Secure Broadcasting Over Fading Channels

Ashish Khisti, *Student Member, IEEE*, Aslan Tchamkerten, *Member, IEEE*, and Gregory W. Wornell, *Fellow, IEEE*

***Abstract*—We study a problem of broadcasting confidential messages to multiple receivers under an information-theoretic secrecy constraint. Two scenarios are considered: 1) all receivers are to obtain a common message; and 2) each receiver is to obtain an independent message. Moreover, two models are considered: parallel channels and fast-fading channels.**

For the case of reversely degraded parallel channels, one eavesdropper, and an arbitrary number of legitimate receivers, we determine the secrecy capacity for transmitting a common message, and the secrecy sum-capacity for transmitting independent messages. For the case of fast-fading channels, we assume that the channel state information of the legitimate receivers is known to all the terminals, while that of the eavesdropper is known only to itself. We show that, using a suitable binning strategy, a common message can be reliably and securely transmitted at a rate independent of the number of receivers. We also show that a simple opportunistic transmission strategy is optimal for the reliable and secure transmission of independent messages in the limit of large number of receivers.

***Index Terms*—Confidential messages, cryptography, fading channels, information-theoretic secrecy, key distribution, multicasting, multiuser diversity, parallel channels, wiretap channel.**

I. Introduction

A NUMBER of existing and emerging applications require a key distribution mechanism to selectively broadcast confidential messages to legitimate receivers. For example, in pay-TV systems, a content provider wishes to selectively broadcast certain content to a subset of customers who have subscribed to it. An online key distribution mechanism enables the service provider to distribute a decryption key to these legitimate receivers while securing it from potential eavesdroppers. The content can be encrypted via standard cryptographic protocols, so that only customers who have access to the decryption key can view it. In the absence of such a mechanism, current solutions rely on variants of traditional public key cryptography (see, e.g., [7]) and are vulnerable to attacks such as piracy [9].

The problem of broadcasting confidential messages in an information-theoretic setting was formulated by Wyner [25]. The so-called wiretap channel model introduced by Wyner in his work has three terminals: a sender, a legitimate receiver, and an eavesdropper. For this formulation, Wyner investigated the fundamental tradeoff between the rate to the legitimate receiver and the eavesdropper's equivocation (the number of bits the eavesdropper must correctly guess to decode the message, given its observations), and characterized the associated rate-equivocation region when the eavesdropper has a degraded channel compared to the legitimate receiver. This formulation is generalized for nondegraded broadcast channels in [5], and applied to Gaussian channels in [13].

Recently, the wiretap channel has received renewed interest for secure communication in wireless environments [20], [2], [12], [15], [17], [11]. The approach in these works is to exploit the channel variations experienced by the receivers to enable secure communication even when the eavesdropper has, on average, a channel stronger than that of the receiver. Some treatments [11], [12], [20] observe that for secure communication over ergodic fading channels, it is sufficient to have only statistical knowledge of the eavesdropper's channel, and the proposed strategies carefully adapt to the channel variations of the legitimate receiver.

In this paper, motivated by the key-distribution application, we further investigate physical-layer security within Wyner's wiretap channel framework by extending it to broadcast scenarios in which there are multiple receivers.

We begin by extending the wiretap model to the case of parallel broadcast channels with one sender, multiple legitimate receivers, and one eavesdropper. We consider two scenarios: 1) there is a common message to be delivered to all legitimate receivers; and 2) there are individual messages to be delivered to each legitimate receiver. For the first scenario, we first derive upper and lower bounds on the common-message secrecy capacity. These bounds coincide when the receivers are reversely degraded. For the second scenario, we establish the secrecy sum-capacity for the reversely degraded case. The capacity-achieving scheme is simple: transmit to the strongest receiver on each parallel channel and use independent codebooks across the subchannels. Our results can be viewed as generalizations of the results in [8], which considers a similar setup without the presence of an eavesdropper. Interestingly, however, the specializations of our capacity-achieving schemes to the case of no eavesdropper are different from those in [8].

We then extend our results for the case of parallel channels to the case of fast-fading channels, emphasizing Rayleigh fading. In our problem formulation, we assume that the channel state information (CSI) for all legitimate receivers is revealed to all communicating parties—including the eavesdropper—while only the eavesdropper knows its own CSI.

Again, we consider both common and independent message transmission over such fading channels. For the common message case, we describe a scheme that achieves a nonvanishing rate in the limit of many legitimate receivers. In our construction, transmitter CSI is required and plays an important role. By

Manuscript received November 27, 2006; revised February 26, 2008. This work was supported in part by the National Science, Foundation under Grant CCF-0515109, Draper Laboratory, the Natural, Sciences Engineering Research Council of Canada, and a Hewlett-Packard Graduate Fellowship. The material in this paper was presented in part at the 44th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, September 2006.

The authors are with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: khisti@mit.edu; tcham@mit.edu; gww@mit.edu).

Communicated by U. Maurer, Guest Editor for Special Issue on Information Theoretic Security.

Color versions Figures 2 and 5 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2008.921861

0018-9448/$25.00 © 2008 IEEE

contrast, when there is no secrecy constraint, transmitter CSI has a more limited impact on the multicasting rate over ergodic channels. Indeed, the regular (nonsecrecy) capacity appears to be not too far from the maximum rate achievable using schemes with a nonadaptive (flat) power allocation.

For the case of independent messages, we develop an opportunistic scheme that selects the receiver with the strongest channel at each time. With Gaussian wiretap codebooks for each legitimate receiver, we show that this scheme achieves the sum-capacity in the limit of large number of receivers. Our results can be interpreted as the wiretap analog of the multiuser diversity results in settings without secrecy constraint (see, e.g., [24]).

The paper is organized as follows. Section II provides some notation for the paper. Section III formally describes the channel models of interest. The main results are summarized in Section IV. Details of the analysis of the scenario of a common message are presented in Sections V and VII for the cases of parallel and fading channels, respectively. In turn, Sections VI and VIII provide the analysis for the scenario of independent messages for the cases of parallel and fading channels, respectively. Finally, Section IX contains some concluding remarks.

II. Notation

A summary of some notation used in the paper is as follows. First, upper case letters are used for random variables and the lower case for their realizations. Also, sequences are denoted using superscripts and sequence elements with parentheses; e.g., $s^n = (s(1), s(2), \dots, s(n))$.

The entropy of a discrete random variable X is denoted by $H(X)$, and the mutual information between random variables X and Y is denoted by $I(X;Y)$. Following this convention, $p(X)$ denotes the probability mass function of random variable X. In addition, we use $E[\cdot]$ to denote expectation, and, when not clear from context, we use a subscript to indicate the distribution with respect to which the expectation is being taken; e.g., $E_V[\cdot]$ denotes expectation with respect to the distribution for V.

We also use $\mathcal{CN}(0, \sigma^2)$ to denote the distribution of a circularly symmetric complex-valued Gaussian random variable with zero-mean and variance σ^2, and define $\{v\}^+ \triangleq \max\{0, v\}$ for any v. Finally, we use "bar" notation (e.g., $\bar{R}$ and $\bar{C}$) to denote rates associated with common message transmission, to distinguish them from (sum) rates for transmission of independent messages (e.g., R and C).

III. Problem and Channel Models

In this section, we formally define the problem and broadcast channel models of interest.

A. Problem Model

We formulate the problems of interest as extensions of the wiretap channel model introduced by Wyner [25] for studying reliable and secure communication in an information-theoretic framework. As such, we emphasize that in our models there is no prior key shared between the sender and legitimate receivers, and both the encoding and decoding functions, and the codebook itself, are public.

Within this framework, we emphasize Wyner's notion of secrecy capacity, which is the maximum rate of reliable communication to the intended receivers subject to the constraint of vanishing mutual information at the eavesdropper. Moreover, we adopt Wyner's definition of "perfect secrecy" as the scenario in which the block-length-normalized mutual information at the eavesdropper vanishes in the limit of long block lengths, which is sufficient for a variety of applications. However, we note that this is significantly weaker than both the notion considered by Shannon [21], which requires that the mutual information be zero regardless of the block length, and the notion by Maurer and Wolf [19] which requires that the (unnormalized) mutual information approach zero with the block length. In our concluding remarks, we comment further on such issues.

Finally, we restrict our attention to the secrecy capacity in this paper, rather than the entire rate-equivocation region described by Wyner. This is because in the motivating key-distribution application of interest, the key length is limited by the equivocation rate, which is effectively the minimum number of bits the eavesdropper needs to correctly guess to decode the message. Accordingly, the secrecy capacity is of primary interest.

B. Parallel Channels

In this broadcast model, there are M parallel subchannels connecting a single sender to each of K legitimate receivers and an eavesdropper, where M and K are parameters.

Definition 1: A *product* broadcast channel is one in which the constituent subchannels have finite input and output alphabets, are memoryless and independent of each other, and are characterized by their transition probabilities

$$\Pr\left(\{y_{1m}^n, \dots, y_{Km}^n, y_{\mathrm{e}m}^n\}_{m=1,\dots,M} \mid \{x_m^n\}_{m=1,\dots,M}\right) = \prod_{m=1}^{M}\prod_{t=1}^{n} \Pr(y_{1m}(t), \dots, y_{Km}(t), y_{\mathrm{e}m}(t) \mid x_m(t)) \quad (1)$$

where $x_m^n = (x_m(1), x_m(2), \dots, x_m(n))$ denotes the sequence of symbols transmitted on subchannel m, where $y_{km}^n = (y_{km}(1), y_{km}(2), \dots, y_{km}(n))$ denotes the sequence of symbols obtained by receiver k on subchannel m, and where $y_{\mathrm{e}m}^n = (y_{\mathrm{e}m}(1), y_{\mathrm{e}m}(2), \dots, y_{\mathrm{e}m}(n))$ denotes the sequence of symbols received by the eavesdropper on subchannel m. The alphabet of X_m is $\mathcal{X}$, and the alphabet for both Y_{km} and $Y_{\mathrm{e}m}$ is $\mathcal{Y}$.

A special class of product broadcast channels, known as the reversely degraded broadcast channel [8] are of particular interest.

Definition 2: A product broadcast channel is *reversely degraded* when each of the M constituent subchannels is degraded in a prescribed order. In particular, for each subchannel m, there exists a permutation $\{\pi_m(1), \pi_m(2), \dots, \pi_m(K+1)\}$ of the set $\{1, 2, \dots, K, \mathrm{e}\}$ such that the following Markov chain is satisfied, i.e.,

$$X_m \to Y_{\pi_m(1)} \to Y_{\pi_m(2)} \to \cdots \to Y_{\pi_m(K+1)}.$$

With this definition, $Y_{\pi_m(1)}, Y_{\pi_m(2)}, \dots, Y_{\pi_m(K+1)}$ is an ordering of the receivers from strongest to weakest in the mth subchannel, and we will at times find it convenient to adopt the ad-

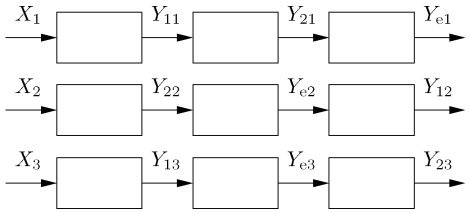


Fig. 1. An example of reversely degraded parallel broadcast channel, in which there are $M = 3$ subchannels connecting a single sender to each of $K = 2$ legitimate receivers and an eavesdropper. The input symbols to the subchannels are (X_1, X_2, X_3). The output symbols at the kth intended receiver are (Y_{k1}, Y_{k2}, Y_{k3}), and at the eavesdropper are $(Y_{\mathrm{e}1}, Y_{\mathrm{e}2}, Y_{\mathrm{e}3})$. Note that the order of degradation is not the same for all subchannels.

ditional notation $\pi_m \triangleq \pi_m(1)$. Also, we stress that in Definition 2 the order of degradation need not be the same for all subchannels, so the overall channel need not be degraded. An example of reversely degraded parallel broadcast channel is depicted in Fig. 1.

