le \mathrm{Re}(M\mathbf{H}_{i,j} - M\bar{\mathbf{H}}_{i,j}) \le 1\\ & 0 \le \mathrm{Im}(M\mathbf{H}_{i,j} - M\bar{\mathbf{H}}_{i,j}) \le 1\end{array}\right\}. \qquad (65)$$

The expression $\bigcup_{\bar{\mathbf{H}}} \text{cube}_{\bar{\mathbf{H}}}$ contains all matrices whose elements' real and imaginary parts are within interval $[-1, 1]$. For each $\bar{\mathbf{H}}$, we choose *any single* matrix from $\text{cube}_{\bar{\mathbf{H}}}$ that satisfies (64) *if it exists* and include it in S_M. Since there are at most $(2M+1)^{2N_T N_E}$ hyper-cubes $\text{cube}_{\bar{\mathbf{H}}}$, S_M is a finite set with at most $(2M+1)^{2N_T N_E}$ elements.[10]

Then from (63), we have:

$$\sum_{\tilde{\mathbf{h}} \in S_M} \mathrm{E}_{\mathcal{C}}\left[d_{\tilde{\mathbf{h}}^n,\mathcal{C}}\right] \le (2M+1)^{2N_T N_E}\exp(-c'n). \qquad (66)$$

Remark 7: Note that this is the same strategy used in proving the compound channel coding theorem in [38]. Reference [38] considered a discrete memoryless channel which is taken from a potentially infinite set and S_M is constructed by quantizing the channel transition probability matrix. Here, since we are considering the Gaussian channel, doing so will not lead to a finite set. The remedy we have is that we construct S_M by quantizing the channel gains instead, which leads to a finite set owing to the fact that the channel gains are bounded[11]. □

Since the codebook ensemble is constructed as in [34], for some n_0, we have [34]

$$\lambda \le 5\exp(-nE(R(\delta'))), \forall n > n_0 \qquad (67)$$

where λ, defined in (54), is the average probability of the decoding error for the codebook ensemble.

Hence, as in [30, Appendix II, Section E], from Markov inequality and (67), we know there must exist one codebook such that

1) The probability of decoding error of the intended receiver vanishes as $n \to \infty$.
2) For each $\tilde{\mathbf{h}} \in S_M$, we have

$$\begin{aligned}&d_{\tilde{\mathbf{h}}^n,\mathcal{C}} \le\\&2\sum_w p_W(w)\int_{y^n}\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|w)\right|dy^n \qquad (68)\\&\le 3 \times 2(2M+1)^{2N_T N_E}e^{-c'n}. \qquad (69)\end{aligned}$$

Also, we observe, by our definition of the codebook ensemble that, for this fixed codebook, the average power of each codeword must be less than or equal to P.

[9] We thank the anonymous reviewer for providing the proof in Appendix D in during the second revision cycle of the paper.

[10] S_M is not empty since it at least contains the matrix $[\mathbf{I}_{N_E \times N_E}, \mathbf{0}_{N_E \times (N_T - N_E)}]$.

[11] We thank the reviewer who suggested we include this insight.

This concludes the second step in the proof outline. From here onward, all the discussion is for this fixed codebook.

We next evaluate $d_{\tilde{\mathbf{h}}^n,\mathcal{C}}$ when $\tilde{\mathbf{h}} \notin S_M$.

Let $\mathbf{A}_i$ denote the ith row of matrix $\mathbf{A}$. We know there must exist a $\mathbf{h}' \in S_M$, such that for the ith row of $\mathbf{h}_\Delta(k) = \tilde{\mathbf{h}}(k) - \mathbf{h}'(k)$, denoted by $\mathbf{h}_{\Delta,i}(k)$, we have:[12]

$$\|\mathbf{h}_{\Delta,i}(k)\|^2 < 2N_T/M^2, i = 1, ..., N_E, k = 1, ..., n. \tag{70}$$

Let $\mathbf{h}_\Delta^n = \tilde{\mathbf{h}}^n - \mathbf{h}'^n$. With the notation (33), we define $x_\Delta^n = \mathbf{h}_\Delta^n x^n, x^n \in \mathcal{C}$. Note that x^n is an $N_T \times n$ matrix and $\mathrm{Tr}[x^n(x^n)^H] < nP, \forall x^n \in \mathcal{C}$.

Let $\lambda_{\max}(\mathbf{A})$ be the largest eigenvalue of matrix $\mathbf{A}$. Then for the ith row of x_Δ^n, $x_{\Delta,i}^n$, we have:

$$\begin{aligned}
&\frac{1}{n}\left\|x_{\Delta,i}^n\right\|^2\\
=&\frac{1}{n}\left\|\mathbf{h}_{\Delta,i}^n x^n\right\|^2 && (71)\\
=&\frac{1}{n}\sum_{k=1}^n \|\mathbf{h}_{\Delta,i}(k)x(k)\|^2 && (72)\\
\leq&\frac{1}{n}\sum_{k=1}^n \lambda_{\max}\left(x(k)\left(x(k)\right)^H\right)\|\mathbf{h}_{\Delta,i}(k)\|^2 && (73)\\
\leq&\frac{1}{n}\sum_{k=1}^n \lambda_{\max}\left(x(k)\left(x(k)\right)^H\right)\frac{2N_T}{M^2} && (74)\\
\leq&\frac{1}{n}\sum_{k=1}^n \mathrm{Tr}\left(x(k)\left(x(k)\right)^H\right)\frac{2N_T}{M^2} && (75)\\
\leq&\mathrm{Tr}\left(\frac{1}{n}\sum_{k=1}^n x(k)\left(x(k)\right)^H\right)\frac{2N_T}{M^2} && (76)\\
\leq&\frac{2N_T P}{M^2}. && (77)
\end{aligned}$$

In (74), we use (70). In (75), we use the fact that the eigenvalues of $\frac{1}{n}x(k)\left(x(k)\right)^H$ are nonnegative.

It follows then that

$$\frac{1}{n}\|x_\Delta^n\|^2 \leq \frac{2N_T N_E P}{M^2}. \tag{78}$$

For $\varepsilon > 0$, define r' and r as:

$$(r')^2 = \frac{2N_T N_E P}{M^2\sigma^2}, \tag{79}$$

$$r = r' + \sqrt{N_E(1+\varepsilon)}. \tag{80}$$

With $\bar{\mathbf{X}}^n$ and $\tilde{\mathbf{Y}}^n$ being the inputs and outputs of the eavesdropper channel with states $\tilde{\mathbf{H}}^n = \tilde{\mathbf{h}}^n$, $\tilde{\mathbf{Y}}^n - \tilde{\mathbf{h}}^n\bar{\mathbf{X}}^n$ is a zero mean rotationally invariant Gaussian distribution whose covariance matrix is equal to $\sigma^2\mathbf{I}_{N_E\times N_E}$. Since from (80), it follows that $r^2 > N_E(1+\varepsilon) > N_E$, there exists a positive $\alpha(\varepsilon)$, such that [33, (B2)]:

$$\Pr\left(\frac{1}{n\sigma^2}\left\|\tilde{\mathbf{Y}}^n - \tilde{\mathbf{h}}^n\bar{\mathbf{X}}^n\right\|^2 \geq r^2 \left|\bar{\mathbf{X}}^n = x^n\right.\right) < e^{-n\alpha(\varepsilon)}. \tag{81}$$

Note that this bound is uniform regardless of the value of $\tilde{\mathbf{h}}^n$.

Let S_M^n denote the n-fold Cartesian product of S_M.

For $\varepsilon > 0$ and r, r' given in (80) and (79), define $g(r, r')$ as

$$g(r, r') = r'(2r + r'). \tag{82}$$

Then we have the following lemma.

Lemma 5: If we can choose M with respect to n such that

$$ng(r, r') < 1 \tag{83}$$

then for any $\tilde{\mathbf{h}}^n$ there must exist $\mathbf{h}'^n \in S_M^n$ such that

$$\begin{aligned}
&d_{\tilde{\mathbf{h}}^n,\mathcal{C}} \leq\\
&2\sum_w p_W(w)\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_\mathcal{C}^n|W}(y^n|w)\right|dy^n\\
&+8e^{-n\alpha(\bar{\epsilon})} && (84)\\
&\leq 2\sum_w p_W(w)\int\left|f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}_G^n}(y^n) - f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}_\mathcal{C}^n|W}(y^n|w)\right|dy^n\\
&+8e^{-n\alpha(\bar{\epsilon})} + 4e^{-n\alpha(\varepsilon)} + 4ng(r, r') && (85)\\
&\leq 12(2M+1)^{2N_T N_E}e^{-c'n} + 8e^{-n\alpha(\bar{\epsilon})}\\
&+4e^{-n\alpha(\varepsilon)} + 4ng(r, r'). && (86)
\end{aligned}$$

Proof: The proof is provided in Appendix G. The inequality (84) is proved with Lemma 1, (58) and Lemma 2. The inequality (85) is proved by evaluating the integral for the two cases $\frac{1}{n\sigma^2}\left\|y^n - \tilde{\mathbf{h}}^n x^n\right\|^2 \geq r^2$ and $\frac{1}{n\sigma^2}\left\|y^n - \tilde{\mathbf{h}}^n x^n\right\|^2 < r^2$ separately. Bounding the second case requires the property that the average energy of each codeword in the codebook does not exceed P. This is the reason we need to sample from distribution (43) when we construct the codebook ensemble. The inequality (86) follows by applying (68)-(69) to (85). ■

Lemma 6: There exists a codebook, such that for $c_0 > 0$, we have

$$d_{\tilde{\mathbf{h}}^n,\mathcal{C}} < \exp(-c_0 n), \quad \forall\tilde{\mathbf{h}}^n. \tag{87}$$

Proof: $g(r, r')$ decrease at the rate of $1/M$. Hence there must exist a positive constant $c_M > 0$, such that $M = \exp(nc_M)$ and both $2(2M+1)^{2N_T N_E}\exp(-c'n)$ and $ng(r, r')$ decrease exponentially fast to 0 with respect to n. Applying it to Lemma 5, we have Lemma 6. ■

Let $c_4 = \delta' + \delta$. From (48), (49) we observe the codebook rate is given by

$$\lim_{n\to\infty}\frac{1}{n}H(W) \geq I\left(\bar{\mathbf{X}}_G; \mathbf{Y}_G\right) - N_E C\left(\frac{P(1-\bar{\epsilon})}{N_{T,R}\sigma^2}\right) - c_4, \tag{88}$$

where $C(x) = \log_2(1+x)$. From (47), we notice that (88) can be made arbitrarily close to

$$I\left(\bar{\mathbf{X}}_G; \mathbf{Y}_G\right) - N_E C\left(\frac{P}{N_{T,R}\sigma^2}\right). \tag{89}$$

To prove that (89) is an achievable secrecy rate, we need the following lemma from [35], which relates $d_{\tilde{\mathbf{h}}^n,\mathcal{C}}$ to the mutual information $I\left(W; \tilde{\mathbf{Y}}^n|\tilde{\mathbf{H}}^n = \tilde{\mathbf{h}}^n\right)$:

[12] $\tilde{\mathbf{h}}$ must be contained in a certain cube defined in (65) which must contain at least one element that satisfies (64) and is included S_M, which is $\mathbf{h}'$.

Lemma 7: [35, Lemma 1] Let $|\mathcal{W}|$ be the cardinality of the message set $\mathcal{W}$. Then we have:

$$I\left(W;\tilde{\mathbf{Y}}^n|\tilde{\mathbf{H}}^n=\tilde{\mathbf{h}}^n\right)\leq d_{\tilde{\mathbf{h}}^n,\mathcal{C}}\log_2\frac{|\mathcal{W}|}{d_{\tilde{\mathbf{h}}^n,\mathcal{C}}}. \tag{90}$$

As shown by Lemma 6, the variational distance $d_{\tilde{\mathbf{h}}^n,\mathcal{C}}$ decreases to 0 exponentially fast with respect to n. $|\mathcal{W}|$ equals N_B from (49). $\frac{1}{n}\log_2|N_B|$ is the right hand side of (88), therefore $\log_2|\mathcal{W}|$ increases linearly with n. Hence from Lemma 7, $I\left(W;\tilde{\mathbf{Y}}^n|\tilde{\mathbf{H}}^n=\tilde{\mathbf{h}}^n\right)$ decreases to 0 exponentially fast with respect to n, and the exponent does not depend on $\tilde{\mathbf{h}}^n$.

This, along with the fact that W is received reliably by the intended receiver and the average power constraint is satisfied by each codeword in the codebook, shows that the rate of the codebook given by (89) is indeed an achievable secrecy rate.

The achieved secrecy rate can then be found by evaluating (89) based on (36) and (38), which leads to (25). This concludes the proof for the static case.

We next extend the result we derived for the static channel to the case where the eavesdropper channel varies from one channel use to the next.

E. Unknown and Varying Eavesdropper Channel

When the eavesdropper channel is varying from one channel use to the next, the second step in the proof for the static case must be modified. This is because, even though the variational distance decreases exponentially fast, the size of the subset of the quantized eavesdropper channel state sequences also increases exponentially fast. In this case, Markov inequality is not sufficient to guarantee the existence of a good codebook and the correlation elimination argument from [40] must be used. The proof outline is as follows:

1) The first step is the same as the static case. We prove for any given sequence of the eavesdropper channel states, the variational distance averaged over an ensemble of wiretap codebooks decreases uniformly and exponentially fast with respect to the code length n.
2) Then, for a finite subset of quantized eavesdropper channel state sequences, we use the correlation elimination argument from [40] to show that there exists a small number of codebooks in the codebook ensemble[13] such that the variational distance averaged over these codebooks is small when the eavesdropper channel state sequence is within the finite set. This is proved by showing that the probability that the variational distance averaged over these codebooks exceeds any given constant is *super*-exponentially small with respect to n for an eavesdropper channel state sequence within the finite subset.
3) The third step is the same as the static case. We show that when the eavesdropper channel state sequence is outside the finite set, the variational distance averaged over this small set of codebooks can be approximated by the variational distance when the eavesdropper channel state sequence is in the finite set and hence is also small. As in the static case proof, a small variational distance implies that the secrecy constraint is satisfied.
4) We then use the small set of codebooks to construct the coding scheme using a two stage transmission scheme introduced in [40].

We next start the proof by defining a normalized version of the variational distance. For a given codebook $\mathcal{C}$, and a given eavesdropper channel state sequence $\{\tilde{\mathbf{H}}(1),...,\tilde{\mathbf{H}}(n)\}=\tilde{\mathbf{h}}^n$, the normalized variational distance $d'_{\tilde{\mathbf{h}}^n,\mathcal{C}}$ is defined as:

$$d'_{\tilde{\mathbf{h}}^n,\mathcal{C}}= \\ \frac{1}{2}\sum_{w}p_W(w)\int_{y^n}\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}\left(y^n\right)-f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}\left(y^n|w\right)\right|dy^n. \tag{91}$$

Clearly, we have

$$0\leq d'_{\tilde{\mathbf{h}}^n,\mathcal{C}}\leq 1. \tag{92}$$

Also, from (84) in Lemma 5, we have:

$$d_{\tilde{\mathbf{h}}^n,\mathcal{C}}\leq 4d'_{\tilde{\mathbf{h}}^n,\mathcal{C}}+8e^{-n\alpha(\bar{\epsilon})}. \tag{93}$$

We then use Lemma 4 to bound $d'_{\tilde{\mathbf{h}}^n,\mathcal{C}}$. Note that Lemma 4 still holds when the eavesdropper channel is varying.

From (62) in Lemma 4, there must exist a constant c', which only depends on ε and $\bar{\epsilon}$ such that

$$\mathrm{E}_{\mathcal{C}}\left[d'_{\tilde{\mathbf{h}}^n,\mathcal{C}}\right]\leq\exp(-c'n). \tag{94}$$

Applying (94) to (93), we complete the first step in the proof outline.