We also emphasize that in any subchannel the K receivers and eavesdropper are *physically* degraded. Our capacity results, however, only depend on the marginal distribution of receivers in each subchannel. Accordingly, our results in fact hold for the larger class of channels in which there is only stochastic degradation in the subchannels.

Finally, we obtain further results when the channel is Gaussian.

Definition 3: A reversely degraded product broadcast channel is *Gaussian* when it takes the form

$$\begin{aligned} Y_{km} &= X_m + Z_{km}, \quad m = 1, \dots, M, \quad k = 1, \dots, K \\ Y_{\mathrm{e}m} &= X_m + Z_{\mathrm{e}m} \end{aligned} \tag{2}$$

where the noise variables are all mutually independent, and $Z_{km} \sim \mathcal{CN}(0, \sigma^2_{km})$ and $Z_{\mathrm{e}m} \sim \mathcal{CN}(0, \sigma^2_{\mathrm{e}m})$. For this channel, there is also an average power constraint

$$E\left[\sum_{m=1}^{M} |X_m|^2\right] \le P.$$

We now provide the formal definitions of the common-message secrecy capacity and the sum-secrecy capacity for independent messages.

Definition 4: An $(n, 2^{nR})$ code consists of a message set $\mathcal{W} = \{1, 2, \dots, 2^{nR}\}$, a (possibly stochastic) encoder

$$\omega_n : \mathcal{W} \to \mathcal{X}^n \times \mathcal{X}^n \times \overset{(M\text{-fold})}{\cdots\cdots\cdots} \times \mathcal{X}^n$$

mapping the message set to the codewords for the M subchannels, and a decoder

$$\Phi_{k,n} : \mathcal{Y}^n \times \mathcal{Y}^n \times \overset{(M\text{-fold})}{\cdots\cdots\cdots} \times \mathcal{Y}^n \to \mathcal{W}$$

for $k = 1, 2, \dots, K$ at each receiver. Using $\hat{W}_k$ to denote message estimate at decoder k, a common-message-secrecy-rate R is said to be achievable if, for any $\epsilon > 0$, there exists a length n code such that $\Pr(W \neq \hat{W}_k) \le \epsilon$ for $k = 1, 2, \dots, K$, while

$$\frac{1}{n} H(W | Y^n_{\mathrm{e}1}, Y^n_{\mathrm{e}2}, \dots, Y^n_{\mathrm{e}K}) \ge R - \epsilon. \tag{3}$$

The common-message secrecy capacity is the supremum over all achievable rates.

Definition 5: A $(2^{nR_1}, 2^{nR_2}, \dots, 2^{nR_K}, n)$ code for the product broadcast channel in Definition 1 consists of a message set $\mathcal{W} = \{1, 2, \dots 2^{nR}\}$, an encoder

$$\omega_n : \mathcal{W}_1 \times \mathcal{W}_2 \times \cdots \times \mathcal{W}_K \to \mathcal{X}^n \times \mathcal{X}^n \overset{(M\text{-fold})}{\cdots\cdots\cdots} \times \mathcal{X}^n$$

mapping the messages for the K receivers to the M subchannel inputs, and K decoding functions

$$\phi_{k,n} : \mathcal{Y}^n \times \mathcal{Y}^n \times \overset{(M\text{-fold})}{\cdots\cdots\cdots} \times \mathcal{Y}^n \to \mathcal{W}_k$$

one at each legitimate receiver. We denote the message estimate at decoder k by $\hat{W}_k$. A secrecy rate-tuple $(R_1, R_2, \dots, R_K)$ is achievable if, for every $\epsilon > 0$, there is a code of length n such that $\Pr(W_k \neq \hat{W}_k) \le \epsilon$ for all $k = 1, 2, \dots, K$, and such that

$$\begin{aligned} \frac{1}{n} H(W_k | W_1, \dots, W_{k-1}, W_{k+1}, \dots, W_K, Y^n_{\mathrm{e}1}, \dots, Y^n_{\mathrm{e}M}) \\ \ge \frac{1}{n} H(W_k) - \epsilon, \qquad k = 1, 2, \dots, K \end{aligned} \tag{4}$$

with W_k uniformly distributed in $\{1, 2, \dots, 2^{nR_k}\}$. The secrecy sum-capacity is the supremum of $R_1 + R_2 + \cdots + R_K$ over the achievable rate tuples $(R_1, R_2, \dots, R_K)$.

We remark that our constraint (4) provides perfect equivocation for each message, even if all the other messages are revealed to the eavesdropper.

C. Fading Channels

Definition 6: Our fast-fading broadcast model of interest has the following properties. The received sequences $Y_1^n, Y_2^n, \dots, Y_K^n$ and Y_{e}^n at the legitimate receivers and eavesdropper, respectively, are of the form

$$\begin{aligned} Y_k(t) &= H_k(t)\, X(t) + Z_k(t), \qquad k = 1, 2, \dots, K \\ Y_{\mathrm{e}}(t) &= H_{\mathrm{e}}(t)\, X(t) + Z_{\mathrm{e}}(t) \end{aligned} \tag{5}$$

where X^n is the transmitted sequence, and $Z_k(t) \sim \mathcal{CN}(0, 1)$. The channel gains and noises among all receivers (including the eavesdropper) are all mutually independent of one another, and all vary in an independent and identically distributed (i.i.d.) manner with time, corresponding to fast fading.[1] Finally, the input must satisfy an average power constraint $E[|X(t)|^2] \le P$.

In parts of our development, we explicitly restrict our attention to the special case of Definition 6 corresponding to Rayleigh fading, in which case $H_k(t) \sim \mathcal{CN}(0, \mu_k)$ and $H_{\mathrm{e}}(t) \sim \mathcal{CN}(0, \mu_{\mathrm{e}})$ as well.

In addition, in our model the $H_1(t), \dots, H_K(t)$ are revealed to the transmitter, the K legitimate receivers and the eavesdropper in a causal manner. Implicitly we assume that there is an authenticated public feedback link from the receivers to the transmitter. The channel coefficients of the eavesdropper $\{H_{\mathrm{e}}^n\}$ are known only to the eavesdropper, but the transmitter and the legitimate receivers know the probability distribution of the eavesdropper's channel gains.

[1] In practice, the fast fading model (5) applies when the codebooks are interleaved such that each symbol sees an independent fade.

Note that for such channels, the transmitter must exploit the CSI of legitimate receivers. Indeed, any scheme that does not would reveal the message to any eavesdropper that has a channel statistically equivalent to the intended receiver(s).

We now provide the formal definitions of the common-message secrecy capacity and the sum-secrecy capacity for independent messages.

Definition 7: An $(n, 2^{nR})$ code for the channel consists of an encoding function that maps from the message $w \in \{1, 2, \dots, 2^{nR}\}$ into transmitted symbols

$$x(t) = f_t(w; h_1^t, h_2^t, \dots, h_K^t), \qquad \text{for } t = 1, 2, \dots, n$$

and a decoding function $\hat{w}_k = \phi_k(y_k^n; h_1^n, h_2^n, \dots, h_K^n)$ at each receiver k. A rate R is achievable if, for every $\epsilon > 0$, there exists a sequence of length n codes such that $\Pr(\hat{W}_k \neq W) \le \epsilon$ for any $k = 1, 2, \dots, K$ such that

$$\frac{1}{n} H\big(W \mid Y_{\rm e}^n, H_{\rm e}^n, H_1^n, \dots, H_K^n\big) \ge R - \epsilon. \tag{6}$$

Definition 8: An $(n, 2^{nR_1}, \dots, 2^{nR_K})$ code consists of an encoding function from the messages $w_1, \dots, w_K$ with $w_k \in \{1, 2, \dots, 2^{nR_k}\}$ to transmitted symbols

$$x(t) = f_t(w_1, w_2, \dots, w_K; h_1^t, h_2^t, \dots, h_K^t), \text{ for } t = 1, 2, \dots, n$$

and a decoding function $\hat{w}_k = \phi_k(y_k^n; h_1^n, h_2^n, \dots, h_K^n)$ at each receiver. A secrecy rate-tuple $(R_1, R_2, \dots, R_K)$ is achievable if, for any $\epsilon > 0$, there exists a length n code such that, for each $k = 1, 2, \dots, K$, with W_k uniformly distributed over $\{1, 2, \dots, 2^{nR_k}\}$, we have $\Pr(\hat{W}_k \neq W_k) \le \epsilon$ and

$$\frac{1}{n} H\big(W_k \mid W_1, \dots, W_{k-1}, W_{k+1}, \dots \\ \dots, W_K, Y_{\rm e}^n, H_{\rm e}^n, H_1^n, \dots, H_K^n\big) \\ \ge R_k - \epsilon. \tag{7}$$

The secrecy sum-capacity is the supremum value of $R_1 + R_2 + \dots + R_K$ among all achievable rate tuples.

Note that the entropy term in both (6) and (7) is conditioned on $H_1^n, \dots, H_K^n$ as these channel gains of the K receivers are assumed to be known to the eavesdropper. However, the encoding and decoding functions do not depend on $h_{\rm e}^n$ as this realization is not known to the sender and the receivers.

An immediate consequence of this formulation is that the secrecy capacity depends only on the distribution of $H_{\rm e}(t)$ and not on the actual realized sequence of these eavesdropper gains. Indeed, since the transmitter and the legitimate receivers do not have the eavesdropper's CSI, the encoding and decoding functions cannot depend on this information. From this perspective, in our formulation a message that is secure with respect to any given eavesdropper is also secure against any statistically equivalent eavesdropper.

IV. Main Results

In this section, we summarize our results on the secrecy capacity of broadcast channels. The detailed development of these results is provided in subsequent sections.

A. Parallel Channels and a Common Message

We have the following upper and lower bounds on the common-message secrecy capacity for the product broadcast channel of Definition 1.

Proposition 1: For the product broadcast channel model, an upper bound on the secrecy capacity is given by

$$\bar{C}_{K,M} \le \\ \bar{R}^{+}_{K,M} \triangleq \min_{\mathcal{P}} \max_{\prod_{m=1}^{M} p(X_m)} \min_{k \in \{1,\dots,K\}} \sum_{m=1}^{M} I(X_m; Y_{km} | Y_{{\rm e}m}) \tag{8}$$

where the set $\mathcal{P} = \mathcal{P}_1 \times \dots \times \mathcal{P}_M$ is Cartesian product of the sets $\{\mathcal{P}_m\}_{m=1}^{M}$, and where each $\mathcal{P}_m$ is the collection of all joint distributions $p'(Y_{1m}, \dots, Y_{Km}, Y_{{\rm e}m} | X_m)$ having the same marginal distribution as $p(Y_{1m}|X_m), \dots, p(Y_{Km}|X_m)$ and $p(Y_{{\rm e}m}|X_m)$, and where the maximum is over all marginal distributions $p(X_1), \dots, p(X_M)$.