We next use the correlation elimination argument from [40] and consider K codebooks, each generated as described in Section IV-C. Denote the kth randomly generated codebook with $\mathcal{C}_k$. Since for different k, $\mathcal{C}_k$ are i.i.d., $d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k}$ are also i.i.d.. These facts, along with (92), mean that the derivation in [40, (4.1)-(4.5)] can be applied here. In particular, the j, T_j, ε and R in [40] corresponds to k, $d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k}$, c' and K here respectively. Consider a positive sequence $\{\epsilon_n\}$. Reference [40, (4.1)-(4.5)] shows that if (94) holds, then for $\alpha'>0$ and for n such that:

$$1+e^{\alpha'}e^{-c'n}\leq e^{\epsilon_n}, \tag{95}$$

we have:

$$\Pr\left(\frac{1}{K}\sum_{k=1}^{K}d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k}\geq\epsilon_n\right)\leq e^{-(\alpha'-1)K\epsilon_n}. \tag{96}$$

Let $\alpha'=2$. Then we have

$$\Pr\left(\frac{1}{K}\sum_{k=1}^{K}d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k}\geq\epsilon_n\right)\leq e^{-\epsilon_n K}. \tag{97}$$

Recall that S_M^n is the n-fold Cartesian product of the set S_M defined in Section IV-D. Therefore S_M^n has at most $(2M+1)^{2N_T N_E n}$ components. Let $|S_M^n|$ denote the size of the set S_M^n. Then we have:

$$\Pr\left(\frac{1}{K}\sum_{k=1}^{K}d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k}<\epsilon_n,\forall\tilde{\mathbf{h}}^n\in S_M^n\right)$$

[13] In our case, we use $K=e^{\varepsilon' n}$ codebooks, where ε' is a positive constant that can be made arbitrarily small.

$$\geq 1 - |S_M^n| \Pr\left(\frac{1}{K}\sum_{k=1}^{K} d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k} \geq \epsilon_n\right) \tag{98}$$

$$\geq 1 - |S_M^n| e^{-\epsilon_n K}. \tag{99}$$

Recall that r and r' were defined in (80) and (79) respectively. When $\tilde{\mathbf{h}}^n \notin S_M^n$, from (84) being upper bounded by (85) in Lemma 5, we have that if

$$ng(r,r') < 1, \tag{100}$$

then there must exist $\mathbf{h}'^n \in S_M^n$, such that

$$d'_{\tilde{\mathbf{h}}^n,\mathcal{C}} \leq d'_{\mathbf{h}'^n,\mathcal{C}} + e^{-n\alpha(\varepsilon)} + ng(r,r'). \tag{101}$$

Therefore

$$\frac{1}{K}\sum_{k=1}^{K} d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k} \leq \frac{1}{K}\sum_{k=1}^{K} d'_{\mathbf{h}'^n,\mathcal{C}_k} + e^{-n\alpha(\varepsilon)} + ng(r,r'). \tag{102}$$

From Markov inequality and (67), we have:

$$\Pr\left(\lambda_{\mathcal{C}_k} > 5nKe^{-Np\left(R(\delta')\right)}\right) \leq \frac{1}{nK}. \tag{103}$$

Therefore:

$$\Pr\left(\exists k : \lambda_{\mathcal{C}_k} > 5nKe^{-nE\left(R(\delta')\right)}\right)$$

$$\leq \sum_{k=1}^{K} \Pr\left(\lambda_{\mathcal{C}_k} > 5nKe^{-nE\left(R(\delta')\right)}\right) \tag{104}$$

$$\leq \frac{1}{n}. \tag{105}$$

Or equivalently

$$\Pr\left(\lambda_{\mathcal{C}_k} \leq 5nKe^{-nE\left(R(\delta')\right)}, k = 1, ..., K\right) \geq 1 - \frac{1}{n}. \tag{106}$$

We next choose ϵ_n, the number of codebooks K and the variable M, which controls the size of the set S_M^n carefully such that for sufficiently large n,

1) (95) is satisfied for $\alpha' = 2$.
2) $\frac{1}{K}\sum_{k=1}^{K} d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k}$ in (98) vanishes with high probability for $\tilde{\mathbf{h}}^n \in S_M^n$.
3) $\frac{1}{K}\sum_{k=1}^{K} d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k}$ on the left hand side of (102) vanishes for $\tilde{\mathbf{h}}^n \notin S_M^n$. Note that in order to use the bound (102) to prove this result, (100) must be satisfied and the right hand side of (102) must vanish as well, which relies on 2).
4) $\lambda_{\mathcal{C}_k}, k = 1, ..., K$ in (106) vanishes with high probability.

Satisfying these conditions leads to the claim that there exists K good codebooks.

A proper choice for ϵ_n, K and M is as follows: Recall that $\alpha(\varepsilon)$ was defined in (81). $\alpha(\bar{\epsilon})$ was defined in (46) and (47). c' was given in (94). For a positive constant ε' such that

$$\varepsilon' < c' \tag{107}$$

$$\varepsilon' < \alpha(\varepsilon) \tag{108}$$

$$\varepsilon' < \alpha(\bar{\epsilon}) \tag{109}$$

$$2\varepsilon' < E(R(\delta')), \tag{110}$$

the variables ϵ_n, K and M are chosen to be:

$$\epsilon_n = e^{-n\varepsilon'} \tag{111}$$

$$K = e^{2\varepsilon' n} \tag{112}$$

$$M = e^{2\varepsilon' n}. \tag{113}$$

We first check if these choices satisfy (95). We observe that, since $\epsilon_n > 0$, the right hand side of (95) is lower bounded as:

$$e^{\epsilon_n} \geq 1 + \epsilon_n \tag{114}$$

which, due to (111), equals:

$$1 + e^{-\varepsilon' n}. \tag{115}$$

Due to (107), we find (115) is greater than the left hand side of (95) for sufficiently large n such that

$$1 + e^2 e^{-c'n} < 1 + e^{-\varepsilon' n}. \tag{116}$$

Hence, (95) is satisfied.

Next we observe from (111) and (112) that

$$e^{-\epsilon_n K} = e^{-e^{-n\varepsilon'} e^{2n\varepsilon'}} = e^{-e^{n\varepsilon'}}. \tag{117}$$

We also observe that, due to (113), for sufficiently large n,

$$2M + 1 \leq e^{4\varepsilon' n}. \tag{118}$$

Hence,

$$|S_M^n| = (2M+1)^{2N_T N_E n} < e^{8N_T N_E \varepsilon' n^2}. \tag{119}$$

Therefore, from (117) and (119), we have:

$$\lim_{n\to\infty} |S_M^n| e^{-\epsilon_n K} = 0. \tag{120}$$

This means (99) will converge to 1 when n goes to ∞. Since ϵ_n is shown by (111) to converge to 0 when n goes to ∞, we observe, from (98)-(99), that $\frac{1}{K}\sum_{k=1}^{K} d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k}$ in (98) vanishes with high probability.

We next examine (102). We observe from (79), (80) and (82) that $g(r,r')$ decreases at the rate of $1/M$ which, according to (113), equals $e^{-2\varepsilon' n}$. Hence for sufficiently large n, we observe that

$$ng(r,r') < e^{-1.5\varepsilon' n} < e^{-\varepsilon' n} = \epsilon_n \tag{121}$$

holds and (100) is satisfied.

Also, due to (108), we have

$$e^{-n\alpha(\varepsilon)} < e^{-n\varepsilon'} = \epsilon_n. \tag{122}$$

If, for the $\mathbf{h}'^n$ in (102),

$$\frac{1}{K}\sum_{k=1}^{K} d'_{\mathbf{h}'^n,\mathcal{C}_k} < \epsilon_n, \tag{123}$$

then, from (121), (122) and (102), we have

$$\frac{1}{K}\sum_{k=1}^{K} d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k} < 3\epsilon_n. \tag{124}$$

Hence, from (98)-(99), (120) and (123)-(124), we observe there must exist K codebooks, where K is given by (112), such that for any k and $\tilde{\mathbf{h}}^n$,

$$\lambda_{\mathcal{C}_k} \leq 5nKe^{-nE\left(R(\delta')\right)}, \tag{125}$$

$$\frac{1}{K}\sum_{k=1}^{K} d'_{\tilde{\mathbf{h}}^n,\mathcal{C}_k} < 3\epsilon_n. \tag{126}$$

The variable ϵ_n is given by (111).

We next check if our choice of K leads to a vanishing $\lambda_{\mathcal{C}_k}$. By applying (112) to the right hand side of (125), we find it equals:

$$5ne^{-n(E(R(\delta'))-2\varepsilon')}. \tag{127}$$

Due to (110), we find that the right hand side of (125) converges to 0 when n goes to ∞. This means:

$$\lim_{n\to\infty} \lambda_{\mathcal{C}_k} = 0, \quad \forall k. \tag{128}$$

We next express (126) in terms of $d_{\tilde{\mathbf{h}}^n,\mathcal{C}_k}$. Due to (109), we have, from (93), for sufficiently large n:

$$12\epsilon_n = 12e^{-n\varepsilon'} > 8e^{-n\alpha(\bar{\epsilon})}. \tag{129}$$

Hence (126) implies

$$\frac{1}{K}\sum_{k=1}^{K} d_{\tilde{\mathbf{h}}^n,\mathcal{C}_k} < 24\epsilon_n, \quad \forall \tilde{\mathbf{h}}^n. \tag{130}$$

Equation (130) concludes the second and third steps in the proof outline.

We next use the K codebooks to construct the coding scheme. Let the confidential message W be uniformly distributed over the set of $\{1,...,N_B\}$. The encoder f_n used by the transmitter is described as follows:

1) In the first stage, the transmitter chooses the value for an integer K' from $\{1,...,K\}$ according to a uniform distribution. Given $W=i$, f_n outputs the label (i,j) computed by $\mathrm{f}_{n,\mathcal{C}_{K'}}$.
2) In the second stage, K' is transmitted to the intended receiver using a good channel codebook for the main channel.

The decoder of the intended receiver first decodes K', then decodes the confidential message using $\psi_{n,\mathcal{C}_{K'}}$.

Let $\hat{K}'$ be the result decoded by the intended receiver for K'. Then

$$\begin{aligned}&\Pr\left(W \neq \hat{W}\right)\\ &\leq \Pr\left(K' \neq \hat{K}'\right) + \Pr\left(W \neq \hat{W}|K' = \hat{K}'\right) && (131)\\ &= \Pr\left(K' \neq \hat{K}'\right) + \frac{1}{K}\sum_{k=1}^{K}\lambda_{\mathcal{C}_k}. && (132)\end{aligned}$$

Since

$$\lim_{n\to\infty} \Pr\left(K' \neq \hat{K}'\right) = 0 \tag{133}$$

and (128) holds, we have $\lim_{n\to\infty}\Pr\left(W \neq \hat{W}\right) = 0$.

The variational distance for this coding scheme, $d_{\tilde{\mathbf{h}}^n}$, is given by

$$\begin{aligned}d_{\tilde{\mathbf{h}}^n} &= d\left(p_W p_{K'} f_{\tilde{\mathbf{h}}^n,\tilde{Y}^n_{\mathcal{C}_{K'}}}, p_W p_{K'} f_{\tilde{\mathbf{h}}^n,\tilde{Y}^n_{\mathcal{C}_{K'}}|W}\right) && (134)\\ &= \sum_{k,w}\left(\begin{array}{l} p_W(w)\,p_{K'}(k)\times \\ \int\left|f_{\tilde{\mathbf{h}}^n,\tilde{Y}^n_{\mathcal{C}_k}}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{Y}^n_{\mathcal{C}_k}|W}(y^n|w)\right| dy^n \end{array}\right) && (135)\\ &= \frac{1}{K}\sum_{k=1}^{K} d_{\tilde{\mathbf{h}}^n,\mathcal{C}_k}. && (136)\end{aligned}$$

From (130) and (111), we observe that (136) decreases at the speed of $e^{-\varepsilon' n}$. Then from Lemma 7, we have:

$$\lim_{n\to\infty}\sup_{\tilde{\mathbf{h}}^n} I\left(W; K', \tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}^n}\right) = 0. \tag{137}$$

The limit in (137) converges *exponentially fast* with respect to n. Let n' denote the total number of channel uses. Then the second stage takes n_2 channel uses with n_2 given by:

$$n_2 = \frac{1}{R_0}\log_2 K = \frac{2\varepsilon'\log_2 e}{R_0}n, \tag{138}$$

where $R_0 > 0$ is the rate of the conventional channel codebook $\mathcal{C}_0$. The first stage takes n channel uses. Therefore

$$n' = n + n_2 = \left(\frac{2\varepsilon'\log_2 e}{R_0} + 1\right)n. \tag{139}$$

Define $c(\varepsilon')$ as

$$c(\varepsilon') = \left(\frac{2\varepsilon'\log_2 e}{R_0} + 1\right)^{-1} \tag{140}$$

which can be made arbitrarily close to 1 by making ε' small.

Let $\tilde{\mathbf{Y}}^{n_2}_{\tilde{\mathbf{h}}^n}$ denote the signals received by the eavesdropper during the second stage. Then

$$\begin{aligned}&\lim_{n'\to\infty}\sup_{\tilde{\mathbf{h}}^{n'}} I\left(W; \tilde{\mathbf{Y}}^{n'}_{\tilde{\mathbf{h}}^n}\right)\\ &= \lim_{n'\to\infty}\sup_{\tilde{\mathbf{h}}^{n'}} I\left(W; \tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}^n}, \tilde{\mathbf{Y}}^{n_2}_{\tilde{\mathbf{h}}^n}\right) && (141)\\ &\leq \lim_{n'\to\infty}\sup_{\tilde{\mathbf{h}}^n} I\left(W; K', \tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}^n}\right) && (142)\\ &= \lim_{n\to\infty}\sup_{\tilde{\mathbf{h}}^n} I\left(W; K', \tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}^n}\right) = 0. && (143)\end{aligned}$$

Let $c_4 = \delta' + \max\{2\varepsilon, \varepsilon + \frac{\alpha(\bar{\epsilon})}{2}\log_2 e\}$. The secrecy rate is then given by:

$$\begin{aligned}&\lim_{n'\to\infty}\frac{1}{n'}H(W)\\ &\geq \left\{I\left(\bar{\mathbf{X}}_G; \mathbf{Y}_G\right) - N_E C\left(\frac{P(1-\bar{\epsilon})}{N_{T,R}\sigma^2}\right) - c_4\right\}c(\varepsilon'). && (144)\end{aligned}$$

From (47), we notice (144) can be made arbitrarily close to

$$I\left(\bar{\mathbf{X}}_G; \mathbf{Y}_G\right) - N_E C\left(\frac{P}{N_{T,R}\sigma^2}\right). \tag{145}$$

Therefore, the same secrecy rate as given in (25) is achievable even when the eavesdropper channel varies from one channel use to the next.

Remark 8: In order to use the correlation elimination argument from [40], we made three modifications to its proof:

1) Instead of using average error probability as in [40], we use the normalized variational distance defined in (91).
2) In [40], only $K = n^2$ codebooks are used. Here, in order to use Lemma 7 to bound the mutual information with the variational distance, we use $K = e^{\varepsilon' n}$ codebooks.
3) In [40], the index of the codebook used at the transmitter, i.e., K', needs to be reliably communicated to the receiver over an arbitrarily varying channel. In this work, K' is transmitted using a good channel codebook for the main channel which is static. On the other hand, K' may or may not be reliably received over the varying eavesdropper channel. We simply assume K' is revealed to the eavesdropper in order to compute the lower bound on the achievable secrecy rate which is our goal.

□

Remark 9: Recall that the eavesdropper channel state sequence does not adopt in an adversarial manner in accordance with transmitted signals. This means the eavesdropper channel state sequence is chosen without the knowledge of K'. □

Remark 10: The actual distribution of the message $p_W(w)$ is not needed to prove the secrecy constraint in (7). The assumption that W is uniformly distributed is required only when calculating the achievable transmission rate. □

F. Converse for Theorem 1

In this section, we establish the result in Theorem 1, by providing the converse for the high SNR characterization of the secrecy rate found in (145).

Since $\tilde{\mathbf{H}}$ can be arbitrary, when $N_E \geq N_T$, we can choose $\tilde{\mathbf{H}}$ as $[\mathbf{I}_{N_T \times N_T}, \mathbf{0}_{N_T \times (N_E - N_T)}]^T$. The eavesdropper in this case has perfect knowledge of the transmitted signal. Clearly, the secrecy capacity is 0.