Proposition 2: For the product broadcast channel model, an achievable lower bound on the secrecy capacity is given by

$$\bar{C}_{K,M} \ge \bar{R}^{-}_{K,M} = \\ \max_{\substack{\{p(U_m)\}_{m=1}^{M} \\ \{X_m = f_m(U_m)\}_{m=1}^{M}}} \min_{k \in \{1,..,K\}} \sum_{m=1}^{M} \{I(U_m; Y_{km}) - I(U_m; Y_{{\rm e}m})\}^{+} \tag{9}$$

where the random variables $U_1, \dots, U_M$ are independent over some alphabet $\mathcal{U}$, and each $f_m(\cdot)$ for $m = 1, \dots, M$ is a mapping from $\mathcal{U}$ to $\mathcal{X}$.

For the special case of a product broadcast channel that is reversely degraded, our upper and lower bounds above coincide, yielding the following common-message secrecy capacity.

Theorem 1: The common-message secrecy capacity for the reversely degraded channel model is

$$\bar{C}_{K,M} = \max_{\prod_{m=1}^{M} p(X_m)} \min_{k \in \{1,2,\dots,K\}} \sum_{m=1}^{M} I(X_m; Y_{km} | Y_{{\rm e}m}). \tag{10}$$

We remark that [8] considers the problem of broadcasting common and independent messages over reversely degraded channels, but without a secrecy constraint. It is worth noting that the coding scheme we construct that achieves the secrecy capacity (10), when specialized to the case of no eavesdropper, yields a different capacity-achieving scheme than that of [8]. Moreover, an obvious random binning extension of the scheme presented in [8] does not achieve the secrecy capacity (10).

Finally, for the Gaussian parallel-channel model of Definition 3, we have the following straightforward extension of Theorem 1.

Corollary 1: The common-message secrecy capacity for the Gaussian parallel broadcast channel is

$$\bar{C}^{\rm G}_{K,M} \\ = \max_{(P_1,\dots,P_M) \in \mathcal{F}} \min_{1 \le k \le K} \sum_{m=1}^{M} \left\{ \log \left(\frac{1 + P_m/\sigma_{km}^2}{1 + P_m/\sigma_{{\rm e}m}^2} \right) \right\}^{+} \tag{11}$$

where $\mathcal{F}$ is the set of all feasible power allocations, i.e.,

$$\mathcal{F} = \left\{ (P_1, \dots, P_M) \,\middle|\, P_m \ge 0, \sum_{m=1}^{M} P_m \le P \right\}. \tag{12}$$

B. Parallel Channels and Independent Messages

In absence of the secrecy constraint, the sum-capacity for the reversely degraded broadcast channel is maximized when only the strongest receiver on each parallel channel is served [23]. We show that the same scheme is also optimal with the secrecy constraint. In particular, we establish the following result.

Theorem 2: The secrecy sum-capacity for the reversely degraded product broadcast channel is

$$C_{K,M} = \max_{\prod_{m=1}^{M} p(X_m)} \sum_{m=1}^{M} I(X_m; Y_{\pi_m} | Y_{\mathrm{e}m}) \tag{13}$$

where π_m denotes the index of the strongest receiver on channel m. Furthermore, the right-hand side of (13) is an upper bound on the secrecy sum-capacity when the set of legitimate receivers are reversely degraded, but the set of these receivers taken together with the eavesdropper are collectively not reversely degraded.

Finally, for the Gaussian parallel-channel model of Definition 3, we have the following straightforward extension of Theorem 2.

Corollary 2: The secrecy sum-capacity for the Gaussian parallel broadcast channel is

$$C_{K,M}^{\mathrm{G}}(P) = \max_{(P_1,\ldots,P_M)\in\mathcal{F}} \sum_{m=1}^{M} \log\left(\frac{1+P_m/\sigma_{\pi_m}^2}{1+P_m/\sigma_{\mathrm{e}m}^2}\right) \tag{14}$$

where the feasible set of power distributions as defined in (12), and where $\sigma_{\pi_m}^2$ denotes the variance of the noise of the strongest receiver on subchannel m.

C. Fading Channels and a Common Message

Several recent works [2], [11], [12], [15] have observed that secure communication is possible over fading channels even when the eavesdropper's channel is on an average stronger than a legitimate receiver's channel. This is accomplished by adapting the rate and transmit power to the channel of the intended receiver.

We develop additional insight into the robustness of such schemes by considering the case when a common message has to be delivered to multiple receivers, while keeping it secret from potential eavesdroppers. The common message constraint requires us to adapt rate and power to the channel gains of several legitimate receivers simultaneously. Despite such a stringent requirement, we demonstrate that it is possible to broadcast at a strictly positive rate independent of the number of legitimate receivers. In particular, we have the following theorem.

Theorem 3: The common-message secrecy rate for the fast-fading broadcast channel is bounded by

$$\bar{R}^-(P) \le \bar{C}_K(P) \le \bar{R}^+(P) \tag{15}$$

where

$$\bar{R}^-(P) = \min_{1\le k\le K} E_{H_k}\left[\left\{\log\left(\frac{1+|H_k|^2P}{\exp\{E_{H_\mathrm{e}}[\log(1+|H_\mathrm{e}|^2P)]\}}\right)\right\}^+\right] \tag{16a}$$

and

$$\bar{R}^+(P) = \min_{1\le k\le K} \max_{\substack{\rho(H_k):\\ E[\rho(H_k)]\le P}} E\left[\left\{\log\left(\frac{1+|H_k|^2\rho(H_k)}{1+|H_\mathrm{e}|^2\rho(H_k)}\right)\right\}^+\right]. \tag{16b}$$

When the channel gains H_k are identically distributed across the users, note that both lower and upper bounds in (16) are independent of the number of receivers K. The fact that the common-message secrecy capacity does not vanish with the number of users is surprising. Simple schemes such as transmitting when all the users have a channel gain above a threshold or time-sharing between the users only achieve a rate that vanishes with the number of users. In contrast, our lower bound is achieved by a scheme that simultaneously adapts to the time variations of all the legitimate users.

In the high signal-to-noise ratio (SNR) regime, the bounds Theorem 3 specialize as follows.

Corollary 3: When the channel gains of all the receivers are distributed as $\mathcal{CN}(0,1)$, the bounds in (16) are asymptotically

$$\lim_{P\to\infty} \bar{R}^+(P) = E\left[\left\{\log\frac{|H|^2}{|H_\mathrm{e}|^2}\right\}^+\right] \tag{17a}$$

$$\lim_{P\to\infty} \bar{R}^-(P) = E\left[\left\{\log|H|^2 + \frac{\gamma}{\log 2}\right\}^+\right] \tag{17b}$$

where γ is the Euler-Gamma constant ($\gamma \approx 0.5772$).

Evaluating (17) at high SNR, when $E[|H_k|^2] = 1$, gives

$$0.709 \le \lim_{P\to\infty} \bar{C}_K(P) \le 1 \quad \text{in b/s/Hz.} \tag{18}$$

We remark that since this scheme achieves a rate independent of the number of receivers, it achieves the best possible scaling with the number of receivers. However, it is not known whether the scheme is capacity achieving. Indeed, even for the special case corresponding to a single legitimate receiver ($K=1$), the fast-fading secrecy capacity is not yet known [12], [16].

D. Fading Channels and Independent Messages

The problem of broadcasting independent messages to multiple receivers over ergodic fading channels has been well studied when there is no security constraint; see, e.g., [14], [23]. For such scenarios, an opportunistic transmission scheme is shown to attain the largest sum-capacity. We establish the following analogous result for secure transmission.

Proposition 3: For the fast-fading broadcast channel, the secrecy sum-capacity is bounded by

$$R_K^-(P) \le C_K(P) \le R_K^+(P) \tag{19}$$

where

$$R^+(P) = \max_{\substack{\rho(H_{\max}):\\ E[\rho(H_{\max})]\le P}} E\left[\left\{\log\left(\frac{1+|H_{\max}|^2\rho(H_{\max})}{1+|H_\mathrm{e}|^2\rho(H_{\max})}\right)\right\}^+\right] \tag{20a}$$

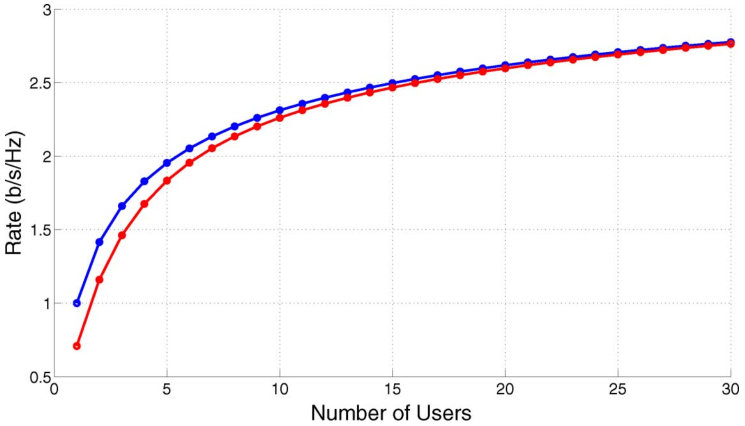


Fig. 2. Upper and lower bounds on the secrecy sum-capacity in (20) for the broadcasting of independent messages in Rayleigh fast-fading environments in the high-SNR regime, as a function of the number of legitimate receivers.

and

$$R_K^-(P) = \max_{\substack{\rho(H_{\max}): \\ E[\rho(H_{\max})] \le P}} E\left[\log\left(\frac{1+|H_{\max}|^2\rho(H_{\max})}{1+|H_{\mathrm{e}}|^2\rho(H_{\max})}\right)\right] \tag{20b}$$

with $H_{\max}$ denoting the gain of the strongest of the K legitimate receivers (at any instant).

Our upper and lower bounds in (20) are distinguished by the inclusion of the operator $\{\cdot\}^+$ is inside the expectation of the former. Hence, the arguments of the expectation differ whenever $|H_{\max}|^2 \le |H_{\mathrm{e}}|^2$, and so an upper bound on the rate gap is

$$\begin{aligned} R_K^+(P) &- R_K^-(P) \\ &\le \Pr(|H_{\mathrm{e}}|^2 \ge |H_{\max}|^2) \\ &\quad \cdot E\left[\log\left(|H_{\mathrm{e}}|^2/|H_{\max}|^2\right) \mid |H_{\mathrm{e}}|^2 \ge |H_{\max}|^2\right]. \end{aligned} \tag{21}$$

As the number of legitimate receivers grows, the event $\{|H_{\max}|^2 \le |H_{\mathrm{e}}|^2\}$ happens increasingly rarely and for the case of identical Rayleigh distributed fading, the gap between the bounds vanishes. As a result, we obtain the following theorem.

Theorem 4: For the fast-fading broadcast channel with identical Rayleigh-distributed fading and large K, the secrecy capacity scales according to

$$C_K(P) = \max_{\substack{\rho(H_{\max}): \\ E[\rho(H_{\max})] \le P}} E\left[\log\left(\frac{1+|H_{\max}|^2\rho(H_{\max})}{1+|H_{\mathrm{e}}|^2\rho(H_{\max})}\right)\right] + o(1). \tag{22}$$

where we use $o(1)$ to denote terms that approach zero as $K \to \infty$.

Theorem 4 establishes that an architecture that uses single-user Gaussian wiretap base codes in conjunction with opportunistic transmission achieves the secrecy sum-capacity in the limit of a large number of receivers.

For finite values of K, incorporating synthesized noise into the transmission as a masking technique yields still higher rates [12], [16]. However, even with such refinements, there remains a gap between the upper and lower bounds. Fig. 2 illustrates the upper and lower bounds in (20) in the high-SNR regime for identically distributed Rayleigh-fading distribution. We note that even for a moderate number of users, these bounds are close and further improvements will only provide diminishing gains in this regime.