We next consider the case when $N_E < N_T$. We use X_i^j to denote the ith to the jth component in a vector $\mathbf{X}$. The secrecy rate is upper bounded by [4]:

$$R_s \leq I\left(\mathbf{X}; \mathbf{Y} | \tilde{\mathbf{Y}}\right). \tag{146}$$

When $N_T \geq N_R$, we assume $\mathbf{H} = [\mathbf{D}_{N_R \times N_R}, \mathbf{0}_{N_R \times (N_T - N_R)}]$ for a diagonal matrix $\mathbf{D}_{N_R \times N_R}$[14]. Since $\tilde{\mathbf{H}}$ is arbitrary, we choose $\tilde{\mathbf{H}}$ as $[\mathbf{I}_{N_E \times N_E}, \mathbf{0}_{N_E \times (N_T - N_E)}]$. Then (146) equals:

$$\begin{aligned}
&I\left(\mathbf{X}; \mathbf{D}_{N_R \times N_R} X_1^{N_R} + \mathbf{Z} | X_1^{N_E}\right) \\
=&I\left(X_1^{N_R}, X_{N_R+1}^{N_T}; \mathbf{D}_{N_R \times N_R} X_1^{N_R} + \mathbf{Z} | X_1^{N_E}\right) &(147)\\
=&I\left(X_1^{N_R}; \mathbf{D}_{N_R \times N_R} X_1^{N_R} + \mathbf{Z} | X_1^{N_E}\right) \\
&\quad + I\left(X_{N_R+1}^{N_T}; \mathbf{D}_{N_R \times N_R} X_1^{N_R} + \mathbf{Z} | X_1^{N_E}, X_1^{N_R}\right) &(148)\\
=&I\left(X_1^{N_R}; \mathbf{D}_{N_R \times N_R} X_1^{N_R} + \mathbf{Z} | X_1^{N_E}\right). &(149)
\end{aligned}$$

When $N_T < N_R$, we assume $\mathbf{H} = [\mathbf{D}_{N_T \times N_T}, \mathbf{0}_{N_T \times (N_R - N_T)}]^T$ for a diagonal matrix $\mathbf{D}_{N_T \times N_T}$. We use the same $\tilde{\mathbf{H}}$ as we did in the previous case. Then (146) equals:

$$\begin{aligned}
&I\left(\mathbf{X}; \mathbf{D}_{N_T \times N_T} X_1^{N_T} + Z_1^{N_T}, Z_{N_T+1}^{N_R} | X_1^{N_E}\right) \\
=&I\left(\mathbf{X}; \mathbf{D}_{N_T \times N_T} X_1^{N_T} + Z_1^{N_T} | X_1^{N_E}\right) + \\
&I\left(\mathbf{X}; Z_{N_T+1}^{N_R} | \mathbf{D}_{N_T \times N_T} X_1^{N_T} + Z_1^{N_T}, X_1^{N_E}\right) &(150)\\
=&I\left(\mathbf{X}; \mathbf{D}_{N_T \times N_T} X_1^{N_T} + Z_1^{N_T} | X_1^{N_E}\right). &(151)
\end{aligned}$$

Define $N_m = \min\{N_T, N_R\}$. Then, in both cases, (146) can be written as:

$$\begin{aligned}
&I\left(X_1^{N_m}; \mathbf{D}_{N_m \times N_m} X_1^{N_m} + Z_1^{N_m} | X_1^{N_E}\right) \\
=&I\left(X_1^{N_m}; Y_1^{N_m} | X_1^{N_E}\right) &(152)
\end{aligned}$$

which equals:

$$\begin{aligned}
&I(X_{N_E+1}^{N_m}; Y_1^{N_m} | X_1^{N_E}) \\
=&h(Y_{N_E+1}^{N_m} | X_1^{N_E}) + h(Y_1^{N_E} | X_1^{N_E}, Y_{N_E+1}^{N_m}) - h(Y_1^{N_m} | X_1^{N_m}) &(153)\\
\leq&h(Y_{N_E+1}^{N_m}) + h(Y_1^{N_E} | X_1^{N_E}, Y_{N_E+1}^{N_m}) - h(Y_1^{N_m} | X_1^{N_m}) &(154)\\
\leq&h(Y_{N_E+1}^{N_m}) + h(Y_1^{N_E} | X_1^{N_E}) - h(Y_1^{N_m} | X_1^{N_m}) &(155)\\
=&h(Y_{N_E+1}^{N_m}) + h(Z_1^{N_E} | X_1^{N_E}) - h(Z_1^{N_m} | X_1^{N_m}) &(156)\\
=&h(Y_{N_E+1}^{N_m}) + h(Z_1^{N_E}) - h(Z_1^{N_m}) &(157)\\
=&h(Y_{N_E+1}^{N_m}) - h(Z_{N_{E+1}}^{N_m}) &(158)\\
=&h(Y_{N_E+1}^{N_m}) - h(Y_{N_{E+1}}^{N_m} | X_{N_{E+1}}^{N_m}) &(159)\\
=&I(X_{N_E+1}^{N_m}; Y_{N_E+1}^{N_m}). &(160)
\end{aligned}$$

Since we assume $\mathbf{H}$ of the original MIMO wiretap channel has a full rank, $\mathbf{D}_{N_m \times N_m}$ also has full rank. Hence the elements on the diagonal line of $\mathbf{D}$ are all positive. This means equation (160) increases at a rate of $O((\min\{N_T, N_R\} - N_E) C(\bar{P}))$. Hence we have proved the converse of Theorem 1.

V. Discussion

A. Interpretation of the Model and the Results

In our model, we assume that the eavesdropper channel state can take any arbitrary value in each channel use, but is not chosen adapting to the signals sent by the transmitter. It is important to note that the secrecy proof relies on this assumption.

In our achievability scheme, at the beginning of each code block, the encoder randomly chooses a codebook from a set of codebooks. It is possible that there exists a small number of eavesdropper channel state sequences for which this codebook does not secure this message. The key observation is that the fact that the codebook choice by the encoder varies in each block leads to the set of these detrimental eavesdropper channel state sequences to vary from block to block. On the other hand, since the eavesdropper channel state sequence is not chosen with the knowledge of signals sent by the transmitter, it is not changing adaptively to remain in this set. This leads us to prove strong secrecy in this set up.

[14] Else, we can perform SVD on $\mathbf{H}$ and transform it into this form.

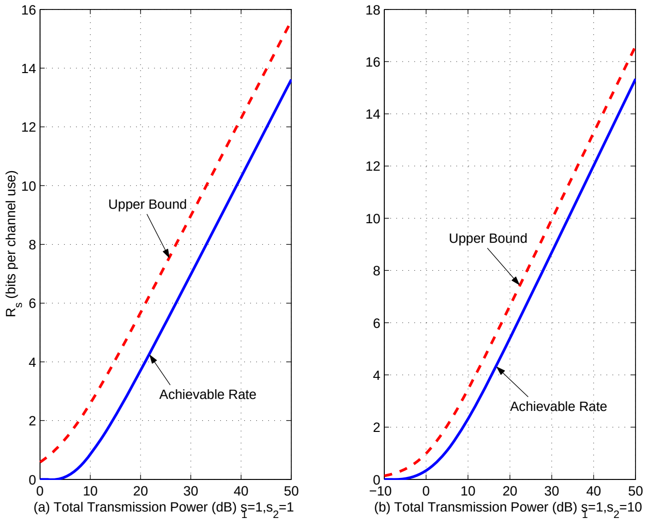

Fig. 6. Secrecy Rate for different total transmission power $\bar{P}$. $N_T = N_R = 2$, $N_E = 1$.

B. Numerical Results

In Figure 6, we plot the achievable secrecy rate given by (25) for different values of $\bar{P}$. The secrecy rate is computed when $N_T = N_R = 2, N_E = 1$. In Figure 6(a), the singular values of the main channel state matrix are $s_1 = 1, s_2 = 1$. In Figure 6(b), the singular values of the main channel state matrix are $s_1 = 1, s_2 = 10$. As shown in Appendix H, the secrecy rate for this antenna configuration can be upper bounded by

$$\min\left\{\log_2(1+|s_1|^2P_1), \log_2(1+|s_2|^2P_2)\right\} \tag{161}$$

subject to

$$P_1 + P_2 \leq \bar{P}, P_1 \geq 0, P_2 \geq 0. \tag{162}$$

In Figure 6, we evaluate this bound and compare it with the achievable rate. The gap is between 1 to 2 bit per channel use.

C. Methods of Proving Strong Secrecy

As mentioned in Section I, in this work we prove strong secrecy directly, as opposed to the indirect approach in [28] which proves strong secrecy using an intermediate step involving weak secrecy. There is other existing work that prove strong secrecy directly, see for example [35], [41]. An indirect proof in [28] has its advantage in that the proof technique, i.e., privacy amplification [28], works with any weak secrecy scheme irrespective of the channel model. Hence it is beneficial to discuss the reason why this indirect approach is not used in this work.

Given a weak secrecy coding scheme that spans over n channel uses, reference [28] considers an equivalent channel for which every n inputs to the weak secrecy scheme is viewed as a single input to the equivalent channel. It then designs a strong secrecy scheme for this equivalent channel. If this approach were followed here, then the unknown and varying channel would be encapsulated inside a universal weak secrecy coding scheme to form the equivalent channel. However, this does not change the fact that the equivalent channel still has a varying joint distribution for its inputs and outputs. It can be shown that the resulting coding scheme does satisfy the strong secrecy constraint in (7) for all possible eavesdropper channel state sequences. However, the convergence speed of the limit in (7) is not uniform over these sequences. Due to this subtlety, the approach of [28] is not used prove strong secrecy in this work.

Comparing with existing contributions that prove strong secrecy directly [26], [27], [30], our proof method is different in the following ways:

- Reference [30]: Both [30] and this work utilize the information spectrum method [37] to prove secrecy results. The proof in this work differs from [30] in the following aspects:
  1) To prove the results in [30], it is sufficient to prove that the left hand side of (61) converges to zero as n goes to ∞. Here, we prove it converges exponentially fast to zero with respect to n to leverage Lemma 7 to bound $I\left(W; \tilde{\mathbf{Y}}^n|\tilde{\mathbf{H}}^n = \tilde{\mathbf{h}}^n\right)$.
  2) In order to use the approximation argument from [38], we need a uniform upper bound on the average power of each codeword in the codebook. This was used in obtaining (77) from (76). To obtain such a bound, we have to sample from a truncated n-letter Gaussian distribution, as shown in (43) and (45), which complicates the analysis of the information density. In contrast, in the setting of [30], it is sufficient to sample from a single letter Gaussian distribution.

  The model considered by [30] is also different from the model considered in this work in the sense that the sequences of eavesdropper channel states in [30] must have a known n letter distribution.
- References [26], [27]: Recently, [26] considered a general varying *discrete* memoryless wiretap channel and proved the existence of a universal coding scheme for this channel for *weak* secrecy. Our work differs from [26] in that we consider a Gaussian MIMO channel model. We find it is possible to prove *strong* secrecy for this settings despite the inputs and outputs being *continuous*. We stress that for the unknown and varying channel setting, the analysis for continuous alphabets do not follow from its discrete counterpart with finite alphabets [26, Lemma 3, Lemma 4]. Thus, care must be exercised in establishing the results from scratch as evidenced by Section IV.

VI. Conclusion

In this work, we have considered secure communication in the presence of eavesdroppers whose channel state varies from one channel use to the next is completely unknown to the legitimate parties. We have shown that when the eavesdropper has fewer antennas than the transmitter and its intended receiver, there exists a universal coding scheme that can guarantee positive secrecy rates irrespective of the channel state sequence of the eavesdropper. The proof utilizes artificial noise, the information spectrum method and the correlation elimination argument.

We have derived achievable secrecy rates for the MIMO wiretap channel, and achievable secrecy rate regions for the two-user MIMO MAC wiretap channel and the two-user MIMO broadcast wiretap channel where the transmitter(s) and the intended receiver(s) have the same number of antennas. We have also derived the secure degrees of freedom, and the secure degrees of freedom regions for these channels by matching the converse to the achievable rates in high SNR. These results are derived in the sense of strong secrecy.

As future work, it is of interest to consider MIMO MAC and MIMO BC wiretap channels with asymmetric number of antennas, for which a time sharing scheme is unlikely to be optimal and the effort in this paper could provide a foundation. In this work, we considered the case where conditioned on a given sequence of channel states, the eavesdropper channel is memoryless. The corresponding channel model with memory deserves further investigation. Finally, we note that in this work the eavesdropper channel sequence is ‘arbitrary’ but is not chosen adaptively, in an adversarial manner by the eavesdropper based on the previous signals received by the eavesdropper. The multiple antenna channel where the eavesdropper can adaptively choose its channel state in an adversarial manner is future work.

ACKNOWLEDGEMENT

The authors wish to thank the Associate Editor and the anonymous reviewers for several constructive suggestions that have improved the overall presentation of the paper.

APPENDIX A
PROOF OF LEMMA 2

Through data processing inequality on variational distance (see Lemma 11), we have:

$$\int \left| f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_G^n}\left(y^n\right) - f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_T^n}\left(y^n\right) \right| dy^n$$

$$\leq \int \left| f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) - f_{\mathbf{X}_T^n}\left(x^n\right) \right| dx^n \tag{163}$$

whose right hand side can be upper bounded as:

$$\int \left| f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) - f_{\mathbf{X}_T^n}\left(x^n\right) \right| dx^n \tag{164}$$

$$= \int_{\frac{1}{n}\|x^n\|^2 > P} \left| f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) - f_{\mathbf{X}_T^n}\left(x^n\right) \right| dx^n +$$

$$\int_{\frac{1}{n}\|x^n\|^2 < P} \left| f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) - f_{\mathbf{X}_T^n}\left(x^n\right) \right| dx^n \tag{165}$$

$$\leq \int_{\frac{1}{n}\|x^n\|^2 > P} f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) dx^n + \int_{\frac{1}{n}\|x^n\|^2 > P} f_{\mathbf{X}_T^n}\left(x^n\right) dx^n$$

$$+ \int_{\frac{1}{n}\|x^n\|^2 < P} \left| f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) - f_{\mathbf{X}_T^n}\left(x^n\right) \right| dx^n \tag{166}$$

$$\leq (1 - \mu_{n,\bar{\epsilon}}) +$$

$$\int_{\frac{1}{n}\|x\|^2 < P} \left| f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) - \mu_{n,\bar{\epsilon}}^{-1} f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) \right| dx^n \tag{167}$$

$$\leq (1 - \mu_{n,\bar{\epsilon}}) + \mu_{n,\bar{\epsilon}}^{-1} - 1 \tag{168}$$

$$= \mu_{n,\bar{\epsilon}}^{-1} - \mu_{n,\bar{\epsilon}}. \tag{169}$$

From (46), we can choose sufficiently large n, such that $\mu_{n,\bar{\epsilon}} > 1/2$. For such n, we have

$$\mu_{n,\bar{\epsilon}}^{-1} - \mu_{n,\bar{\epsilon}}$$

$$= \mu_{n,\bar{\epsilon}}^{-1}\left(1 - \mu_{n,\bar{\epsilon}}^2\right) \tag{170}$$

$$\leq 2\left(1 - \mu_{n,\bar{\epsilon}}^2\right) \tag{171}$$

$$= 2\left(1 - \mu_{n,\bar{\epsilon}}\right)\left(1 + \mu_{n,\bar{\epsilon}}\right) \tag{172}$$

$$\leq 4\left(1 - \mu_{n,\bar{\epsilon}}\right) \leq 4e^{-n\alpha(\varepsilon_P)}. \tag{173}$$

Therefore (163) is upper bounded by $4e^{-n\alpha(\varepsilon_P)}$. This concludes the proof of the lemma.

APPENDIX B
PROOF OF LEMMA 3

Recall that P was defined in (24). Define P' and P'' as

$$P' = \frac{P(1-\bar{\epsilon})}{N_T} \tag{174}$$

$$P'' = P' + \sigma^2. \tag{175}$$

We prove Lemma 3 when $\tilde{\mathbf{H}}^n$ is a sequence such that $\tilde{\mathbf{H}}(i)$ is given by (34). The case where $\tilde{\mathbf{H}}(i)$ is invariant with respect to i is a special case, and does not require a separate proof.

We begin with:

$$\prod_{i=1}^{n} f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}|\bar{\mathbf{X}}}\left(\tilde{\mathbf{Y}}_i | \bar{\mathbf{X}}_i\right)$$

$$= \frac{1}{\left(\pi\sigma^2\right)^{nN_E}} \exp\left(-\frac{\left\|\tilde{\mathbf{Y}}^n - \tilde{\mathbf{h}}^n\bar{\mathbf{X}}^n\right\|^2}{\sigma^2}\right) \tag{176}$$

where $\tilde{\mathbf{h}}^n\bar{\mathbf{X}}^n$ is written using the notation defined in (32). If we choose $\bar{\mathbf{X}}^n \triangleq \bar{\mathbf{X}}_G^n$ and the channel matrix sequence $\tilde{\mathbf{H}}^n = \tilde{\mathbf{h}}^n$ given by (34), $\tilde{\mathbf{Y}}_G^n$ is a rotationally invariant complex Gaussian random vector with zero mean and covariance matrix $P''\mathbf{I}$:

$$f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_G^n}\left(\tilde{\mathbf{Y}}^n\right) = \frac{1}{\left(\pi P''\right)^{nN_E}} \exp\left(-\frac{\left\|\tilde{\mathbf{Y}}^n\right\|^2}{P''}\right). \tag{177}$$

Therefore

$$\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_G^n, \tilde{\mathbf{Y}}_G^n\right)$$

$$= N_E \log_2\left(\frac{P''}{\sigma^2}\right)$$

$$+ \left\{ \frac{1}{n}\left(\frac{\left\|\tilde{\mathbf{Y}}_G^n\right\|^2}{P''}\right) - \frac{1}{n}\left(\frac{\left\|\tilde{\mathbf{Y}}_G^n - \tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n\right\|^2}{\sigma^2}\right) \right\} \log_2 e. \tag{178}$$

Define $\tilde{\mathbf{N}}^n$ as

$$\tilde{\mathbf{N}}^n = \tilde{\mathbf{Y}}_G^n - \tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n. \tag{179}$$

Then we have:

$$\Pr\left(\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}(\bar{\mathbf{X}}_G^n, \tilde{\mathbf{Y}}_G^n) > I(\bar{\mathbf{X}}_G; \tilde{\mathbf{Y}}_G) + \varepsilon\right)$$

$$=\Pr\left(\begin{array}{l} N_E\log_2\left(\frac{P''}{\sigma^2}\right) \\ +\left\{\frac{1}{n}\frac{\left\|\tilde{\mathbf{Y}}_G^n\right\|^2}{P''}-\frac{1}{n}\frac{\left\|\tilde{\mathbf{N}}^n\right\|^2}{\sigma^2}\right\}\log_2 e \\ > I(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G)+\varepsilon \end{array}\right) \tag{180}$$

$$=\Pr\left(\begin{array}{l} N_E\log_2\left(\frac{P''}{\sigma^2}\right)+\frac{\log_2 e}{n}\frac{\left\|\tilde{\mathbf{Y}}_G^n\right\|^2}{P''}> \\ I(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G)+\frac{\log_2 e}{n}\frac{\left\|\tilde{\mathbf{N}}^n\right\|^2}{\sigma^2}+\varepsilon \end{array}\right). \tag{181}$$

Note that

$$\begin{aligned} &I(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G)-N_E\log_2\left(\frac{P''}{\sigma^2}\right) \\ &=N_E\log_2\left(1+\frac{P(1-\bar{\epsilon})}{N_T\sigma^2}\right)-N_E\log_2\left(1+\frac{P(1-\bar{\epsilon})}{N_T\sigma^2}\right) \\ &=0. \end{aligned} \tag{182}$$

Define $\varepsilon'=\varepsilon/\log_2 e$ and $\mathbf{N}'=\tilde{\mathbf{N}}/\sigma$. Then (181) can be written as:

$$\begin{aligned} &\Pr\left(\frac{1}{n}\frac{\left\|\tilde{\mathbf{Y}}_G^n\right\|^2}{P''}>\frac{1}{n}\left\|\mathbf{N}'^n\right\|^2+\varepsilon'\right) \\ &\le\Pr\left(\frac{1}{n}\frac{\left\|\tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n\right\|^2+\left\|\tilde{\mathbf{N}}^n\right\|^2}{P''}>\frac{1}{n}\left\|\mathbf{N}'^n\right\|^2+\varepsilon'\right) &(183) \\ &=\Pr\left(\frac{1}{n}\frac{\left\|\tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n\right\|^2}{P''}>\frac{P'}{nP''}\left\|\mathbf{N}'^n\right\|^2+\varepsilon'\right) &(184) \\ &=\Pr\left(\frac{1}{n}\frac{\left\|\tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n\right\|^2}{P'}>\frac{1}{n}\left\|\mathbf{N}'^n\right\|^2+\frac{P''}{P'}\varepsilon'\right). &(185) \end{aligned}$$

Define $\varepsilon''=\frac{P''}{P'}\varepsilon'$. Then for a fixed positive constant ε_2, (185) can be upper bounded by:

$$\begin{aligned} &\Pr\left(\frac{1}{n}\left\|\mathbf{N}'^n\right\|^2<N_E(1-\varepsilon_2)\right) \\ &+\Pr\left(\frac{1}{n}\left\|\mathbf{N}'^n\right\|^2\ge N_E(1-\varepsilon_2)\right)\times \\ &\Pr\left(\begin{array}{c} \frac{1}{n}\frac{\left\|\tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n\right\|^2}{P'}>\frac{1}{n}\left\|\mathbf{N}'^n\right\|^2+\varepsilon'' \\ \left|\frac{1}{n}\left\|\mathbf{N}'^n\right\|^2\ge N_E(1-\varepsilon_2)\right. \end{array}\right). \end{aligned} \tag{186}$$

The first term in (186) is negligible. This is shown by noting $\mathbf{N}'^n$ is a zero mean Gaussian random vector whose covariance matrix is $\mathbf{I}$. Hence, from [33, (B1)], there exists $\alpha(\varepsilon_2)$, such that

$$\Pr\left(\frac{1}{nN_E}\left\|\mathbf{N}'^n\right\|^2<1-\varepsilon_2\right)<e^{-n\alpha(\varepsilon_2)}. \tag{187}$$

The second term in (186) is upper bounded by:

$$\begin{aligned} &\Pr\left(\begin{array}{c} \frac{1}{n}\frac{\left\|\tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n\right\|^2}{P'}>\frac{1}{n}\left\|\mathbf{N}'^n\right\|^2+\varepsilon'' \\ \left|\frac{1}{n}\left\|\mathbf{N}'^n\right\|^2\ge N_E(1-\varepsilon_2)\right. \end{array}\right) \\ &\le\Pr\left(\begin{array}{c} \frac{1}{n}\frac{\left\|\tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n\right\|^2}{P'}>N_E(1-\varepsilon_2)+\varepsilon'' \\ \left|\frac{1}{n}\left\|\mathbf{N}'^n\right\|^2\ge N_E(1-\varepsilon_2)\right. \end{array}\right) &(188) \\ &=\Pr\left(\frac{1}{n}\frac{\left\|\tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n\right\|^2}{P'}>N_E(1-\varepsilon_2)+\varepsilon''\right) &(189) \\ &=\Pr\left(\frac{1}{nN_E}\left\|\frac{\tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n}{\sqrt{P'}}\right\|^2>(1-\varepsilon_2)+\frac{\varepsilon''}{N_E}\right). &(190) \end{aligned}$$

Since $\tilde{\mathbf{h}}^n$ takes the special form given by (34), each component of $\frac{\tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n}{\sqrt{P'}}$ is a rotationally invariant zero mean complex Gaussian random variable with unit variance, regardless of the value of $\tilde{\mathbf{h}}^n$ and these components are independent.

Therefore, for $\varepsilon_3>0$, we have: if $1-\varepsilon_2+\frac{\varepsilon''}{N_E}\ge 1+\varepsilon_3$, i.e., $\varepsilon''\ge N_E(\varepsilon_3+\varepsilon_2)$, there must exist $\alpha(\varepsilon_3)$, such that [33, (B2)]:

$$\Pr\left(\frac{1}{nN_E}\left\|\frac{\tilde{\mathbf{h}}^n\bar{\mathbf{X}}_G^n}{\sqrt{P'}}\right\|^2>(1-\varepsilon_2)+\frac{\varepsilon''}{N_E}\right)<e^{-n\alpha(\varepsilon_3)}. \tag{191}$$

We have obtained an exponential bound on both terms of (186). Finally, we let $\varepsilon''=N_E(\varepsilon_3+\varepsilon_2)$ and $\varepsilon_2=\varepsilon_3$, which implies $\varepsilon=2N_E\varepsilon_2\log_2 e\frac{P'}{P''}$ and

$$\Pr\left(\begin{array}{l} \frac{1}{n}i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}(\bar{\mathbf{X}}_G^n,\tilde{\mathbf{Y}}_G^n)> \\ I(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G)+2N_E\log_2 e\frac{P'}{P''}\varepsilon_2 \end{array}\right)\le 2e^{-n\alpha(\varepsilon_2)}. \tag{192}$$

Hence we obtain Lemma 3.

Appendix C
Proof of Lemma 4

A. Supporting Results

As in [30, Appendix II, Section D], we can use the symmetry property of the random codebook ensemble and write:

$$\begin{aligned} &\mathrm{E}_{\mathcal{C}}\left[\sum_w p_W(w)\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n)-f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|w)\right|dy^n\right] \\ =&\mathrm{E}_{\mathcal{C}}\left[\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n)-f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|1)\right|dy^n\right]. \end{aligned} \tag{193}$$

For the sake of completeness, we provide Appendix E for steps to obtain (193).

For (193), we have the following lemma.

Lemma 8: [37, Lemma 5] [30, Lemma 6] For any positive sequence $\{\mu_n>0\}$ and a fixed codebook in the ensemble, we have:[15]

$$\begin{aligned} &\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n)-f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|1)\right|dy^n \\ &\le\frac{2}{\log_2 e}\mu_n+2\Pr\left[\log_2\frac{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}\left(\tilde{\mathbf{Y}}_{\mathcal{C},W=1}^n|1\right)}{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}\left(\tilde{\mathbf{Y}}_{\mathcal{C},W=1}^n\right)}>\mu_n\right] \end{aligned} \tag{194}$$

where $\tilde{\mathbf{Y}}_{\mathcal{C},W=1}^n$ denotes the random variable whose p.d.f. is $f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|1)$.

[15] The dependence of μ_n on n is useful when proving Lemma 3. See (213).

The second term on the right hand side of (194) can be bound with the following lemma.

Lemma 9: For any positive μ_n, let τ_n be

$$\tau_n = \frac{e^{\mu_n \ln 2} - \mu_{n,\bar{\epsilon}}^{-1}}{2}. \tag{195}$$

If $\tau_n > 0$, then we have:

$$\begin{aligned}&\mathrm{E}_{\mathcal{C}}\left[\Pr\left[\log_2 \frac{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}\left(\tilde{\mathbf{Y}}^n_{\mathcal{C},W=1}|1\right)}{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_G}\left(\tilde{\mathbf{Y}}^n_{\mathcal{C},W=1}\right)} > \mu_n\right]\right]\\ &\le \mu_{n,\bar{\epsilon}}^{-1}\{\mathrm{A}+\mathrm{B}+\mathrm{C}\}\end{aligned} \tag{196}$$

where

$$\mathrm{A} \triangleq \Pr\left[\begin{array}{c}\frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}^n_G\tilde{\mathbf{Y}}^n_G}\left(\bar{\mathbf{X}}^n_G,\tilde{\mathbf{Y}}^n_G\right)\\ > I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\delta+\frac{1}{n}\log_2\tau_n\end{array}\right] \tag{197}$$

$$\mathrm{B} \triangleq \Pr\left[\frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}^n_G\tilde{\mathbf{Y}}^n_G}\left(\bar{\mathbf{X}}^n_G,\tilde{\mathbf{Y}}^n_G\right) > I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\delta\right] \tag{198}$$

$$\mathrm{C} \triangleq \frac{1}{\tau_n^2}\left(\Pr\left[\begin{array}{l}\frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}^n_G\tilde{\mathbf{Y}}^n_G}\left(\bar{\mathbf{X}}^n_G,\tilde{\mathbf{Y}}^n_G\right) >\\ I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\frac{\delta}{2}\end{array}\right]+2^{-n\delta/2}\right). \tag{199}$$

The variable δ is the codebook parameter used in (49)-(50).

Remark 11: The proof of Lemma 9 is adapted from [37, Proof of Theorem 4] [16]. The difference is that, [37, Proof of Theorem 4] would require the expectation to be taken over an ensemble whose codewords are sampled in an i.i.d. fashion from a Gaussian distribution, since (196) is evaluated for this distribution. In Lemma 9, the expectation is over the ensemble whose codewords are sampled from $Q_{\bar{\mathbf{X}}^n_T}$, which is close to but not equal to a Gaussian distribution. This difference leads to the term $\mu_{n,\bar{\epsilon}}^{-1}$ in front of the upper bound given by (196). □

Proof: The proof is provided in Appendix F. Combining (234) and (256) yields (195). ■

B. Proof

Since $\mu_n > 0$, we find that (195) is lower bounded by:

$$\frac{\mu_n \ln 2 + 1 - \mu_{n,\bar{\epsilon}}^{-1}}{2}. \tag{200}$$

From (46), we can choose a sufficiently large n such that $\mu_{n,\bar{\epsilon}} > 1/2$. This means $\mu_{n,\bar{\epsilon}}^{-1} < 2$. Since from (46), $1-\mu_{n,\bar{\epsilon}} \le e^{-n\alpha(\bar{\epsilon})}$, we have

$$\mu_{n,\bar{\epsilon}}^{-1} - 1 < 2e^{-n\alpha(\bar{\epsilon})}. \tag{201}$$

Applying (201) to (200), we have

$$\tau_n \ge \frac{\mu_n \ln 2 - 2e^{-n\alpha(\bar{\epsilon})}}{2}. \tag{202}$$

The remainder of the proof entails finding an upper bound for (196), which will lead to an upper bound on (62) via Lemma 8. (196) can be bounded using Lemma 3 if its conditions are satisfied. This means that, for a given $\varepsilon > 0$, we require the following three conditions to be satisfied for all n.

$$\frac{\delta}{2} \ge \varepsilon, \tag{203}$$

$$\delta + \frac{1}{n}\log_2 \tau_n \ge \varepsilon, \tag{204}$$

$$\tau_n > 0. \tag{205}$$

Suppose all three conditions are fulfilled, then the three terms in (196) can be bounded as follows. The third term, as shown in Lemma 3 and using (195)-(200), can be bounded as

$$\begin{aligned}&\frac{1}{\tau_n^2}\left(\begin{array}{l}\Pr\left[\frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}^n_G\tilde{\mathbf{Y}}^n_G}\left(\bar{\mathbf{X}}^n_G,\tilde{\mathbf{Y}}^n_G\right) > I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\frac{\delta}{2}\right]\\ +2^{-n\delta/2}\end{array}\right)\\ \le&\frac{4}{(\mu_n\ln 2+1-\mu_{n,\bar{\epsilon}}^{-1})^2}(e^{-n\alpha'(\varepsilon)}+2^{-n\delta/2}).\end{aligned} \tag{206}$$

On the other hand, if (203)-(205) hold, the first two terms in Lemma 9 are all bounded by $e^{-n\alpha'(\varepsilon)}$. Therefore, from Lemma 9, we find

$$\begin{aligned}&\mathrm{E}_{\mathcal{C}}\left[\Pr\left[\log_2 \frac{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}\left(\tilde{\mathbf{Y}}^n_{\mathcal{C}}|1\right)}{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_G}\left(\tilde{\mathbf{Y}}^n_{\mathcal{C}}\right)} > \mu_n\right]\right] \le\\ &\mu_{n,\bar{\epsilon}}^{-1}\left\{2e^{-n\alpha'(\varepsilon)}+\frac{4}{(\mu_n\ln 2+1-\mu_{n,\bar{\epsilon}}^{-1})^2}(e^{-n\alpha'(\varepsilon)}+2^{-n\delta/2})\right\}.\end{aligned} \tag{207}$$

Then, from Lemma 8, we have

$$\begin{aligned}&\mathrm{E}_{\mathcal{C}}\left[\int_{y^n}|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_G}(y^n)-f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}(y^n|1)|dy^n\right]\\ \le&\frac{2}{\log_2 e}\mu_n+\\ &2\mu_{n,\bar{\epsilon}}^{-1}\left(\begin{array}{l}2e^{-n\alpha'(\varepsilon)}+\\ \frac{4}{(\mu_n\ln 2+1-\mu_{n,\bar{\epsilon}}^{-1})^2}(e^{-n\alpha'(\varepsilon)}+2^{-n\delta/2})\end{array}\right).\end{aligned} \tag{208}$$

From Lemma 2, we have

$$\int|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_G}(y^n)-f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_T}(y^n)|dy^n < 4e^{-n\alpha(\bar{\epsilon})}. \tag{209}$$

Finally, from Lemma 1, we have

$$\begin{aligned}&\mathrm{E}_{\mathcal{C}}[d_{\tilde{\mathbf{h}}^n,\mathcal{C}}] \le\\ &\mathrm{E}_{\mathcal{C}}\left[2\sum_w p_W(w)\int_{y^n}\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_T}(y^n)-f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}(y^n|w)\right|dy^n\right]\end{aligned} \tag{210}$$

$$\begin{aligned}\le&\mathrm{E}_{\mathcal{C}}\left[2\sum_w p_W(w)\int_{y^n}\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_G}(y^n)-f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}(y^n|w)\right|dy^n\right]\\ &+2\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_G}(y^n)-f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_T}(y^n)\right|dy^n\end{aligned} \tag{211}$$

$$\begin{aligned}\le&\frac{4}{\log_2 e}\mu_n+\\ &4\mu_{n,\bar{\epsilon}}^{-1}\left(\begin{array}{l}2e^{-n\alpha'(\varepsilon)}+\\ \frac{4}{(\mu_n\ln 2+1-\mu_{n,\bar{\epsilon}}^{-1})^2}(e^{-n\alpha'(\varepsilon)}+e^{-n\delta\ln 2/2})\end{array}\right)\\ &+8e^{-n\alpha(\bar{\epsilon})}.\end{aligned} \tag{212}$$

[16] See also [30, Lemma 7].

We then shown that if $\delta > 2\varepsilon$, then (203)-(205) can be satisfied. In order for the bound given by (212) to be small, we choose μ_n such that it decreases exponentially fast with respect to n. In order for the second term in (212) to decrease exponentially fast with respect to n, we can choose

$$\mu_n = e^{-n\min\left\{\frac{\alpha'(\varepsilon)}{4}, \frac{\delta \ln 2}{8}\right\}} + e^{-n\frac{\alpha(\bar{\epsilon})}{2}}. \tag{213}$$

The term $e^{-n\frac{\alpha(\bar{\epsilon})}{2}}$ ensures by (202) that τ_n stays positive, which is required by (205). As shown by (202), this means for sufficiently large n, whose lower bound only depends on $\bar{\epsilon}$, we have:

$$\tau_n \geq \frac{1}{4} e^{-n\min\left\{\frac{\alpha'(\varepsilon)}{4}, \frac{\delta \ln 2}{8}\right\}} \geq \frac{1}{4} e^{-n\frac{\delta \ln 2}{8}} \tag{214}$$

or equivalently for sufficiently large n:

$$-\frac{1}{n}\log_2 \tau_n \leq \frac{\delta \ln 2}{8}\log_2 e = \frac{\delta}{8}. \tag{215}$$

Hence, we can choose δ such that $\delta \geq \max\{2\varepsilon, \varepsilon + \frac{\delta}{8}\}$, then (203) and (204) hold. Note that $\delta \geq 2\varepsilon$ implies $\delta \geq \varepsilon + \frac{\delta}{8}$.

Finally, for the above choices of μ_n and δ, we note that both (212) and (208) decrease exponentially fast with respect to n. Hence, we have the two inequalities stated in Lemma 4.

APPENDIX D
ALTERNATIVE PROOF OF LEMMA 4

A. Supporting Results

Define the total variance between two probability distributions f and g as

$$\|f - g\|_{TV} = \frac{1}{2}\int_x |f(x) - g(x)| dx \tag{216}$$

$$= \sup_A \int_{x \in A} (f(x) - g(x)) dx. \tag{217}$$

Note that the two definitions, (216) and (217), are equivalent [42, Section 4.1]. Define $i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_T^n \tilde{\mathbf{Y}}_T^n}$ as in (60) except that G is replaced by T therein.