We also remark that Theorem 4 more generally guarantees an arbitrarily small gap between upper and lower bounds on the secrecy sum-capacity for Rayleigh-fading channels of fixed coherence time, provided the number of receivers is large enough.

In [11] *variable-rate* and *fixed-rate* schemes are developed for the case of a single receiver in a slow fading environment. Straightforward extensions of these schemes for multiple receivers reveals the following insights. The variable-rate scheme achieves our upper bound (20a), whereas the fixed-rate scheme achieves our lower bound (20b). Since these two expressions coincide as the number of receivers tends to infinity, it follows that the gains of variable-rate schemes become negligible in this limit.

As a final remark, we comment on collusion attacks. As noted earlier, any number of statistically equivalent eavesdroppers does not affect our capacity—as long as they do not collude. However, if the eavesdroppers collude, they can combine the received signals and attempt to decode the message. In such scenarios, the upper and lower bounds in Proposition 3 can be extended by replacing the term $|H_{\mathrm{e}}|^2$ with $||\boldsymbol{H}_{\mathrm{e}}||^2$, where $\boldsymbol{H}_{\mathrm{e}}$ is the vector of channel gains of the colluding eavesdroppers. One interesting implication of the resulting bounds is that the secrecy capacity is positive unless the colluding eavesdropper population grows as $\log K$.

V. Parallel Channels and a Common Message

In this section, we establish our results concerning the transmission of a common message over parallel channels. In particular, we prove Propositions 1 and 2 and Theorem 1 and Corollary 1 stated in Section IV-A.

A. Upper Bound on Capacity

Proof of Proposition 1:

Suppose there exists a sequence of $(n, 2^{nR})$ codes such that, for every $\epsilon > 0$ and sufficiently large n, we have that for all $k = 1, 2, \ldots, K$

$$\Pr(W \neq \hat{W}_k) \leq \epsilon$$
$$\frac{1}{n} I(W; Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) \leq \epsilon. \tag{23}$$

We first note that from Fano's inequality we have

$$\frac{1}{n} H(W | Y^n_{k1}, Y^n_{k2}, \ldots, Y^n_{kM}) \leq \frac{1}{n} + \epsilon R. \tag{24}$$

Combining (23) and (24) we have, for all $k = 1, 2, \ldots, K$ and $\epsilon' = \epsilon + \frac{1}{n} + \epsilon R$

$$\begin{aligned}
nR &\leq I(W; Y^n_{k1}, \ldots, Y^n_{kM}) - I(W; Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) + n\epsilon' \\
&\leq I(W; Y^n_{k1}, \ldots, Y^n_{kM} | Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) + n\epsilon' \\
&= h(Y^n_{k1}, \ldots, Y^n_{kM} | Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) \\
&\quad - h(Y^n_{k1}, \ldots, Y^n_{kM} | Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}, W) + n\epsilon' \\
&\leq h(Y^n_{k1}, \ldots, Y^n_{kM} | Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) \\
&\quad - h(Y^n_{k1}, \ldots, Y^n_{kM} | Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}, X^n_1, \ldots, X^n_M, W) + n\epsilon' \\
&= h(Y^n_{k1}, \ldots, Y^n_{kM} | Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) \\
&\quad - h(Y^n_{k1}, \ldots, Y^n_{kM} | Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}, X^n_1, \ldots, X^n_M) + n\epsilon' \qquad (25) \\
&= h(Y^n_{k1}, \ldots, Y^n_{kM} | Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) \\
&\quad - \sum_{m=1}^{M} h(Y^n_{km} | X^n_m, Y^n_{em}) + n\epsilon' \qquad (26) \\
&\leq \sum_{m=1}^{M} h(Y^n_{km} | Y^n_{em}) - \sum_{m=1}^{M} h(Y^n_{km} | X^n_m, Y^n_{em}) + n\epsilon' \\
&\leq \sum_{m=1}^{M} I(X^n_m; Y^n_{km} | Y^n_{em}) + n\epsilon' \qquad (27)
\end{aligned}$$

where (25) follows from the fact that

$$W \rightarrow (X^n_1, \ldots, X^n_M, Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) \rightarrow (Y^n_{k1}, \ldots, Y^n_{kM})$$

forms a Markov chain, and (26) holds because the parallel subchannels in Definition 1 are mutually independent so that

$$h(Y^n_{k1}, \ldots, Y^n_{kM} | Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}, X^n_1, \ldots, X^n_M) = \sum_{m=1}^{M} h(Y^n_{km} | X^n_m, Y^n_{em}). \tag{28}$$

We now upper-bound each term in the summation (27). We have

$$\begin{aligned}
&I(X^n_m; Y^n_{km} | Y^n_{em}) \\
&\leq \sum_{k=1}^{n} I(X_m(k); Y_{km}(k) | Y_{em}(k)) \qquad (29) \\
&= \sum_{k=1}^{n} I(X_m(k); Y_{km}(k), Y_{em}(k)) - I(X_m(k); Y_{em}(k)) \\
&= nI(X_m; Y_{km}, Y_{\mathrm{em}} | Q) - nI(X_m; Y_{\mathrm{em}} | Q) \qquad (30) \\
&= nI(X_m; Y_{km} | Y_{\mathrm{em}}, Q) \\
&\leq nI(X_m; Y_{km} | Y_{\mathrm{em}}) \qquad (31)
\end{aligned}$$

where (29) follows from the fact that the channel is memoryless, and (30) is obtained by defining Q to be a (time-sharing) random variable uniformly distributed over $\{1, 2, \ldots, n\}$ independent of all other variables. The random variables $(X_m, Y_{km}, Y_{\mathrm{em}})$ are such that, conditioned on $Q = q$, they have the same joint distribution as $(X_m(q), Y_{km}(q), Y_{\mathrm{em}}(q))$. Finally, (31) follows from the fact that the mutual information is concave with respect to the input distribution $p(X_m)$, a property that is verified in Appendix A.

Combining (31) and (27) we have

$$\begin{aligned}
R &\leq \sum_{m=1}^{M} I(X_m; Y_{km} | Y_{\mathrm{em}}) + \epsilon', \qquad k = 1, 2, \ldots, K \\
&= \min_{1 \leq k \leq K} \sum_{m=1}^{M} I(X_m; Y_{km} | Y_{\mathrm{em}}) + \epsilon' \\
&\leq \max_{\prod_{m=1}^{M} p(X_m)} \min_{1 \leq k \leq K} \sum_{m=1}^{M} I(X_m; Y_{km} | Y_{\mathrm{em}}) + \epsilon' \qquad (32)
\end{aligned}$$

where the last step follows from that fact that for any input distribution $p(X_1, X_2, \ldots, X_M)$, the objective function

$$\min_{1 \leq k \leq K} \sum_{m=1}^{M} I(X_m; Y_{km} | Y_{\mathrm{em}})$$

only depends on the marginal distributions $p(X_1), \ldots, p(X_M)$. Finally, note that (32) depends on the joint distribution across the subchannels while the secrecy capacity only depends on the marginal distribution. Accordingly, we tighten the upper bound by considering the worst distribution in $\mathcal{P} = \mathcal{P}_1 \times \mathcal{P}_2 \times \cdots \times \mathcal{P}_M$, yielding (8). □

B. Lower Bound on Capacity

We now present a coding scheme that achieves the our lower bound.

We first discuss the structure of the coding scheme informally. We construct M independent random codebooks $\mathcal{C}_1, \ldots, \mathcal{C}_M$, one for each subchannel. Codebook $\mathcal{C}_m$ has nearly $2^{n(R + I(U_m; Y_{\mathrm{em}}))}$ codewords, randomly partitioned into 2^{nR} bins, one for each possible message. Hence, there are nearly $Q_m = 2^{nI(U_m; Y_{\mathrm{em}})}$ codewords per bin. Given a particular message $W \in \{1, 2, \ldots, 2^{nR}\}$ to be sent, the encoder selects M codewords, one for each subchannel. Specifically, if the message is w, then for each subchannel m the encoder randomly selects for transmission one of the Q_m codewords from the wth bin in $\mathcal{C}_m$. This bin structure of the codebooks is depicted in Fig. 3 for the case of $M = 2$ subchannels.

To decode, each legitimate receiver attempts to find a message that is jointly typical with its set of M received sequences. As we now show, the rate R of the code can be chosen arbitrarily close to $\bar{R}^{-}_{K,M}$ as defined in (9) and guarantees both successful decoding with high probability for each legitimate receiver, and near-perfect equivocation at the eavesdropper.

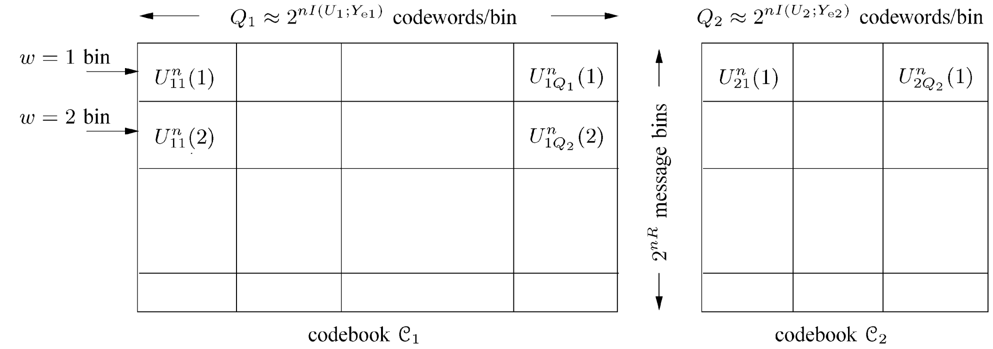


Fig. 3. Binning encoder for the secure product broadcast channel, for the case of $M = 2$ subchannels. The set of codewords for representing a particular message $w \in \{1, \ldots, 2^{nR}\}$ in the mth subchannel are denoted by $U_{m1}^n(w), \ldots, U_{mQ_m}^n(w)$. To encode a particular message w, the encoder randomly selects one of the Q_m codewords in the associated bin for transmission in the mth subchannel, for $m = 1, \ldots, M$.

Before presenting our proof, we make some remarks. As mentioned earlier, when specialized to the case in which there is no eavesdropper (and hence no secrecy constraint), our construction is different from that developed by El Gamal [8] for such product broadcast channels. In particular, as illustrated in Fig. 4 for the case of $M = 3$ subchannels, our construction has the distinguishing feature that independent codebooks are used for the different subchannels. By comparison, with the scheme in [8], each message is mapped to a $M \times n$-dimensional codeword and the mth component of the codeword is transmitted on subchannel m. This corresponds to a single-codebook scheme. By extending this scheme to provide secrecy by incorporating random binning, one can achieve, again for the reversely degraded channel,

$$R^{\text{single}} = \max_{p(X_1,\ldots,X_M)} \min_{k \in \{1,\ldots,K\}} \big\{ I(X_1, X_2 \ldots, X_K; Y_{k1}, \ldots, Y_{kK}) - I(X_1, X_2 \ldots, X_K; Y_{e1}, \ldots, Y_{eK}) \big\} \quad (33)$$

which we observe is in general smaller than that achieved by our construction, viz., (10). Ultimately, allowing the sizes of bins to depend on the mutual information at the eavesdropper on each particular subchannel makes it possible to confuse the eavesdropper on each subchannel, and thereby achieve higher secrecy rates than (33).