We will need the following three lemmas.

1) *Lemma 10:* Corollary VII.2 from [39]:

$$\begin{aligned} &\mathrm{E}_{\mathcal{C}} \left\| f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_T^n}(y^n) - f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_{\mathcal{C}}^n | W}(y^n|w) \right\|_{TV} \\ &\leq \Pr\left(i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_T^n \tilde{\mathbf{Y}}_T^n}(\bar{\mathbf{X}}_T^n, \tilde{\mathbf{Y}}_T^n) > \tau \right) + \frac{1}{2}\sqrt{\frac{2^\tau}{N_C}} \end{aligned} \tag{218}$$

where

$$\tau = n\left(I(\bar{\mathbf{X}}_G; \tilde{\mathbf{Y}}_G) + \frac{\delta}{2} \right). \tag{219}$$

2) *Lemma 11: Data Processing Inequality for variational distance* [43, Lemma 2] [44, Lemma 1] [45, Lemma 8] [42, Problem 4.3] Let D, D' be two distributions over a domain Ω . Fix any randomized function F on Ω, and let $F(D)$ be the distribution such that a draw from $F(D)$ is obtained by drawing independently x from D and f from F and then outputting $f(x)$ (likewise for $F(D')$). Then we have

$$\|F(D) - F(D')\|_{TV} \leq \|D - D'\|. \tag{220}$$

3) *Lemma 12:* [46, Lemma 3.2.1] Let $\{U_n\}$ and $\{V_n\}$ be arbitrary sequences of random variables taking value in a source alphabets $\{\mathcal{Z}_n\}$. Let $\gamma > 0$ be an arbitrary constant. Then, for all $n = 1, 2, ...$ it holds that

$$\Pr\left\{ \frac{1}{n}\log\frac{\Pr_{U_n}(U_n)}{\Pr_{V_n}(U_n)} \leq -\gamma \right\} \leq e^{-n\gamma}. \tag{221}$$

B. Proof

Our goal is to show that for any message value w,

$$\mathrm{E}_{\mathcal{C}}\left[\left\| f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_T^n}(y^n) - f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_{\mathcal{C}}^n | W}(y^n|w) \right\|_{TV} \right] \leq e^{-c'n} \tag{222}$$

for some $c' > 0$ which does not depend on w. This yields (63). (62) follows by applying Lemma 2 to (63).

To prove (222), first, we note that

$$\begin{aligned} \left\| f_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n} - f_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_T^n \tilde{\mathbf{Y}}_T^n} \right\|_{TV} &\overset{(a)}{=} \left\| f_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n} - f_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_T^n} \right\|_{TV} \\ &\overset{(b)}{\leq} 2e^{-n\alpha(\bar{\epsilon})} \end{aligned} \tag{223}$$

where (a) follows because $f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_G^n | \bar{\mathbf{X}}_G^n} = f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_T^n | \bar{\mathbf{X}}_T^n}$ and (b) follows from (164)-(173). By Lemma 11 where F is chosen to be a deterministic function $\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}(\cdot)$, this implies the total variance between the distributions of random variable $\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}(\bar{\mathbf{X}}_T^n, \tilde{\mathbf{Y}}_T^n)$ and $\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}(\bar{\mathbf{X}}_G^n, \tilde{\mathbf{Y}}_G^n)$ is upper bounded by $4e^{-n\alpha(\bar{\epsilon})}$. Due to the equivalent definition of total variance given by (217), we have for any α

$$\left| \begin{array}{l} \Pr\left(\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}(\bar{\mathbf{X}}_T^n, \tilde{\mathbf{Y}}_T^n) > \alpha \right) \\ - \Pr\left(\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}(\bar{\mathbf{X}}_G^n, \tilde{\mathbf{Y}}_G^n) > \alpha \right) \end{array} \right| \leq 2e^{-n\alpha(\bar{\epsilon})}. \tag{224}$$

Note that for any $\alpha, \gamma > 0$, we have

$$\begin{aligned} &\Pr\left(\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_T^n \tilde{\mathbf{Y}}_T^n}(\bar{\mathbf{X}}_T^n, \tilde{\mathbf{Y}}_T^n) > \alpha \right) \\ &= \Pr\left(\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}(\bar{\mathbf{X}}_T^n, \tilde{\mathbf{Y}}_T^n) - \frac{1}{n}\log\frac{f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_T^n}(\tilde{\mathbf{Y}}_T^n)}{f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_G^n}(\tilde{\mathbf{Y}}_T^n)} > \alpha \right) \\ &\leq \Pr\left(\begin{array}{l} \frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}(\bar{\mathbf{X}}_T^n, \tilde{\mathbf{Y}}_T^n) - \frac{1}{n}\log\frac{f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_T^n}(\tilde{\mathbf{Y}}_T^n)}{f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_G^n}(\tilde{\mathbf{Y}}_T^n)} > \alpha \\ \text{and } \frac{1}{n}\log\frac{f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_T^n}(\tilde{\mathbf{Y}}_T^n)}{f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_G^n}(\tilde{\mathbf{Y}}_T^n)} > -\gamma \end{array} \right) \\ &\qquad + \Pr\left(\log\frac{f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_T^n}(\tilde{\mathbf{Y}}_T^n)}{f_{\tilde{\mathbf{h}}^n, \tilde{\mathbf{Y}}_G^n}(\tilde{\mathbf{Y}}_T^n)} \leq -\gamma \right) \\ &\leq \Pr\left(\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}(\bar{\mathbf{X}}_T^n, \tilde{\mathbf{Y}}_T^n) > \alpha - \gamma \right) + e^{-n\gamma} \\ &\leq \Pr\left(\frac{1}{n} i_{\tilde{\mathbf{h}}^n, \bar{\mathbf{X}}_G^n \tilde{\mathbf{Y}}_G^n}(\bar{\mathbf{X}}_G^n, \tilde{\mathbf{Y}}_G^n) > \alpha - \gamma \right) + e^{-n\gamma} + 2e^{-n\alpha(\bar{\epsilon})} \end{aligned} \tag{225}$$

where the second inequality follows from Lemma 12 and the last inequality follows from Equation (224). In particular, Choosing $\alpha = I(\bar{\mathbf{X}}_G; \tilde{\mathbf{Y}}_G) + \frac{\delta}{2}$ and $\gamma = \frac{\delta}{4}$ in Equation (225), and using Lemma 3, we see that the first term on the right hand side of Equation (218) vanishes exponentially fast. The second term of Equation (218) vanishes exponentially fast by definition of N_C.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2014.2359192, IEEE Transactions on Information Theory

Remark 12: Define [46]

$$\bar{I}(\bar{\mathbf{X}}_T^n;\tilde{\mathbf{Y}}_T^n) = \\ \inf\{\alpha| \lim_{n\to\infty} \Pr\left(\frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_T^n\tilde{\mathbf{Y}}_T^n}(\bar{\mathbf{X}}_T^n,\tilde{\mathbf{Y}}_T^n) > \alpha\right) = 0\}. \quad (226)$$

Then (225) implies that $\bar{I}(\bar{\mathbf{X}}_T^n;\tilde{\mathbf{Y}}_T^n) \le I(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G)$.

□

APPENDIX E
PROOF OF (193)

Let $\bar{\mathbf{X}}_{\mathcal{C}}^n$ denote the random variable that is uniformly distributed over the codebook $\mathcal{C}$. Recall that $\mathrm{E}_B[A]$ denotes the expectation of A averaged over B. Recall that $x_{i,j}^n$ denotes the codeword in the codebook that is labeled with (i,j). N_C, defined in (50), is the number of codewords mapped to the same message value. Since j is uniformly distributed, we have:

$$\mathrm{E}_{\mathcal{C}}\left[\sum_w p_W(w)\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|w)\right|dy^n\right]$$

$$=\mathrm{E}_{\mathcal{C}}\left[\sum_w p_W(w)\int\left|\begin{array}{l} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - \\ \frac{1}{N_C}\sum_{j=1}^{N_C} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|\bar{\mathbf{X}}_{\mathcal{C}}^n}\left(y^n|x_{w,j}^n\right)\end{array}\right|dy^n\right] \quad (227)$$

$$=\sum_w p_W(w)\int\prod_{j=1}^{N_C} f_{\bar{\mathbf{X}}_T^n}\left(x_{w,j}^n\right) \\ \int\left|\begin{array}{l} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - \\ \frac{1}{N_C}\sum_{j=1}^{N_C} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|\bar{\mathbf{X}}_{\mathcal{C}}^n}\left(y^n|x_{w,j}^n\right)\end{array}\right|dy^n dx_{w,j,j=1\dots N_C}^n \quad (228)$$

$$=\sum_w p_W(w)\int\prod_{j=1}^{N_C} f_{\bar{\mathbf{X}}_T^n}\left(x_{1,j}^n\right) \\ \int\left|\begin{array}{l} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - \\ \frac{1}{N_C}\sum_{j=1}^{N_C} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|\bar{\mathbf{X}}_{\mathcal{C}}^n}\left(y^n|x_{1,j}^n\right)\end{array}\right|dy^n dx_{1,j,j=1\dots N_C}^n \quad (229)$$

$$=\int\prod_{j=1}^{N_C} f_{\bar{\mathbf{X}}_T^n}\left(x_{1,j}^n\right) \\ \int\left|\begin{array}{l} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - \\ \frac{1}{N_C}\sum_{j=1}^{N_C} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|\bar{\mathbf{X}}_{\mathcal{C}}^n}\left(y^n|x_{1,j}^n\right)\end{array}\right|dy^n dx_{1,j,j=1\dots N_C}^n \quad (230)$$

$$=\mathrm{E}_{\mathcal{C}}\left[\int\left|\begin{array}{l} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - \\ \frac{1}{N_C}\sum_{j=1}^{N_C} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|\bar{\mathbf{X}}_{\mathcal{C}}^n}\left(y^n|x_{1,j}^n\right)\end{array}\right|dy^n\right] \quad (231)$$

$$=\mathrm{E}_{\mathcal{C}}\left[\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}(y^n) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}(y^n|1)\right|dy^n\right] \quad (232)$$

which is (193).

APPENDIX F
PROOF OF LEMMA 9

Recall that N_C was defined in (50). As in [37, Proof of Theorem 4], we begin by defining random variables $\bar{\mathbf{X}}^n$, $\bar{\mathbf{X}}_1^n,\dots,\bar{\mathbf{X}}_{N_C}^n$ and $\tilde{\mathbf{Y}}^n$ such that

1) The distribution of $\bar{\mathbf{X}}^n$ is given by $Q_{\bar{\mathbf{X}}_T^n}$.
2) The distribution of $\tilde{\mathbf{Y}}^n$ conditioned on $\bar{\mathbf{X}}^n$ is determined by the eavesdropper channel.
3) $\bar{\mathbf{X}}_1^n,\dots,\bar{\mathbf{X}}_{N_C}^n$ are i.i.d. and the distribution of $\bar{\mathbf{X}}_j^n, j = 1,\dots,N_C$ is given by $Q_{\bar{\mathbf{X}}_T^n}$. $\bar{\mathbf{X}}_1^n,\dots,\bar{\mathbf{X}}_{N_C}^n$ are independent from $\tilde{\mathbf{Y}}^n$.

Then, we have [37, Proof of Theorem 4, (4.2) (4.3)]:

$$\mathrm{E}_{\mathcal{C}}\left[\Pr\left[\log\frac{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}\left(\tilde{\mathbf{Y}}_{\mathcal{C},W=1}^n|1\right)}{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}\left(\tilde{\mathbf{Y}}_{\mathcal{C},W=1}^n\right)} > \mu_n\right]\right] \\ \le\Pr\left[\frac{1}{N_C}2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}^n,\tilde{\mathbf{Y}}^n\right)} > \tau_n\right] + \\ \Pr\left[\frac{1}{N_C}\sum_{j=1}^{N_C}2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_j^n,\tilde{\mathbf{Y}}^n\right)} > c_{2,n}+\tau_n\right] \quad (233)$$

where $c_{2,n}$ and $\tau_n > 0$ satisfy

$$2\tau_n + c_{2,n} = 2^{\mu_n}. \quad (234)$$

The values of $c_{2,n}$ and τ_n will be specified later.

Let $1\{a > b\}$ denote the indicator function that equals 1 if $a > b$, and 0 otherwise. For the first term in (233), we can write:

$$\Pr\left[\frac{1}{N_C}2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}^n,\tilde{\mathbf{Y}}^n\right)} > \tau_n\right]$$

$$=\int\begin{array}{l} f_{\bar{\mathbf{X}}^n}(x^n)\, f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}(y^n|x^n) \\ 1\left\{\frac{1}{N_C}2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}(x^n,y^n)} > \tau_n\right\} dx^n dy^n\end{array} \quad (235)$$

$$\le\mu_{n,\bar{\epsilon}}^{-1}\int\begin{array}{l} f_{\bar{\mathbf{X}}_G^n}(x^n)\, f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n|\bar{\mathbf{X}}_G^n}(y^n|x^n) \\ 1\left\{\frac{1}{N_C}2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}(x^n,y^n)} > \tau_n\right\} dx^n dy^n\end{array} \quad (236)$$

$$\le\mu_{n,\bar{\epsilon}}^{-1}\Pr\left[\frac{1}{N_C}2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_G^n,\tilde{\mathbf{Y}}_G^n\right)} > \tau_n\right] \quad (237)$$

$$=\mu_{n,\bar{\epsilon}}^{-1}\Pr\left[\begin{array}{l} \frac{1}{n}i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_G^n,\tilde{\mathbf{Y}}_G^n\right) > \\ I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\delta+\frac{1}{n}\log_2\tau_n\end{array}\right] \quad (238)$$

where (238) follows from (237) by applying the expression of N_C in (50).

For the second term in (233), we follow [37, (4.4),(4.5)] and define the following random variables conditioned on $\tilde{\mathbf{Y}}^n = y^n$:

$$V_{n,j}(y^n) = 2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_j^n,y^n\right)} \quad (239)$$

$$Z_{n,j}(y^n) = V_{n,j}(y^n)\,1\{V_{n,j}(y^n) \le N_C\} \quad (240)$$

$$U_{N_C}(y^n) = \frac{1}{N_C}\sum_{j=1}^{N_C}V_{n,j}(y^n) \quad (241)$$

$$T_{N_C}(y^n) = \frac{1}{N_C}\sum_{j=1}^{N_C}Z_{n,j}(y^n). \quad (242)$$

With these notations, as in [37, (4.6)], we can write:

$$\begin{aligned}
&\Pr\left[\frac{1}{N_C}\sum_{j=1}^{N_C} 2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_j^n,\tilde{\mathbf{Y}}^n\right)} > c_{2,n}+\tau_n \,\middle|\, \tilde{\mathbf{Y}}^n = y^n\right]\\
&=\Pr\left[U_{N_C}\left(\tilde{\mathbf{Y}}^n\right) > c_{2,n}+\tau_n \,\middle|\, \tilde{\mathbf{Y}}^n = y^n\right] \qquad (243)\\
&\le \Pr\left(T_{N_C}\left(\tilde{\mathbf{Y}}^n\right) \neq U_{N_C}\left(\tilde{\mathbf{Y}}^n\right) \,\middle|\, \tilde{\mathbf{Y}}^n = y^n\right)\\
&\quad + \Pr\left(T_{N_C}\left(\tilde{\mathbf{Y}}^n\right) > c_{2,n}+\tau_n \,\middle|\, \tilde{\mathbf{Y}}^n = y^n\right). \qquad (244)
\end{aligned}$$

For the first term in (244), we can write [37]:

$$\begin{aligned}
&\Pr\left(T_{N_C}\left(\tilde{\mathbf{Y}}^n\right) \neq U_{N_C}\left(\tilde{\mathbf{Y}}^n\right) \,\middle|\, \tilde{\mathbf{Y}}^n = y^n\right)\\
&\le \sum_{j=1}^{N_C}\Pr\left[Z_{n,j}\left(\tilde{\mathbf{Y}}^n\right) \neq V_{n,j}\left(\tilde{\mathbf{Y}}^n\right) \,\middle|\, \tilde{\mathbf{Y}}^n = y^n\right] \qquad (245)\\
&= N_C \Pr\left[V_{n,1}\left(\tilde{\mathbf{Y}}^n\right) > N_C \,\middle|\, \tilde{\mathbf{Y}}^n = y^n\right]. \qquad (246)
\end{aligned}$$

Equation (245)-(246) implies

$$\begin{aligned}
&\Pr\left(T_{N_C}\left(\tilde{\mathbf{Y}}^n\right) \neq U_{N_C}\left(\tilde{\mathbf{Y}}^n\right)\right) \le\\
&N_C \Pr\left[V_{n,1}\left(\tilde{\mathbf{Y}}^n\right) > N_C\right]. \qquad (247)
\end{aligned}$$

On the other hand, we have:

$$\begin{aligned}
&N_C\left[\Pr\left[V_{n,1}\left(\tilde{\mathbf{Y}}^n\right) > N_C\right]\right]\\
&= N_C\left[\Pr\left[2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}^n,\tilde{\mathbf{Y}}^n\right)} > N_C\right]\right] \qquad (248)\\
&= N_C\int\left\{\begin{array}{l} f_{\bar{\mathbf{X}}^n}\left(x^n\right) f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n\right)\\ 1\left\{2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(x^n,y^n\right)} > N_C\right\}\end{array}\right\} dx^n dy^n \qquad (249)\\
&\le \mu_{n,\bar{\epsilon}}^{-1} N_C\int\left\{\begin{array}{l} f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n|\bar{\mathbf{X}}_G^n}\left(y^n|x^n\right)\\ 1\left\{2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(x^n,y^n\right)} > N_C\right\}\end{array}\right\} dx^n dy^n. \qquad (250)
\end{aligned}$$