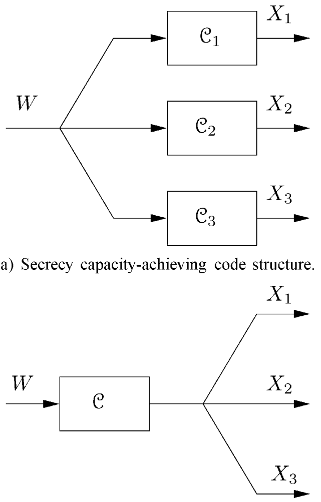
We now provide the formal details and analysis of the coding scheme.

Proof of Proposition 2: First, fix the distributions $p(U_1), p(U_2), \ldots, p(U_M)$ and the (possibly stochastic) functions $f_1(\cdot), \ldots, f_M(\cdot)$. Let η_2 and η_1 be positive constants, to be quantified later. With respect to these quantities, define

$$R = \min_{1 \le k \le K} \sum_{m=1}^{M} \{ I(U_m; Y_{km}) - I(U_m; Y_{em}) \}^+ - \eta_1 \quad (34)$$

and

$$R_{em} = I(U_m; Y_{em}) - \eta_2, \quad m = 1, 2, \ldots, M. \quad (35)$$



(a) Secrecy capacity-achieving code structure.

(b) Nonsecrecy capacity-achieving code structure of [8].

Fig. 4. Structure of two coding schemes for common message transmission over reversely degraded product broadcast channels, for the case of $K = 2$ legitimate receivers and one eavesdropper. To obtain secrecy, separate codebooks are required for each subchannel, so that separate binning can be performed on each. A single codebook is sufficient when there is no secrecy requirement.

The set $T(U_m)$ denotes the set of all sequences that are typical[2] with respect to distribution $p(U_m)$ and the set $T(X_m, U_m)$ denotes the set of all jointly typical sequences (x_m^n, u_m^n) with respect to the distribution $p(X_m, U_m)$. In turn, $T_{u_m^n}(X_m|U_m)$ denotes the set of all sequences x_m^n conditionally typical with respect to a given sequence u_m^n according to $p(X_m|U_m)$.

The details of our construction are as follows.

1) Codebook Generation:

- Codebook $\mathcal{C}_m$ for $m = 1, 2, \ldots, M$ has a total of $M_m = 2^{n(R+R_{em})}$ length n codeword sequences. Each sequence is selected uniformly and independently from the set $T(U_m)$.
- We randomly partition the M_m sequences into 2^{nR} message bins so that there are $Q_m = 2^{nR_{em}}$ codewords per bin.
- The set of codewords associated with bin w in codebook $\mathcal{C}_m$ is denoted as

$$\mathcal{C}_m(w) = \{ u_{m1}^n(w), u_{m2}^n(w), \ldots, u_{mQ_m}^n(w) \} \quad (36)$$

[2] Throughout our development, we mean typicality in the ϵ-weak sense; see, e.g., [4, Ch. 3].

for $w = 1, 2, \ldots, 2^{nR}$ and $m = 1, 2, \ldots, M$. Note that $\mathcal{C}_m = \bigcup_{w=1}^{2^{nR}} \mathcal{C}_m(w)$ is the codebook on subchannel m.

2) Encoding: To encode message w, the encoder randomly and uniformly selects a codeword in the set $\mathcal{C}_m(w)$ for all $1 \le m \le M$. Specifically

- Select M integers $q_1, q_2, \ldots, q_M$, where q_m is selected independently and uniformly from the set $\{1, 2, \ldots Q_m\}$.
- Given a message w, select a codeword $u^n_{mq_m}(w)$ from codebook $\mathcal{C}_m(w)$ for $m = 1, 2, \ldots, M$.
- The transmitted sequence on subchannel m is denoted by $x^n_m = x_m(1), x_m(2), \ldots, x_m(n)$. The symbol $x_m(t)$ is obtained by taking the (possibly stochastic) function $f_m(\cdot)$ of each element of the codeword $u^n_{mq_m}(w)$.

3) Decoding: Receiver k, based on its observations $(y^n_{k1}, y^n_{k2}, \ldots, y^n_{kM})$ from the M parallel subchannels, declares message w according to the following rule.

- Let
  $$\mathcal{S}_k = \{m | 1 \le m \le M, I(U_m; Y_{km}) > I(U_m; Y_{\mathrm{e}m})\}$$
  denote the set of subchannels where receiver k has larger mutual information than the eavesdropper. The receiver only considers the outputs y^n_{km} from these subchannels.
- Receiver k searches for a message w such that, for each $m \in \mathcal{S}_k$, there is an index l_m such that $(u^n_{ml_m}(w), y^n_{km}) \in T(U_m, Y_{km})$. If a unique w has this property, the receiver declares it as the transmitted message. Otherwise, the receiver declares an arbitrary message.

We now analyze the properties of this code.

4) Error Probability: We show that, averaged over the ensemble of codebooks, the error probability is smaller than a constant ϵ' (to be specified). This demonstrates the existence of a codebook with error probability less than ϵ'. We do the analysis for receiver k and, without loss of generality, assume that message w_1 is transmitted.

We first analyze the false-reject event. Let $\mathcal{E}^c_{1m}$ be the event $\{(U^n_{mq_m}(w_1), Y^n_{km}) \notin T(U_m, Y_{km})\}$. Since $U^n_{mq_m} \in T(U_m)$ by construction and Y_{km} is obtained by passing U_m through a discrete memoryless channel, it follows that [4, p. 72, Theorem 3.1.2], $\Pr(\mathcal{E}^c_{1m}) \le \epsilon$. Accordingly, if $\mathcal{E}^c_1$ denotes the event that message w_1 does not appear typical, then we have

$$\Pr(\mathcal{E}^c_1) = \Pr\left(\bigcup_{m=1}^{M} \mathcal{E}^c_{1m}\right) \le M\epsilon. \tag{37}$$

We next analyze the false-accept event. As before, let $\mathcal{S}_k \subseteq \{1, 2, \ldots, M\}$ denote the subset of subchannels for which $I(U_m; Y_{km}) > I(U_m; Y_{\mathrm{e}m})$. In what follows, the index m refers only to subchannels in $\mathcal{S}_k$.

For each $m \in \mathcal{S}_k$, let $\mathcal{E}_{im}$ denote the event that there is a codeword in the set $\mathcal{C}_m(w_i)$ $(i > 1)$ typical with Y^n_{km}. Then

$$\begin{aligned}\Pr(\mathcal{E}_{im}) &= \Pr\big(\exists l \in \{1, \ldots, Q_m\} : \\ &\qquad (U^n_{ml}(w_i), Y^n_{km}) \in T(U_m, Y_{km})\big) \\ &\le \sum_{l=1}^{Q_m} \Pr((U^n_{ml}(w_i), Y^n_{km}) \in T(U_m, Y_{km})) \\ &\le \sum_{l=1}^{Q_m} 2^{-n(I(U_m;Y_{km})-3\epsilon)} && (38) \\ &\le 2^{-n(I(U_m;Y_{km})-I(U_m;Y_{\mathrm{e}m})-3\epsilon+\eta_2)} && (39)\end{aligned}$$

where (38) follows from the fact that since the sequences $(U^n_{ml}(w_i), Y^n_{km})$ are drawn independently, the results in [4, p. 216, Theorem 8.6.1] apply and (39) follows by noting that $Q_m = 2^{n(I(U_m;Y_{\mathrm{e}m})-\eta_2)}$.

In turn, let $\mathcal{E}_i$ denote the event that message w_i has a codeword typical on every subchannel. Then

$$\begin{aligned}\Pr(\mathcal{E}_i) &= \Pr\left(\bigcap_{m \in \mathcal{S}_k} \mathcal{E}_{im}\right) \\ &= \prod_{m \in \mathcal{S}_k} \Pr(\mathcal{E}_{im}) && (40) \\ &= 2^{-n\sum_{m \in \mathcal{S}_k}(I(U_m;Y_{km})-I(U_m;Y_{\mathrm{e}m})-3\epsilon+\eta_2)} \\ &= 2^{-n\sum_{m=1}^{M}(\{I(U_m;Y_{km})-I(U_m;Y_{\mathrm{e}m})\}^+-3\epsilon+\eta_2)}\end{aligned}$$

where (40) follows by independence of codebooks and subchannels.

Finally, the probability of false accept event $\mathcal{E}_{\mathrm{F}}$ is given by

$$\begin{aligned}\Pr(\mathcal{E}_{\mathrm{F}}) &= \Pr\left(\bigcup_{i=2}^{2^{nR}} \mathcal{E}_i\right) \\ &\le 2^{nR} 2^{-n\sum_{m=1}^{M}(\{I(U_m;Y_{km})-I(U_m;Y_{\mathrm{e}m})\}^+-3\epsilon+\eta_2)} \\ &\le 2^{-n(-3M\epsilon+M\eta_2+\eta_1)}\end{aligned}$$

which vanishes with increasing n by selecting the code parameters such that $\eta_1 + M\eta_2 - 3M\epsilon > 0$.

Thus, the probability of error averaged over the ensemble of codebooks is less than

$$\epsilon' = \max\left(M\epsilon, 2^{-n(-3M\epsilon+M\eta_2+\eta_1)}\right)$$

which demonstrates the existence of a codebook with error probability less than ϵ'.

5) Secrecy Analysis: We now show that for any typical code in the ensemble the "perfect equivocation" condition is satisfied, i.e., the normalized mutual information between the message and the output of the eavesdropper is vanishing in the block length. We establish this in two steps. First, our construction of codebooks is such that an eavesdropper who observes only the output of channel m has near-perfect equivocation, i.e., $(1/n)I(W; Y^n_{\mathrm{e}m}) = o_n(1)$.[3] Second, as we show below, the eavesdropper's mutual information only increases by a factor of M even when all the channel outputs are observed

$$\begin{aligned}&\frac{1}{n} I(W; Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) \\ &= \frac{1}{n} h(Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) - \frac{1}{n} h(Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M} | W) \\ &= \frac{1}{n} h(Y^n_{\mathrm{e}1}, \ldots, Y^n_{\mathrm{e}M}) - \sum_{m=1}^{M} \frac{1}{n} h(Y^n_{\mathrm{e}m} | W) && (41) \\ &\le \sum_{m=1}^{M} \frac{1}{n} I(W; Y^n_{\mathrm{e}m}) = o_n(1) && (42)\end{aligned}$$

where (41) follows from the fact that the codewords in the sets $\mathcal{C}_1(W), \mathcal{C}_2(W), \ldots, \mathcal{C}_M(W)$ are independently selected.

[3] We will use $o_n(1)$ to refer to a function that approaches zero as $n \to \infty$.

It remains only to formally establish that for all $m = 1,\dots,M$, we have that

$$\frac{1}{n}I(W;Y_{\text{em}}^n) = o_n(1) \tag{43}$$

which we now do.

Since there are there are $Q_m = 2^{n(I(U_m;Y_{\text{em}})-\eta_2)}$ codewords in each codebook $\mathcal{C}_m(w)$ we have that

$$\frac{1}{n}H(U_m^n|W) = I(U_m;Y_{\text{em}}) - \eta_2 \tag{44}$$

$$\frac{1}{n}H(U_m^n|W,Y_{\text{em}}^n) \le \gamma \triangleq \frac{1}{n} + \eta_2 \, R_{\text{em}} \tag{45}$$

where (44) follows from the fact that the codewords in each bin are selected uniformly, while (45) follows from the fact that a typical codebook $\mathcal{C}_m(w)$ satisfies Fano's inequality. Furthermore, following [25], we can show that for our codebook $\mathcal{C}_m$, all of whose codewords are equally likely to be transmitted, we have that

$$\frac{1}{n}I(U_m^n;Y_{\text{em}}^n) \le I(U_m;Y_{\text{em}}) + |\mathcal{U}|\Pr(U_m^n \notin T(U)) + o_n(1). \tag{46}$$

The equivocation at the eavesdropper can then be lower-bounded using (44)– (46)

$$\begin{aligned} H(W|Y_{\text{em}}^n) &= H(W,U_m^n|Y_{\text{em}}^n) - H(U_m^n|W,Y_{\text{em}}^n) \\ &\ge H(U_m^n|Y_{\text{em}}^n) - n\gamma \qquad (47) \\ &= H(U_m^n) - I(U_m^n;Y_{\text{em}}^n) - n\gamma \\ &= H(U_m^n,W) - I(U_m^n;Y_{\text{em}}^n) - n\gamma \qquad (48) \\ &= H(W) + H(U_m^n|W) - I(U_m^n;Y_{\text{em}}^n) - n\gamma \\ &\ge H(W) + nI(U_m;Y_{\text{em}}) \\ &\quad - I(U_m^n;Y_{\text{em}}^n) - n\gamma - n\eta_2 \qquad (49) \\ &\ge H(W) - n\gamma - n\eta_2 - no_n(1) - n|\mathcal{U}|\epsilon \qquad (50) \end{aligned}$$

where (47) follows from (45), where (48) follows from the fact that W is deterministic given U_m^n, and where (49) and (50) follow from (44) and (46), respectively, and the fact that $\Pr(U_m^n \notin T(U)) \le \epsilon$. Since γ, η_2, and ϵ can be selected to be arbitrarily small, provided n is sufficiently large, we establish (43). □

C. Capacity for Reversely Degraded Channels

We observe that the upper and lower bounds in Proposition 1 and 2, respectively, coincide when the underlying channel is reversely degraded.

Proof of Theorem 1: By selecting $U_m = X_m$ for each $m = 1,2,\dots,M$, in the achievable rate expression (9) in Proposition 2, we have that

$$\bar{R}_{K,M}^{-} = \min_{k\in\{1,\dots,K\}} \sum_{m=1}^{M} \{I(X_m;Y_{km}) - I(X_m;Y_{\text{em}})\}^{+}$$

is an achievable rate. For the reversely degraded channel, for each $k = 1,2,\dots,K$, and $m = 1,2,\dots,M$, we have that either $X_m \to Y_{km} \to Y_{\text{em}}$ or $X_m \to Y_{\text{em}} \to Y_{km}$ holds. In either case, note that

$$\{I(X_m;Y_{km}) - I(X_m;Y_{\text{em}})\}^{+} = I(X_m;Y_{km}|Y_{\text{em}})$$

holds, and hence the lower bound above coincides with (8) in Proposition 1. □

D. Gaussian Channel Capacity

We extend the secrecy capacity in Theorem 1 to Gaussian parallel channels. Since the extension is based on standard techniques, we will only sketch the key steps in the proof.

Proof of Corollary 1: Note that the channel of Definition 3 has the same capacity as another (M,K) reversely degraded broadcast channel in which the sequence obtained at receiver $\pi_m(k+1)$ on subchannel m is

$$\hat{Y}_{\pi_m(k+1)m} = \hat{Y}_{\pi_m(k)m} + \hat{Z}_{\pi_m(k)m}, \qquad k = 0,1,\dots,K$$

where $\pi_m(1),\dots,\pi_m(K+1)$ denotes the ordering of the eavesdropper and legitimate receivers from strongest to weakest, where $\hat{Y}_{\pi_m(0)m} \triangleq X_m$ and $\sigma_{\pi_m(0)m}^2 \triangleq 0$, and where the noises $\hat{Z}_{\pi_m(k)m} \sim \mathcal{CN}(0,\sigma_{\pi_m(k+1)m}^2 - \sigma_{\pi_m(k)m}^2)$ are mutually independent.

With the appropriate Fano's inequality, the converse for Theorem 1 extends to continuous alphabets. The achievability argument relies on weak typicality and also extends to the Gaussian case. Furthermore, the power constraint can be incorporated in the capacity expression, since the objective function is concave in the input distribution (cf. Fact 2 in Appendix A), which gives

$$\bar{C}_{K,M}(P) = \max_{\substack{\prod_{m=1}^{M} p(X_m), \\ E\left[\sum_{m=1}^{M} X_m^2\right] \le P}} \min_{k\in\{1,\dots,K\}} \sum_{m=1}^{M} I(X_m;\hat{Y}_{km}|\hat{Y}_{\text{em}}). \tag{51}$$

Next observe that

$$\max_{p(X_m),E[X_m^2]\le P_m} I(X_m;\hat{Y}_{km}|\hat{Y}_{\text{em}})$$

denotes the capacity of a Gaussian wiretap channel [13]. Accordingly, for each $m = 1,2,\dots,M$

$$\max_{p(X_m),E[X_m^2]\le P_m} I(X_m;\hat{Y}_{km}|\hat{Y}_{\text{em}}) = \left\{\log\left(\frac{1+P_m/\sigma_{km}^2}{1+P_m/\sigma_{\text{em}}^2}\right)\right\}^{+}. \tag{52}$$

Now if $(P_1^*,\dots,P_M^*)$ denotes an optimal power allocation in (51), then via (52), we have that

$$\bar{C}_{K,M}(P) = \min_{k\in\{1,\dots,K\}} \sum_{m=1}^{M} \left\{\log\left(\frac{1+P_m^*/\sigma_{km}^2}{1+P_m^*/\sigma_{\text{em}}^2}\right)\right\}^{+}$$

whence (11) follows. □

VI. Parallel Channels and Independent Messages

In this section, we establish the secrecy sum-capacity for the case of independent messages by providing a proof of Theorem 2 and then specialize the result to the Gaussian case stated in Corollary 2.

A. Capacity

Proof of Theorem 2: We establish, in order, the achievability and converse parts of the proof.

To achieve capacity, on each subchannel m, we send information only to the strongest receiver π_m. It follows from the result of the single-user wiretap channel [25] that a rate of $R_m = \max_{p(X_m)} I(X_m; Y_{\pi_m}|Y_{\mathrm{e}_m})$ is achievable on the mth subchannel. Accordingly, a sum-rate of $\sum_m R_m$ is achievable with this scheme, which is the capacity (13).

We establish the converse in two steps. First we consider a single receiver genie-aided channel whose secrecy capacity upper-bounds the secrecy sum-capacity of the original channel. Then we show that the secrecy capacity for this genie-aided channel coincides with (13), thus completing the proof.

1) Construction of Genie-Aided Channel: Our genie-aided channel has only one receiver which we call receiver 1. It observes the output of the strongest receiver π_m, i.e., $Y_{\pi_m}^n$ on subchannel m, and hence its output is $(Y_{\pi_1}^n, \ldots, Y_{\pi_M}^n)$.

To verify that the secrecy capacity of the genie-aided channel upper bounds $C_{M,K}$ it suffices to show the following.

Lemma 1: If a secrecy rate point $(R_1, R_2, \ldots, R_K)$ is achievable for the K-receiver channel in Theorem 2 then a secrecy rate $\sum_{k=1}^{K} R_k$ is achievable on the genie-aided channel.

Proof: Let the messages corresponding to $(R_1, R_2, \ldots, R_K)$ be denoted as $(W_1, W_2, \ldots, W_K)$. This implies that, for any $\epsilon > 0$ and n large enough, there is a code of length n such that $\Pr(\hat{W}_k \neq W_k) \leq \epsilon$ for $k = 1, 2, \ldots, K$, and such that

$$\frac{1}{n} H(W_k | W_1, \ldots, W_{k-1}, W_{k+1}, \ldots, W_K, Y_{\mathrm{e}1}^n, \ldots, Y_{\mathrm{e}M}^n) \geq R_k - \epsilon. \quad (53)$$

We now consider transmitting the message $\hat{W}_1 = (W_1, W_2, \ldots, W_K)$ to receiver 1 on the genie-aided channel, using the same encoding scheme that achieves $(R_1, \ldots, R_K)$ on the original channel. By construction, receiver 1 on the genie-aided channel can use the same decoder as receiver i on the original channel to decode message W_i. So it remains to verify that the secrecy condition is satisfied on the genie-aided channel

$$\begin{aligned}
&\frac{1}{n} H(\hat{W}_1 | Y_{\mathrm{e}1}^n, \ldots, Y_{\mathrm{e}M}^n) \\
&= \frac{1}{n} H(W_1, \ldots, W_K | Y_{\mathrm{e}1}^n, \ldots, Y_{\mathrm{e}M}^n) \\
&\geq \sum_{k=1}^{K} \frac{1}{n} H(W_k | W_1, \ldots, W_{k-1}, W_{k+1}, \ldots, W_K, Y_{\mathrm{e}1}^n, \ldots, Y_{\mathrm{e}M}^n) \\
&\geq \sum_{k=1}^{K} R_k - K\epsilon
\end{aligned}$$

where the last step follows by substituting (53). Since $\epsilon > 0$ can be arbitrarily small, if n is sufficiently large, this establishes our claim. □

2) Sum-Capacity of the Genie-Aided Channel: It remains to show that the secrecy-capacity of the genie-aided channel equals $C_{M,K}$. This however follows immediately via specialization of Theorem 1 to the case of $K = 1$.

It is worth remarking that this genie-aided upper bound continues to hold even if the eavesdropper's channel is not ordered with respect to the legitimate receivers. In general, following Proposition 1, the upper bound can be tightened by considering, for all $1 \leq m \leq M$, the worst joint distribution $p'(Y_{\pi_m}, Y_{\mathrm{e}m}|X_m)$ among all joint distributions with the same marginal distribution as $p(Y_{\pi_m}|X_m)$ and $p(Y_{\mathrm{e}m}|X_m)$, yielding

$$C_{K,M} \leq \min_{\prod_{m=1}^{M} p'(Y_{\pi_m}, Y_{\mathrm{e}m}|X_m)} \max_{\prod_{m=1}^{M} p(X_m)} \sum_{m=1}^{M} I(X_m; Y_{\pi_m}|Y_{\mathrm{e}m}). \quad (54)$$

□

B. Gaussian Channels

Proof of Corollary 2: The achievability of rate (14) follows by using independent Gaussian wiretap codebooks on each subchannel and only transmitting to the strongest receiver on each subchannel. For the converse, we need to show that Gaussian inputs are optimal in (13), which follows from the same reasoning used for the common message case in Section V-D. □

VII. Fading Channels and a Common Message

In this section, we establish the upper and lower bounds on the common message secrecy-capacity for fast fading channels. In particular, we provide proofs for Theorem 3 and Corollary 3.

A. Capacity Bounds

Proof of Theorem 3: We establish, in order, the upper and lower capacity bounds in (15).

To obtain our upper bound, suppose that we only need to transmit the message to receiver k. An upper bound on the secrecy capacity for this single-user channel is obtained by specializing Proposition 3 (see Section IV-D) to the case of $K = 1$ user. Accordingly, we have

$$\bar{R}^{+}(P) \leq \max_{\substack{\rho(H_k): \\ E[\rho(H_k)] \leq P}} E\left[\left\{\log\left(\frac{1 + |H_k|^2 \rho(H_k)}{1 + |H_{\mathrm{e}}|^2 \rho(H_k)}\right)\right\}^{+}\right]. \quad (55)$$

Since k is arbitrary, we tighten the upper bound (55) by minimizing over k, yielding (16b).