As shown in [37, Proof of Theorem 4], (250) is upper bounded by:

$$\mu_{n,\bar{\epsilon}}^{-1}\Pr\left[\frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_G^n,\tilde{\mathbf{Y}}_G^n\right) > I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\delta\right]. \qquad (251)$$

This, along with (247), means

$$\begin{aligned}
&\Pr\left(T_{N_C}\left(\tilde{\mathbf{Y}}^n\right) \neq U_{N_C}\left(\tilde{\mathbf{Y}}^n\right)\right)\\
&\le \mu_{n,\bar{\epsilon}}^{-1}\Pr\left[\frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_G^n,\tilde{\mathbf{Y}}_G^n\right) > I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\delta\right]. \qquad (252)
\end{aligned}$$

For the second term in (244), we can write [37, (4.7)]:

$$\begin{aligned}
&\mathrm{E}\left[T_{N_C}\left(y^n\right)\right] = \mathrm{E}\left[Z_{n,1}\left(y^n\right)\right] =\\
&\int f_{\bar{\mathbf{X}}^n}\left(x^n\right) 2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(x^n,y^n\right)} 1\left\{2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(x^n,y^n\right)} \le N_C\right\} dx^n \qquad (253)\\
&\le \mu_{n,\bar{\epsilon}}^{-1}\int\left\{\begin{array}{l} f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) 2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(x^n,y^n\right)}\\ 1\left\{2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(x^n,y^n\right)} \le N_C\right\}\end{array}\right\} dx^n \qquad (254)\\
&\le \mu_{n,\bar{\epsilon}}^{-1}. \qquad (255)
\end{aligned}$$

We choose $c_{2,n}$ as

$$c_{2,n} = \mu_{n,\bar{\epsilon}}^{-1}. \qquad (256)$$

Then, following [37, (4.8)], we have

$$\begin{aligned}
&\Pr\left(T_{N_C}\left(\tilde{\mathbf{Y}}^n\right) > c_{2,n}+\tau_n \,\middle|\, \tilde{\mathbf{Y}}^n = y^n\right)\\
&= \Pr\left(T_{N_C}\left(y^n\right) > c_{2,n}+\tau_n\right) \qquad (257)\\
&\le \Pr\left(T_{N_C}\left(y^n\right) - \mathrm{E}\left[T_{N_C}\left(y^n\right)\right] > \tau_n\right) \qquad (258)\\
&\le \frac{1}{\tau_n^2}\mathrm{var}\left(T_{N_C}\left(y^n\right)\right) \qquad (259)\\
&\le \frac{1}{\tau_n^2}\mathrm{E}\left[\frac{1}{N_C} Z_{n,1}^2\left(y^n\right)\right]. \qquad (260)
\end{aligned}$$

The expectation in (260) can be upper bounded by:

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{N_C} Z_{n,1}^2\left(y^n\right)\right] =\\
&\frac{1}{N_C}\int f_{\bar{\mathbf{X}}^n}\left(x^n\right) f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n\right) Z_{n,1}^2\left(y^n\right) dx^n dy^n \qquad (261)\\
&= \frac{1}{N_C}\int\left\{\begin{array}{l} f_{\bar{\mathbf{X}}^n}\left(x^n\right) f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n\right)\\ \left(V_{n,1}\left(y^n\right) 1\left\{V_{n,1}\left(y^n\right) \le N_C\right\}\right)^2\end{array}\right\} dx^n dy^n \qquad (262)\\
&= \frac{1}{N_C}\int\left\{\begin{array}{l} f_{\bar{\mathbf{X}}^n}\left(x^n\right) f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n\right)\\ \left(\begin{array}{l} 2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(x^n,y^n\right)}\\ 1\left\{2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(x^n,y^n\right) \le N_C}\right\}\end{array}\right)^2\end{array}\right\} dx^n dy^n \qquad (263)\\
&\le \mu_{n,\bar{\epsilon}}^{-1}\frac{1}{N_C}\int\left\{\begin{array}{l} f_{\bar{\mathbf{X}}_G^n}\left(x^n\right) f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n|\bar{\mathbf{X}}_G^n}\left(y^n|x^n\right)\\ \left(\begin{array}{l} 2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(x^n,y^n\right)}\\ 1\left\{2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(x^n,y^n\right) \le N_C}\right\}\end{array}\right)^2\end{array}\right\} dx^n dy^n. \qquad (264)
\end{aligned}$$

As illustrated in [37, Proof of Theorem 4], equation (264) is upper bounded by:

$$\mu_{n,\bar{\epsilon}}^{-1}\left\{2^{-n\frac{\delta}{2}} + \Pr\left[\begin{array}{l} \frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_G^n,\tilde{\mathbf{Y}}_G^n\right) >\\ I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\frac{\delta}{2}\end{array}\right]\right\}. \qquad (265)$$

This means

$$\begin{aligned}
&\Pr\left(T_{N_C}\left(\tilde{\mathbf{Y}}^n\right) > c_{2,n}+\tau_n\right)\\
&\le \frac{\mu_{n,\bar{\epsilon}}^{-1}}{\tau_n^2}\left\{2^{-n\frac{\delta}{2}} + \Pr\left[\begin{array}{l} \frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_G^n,\tilde{\mathbf{Y}}_G^n\right) >\\ I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\frac{\delta}{2}\end{array}\right]\right\}. \qquad (266)
\end{aligned}$$

Substituting (252) and (266) to (243)-(244), we observe:

$$\begin{aligned}
&\Pr\left[\frac{1}{N_C}\sum_{j=1}^{N_C} 2^{i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}^n,\tilde{\mathbf{Y}}^n\right)} > c_{2,n}+\tau_n\right]\\
&\le \mu_{n,\bar{\epsilon}}^{-1}\Pr\left[\frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_G^n,\tilde{\mathbf{Y}}_G^n\right) > I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\delta\right] +\\
&\frac{\mu_{n,\bar{\epsilon}}^{-1}}{\tau_n^2}\left\{2^{-n\frac{\delta}{2}} + \Pr\left[\begin{array}{l} \frac{1}{n} i_{\tilde{\mathbf{h}}^n,\bar{\mathbf{X}}_G^n\tilde{\mathbf{Y}}_G^n}\left(\bar{\mathbf{X}}_G^n,\tilde{\mathbf{Y}}_G^n\right) >\\ I\left(\bar{\mathbf{X}}_G;\tilde{\mathbf{Y}}_G\right)+\frac{\delta}{2}\end{array}\right]\right\}. \qquad (267)
\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2014.2359192, IEEE Transactions on Information Theory

Applying this result along with (235)-(238) to (233), we obtain Lemma 9.

APPENDIX G
PROOF OF LEMMA 5

As we had for Lemma 3, we prove Lemma 5 when $\tilde{\mathbf{H}}^n$ is a sequence such that $\tilde{\mathbf{H}}(i)$ is given by (34) and the eavesdropper channel is given by (39).

Let $f_{\tilde{\mathbf{h}}^n}$ be the conditional p.d.f. of the eavesdropper channel implied by (39) when the channel matrix sequence is $\tilde{\mathbf{H}}^n = \tilde{\mathbf{h}}^n$. Let $\mathbf{h}'^n$ be its quantized sequence in S_M. Consider the case when $\tilde{\mathbf{Y}}^n = y^n$, $\bar{\mathbf{X}}^n = x^n$ and $\tilde{\mathbf{H}}^n = \tilde{\mathbf{h}}^n$ such that $\frac{1}{n\sigma^2}\left\|y^n - \tilde{\mathbf{h}}^n x^n\right\|^2 < r^2$. Then we have:

$$\begin{aligned}&\left|\log f_{\tilde{\mathbf{h}}^n}\left(y^n|x^n\right) - \log f_{\mathbf{h}'^n}\left(y^n|x^n\right)\right|\\=&\frac{1}{\sigma^2}\left|\left\|y^n - \tilde{\mathbf{h}}^n x^n\right\|^2 - \left\|y^n - \mathbf{h}'^n x^n\right\|^2\right|.\end{aligned}\tag{268}$$

Recall that $\mathbf{h}^n_\Delta = \tilde{\mathbf{h}}^n - \mathbf{h}'^n$ and $\mathbf{A}_i$ denotes the ith row of matrix $\mathbf{A}$. For a matrix $\mathbf{A}$ that has N rows, let $[\mathbf{A}_1, ..., \mathbf{A}_N]$ denote a row vector formed by concatenating all rows of $\mathbf{A}$. Let $\langle x, y\rangle$ denote the inner product operation for the complex vector space. Note that y^n is a $N_E \times n$ matrix here. Hence the term $\left|\left\|y^n - \tilde{\mathbf{h}}^n x^n\right\|^2 - \left\|y^n - \mathbf{h}'^n x^n\right\|^2\right|$ in (268) can be upper bounded by:

$$2\left|\mathrm{Re}\sum_{i=1}^{N_E}\left\langle y_i^n - \tilde{\mathbf{h}}_i^n x^n, \mathbf{h}^n_{\Delta,i}x^n\right\rangle\right| + \sum_{i=1}^{N_E}\left\|\mathbf{h}^n_{\Delta,i}x^n\right\|^2 \tag{269}$$

$$\leq 2\left|\sum_{i=1}^{N_E}\left\langle y_i^n - \tilde{\mathbf{h}}_i^n x^n, \mathbf{h}^n_{\Delta,i}x^n\right\rangle\right| + \sum_{i=1}^{N_E}\left\|\mathbf{h}^n_{\Delta,i}x^n\right\|^2 \tag{270}$$

$$\begin{aligned}=&2\left|\left\langle\begin{bmatrix}(y_1^n - \tilde{\mathbf{h}}_1^n x^n)\\ \vdots \\ (y_{N_E}^n - \tilde{\mathbf{h}}_{N_E}^n x^n)\end{bmatrix}, \begin{bmatrix}\mathbf{h}^n_{\Delta,1}x^n \\ \vdots \\ \mathbf{h}^n_{\Delta,N_E}x^n\end{bmatrix}\right\rangle\right|\\&+\sum_{i=1}^{N_E}\left\|\mathbf{h}^n_{\Delta,i}x^n\right\|^2\end{aligned}\tag{271}$$

$$\leq 2\sqrt{\left\|y^n - \tilde{\mathbf{h}}^n x^n\right\|^2\left\|x^n_\Delta\right\|^2} + \left\|x^n_\Delta\right\|^2 \tag{272}$$

where in (272) we use the Cauchy-Schwartz inequality. By applying (78) to (272), we observe that (268) is upper bounded by:

$$ng(r, r') = nr'(2r + r'). \tag{273}$$

From Lemma 1, (58) and Lemma 2, we have:

$$\begin{aligned}&d_{\tilde{\mathbf{h}}^n,\mathcal{C}}\\&\leq 2\sum_w p_W(w)\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_T}\left(y^n\right) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}\left(y^n|w\right)\right|dy^n\end{aligned}\tag{274}$$

$$\begin{aligned}&\leq 2\sum_w p_W(w)\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_G}\left(y^n\right) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}\left(y^n|w\right)\right|dy^n\\&\quad + 8e^{-n\alpha(\bar{\epsilon})}.\end{aligned}\tag{275}$$

Hence we have obtained (84) in Lemma 5.

Recall that $\tilde{\mathbf{Y}}^n_G$ is the signal received by the eavesdropper if $\bar{\mathbf{X}}^n = \bar{\mathbf{X}}^n_G$. Since $\tilde{\mathbf{H}}(i)$ always has the form given by (34), we have

$$f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_G}\left(y^n\right) = f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n_G}\left(y^n\right). \tag{276}$$

Therefore, the first term in (275) can be written as:

$$\begin{aligned}&2\sum_w p_W(w)\int\left|f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n_G}\left(y^n\right) - f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}\left(y^n|w\right)\right|dy^n\\\leq&2\sum_w p_W(w)\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}\left(y^n|w\right) - f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}\left(y^n|w\right)\right|dy^n\\&+2\sum_w p_W(w)\int\left|f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n_G}\left(y^n\right) - f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}\left(y^n|w\right)\right|dy^n.\end{aligned}\tag{277}$$

Recall that we label each codebook with (i, j). In the encoder, we let $W = i$ and let the distribution over j be p_j, which is uniform. We denote the codeword with label (w, j) by $x^n_{w,j}$. Then, each term inside the sum over w in the first term of (277) can be upper bounded as:

$$\begin{aligned}&\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}\left(y^n|w\right) - f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n_{\mathcal{C}}|W}\left(y^n|w\right)\right|dy^n\\\leq&\sum_j p_j\int\left|\begin{array}{l}f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right) - \\ f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right)\end{array}\right|dy^n.\end{aligned}\tag{278}$$

The term inside the sum over j can be upper bounded by:

$$\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right) - f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right)\right|dy^n \tag{279}$$

$$\begin{aligned}=&\int_{\frac{1}{n\sigma^2}\left\|y^n - \tilde{\mathbf{h}}^n x^n_{w,j}\right\|^2 > r^2}\left|\begin{array}{l}f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right) - \\ f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right)\end{array}\right|dy^n\\&+\int_{\frac{1}{n\sigma^2}\left\|y^n - \tilde{\mathbf{h}}^n x^n_{w,j}\right\|^2 < r^2}\left|\begin{array}{l}f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right) - \\ f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right)\end{array}\right|dy^n\end{aligned}\tag{280}$$

$$\begin{aligned}\leq&\int_{\frac{1}{n\sigma^2}\left\|y^n - \tilde{\mathbf{h}}^n x^n_{w,j}\right\|^2 > r^2} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right)dy^n +\\&\int_{\frac{1}{n\sigma^2}\left\|y^n - \tilde{\mathbf{h}}^n x^n_{w,j}\right\|^2 > r^2} f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right)dy^n\\&+\int_{\frac{1}{n\sigma^2}\left\|y^n - \tilde{\mathbf{h}}^n x^n_{w,j}\right\|^2 < r^2}\left|\begin{array}{l}f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right) - \\ f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x^n_{w,j}\right)\end{array}\right|dy^n.\end{aligned}\tag{281}$$

Note that from the triangular inequality, we have:

$$\begin{aligned}&\left\|y^n - \tilde{\mathbf{h}}^n x^n_{w,j}\right\|\\&\leq\left\|y^n - \mathbf{h}'^n x^n_{w,j}\right\| + \left\|\left(\mathbf{h}'^n - \tilde{\mathbf{h}}^n\right)x^n_{w,j}\right\|\end{aligned}\tag{282}$$

$$\leq\left\|y^n - \mathbf{h}'^n x^n_{w,j}\right\| + r'\sigma\sqrt{n}. \tag{283}$$

The last step follows from (78) and (79).

Therefore $\frac{1}{n\sigma^2}\left\|y^n - \tilde{\mathbf{h}}^n x^n_{w,j}\right\|^2 > r^2$ implies:

$$\frac{1}{n\sigma^2}\left\|y^n - \mathbf{h}'^n x^n_{w,j}\right\|^2 > (r - r')^2 \tag{284}$$

for

$$r > r'. \tag{285}$$

This means that if (285) holds, (281) is upper bounded by:

$$\begin{aligned}&\int_{\frac{1}{n\sigma^2}\left\|y^n-\tilde{\mathbf{h}}^n x_{w,j}^n\right\|^2>r^2} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)dy^n\\&+\int_{\frac{1}{n\sigma^2}\left\|y^n-\mathbf{h}'^n x_{w,j}^n\right\|^2>(r-r')^2} f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)dy^n\\&+\int_{\frac{1}{n\sigma^2}\left\|y^n-\tilde{\mathbf{h}}^n x_{w,j}^n\right\|^2<r^2}\left|\begin{array}{l} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)-\\ f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)\end{array}\right|dy^n.\end{aligned}\tag{286}$$

Hence, if $r > r'$ and

$$(r-r')^2 \geq N_E(1+\varepsilon) \tag{287}$$

then (281) can be upper bounded by [33, (B2)]:

$$\begin{aligned}&2e^{-n\alpha(\varepsilon)}+\\&\int_{\frac{1}{n\sigma^2}\left\|y^n-\tilde{\mathbf{h}}^n x_{w,j}^n\right\|^2<r^2}\left|\begin{array}{l} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)\\ -f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)\end{array}\right|dy^n.\end{aligned}\tag{288}$$

The second term in (288) can be upper bounded by:

$$\begin{aligned}&\int_{\frac{1}{n\sigma^2}\left\|y^n-\tilde{\mathbf{h}}^n x_{w,j}^n\right\|^2<r^2}\left|\begin{array}{l} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)-\\ f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)\end{array}\right|dy^n\\=&\int_{\frac{1}{n\sigma^2}\left\|y^n-\tilde{\mathbf{h}}^n x_{w,j}^n\right\|^2<r^2}\left\{\begin{array}{l} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)\\ \left|1-\frac{f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)}{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)}\right|\end{array}\right\}dy^n.\end{aligned}\tag{289}$$

Recall that when $\frac{1}{n\sigma^2}\left\|y^n-\tilde{\mathbf{h}}^n x_{w,j}^n\right\|^2 < r^2$, from (269)-(273) we have

$$1-e^{ng(r,r')} \leq 1-\frac{f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)}{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)} \leq 1-e^{-ng(r,r')}. \tag{290}$$

Since $g(r,r') > 0$, we have

$$\begin{aligned}0 \leq &\left|1-\frac{f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)}{f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)}\right|\\ \leq &\max\{e^{ng(r,r')}-1, 1-e^{-ng(r,r')}\}.\end{aligned}\tag{291}$$

Note that $1-e^{-x} \leq 1$ when $x \geq 0$. When $0 \leq x < 1$, $e^x - 1 \leq 2x$. Hence as long as

$$ng(r,r') < 1 \tag{292}$$

we have (289) upper bounded by:

$$\begin{aligned}&\int_{\frac{1}{n\sigma^2}\left\|y^n-\tilde{\mathbf{h}}^n x_{w,j}^n\right\|^2<r^2} f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)2ng\left(r,r'\right)dy^n\\&\leq 2ng\left(r,r'\right).\end{aligned}\tag{293}$$

Therefore as long as (285),(287) and (292) are satisfied, (279) is upper bounded by:

$$\begin{aligned}&\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)-f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}^n|\bar{\mathbf{X}}^n}\left(y^n|x_{w,j}^n\right)\right|dy^n\\&\leq 2e^{-n\alpha(\varepsilon)}+2ng\left(r,r'\right).\end{aligned}\tag{294}$$

Applying this result to (277), we have

$$\begin{aligned}&2\sum_w p_W(w)\int\left|f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_G^n}\left(y^n\right)-f_{\tilde{\mathbf{h}}^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}\left(y^n|w\right)\right|dy^n\\&\leq 4e^{-n\alpha(\varepsilon)}+4ng\left(r,r'\right)\\&\quad+2\sum_w p_W(w)\int\left|f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}_G^n}\left(y^n\right)-f_{\mathbf{h}'^n,\tilde{\mathbf{Y}}_{\mathcal{C}}^n|W}\left(y^n|w\right)\right|dy^n\end{aligned}\tag{295}$$

and, we obtain (85) in Lemma 5.