Next, we establish the lower bound (16a) by considering the following probabilistic extension of the parallel broadcast channel [14]. At each time, only one of the subchannels operates. Subchannel m is selected with a probability p_m, independent of the selection at all other times. Also, suppose that there is a total power constraint P on the input.

In this case, a straightforward extension of Proposition 2 provides the following achievable rate:

$$\bar{R}_{K,M}(P) \triangleq \max \min_{i \in \{1, \ldots, K\}} \sum_{m=1}^{M} p_m \left\{ I(U_m; Y_{km}) - I(U_m; Y_{\mathrm{e}m}) \right\}^{+} \quad (56)$$

where $U_1, U_2, \ldots, U_M$ are auxiliary random variables and the maximum is over the product distribution $p(U_1)p(U_2)\ldots p(U_M)$ and the stochastic mappings $X_m = f_m(U_m)$ that satisfy $\sum_{m=1}^{M} p_m E[X_m^2] \leq P$.

To simplify the exposition, we focus on the case of $K = 2$ receivers. The extension to $K > 2$ receivers is analogous and straightforward.

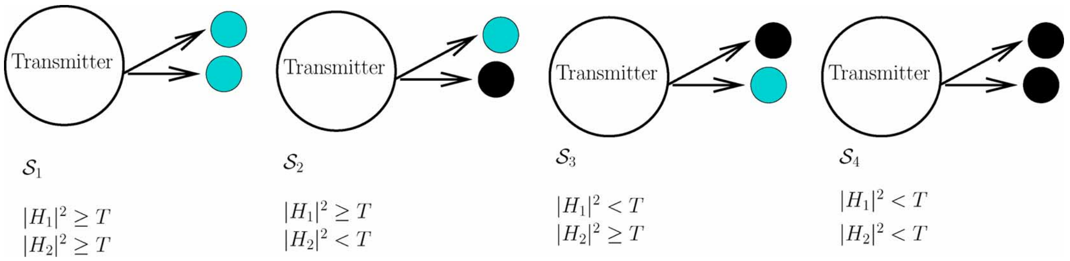


Fig. 5. Decomposition of the system with $K = 2$ receivers into four states, as a function of their channel gains relative to a threshold T. The darkly and lightly shaded circles, respectively, indicate that a channel gain is, respectively, below and above the threshold.

To start, we fix a threshold $T > 0$ and decompose the system into four states as shown in Fig. 5. The transmission takes place over a block of length n, and we classify $t = 1, 2, \ldots, n$ according to

$$\begin{aligned}\mathcal{S}_1 &= \{t \in \{1,n\} \mid |h_1(t)|^2 \geq T, |h_2(t)|^2 \geq T\}\\ \mathcal{S}_2 &= \{t \in \{1,n\} \mid |h_1(t)|^2 \geq T, |h_2(t)|^2 < T\}\\ \mathcal{S}_3 &= \{t \in \{1,n\} \mid |h_1(t)|^2 < T, |h_2(t)|^2 \geq T\}\\ \mathcal{S}_4 &= \{t \in \{1,n\} \mid |h_1(t)|^2 < T, |h_2(t)|^2 < T\}. \qquad (57)\end{aligned}$$

The resulting channel is a probabilistic parallel channel with probabilities of the four channels are then given by

$$\begin{aligned}p(\mathcal{S}_1) &= \Pr(|H_1|^2 \geq T, |H_2|^2 \geq T)\\ p(\mathcal{S}_2) &= \Pr(|H_1|^2 \geq T, |H_2|^2 < T)\\ p(\mathcal{S}_3) &= \Pr(|H_1|^2 < T, |H_2|^2 \geq T)\\ p(\mathcal{S}_4) &= \Pr(|H_1|^2 < T, |H_2|^2 < T).\end{aligned}$$

In turn, with $X_m = U_m \sim \mathcal{CN}(0,P)$ in (56) the achievable rate expression is

$$\bar{R}^-(P) = \min_{k\in\{1,2\}} \left\{ \Pr(|H_k|^2 \geq T) E\left[\log\left(\frac{1+|H_k|^2P}{1+|H_\mathrm{e}|^2P}\right) \middle| \, |H_k|^2 \geq T\right]\right\}. \qquad (58)$$

Finally, optimizing (61) over the threshold, we obtain (16a) as follows (for the case $K = 2$):

$$\begin{aligned}&\bar{R}^-(P) && (59)\\ &= \max_{T>0} \min_{k\in\{1,2\}} \Bigg\{ \Pr(|H_k|^2 \geq T) && (60)\\ &\quad \cdot E\left[\log\left(\frac{1+|H_k|^2P}{1+|H_\mathrm{e}|^2P}\right) \middle| \, |H_k|^2 \geq T\right]\Bigg\}\\ &= \max_{T>0} \min_{k\in\{1,2\}} \Bigg\{\\ &\quad \int_T^\infty \log\left(\frac{1+xP}{\exp\{E_{H_\mathrm{e}}[\log(1+|H_\mathrm{e}|^2P)]\}}\right) p_k(x)\,dx\Bigg\}\\ &\geq \min_{k\in\{1,2\}} \int_{T^*}^\infty \log\left(\frac{1+xP}{\exp\{E_{H_\mathrm{e}}[\log(1+|H_\mathrm{e}|^2P)]\}}\right) p_k(x)\,dx && (61)\\ &= \min_{k\in\{1,2\}} E_{H_k}\left[\left\{\log\left(\frac{1+|H_k|^2P}{\exp\{E_{H_\mathrm{e}}[\log(1+|H_\mathrm{e}|^2P)]\}}\right)\right\}^+\right] && (62)\end{aligned}$$

where T^* in (64) is obtained via

$$\log(1+T^*P) - E_{H_\mathrm{e}}[\log(1+|H_\mathrm{e}|^2P] = 0.$$

For $K > 2$ receivers, we use the straightforward generalization of this scheme to a construction with 2^K states, where each state specifies the subset of receivers that are above the threshold T^*.

An alternative proof based on discretizing the fading coefficients along the lines of [10] is developed in Appendix B. □

It is worth remarking that our code construction more generally suggests a concatenated coding approach for this channel, with an outer erasure code and an inner wiretap code. With this structure, incoming information bits are mapped into a codeword of a $(2^K - 1, 2^{K-1})$ erasure code over a sufficiently large alphabet. Each resulting symbol then forms the message for its corresponding state. Each receiver obtains 2^{K-1} symbols in states where its channel gain is above the threshold and can recover the information symbols. Details of this architecture are developed in [12].

B. High SNR Regime

Proof of Corollary 3: Since the channel gains of all the legitimate receivers are distributed as $\mathcal{CN}(0,1)$, we use a generic variable H to denote the channel gain of any given user.

For the upper bound (17a), it suffices to note that

$$E\left[\left\{\log\frac{1+|H|^2\rho(H)}{1+|H_\mathrm{e}|^2\rho(H)}\right\}^+\right] \leq E\left[\left\{\log\frac{|H|^2}{|H_\mathrm{e}|^2}\right\}^+\right]$$

For the lower bound (17b), first recall that

$$\bar{R}^-(P) = E[\{\log(1+|H|^2P) - E[\log(1+|H_\mathrm{e}|^2P)]\}^+].$$

Since, as established in Appendix C,

$$f_P(x) \triangleq \{\log(1+xP) - E[\log(1+|H_\mathrm{e}|^2P)]\}^+ \qquad (63)$$

satisfies the conditions for the dominated convergence theorem [1], we obtain

$$\begin{aligned}&\lim_{P\to\infty} \bar{R}^-(P)\\ &\quad = E[\lim_{P\to\infty} \{\log(1+|H|^2P) - E[\log(1+|H_\mathrm{e}|^2P)]\}^+] && (64)\\ &\quad = E\left[\left\{\log \lim_{P\to\infty} \frac{1+|H|^2P}{\exp\{E[\log(1+|H_\mathrm{e}|^2P)]\}}\right\}^+\right]\\ &\quad = E\left[\left\{\log|H|^2 + \frac{\gamma}{\log 2}\right\}^+\right] && (65)\end{aligned}$$

where γ is the Euler-Gamma constant ($\gamma \approx 0.5772$). □

VIII. Fading Channels and Independent Messages

In this section, we establish our results for fading channels with independent messages.

A. Capacity Bounds

In what follows, we establish the upper and lower bounds in (19).

Proof of Upper Bound in Proposition 3: Our upper bound is based on introducing a single-user genie-aided channel, as in Section VI-A, whose achievable rate we upper-bound. The result is closely related to an upper bound provided in [11] for the ergodic fading channel with large coherence periods. However, in the interest of completeness, we now provide the full derivation.

To start, consider the following channel with one receiver and one eavesdropper:

$$\begin{aligned} Y(t) &= H_{\max}(t)X(t) + Z(t) \\ Y_{\mathrm{e}}(t) &= H_{\mathrm{e}}(t)X(t) + Z_{\mathrm{e}}(t). \end{aligned} \tag{66}$$

Using reasoning along the lines of the analysis in Section VI-A, we deduce that the secrecy sum-capacity of the channel (5) is upper-bounded by the secrecy capacity of the genie-aided channel (66), and thus it remains only to show that an upper bound on the secrecy capacity of this channel is (20a). Furthermore, the joint distribution of the noise variables $(Z(t), Z_{\mathrm{e}}(t))$ is selected so that if $|h_{\mathrm{e}}(t)| \le |h_{\max}(t)|$ we have the Markov chain $X(t) \to Y(t) \to Y_{\mathrm{e}}(t)$; otherwise, we have the Markov chain $X(t) \to Y_{\mathrm{e}}(t) \to Y(t)$.

We show that for any sequence of length n, rate R codes, as in Definition 8, the upper bound (20a) holds. Recall that the encoding function has the form

$$X(t) = f_t(W, H_{\max}^t), \qquad t = 1, 2, \ldots, n, \tag{67}$$

and for every $\epsilon > 0$, and sufficiently large n, we have, via Fano's inequality and the secrecy condition

$$\frac{1}{n} H(W | H_{\max}^n, Y^n) \le \epsilon \tag{68}$$

$$\frac{1}{n} I(W; Y_{\mathrm{e}}^n, H_{\mathrm{e}}^n | H_{\max}^n) \le \epsilon. \tag{69}$$

An upper bound on the rate is as follows:

$$\begin{aligned} nR &= H(W|H_{\max}^n) \\ &\le I(W; Y^n | H_{\max}^n) - I(W; Y_{\mathrm{e}}^n, H_{\mathrm{e}}^n | H_{\max}^n) + 2n\epsilon && (70) \\ &\le I(X^n; Y^n | H_{\max}^n, H_{\mathrm{e}}^n, Y_{\mathrm{e}}^n) + 2n\epsilon && (71) \\ &= h(Y^n | H_{\max}^n, H_{\mathrm{e}}^n, Y_{\mathrm{e}}^n) - h(Y^n | H_{\max}^n, H_{\mathrm{e}}^n, Y_{\mathrm{e}}^n, X^n) + 2n\epsilon \\ &= h(Y^n | H_{\max}^n, H_{\mathrm{e}}^n, Y_{\mathrm{e}}^n) \\ &\quad - \sum_{t=1}^{n} h(Y(t) | H_{\max}(t), H_{\mathrm{e}}(t), Y_{\mathrm{e}}(t), X(t)) + 2n\epsilon && (72) \\ &\le h(Y^n | H_{\max}^n, H_{\mathrm{e}}^n, Y_{\mathrm{e}}^n) \\ &\quad - \sum_{t=1}^{n} h(Y(t) | H_{\max}^t, H_{\mathrm{e}}(t), Y_{\mathrm{e}}(t), X(t)) + 2n\epsilon \\ &\le \sum_{t=1}^{n} I(X(t); Y(t) | Y_{\mathrm{e}}(t), H_{\max}^t, H_{\mathrm{e}}(t)) + 2n\epsilon && (73) \end{aligned}$$

where (70) follows by substituting (68) and (69), (71) follows from the Markov chain $W \to (X^n, Y_{\mathrm{e}}^n, H_{\max}^n, H_{\mathrm{e}}^n) \to Y^n$, where (72) follows from the fact that the channel is memoryless.