Since $\mathbf{h}'^n \in S_M$, we can apply (68)-(69) to $\mathbf{h}'^n$, and bound (295) by:

$$4e^{-n\alpha(\varepsilon)}+4ng\left(r,r'\right)+12(2M+1)^{2N_T N_E}\exp(-c'n). \tag{296}$$

Applying this result to (274)-(275), we obtain (86) in Lemma 5.

It remains to check that (285), (287) and (292) are satisfied. This is guaranteed by the definitions of r and r' in (79) and (80) and the condition (83) in the Lemma 5. Hence we have completed the proof of Lemma 5.

APPENDIX H
PROOF OF (161)

Let X_1, X_2 denote the main channel inputs and let Y_1, Y_2 denote the main channel outputs after performing SVD decomposition. Hence $Y_k = s_k X_k + Z_k, k = 1, 2$, where Z_k is the channel noise. Then we observe the secrecy rate is upper bounded by

$$\max_{\substack{f_{X_1,X_2}(x_1,x_2)\\ E\left[|X_1|^2+|X_2|^2\right]\leq\bar{P}}} \min\left\{\begin{array}{l} I\left(X_1,X_2;Y_1,Y_2|X_1\right),\\ I\left(X_1,X_2;Y_1,Y_2|X_2\right)\end{array}\right\} \tag{297}$$

where $f_{X_1,X_2}(x_1,x_2)$ denotes the probability density distribution of X_1, X_2. The two mutual information terms in (297) come from assuming the eavesdropper receives X_1 and X_2 respectively. The bound (297) is upper bounded by

$$\begin{aligned}&\max_{P_1+P_2\leq\bar{P},P_k\geq 0,k=1,2}\\&\min\left\{\begin{array}{ll}\max\limits_{\substack{f_{X_1,X_2}(x_1,x_2)\\ E\left[|X_k|^2\right]\leq P_k,k=1,2}} & I\left(X_1,X_2;Y_1,Y_2|X_1\right),\\ \max\limits_{\substack{f_{X_1,X_2}(x_1,x_2)\\ E\left[|X_k|^2\right]\leq P_k,k=1,2}} & I\left(X_1,X_2;Y_1,Y_2|X_2\right)\end{array}\right\}\end{aligned}\tag{298}$$

which is (161).

APPENDIX I
WEAK SECRECY PROOF WITH A STATIC EAVESDROPPER CHANNEL

The proof for weak secrecy differs from that of strong secrecy in Section IV-D. Additionally, different from previous weak secrecy proofs, one has to exercise care in ensuring that the equivocation constraint is satisfied for any (unknown) eavesdropper channel gain. For these reasons, we provide the proof here. We show the existence of a codebook composed of sub-codebooks each of which is a good channel code for the eavesdropper channel[17] whereas the entire codebook is a good channel code for the main channel. These properties allow us to bound the equivocation as in [17].

The 2^{nR} codewords are sampled from (43) as in Section IV-C where R is

$$R = I(\bar{\mathbf{X}};\mathbf{Y}) - 2\delta_n. \tag{299}$$

and $\{\delta_n\}$, which will be specified later, is a positive sequence that converges to 0 when n goes to ∞. The number of bins and the number of codewords per bin however differ from those in Section IV-C, i.e., we now have

$$N_B = 2^{n(R-I(\bar{\mathbf{X}};\tilde{\mathbf{Y}})+\delta_n)} \tag{300}$$

$$N_C = 2^{n(I(\bar{\mathbf{X}};\tilde{\mathbf{Y}})-\delta_n)}. \tag{301}$$

The intended receiver uses the same decoding rule as in (51). We also define a fictitious decoder $\phi_{\tilde{\mathbf{h}}}$ used by the eavesdropper whose channel state matrix is $\tilde{\mathbf{h}}$. The decoder computes j, the codeword index, given the bin index $i = i_0$ and $\tilde{\mathbf{Y}}^n = \tilde{y}^n$, and is a maximum likelihood decoder:

$$\phi_{\tilde{\mathbf{h}}}(\tilde{y}^n) = \arg \min_{j: x^n_{i_0,j} \in \mathcal{C}} \|\tilde{y}^n - \tilde{\mathbf{H}} x^n_{i_0,j}\|. \tag{302}$$

We use $\eta_{\tilde{\mathbf{h}},j|i}$ to denote the error probability for the eavesdropper to reliably decode j given i and $\tilde{\mathbf{Y}}^n$ when the channel state matrix is $\tilde{\mathbf{h}}$.

Let the distribution of i, j as $p_{i,j}$. Then we can define the average probability of decoding error, $\eta_{\tilde{\mathbf{h}}}$, as

$$\eta_{\tilde{\mathbf{h}}} = \sum_{i,j} p_{i,j}(i,j)\eta_{\tilde{\mathbf{h}},j|i}. \tag{303}$$

In (303), $p_{i,j}$ is determined by the encoder f_n used by the transmitter, which we shall specify next. Let the confidential message W be uniformly distributed over the set of $\{i\}$. Given $W = i$, f_n selects a codeword from all the codewords with label i according to a uniform distribution. With this encoder, we note that $p_{i,j}$ is uniform, and therefore

$$H(\bar{\mathbf{X}}^n) = nR \tag{304}$$

For the intended receiver, we simply follow the definitions of the probability of decoding error in (52)-(54).

First, we have the following lemma.

Lemma 13: For the codebook ensemble described above, $\mathrm{E}_{\mathcal{C}}[\eta_{\tilde{\mathbf{h}}}]$ is the same for all $\tilde{\mathbf{h}}$.

Proof: Consider two eavesdroppers, whose respective channel matrices are given below:

$$\tilde{\mathbf{h}} = [\mathbf{I}, \mathbf{0}]\mathbf{U}_1 \tag{305}$$

$$\tilde{\mathbf{h}}' = [\mathbf{I}, \mathbf{0}]\mathbf{U}_2. \tag{306}$$

Let $\mathcal{C}_1$ be any codebook from the ensemble $\{\mathcal{C}\}$ described in Section IV-C. Let $\mathcal{C}_2$ be

$$\mathcal{C}_2 = \mathbf{U}_2^{-1}\mathbf{U}_1\mathcal{C}_1 \tag{307}$$

$$= \{\mathbf{U}_2^{-1}\mathbf{U}_1 x^n, x^n \in \mathcal{C}_1\}. \tag{308}$$

Define the probability density function of a codebook, $f(\mathcal{C})$, as

$$f(\mathcal{C}) = \prod_{i,j} Q_{\bar{\mathbf{X}}_T^n}(x^n_{i,j}). \tag{309}$$

Since a unitary transform does not change the L_2 norm, we have

$$f(\mathcal{C}_1) = f(\mathcal{C}_2). \tag{310}$$

We also observe from the maximum likelihood decoder (302), that the value of $\eta_{\tilde{\mathbf{h}}}$ for a given codebook $\mathcal{C}$, $\eta_{\tilde{\mathbf{h}}}(\mathcal{C})$, only depends on the set $\tilde{\mathbf{h}}\mathcal{C}$, which is defined as:

$$\tilde{\mathbf{h}}\mathcal{C} = \{\tilde{\mathbf{h}}x^n : x^n \in \mathcal{C}\}. \tag{311}$$

Since

$$\tilde{\mathbf{h}}\mathcal{C}_1 = \tilde{\mathbf{h}}'\mathcal{C}_2, \tag{312}$$

we have

$$\eta_{\tilde{\mathbf{h}}}(\mathcal{C}_1) = \eta_{\tilde{\mathbf{h}}'}(\mathcal{C}_2). \tag{313}$$

Let $\mathbf{U}' = \mathbf{U}_2^{-1}\mathbf{U}_1$. Then we have

$$\mathrm{E}_{\mathcal{C}}[\eta_{\tilde{\mathbf{h}}}] = \int \eta_{\tilde{\mathbf{h}}}\left(\mathcal{C}_1\right) f\left(\mathcal{C}_1\right) d\mathcal{C}_1 \tag{314}$$

$$= \int \eta_{\tilde{\mathbf{h}}'}\left(\mathbf{U}'\mathcal{C}_1\right) f\left(\mathbf{U}'\mathcal{C}_1\right) d\mathcal{C}_1 \tag{315}$$

$$= \int \eta_{\tilde{\mathbf{h}}'}\left(\mathcal{C}_2\right) f\left(\mathcal{C}_2\right) d\mathcal{C}_2 \tag{316}$$

$$= \mathrm{E}_{\mathcal{C}}[\eta_{\tilde{\mathbf{h}}'}]. \tag{317}$$

■

We again quantize the channel gains of the eavesdropper channel. Let us construct the same finite set S_M we used in the strong secrecy proof.

From [34, (7.3.22)], we know that there exists an error exponent $E(R) > 0$ such that, for some n_0,

$$\lambda = \mathrm{E}_{\mathcal{C}}[\lambda_{\mathcal{C}}] \leq 5\exp(-nE(R)), \forall n > n_0. \tag{318}$$

By the same argument, for an eavesdropper whose channel matrix $\tilde{\mathbf{h}}$ is in the set S_M, we know there exists an error exponent $E'_{\tilde{\mathbf{h}}}(\tilde{R}) > 0$ such that for some n_0,

$$\mathrm{E}_{\mathcal{C}}[\eta_{\tilde{\mathbf{h}}}] \leq 5\exp(-nE'_{\tilde{\mathbf{h}}}(\tilde{R})), \forall n > n_0, \tag{319}$$

where $\tilde{R} = I(\bar{\mathbf{X}}, \tilde{\mathbf{Y}}) - \delta_n$.

Note that by Lemma 13, $\mathrm{E}_{\mathcal{C}}[\eta_{\tilde{\mathbf{h}}}]$ is not a function of $\tilde{\mathbf{h}}$. Hence, if an error exponent holds for a certain $\tilde{\mathbf{h}}$, it holds for all $\tilde{\mathbf{h}}$. Therefore, we can omit the subscript $\tilde{\mathbf{h}}$ in the error exponent and rewrite it as $E'(\tilde{R})$.

Recall that $2\delta_n = I(\bar{\mathbf{X}}, \mathbf{Y}) - R$. Hence, we can rewrite both $E(R)$ and $E'(\tilde{R})$ as a function of δ_n. From [34], $E(R)$ and $E'(R)$ have the following property: For $R > 0$ and $\tilde{R} > 0$:

1) $E(R)$ and $E'(\tilde{R})$ are both positive.
2) Both $E(R)$ and $E'(\tilde{R})$ are monotonically decreasing functions of δ_n.
3) If $\delta_n \to 0$, then both $E(R)$ and $E'(\tilde{R})$ converge to 0.

Let $\bar{E}(\delta_n) = \max\{E(R), E'(\tilde{R})\}$. Then $\bar{E}(\delta_n)$ also has the three properties listed above. From the linearity of expectation,

[17] That is to say that if the eavesdropper knows the transmitted signals are restricted to a certain sub-codebook, it can decode the transmitted confidential message reliably.

for sufficiently large n that does not depend on $\tilde{\mathbf{h}}$, we can write:

$$\mathrm{E}_{\mathcal{C}}\left[\lambda_{\mathcal{C}}+\sum_{\tilde{\mathbf{h}}\in S_M}\eta_{\tilde{\mathbf{h}}}\right]\leq 5((2M+1)^{2N_TN_E}+1)e^{-\bar{E}(\delta_n)n}. \tag{320}$$

This means there must exist one codebook in the ensemble such that

$$\lambda_{\mathcal{C}}\leq 5((2M+1)^{2N_TN_E}+1)e^{-\bar{E}(\delta_n)n} \tag{321}$$

$$\eta_{\tilde{\mathbf{h}}}\leq 5((2M+1)^{2N_TN_E}+1)e^{-\bar{E}(\delta_n)n}. \tag{322}$$

To prove the secrecy constraint holds, we first consider the equivocation for one eavesdropper whose channel matrix is in the set S_M, with this codebook:

$$H\left(W|\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right) \tag{323}$$

$$\geq H\left(W|\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right)-H\left(W|\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}},\bar{\mathbf{X}}^n\right) \tag{324}$$

$$=H\left(\bar{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right)-H\left(\bar{\mathbf{X}}^n|W,\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right) \tag{325}$$

$$\geq H\left(\bar{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right)-1-\eta_{\tilde{\mathbf{h}}}nR \tag{326}$$

$$\geq H\left(\bar{\mathbf{X}}^n|\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right)-1-5nR\left((2M+1)^{2N_TN_E}+1\right)e^{-\bar{E}(\delta_n)n} \tag{327}$$

$$=H\left(\bar{\mathbf{X}}^n\right)-I\left(\bar{\mathbf{X}}^n;\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right)-n\varepsilon_n \tag{328}$$

$$\geq nR-h\left(\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right)+h\left(\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}|\bar{\mathbf{X}}^n\right)-n\varepsilon_n \tag{329}$$

where (326) follows from Fano's inequality, (327) follows from (322) and ε_n is defined as:

$$\varepsilon_n=\frac{1}{n}+5R\left((2M+1)^{2N_TN_E}+1\right)e^{-\bar{E}(\delta_n)n}. \tag{330}$$

To proceed, we need the following lemma.

Lemma 14:

$$h\left(\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right)\leq N_En\log\pi e(P+\sigma^2),\quad\forall\tilde{\mathbf{h}}. \tag{331}$$

Proof: Let $\tilde{\mathbf{Y}}_i$ denote the part of $\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}$ received during the ith channel use. We have

$$\sum_{i=1}^{n}E\left[\|\tilde{\mathbf{Y}}_i\|^2\right] \tag{332}$$

$$=\sum_{i=1}^{n}E\left[\|\tilde{\mathbf{h}}\bar{\mathbf{X}}_i+\tilde{\mathbf{N}}_i\|^2\right] \tag{333}$$

$$=\sum_{i=1}^{n}E\left[\|\tilde{\mathbf{N}}_i\|^2\right]+E\left[\|\tilde{\mathbf{h}}\bar{\mathbf{X}}_i\|^2\right] \tag{334}$$

$$=nN_E\sigma^2+\sum_{i=1}^{n}E\left[\|\tilde{\mathbf{h}}\bar{\mathbf{X}}_i\|^2\right]. \tag{335}$$

Let $\tilde{\mathbf{h}}_j$ be the jth row of $\tilde{\mathbf{h}}$. From (34), we have $\tilde{\mathbf{h}}_j\tilde{\mathbf{h}}_j{}^H=1$. Using this result, (335) equals

$$nN_E\sigma^2+\sum_{i=1}^{n}\sum_{j=1}^{N_E}E\left[|\tilde{\mathbf{h}}_j\bar{\mathbf{X}}_i|^2\right], \tag{336}$$

which, due to Cauchy-Schwarz inequality, is upper bounded by

$$nN_E\sigma^2+\sum_{i=1}^{n}\sum_{j=1}^{N_E}E\left[\|\tilde{\mathbf{h}}_j\|^2\|\bar{\mathbf{X}}_i\|^2\right] \tag{337}$$

$$=nN_E\sigma^2+\sum_{j=1}^{N_E}\|\bar{\mathbf{X}}^n\|^2 \tag{338}$$

$$\leq nN_E\left(\sigma^2+P\right). \tag{339}$$

The last step follows from the following fact

$$\frac{1}{n}\|\bar{\mathbf{X}}^n\|^2\leq P,\quad\forall\bar{\mathbf{X}}^n\in\mathcal{C}. \tag{340}$$

Note that this is a stronger requirement than the average power constraint, in that it requires the power of *each* codeword not to exceed P. This is guaranteed by our codebook construction described in Section IV-C. Lemma 14 then follows by using the fact that the average power constrained random vector achieves the largest differential entropy when it has Gaussian distribution with i.i.d. components [47]. ■

Using Lemma 14 and the fact that $h\left(\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}|\bar{\mathbf{X}}^n\right)=nN_E\log\pi e\sigma^2$, we arrive at:

$$H\left(W|\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right)\geq nR-nN_E\log_2\left(1+P/\sigma^2\right)-n\varepsilon_n \tag{341}$$

$$=n\left(I\left(\bar{\mathbf{X}};\mathbf{Y}\right)-N_EC\left(P/\sigma^2\right)\right)-n(2\delta_n+\varepsilon_n). \tag{342}$$

Since $H(W)$ is given by:

$$H\left(W\right)=n\left(I\left(\bar{\mathbf{X}};\mathbf{Y}\right)-I\left(\bar{\mathbf{X}};\tilde{\mathbf{Y}}_{\tilde{\mathbf{h}}}\right)\right)-n\delta_n \tag{343}$$

$$=n\left(I\left(\bar{\mathbf{X}};\mathbf{Y}\right)-N_EC\left(P/\sigma^2\right)\right)-n\delta_n \tag{344}$$

we have shown the weak secrecy constraint holds for this first case, i.e., we have

$$I\left(W;\tilde{\mathbf{Y}}^n_{\tilde{\mathbf{h}}}\right)\leq n\left(\delta_n+\varepsilon_n\right). \tag{345}$$

To complete the proof, we next need to also show the secrecy constraint holds when the eavesdropper's channel matrix is not in S_M.

Let $r^2=2N_E$ and $(r')^2=\frac{2N_TN_EP}{M^2}$. Define α as in (81) with $r^2=2N_E$. Let $g_{r,r'}=r'(2r+r')$.

Let $\tilde{\mathbf{h}}$ denote the channel matrix of the eavesdropper. From our construction of S_M, we know that we can quantize $\tilde{\mathbf{h}}$ to an $\tilde{\mathbf{h}}'\in S_M$ such that (70) holds.

We let the eavesdropper with channel matrix $\tilde{\mathbf{h}}$ use the same "fictitious" decoder we designed for the eavesdropper with channel matrix $\tilde{\mathbf{h}}'$. $\eta_{\tilde{\mathbf{h}}}$ be the corresponding probability of decoding error with this decoder. Then we have the following lemma.

Lemma 15:

$$\eta_{\tilde{\mathbf{h}}}\leq e^{-n\alpha}+5((2M+1)^{2N_TN_E}+1)e^{-(\bar{E}(\delta_n)-g(r,r'))n}. \tag{346}$$

Proof: Let B_{i_0,x^n} be the set of values of $\tilde{\mathbf{Y}}^n$ for which the decoder $\varphi_{\tilde{\mathbf{h}}'}$ outputs x^n given the label i_0. Define $\eta_{\tilde{\mathbf{h}}}|x^n$ be the probability of decoding error for the eavesdropper indexed by $\tilde{\mathbf{h}}$ when the codeword x^n is transmitted. Let $B_{x^n} = B_{i_0,x^n}$ with i_0 being the i label of x^n. Let $r^2 = 2N_E$. Then we have

$$\eta_{\tilde{\mathbf{h}}}|x^n = \int_{y^n \notin B_{x^n}} f_{\tilde{\mathbf{h}}}\left(y^n|x^n\right) dy^n \tag{347}$$

$$\leq \int_{\left\|y^n - \tilde{\mathbf{h}}x^n\right\|^2 \geq nr^2} f_{\tilde{\mathbf{h}}}\left(y^n|x^n\right) dy^n + \int_{\substack{\left\|y^n - \tilde{\mathbf{h}}x^n\right\|^2 < nr^2 \\ y^n \notin B_{x^n}}} f_{\tilde{\mathbf{h}}}\left(y^n|x^n\right) dy^n \tag{348}$$

$$\leq e^{-n\alpha} + \int_{\substack{\left\|y^n - \tilde{\mathbf{h}}x^n\right\|^2 < nr^2 \\ y^n \notin B_{x^n}}} f_{\tilde{\mathbf{h}}}\left(y^n|x^n\right) dy^n. \tag{349}$$

Equation (349) follows from [33, (B2)]. Next, we apply (78) and (268)-(272) to the second term of (349) with $r^2 = 2N_E$ and $(r')^2 = \frac{2N_T N_E P}{M^2}$ and find that (349) is upper bounded by

$$e^{-n\alpha} + e^{ng(r,r')} \int_{\substack{\left\|y^n - \tilde{\mathbf{h}}x^n\right\|^2 < nr^2 \\ y^n \notin B_{x^n}}} f_{\tilde{\mathbf{h}}'}\left(y^n|x^n\right) dy^n \tag{350}$$

$$\leq e^{-n\alpha} + e^{ng(r,r')} \int_{y^n \notin B_{x^n}} f_{\tilde{\mathbf{h}}'}\left(y^n|x^n\right) dy^n. \tag{351}$$

Therefore

$$\eta_{\tilde{\mathbf{h}}} \leq e^{-n\alpha} + e^{ng(r,r')}\eta_{\tilde{\mathbf{h}}'} \tag{352}$$

$$\leq e^{-n\alpha} + 5((2M+1)^{2N_T N_E} + 1)e^{-(\bar{E}(\delta_n) - g(r,r'))n} \tag{353}$$

where the last step follows by applying (322). This concludes the proof of Lemma 15. ■

We next repeat the equivocation computation in (323)-(345). This yields:

$$I\left(W; \tilde{\mathbf{Y}}_{\tilde{\mathbf{h}}}^n\right) \leq n\left(\delta_n + \varepsilon'_n\right), \tag{354}$$

where

$$\varepsilon'_n = \frac{1}{n} + R\left(e^{-n\alpha} + 5((2M+1)^{2N_T N_E} + 1)\right) e^{-(\bar{E}(\delta_n) - g(r,r'))n}. \tag{355}$$

The last step of the achievability proof requires choosing δ_n carefully with respect to n, such that ε_n and ε'_n goes to 0 as n goes to ∞. This can be done by choosing $\bar{E}(\delta_n)$ properly as follows:

1) $\bar{E}(\delta_n)$ decreases to 0 at the rate of $n^{-1/2}$, which ensures $n\bar{E}(\delta_n) \to \infty$ as $n \to \infty$.
2) M increases at the rate of n, hence $g(r,r')$ decreases at the rate of n^{-1}. Therefore $n(\bar{E}(\delta_n) - g(r,r')) \to \infty$ at the rate of $\exp(-c_1\sqrt{n})$ for $c_1 > 0$, as $n \to \infty$.

Since $\bar{E}(\delta_n)$ is a monotonically decreasing function of δ_n, this means δ_n converges to 0 as $n \to \infty$. We also observe in this case both ε_n and ε'_n converge uniformly to 0 as $n \to \infty$ for *any eavesdropper channel matrix*.

In summary, for any eavesdropper, with the same codebook $\mathcal{C}$, we always have

$$\lim_{n\to\infty} \sup_{\tilde{\mathbf{h}}} \frac{1}{n} I\left(W; \tilde{\mathbf{Y}}_{\tilde{\mathbf{h}}}^n\right) = 0. \tag{356}$$

The convergence is uniform over all possible values of the eavesdropper channel states. The reliability requirement (6) is fulfilled by (321).

Remark 13: The secrecy rate found (344) is identical to the rate we derived with strong secrecy requirement in (89). □

References

[1] C. E. Shannon, “Communication Theory of Secrecy Systems,” *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, September 1949.

[2] A. D. Wyner, “The Wire-tap Channel,” *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[3] I. Csiszár and J. Körner, “Broadcast Channels with Confidential Messages,” *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[4] S. Leung-Yan-Cheong and M. Hellman, “The Gaussian Wire-tap Channel,” *IEEE Transactions on Information Theory*, vol. 24, no. 4, pp. 451–456, July 1978.

[5] E. Tekin and A. Yener, “The Gaussian Multiple Access Wire-tap Channel,” *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5747–5755, December 2008.

[6] ——, “ The General Gaussian Multiple Access and Two-Way Wire-Tap Channels: Achievable Rates and Cooperative Jamming,” *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2735–2751, June 2008.

[7] R. Liu, I. Maric, P. Spasojevic, and R. D. Yates, “ Discrete Memoryless Interference and Broadcast Channels with Confidential Messages: Secrecy Rate Regions,” *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2493–2507, June 2008.

[8] E. Ekrem and S. Ulukus, “The Secrecy Capacity Region of the Gaussian MIMO Multi-receiver Wiretap Channel,” *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2083–2114, April 2011.

[9] R. Liu, T. Liu, and H. V. Poor, “Multiple-input Multiple-output Gaussian Broadcast Channels with Confidential Messages,” *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4215–4227, September 2010.

[10] E. Ekrem and S. Ulukus, “Capacity-Equivocation Region of the Gaussian MIMO Wiretap Channel,” *IEEE Transactions on Information Theory*, vol. 58, no. 9, pp. 5699–5710, September 2012.

[11] R. Liu, T. Liu, H. V. Poor, and S. Shamai, “MIMO Gaussian Broadcast Channels with Confidential and Common Messages,” in *International Symposium on Information Theory*, June 2010.

[12] X. He and A. Yener, “Cooperation with an Untrusted Relay: A Secrecy Perspective,” *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3801–3827, August 2010.

[13] E. Ekrem and S. Ulukus, “Secrecy in cooperative relay broadcast channels,” *Information Theory, IEEE Transactions on*, vol. 57, no. 1, pp. 137–155, January 2011.

[14] A. Khisti and G. Wornell, “Secure Transmission with Multiple Antennas-I: The MISOME Wiretap Channel,” *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3088–3104, July 2010.

[15] S. Shafiee, N. Liu, and S. Ulukus, “Towards the Secrecy Capacity of the Gaussian MIMO Wire-tap Channel: The 2-2-1 Channel,” *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4033–4039, September 2009.

[16] F. Oggier and B. Hassibi, “The Secrecy Capacity of the MIMO Wiretap Channel,” in *IEEE International Symposium on Information Theory*, July 2008.

[17] Y. Liang, G. Kramer, H. V. Poor, and S. Shamai, “Compound Wiretap Channels,” *Eurasip Journal on Wireless Communication and Networking, Special issue in Wireless Physical Layer Security*, vol. 2009, Article ID 142374, 12 pages, 2009, doi:10.1155/2009/142374.

[18] E. Ekrem and S. Ulukus, “Secrecy Capacity Region of the Degraded Compound Multi-Receiver Wiretap Channel,” in *47th Allerton Conference on Communication, Control, and Computing*, September 2009.

[19] A. Khisti, “Interference Alignment for the Multi-Antenna Compound Wiretap Channel,” *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 2967–2993, May 2011.

[20] M. Kobayashi, Y. Liang, S. Shamai, and M. Debbah, “On the Compound MIMO Broadcast Channels with Confidential Messages,” in *IEEE International Symposium on Information Theory*, June 2009.

[21] P. K. Gopala, L. Lai, and H. El-Gamal, “On the Secrecy Capacity of Fading Channels,” *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 4687–4698, October 2008.

[22] N. Cai and R. W. Yeung, “Secure Network Coding,” in *IEEE International Symposium on Information Theory*, June 2002.

[23] D. Silva and F. R. Kschischang, "Security for wiretap networks via rank-metric codes," in *IEEE International Symposium on Information Theory*, July 2008.

[24] L. H. Ozarow and A. D. Wyner, "Wire-tap channel. II," *AT & T Bell Laboratories technical journal*, vol. 63, no. 10, pp. 2135–2157, 1984.

[25] S. Goel and R. Negi, "Guaranteeing Secrecy using Artificial Noise," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2180–2189, June 2008.

[26] E. MolavianJazi, "Secure Communication Over Arbitrarily Varying Wiretap Channels," *Master Thesis*, December 2009, available online at http://etd.nd.edu/ETD-db/theses/available/etd-12112009-112419/unrestricted/MolavianJaziE122009.pdf.

[27] E. MolavianJazi, M. Bloch, and J. Laneman, "Arbitrary Jamming can Preclude Secure Communication," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*. IEEE, September 2009, pp. 1069–1075.

[28] U. Maurer and S. Wolf, "Information-theoretic Key Agreement: From Weak to Strong Secrecy for Free," *Lecture Notes in Computer Science*, pp. 351–368, 2000.

[29] X. He and A. Yener, "Strong Secrecy and Reliable Byzantine Detection in the Presence of an Untrusted Relay," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 177–192, 2012.

[30] M. Bloch and J. N. Laneman, "On the secrecy capacity of arbitrary wiretap channels," in *46th Allerton Conference on Communication, Control, and Computing*, September 2008.

[31] X. He and A. Yener, "Providing Secrecy With Structured Codes: Two-User Gaussian Channels," *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2121–2138, 2014.

[32] G. Bagherikaram, A. S. Motahari, and A. K. Khandani, "On the Secure DoF of the Single-Antenna MAC," in *IEEE International Symposium on Information Theory*, June 2010.

[33] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 409–417, 1994.

[34] R. G. Gallager, *Information theory and reliable communication*. John Wiley & Sons, Inc. New York, NY, USA, 1968.

[35] I. Csiszár, "Almost Independence and Secrecy Capacity," *Problems of Information Transmission*, vol. 32, no. 1, pp. 48–57, 1996.

[36] M. Bloch and J. Laneman, "Information-spectrum methods for information-theoretic security," in *Information Theory and Applications Workshop, 2009*. IEEE, February 2009, pp. 23–28.

[37] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, 1993.

[38] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1229–1241, 1959.

[39] P. Cuff, "Distributed Channel Synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[40] R. Ahlswede, "Elimination of Correlation in Random Codes for Arbitrarily Varying Channels," *Probability Theory and Related Fields*, vol. 44, no. 2, pp. 159–175, 1978.

[41] A. J. Pierrot and M. R. Bloch, "Strongly Secure Communications Over the Two-Way Wiretap Channel," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 595–605, September 2011.

[42] D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times*. American Mathematical Soc., 2009.

[43] L. Reyzin, "Extractors and the leftover hash lemma," available online at http://www.cs.bu.edu/∼reyzin/teaching/s11cs937/notes-leo-1.pdf.

[44] C. Canonne, D. Ron, and R. A. Servedio, "Testing probability distributions using conditional samples," November 2012, available online at http://arxiv.org/abs/1211.2664.

[45] M. Bloch and J. N. Laneman, "Secrecy from resolvability," *CoRR*, vol. abs/1105.5419, 2011.

[46] T. S. Han, *Information-spectrum Methods in Information Theory*. Springer, 2002.

[47] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience New York, 2006.

Xiang He (S'08, M'10) received B.S. and M.S. degrees in Electrical Engineering from Shanghai Jiao Tong University, Shanghai, China in 2003 and 2006 respectively. His master study is about high speed FPGA implementation of channel encoder, decoder and MIMO detectors. He received his Ph.D. degree in 2010 from the Department of Electrical Engineering at the Pennsylvania State University and joined Microsoft in that year. In 2010, he received Melvin P. Bloom Memorial Outstanding Doctoral Research Award from the Department of Electrical Engineering at the Pennsylvania State University and the best paper award from the Communication Theory Symposium in IEEE International Conference on Communications (ICC). In 2011, he was named as one of the exemplary reviewers by IEEE Communication Letters. His research interests include information theoretic secrecy, coding theory, queuing theory, optimization techniques, distributed detection and estimation.

Aylin Yener Aylin Yener (S'91-M'00-SM'13) received the B.Sc. degree in electrical and electronics engineering, and the B.Sc. degree in physics, from Boğaziçi University, Istanbul, Turkey; and the M.S. and Ph.D. degrees in electrical and computer engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University, New Brunswick, NJ. Commencing fall 2000, for three semesters, she was a P.C. Rossin Assistant Professor at the Electrical Engineering and Computer Science Department, Lehigh University, PA. In 2002, she joined the faculty of The Pennsylvania State University, University Park, PA, where she was an Assistant Professor (2002-2006), then Associate Professor (2006-2010), and is currently Professor of Electrical Engineering since 2010. During the academic year 2008-2009, she was a Visiting Associate Professor with the Department of Electrical Engineering, Stanford University, CA. Her research interests are in information theory, communication theory and network science, with recent emphasis on green communications and information security. She received the NSF CAREER award in 2003.

Dr. Yener previously served as a technical program chair or co-chair for various conferences for the IEEE Communications Society, as an associate editor for the IEEE Transactions on Communications, as an associate editor and an editorial advisory board member for the IEEE Transactions on Wireless Communications. She served as the student committee chair for the IEEE Information Theory Society 2007-2011, and was the co-founder of the Annual School of Information Theory in North America co-organizing the school in 2008, 2009 and 2010. Dr. Yener currently serves on the board of governors of the IEEE Information Theory Society as its treasurer.