From the capacity of the Gaussian wiretap channel [13], we have that

$$\begin{aligned} I(X(t); Y(t) | Y_{\mathrm{e}}(t), H_{\max}^t, H_{\mathrm{e}}(t)) \\ \le E_{H_{\max}^t, H_{\mathrm{e}}(t)} \left[\left\{ \log \frac{1 + |H_{\max}(t)|^2 E[|X(t)|^2]}{1 + |H_{\mathrm{e}}(t)|^2 E[|X(t)|^2]} \right\}^+ \right] \end{aligned} \tag{74}$$

with equality if $X(t)$ is conditionally Gaussian given $(H_{\max}^t, H_{\mathrm{e}}(t))$. Since a Gaussian distribution depends only on its mean and variance and $X(t)$ is independent of $H_{\mathrm{e}}(t)$ (cf. (67)), we can write without loss of generality[4] that

$$X(t) \sim \mathcal{CN}\left(0, \sqrt{\rho_t(H_{\max}^t)}\right) \tag{75}$$

for some sequence of functions $\rho_t(\cdot)$ that satisfy the average power constraint $\frac{1}{n}\sum_{t=1}^{n} E[\rho_t(H_{\max}^t)] \le P$. With this substitution, we have from (73) that

$$\begin{aligned} &nR \le \\ &\sum_{t=1}^{n} E_{H_{\max}^t, H_{\mathrm{e}}(t)} \left[\left\{ \log \frac{1 + |H_{\max}(t)|^2 \rho_t(H_{\max}^t)}{1 + |H_{\mathrm{e}}(t)|^2 \rho_t(H_{\max}^t)} \right\}^+ \right] + 2n\epsilon. \end{aligned} \tag{76}$$

It turns out, as we show below, that the right-hand side in (79) is maximized, for each t, by a function $\gamma_t(\cdot)$ that only depends on $H_{\max}^t$ via $H_{\max}(t)$. The upper bound expression in (20a) then follows, since from (79)

$$\begin{aligned} &nR - 2n\epsilon \\ &\le \sum_{t=1}^{n} E_{H_{\max}, H_{\mathrm{e}}} \left[\left\{ \log \frac{1 + |H_{\max}|^2 \gamma_t(H_{\max})}{1 + |H_{\mathrm{e}}|^2 \gamma_t(H_{\max})} \right\}^+ \right] \\ &\le n E_{H_{\max}, H_{\mathrm{e}}} \left[\left\{ \log \frac{1 + |H_{\max}|^2 \frac{1}{n}\sum_{t=1}^{n} \gamma_t(H_{\max})}{1 + |H_{\mathrm{e}}|^2 \frac{1}{n}\sum_{t=1}^{n} \gamma_t(H_{\max})} \right\}^+ \right] && (77) \\ &= n E_{H_{\max}, H_{\mathrm{e}}} \left[\left\{ \log \frac{1 + |H_{\max}|^2 \gamma(H_{\max})}{1 + |H_{\mathrm{e}}|^2 \gamma(H_{\max})} \right\}^+ \right] && (78) \end{aligned}$$

where (77) follows from the fact $\{\log(1 + ax)/(1 + bx)\}^+$ is concave in $x > 0$ for a fixed a and b, so Jensen's inequality can be applied and where (78) follows by defining $\gamma(H_{\max}) = \frac{1}{n}\sum_{t=1}^{n} \gamma_t(H_{\max})$. Note that the power constraint $E[\gamma(H_{\max})] \le P$ naturally follows from the definition of $\gamma(\cdot)$.

It remains to establish the existence of $\gamma_t(\cdot)$ as we now do. In particular, for any sequence of functions $\rho_t(\cdot)$, we define $\gamma_t(\cdot)$ according to

$$\gamma_t(H_{\max}(t)) \triangleq E_{H_{\max}^{t-1}}[\rho_t(H_{\max}^t) | H_{\max}(t)]$$

and show below that each term in the summation in (76) only increases if we replace $\rho_t(\cdot)$ by $\gamma_t(\cdot)$

$$E_{H_{\max}^t, H_{\mathrm{e}}(t)} \left[\left\{ \log \frac{1 + |H_{\max}(t)|^2 \rho_t(H_{\max}^t)}{1 + |H_{\mathrm{e}}(t)|^2 \rho_t(H_{\max}^t)} \right\}^+ \right] \tag{79}$$

[4] An analogous approach is taken in [3, Sec IV, Proposition 3] for establishing the capacity of fading channels with side information at the transmitter.

$$
\begin{aligned}
&= E_{H_{\max}(t),H_{\mathrm{e}}(t)}\left[E_{H_{\max}^{t-1}}\left[\left\{\log\frac{1+|H_{\max}(t)|^2\rho_t(H_{\max}^t)}{1+|H_{\mathrm{e}}(t)|^2\rho_t(H_{\max}^t)}\right\}^+\right]\right]\\
&\le E_{H_{\max}(t),H_{\mathrm{e}}(t)}\Bigg[\\
&\left\{\log\frac{1+|H_{\max}(t)|^2E_{H_{\max}^{t-1}}[\rho_t(H_{\max}^t)|H_{\max}(t)]}{1+|H_{\mathrm{e}}(t)|^2E_{H_{\max}^{t-1}}[\rho_t(H_{\max}^t)|H_{\max}(t)]}\right\}^+\Bigg] \qquad (80)\\
&= E_{H_{\max}(t),H_{\mathrm{e}}(t)}\left[\left\{\log\frac{1+|H_{\max}(t)|^2\gamma_t(H_{\max}(t))}{1+|H_{\mathrm{e}}(t)|^2\gamma_t(H_{\max}(t))}\right\}^+\right] \qquad (81)\\
&= E_{H_{\max},H_{\mathrm{e}}}\left[\left\{\log\frac{1+|H_{\max}|^2\gamma_t(H_{\max})}{1+|H_{\mathrm{e}}|^2\gamma_t(H_{\max})}\right\}^+\right] \qquad (82)
\end{aligned}
$$

where (80) follows from Jensen's inequality. This completes the proof. □

Proof of Lower Bound in Proposition 3: The lower bound (20b) is achieved by a scheme that, at each time, transmits only to the receiver with the best instantaneous channel gain.

In detail, we first quantize each receiver's channel gain into q levels $A_1 = 0 < A_2 < \cdots < A_q \le A_{q+1} = J$ (if any user's channel gain exceeds J, then this slot is ignored for transmission). Since the channel gains of the K receivers are independent, there are a total of $M = q^K$ different super-states. These are denoted as $S_1, S_2, \ldots, S_M$. Each of the super-states denotes one subchannel.

Our scheme transmits an independent message on each of the M parallel channels. Let $G_m \in \{A_1, A_2, \ldots, A_q\}$ denote the gain of the strongest receiver on channel m. We use a Gaussian codebook with power $\rho(G_m)$ on channel m. The achievable rate on channel m is

$$
\begin{aligned}
R_m &= I(U_m;Y_m) - I(U_m;Y_{\mathrm{e}m},H_{\mathrm{e}m})\\
&= \log(1+G_m\rho(G_m)) - E[\log(1+|H_{\mathrm{e}}|^2\rho(G_m))]
\end{aligned}
$$

where the second equality follows from our choice of $X_m = U_m \sim \mathcal{CN}(0,\rho(G_m))$. The overall achievable sum-rate is

$$
\begin{aligned}
&R_K^-(P)\\
&= \sum_{m=1}^{M}\Pr(S_m)R_m \qquad (83)\\
&= \sum_{m=1}^{M}\Pr(S_m)\log\left(\frac{1+G_m\rho(G_m)}{\exp\{E_{H_{\mathrm{e}}}[\log(1+|H_{\mathrm{e}}|^2\rho(G_m))]\}}\right)\\
&= \sum_{l=1}^{q}\Pr(A_l)\log\left(\frac{1+A_l\rho(A_l)}{\exp\{E_{H_{\mathrm{e}}}[\log(1+|H_{\mathrm{e}}|^2\rho(A_l))]\}}\right) \qquad (84)\\
&= \sum_{l=1}^{q}\Pr(A_l)\left\{\log\left(\frac{1+A_l\rho(A_l)}{\exp\{E_{H_{\mathrm{e}}}[\log(1+|H_{\mathrm{e}}|^2\rho(A_l))]\}}\right)\right\}^+ \qquad (85)
\end{aligned}
$$

where (84) follows by using the fact that $G_m \in \{A_1, A_2, \ldots, A_q\}$ and rewriting the summation over these indices, and where (85) follows from the fact that if for some $\rho(A_l)$ we have

$$
\log(1+A_l\rho(A_l)) - E_{H_{\mathrm{e}}}[\log(1+|H_{\mathrm{e}}|^2\rho(A_l)) < 0
$$

then we can simply replace $\rho(A_l)$ by zero to increase the value. When we fix J and take $q \to \infty$ we show in Appendix D that the summation converges to

$$
\begin{aligned}
&R_K^-(P)\\
&= \int_0^J \left\{\log\left(\frac{1+a\rho(a)}{\exp\{E_{H_{\mathrm{e}}}[\log(1+|H_{\mathrm{e}}|^2\rho(a))]\}}\right)\right\}^+ p(a)\,da\\
&= \int_0^\infty \left\{\log\left(\frac{1+a\rho(a)}{\exp\{E_{H_{\mathrm{e}}}[\log(1+|H_{\mathrm{e}}|^2\rho(a))]\}}\right)\right\}^+ p(a)\,da\\
&\quad -\int_J^\infty \left\{\log\left(\frac{1+a\rho(a)}{\exp\{E_{H_{\mathrm{e}}}[\log(1+|H_{\mathrm{e}}|^2\rho(a))]\}}\right)\right\}^+ p(a)\,da. \qquad (86)
\end{aligned}
$$

Since the integral above is finite, the second term vanishes as $J \to \infty$, hence

$$
\begin{aligned}
R_K^-(P) = E\,[\{\log(1+|H_{\max}|^2\rho(H_{\max})) -{}\\
E_{H_{\mathrm{e}}}[\log(1+|H_{\mathrm{e}}|^2\rho(H_{\max}))]\}^+] \qquad (87)
\end{aligned}
$$

is an achievable rate, whence (20b) follows. □

B. Scaling Law

We now establish (22).

Proof of Theorem 4: Letting $\rho^*(H_{\max})$ denote the power allocation that maximizes $R_K^+(P)$ in (20a), we obtain

$$
\begin{aligned}
&R_K^+(P) - R_K^-(P) \qquad (88)\\
&\le E\left[\left\{\log\left(\frac{1+|H_{\max}|^2\rho^*(H_{\max})}{1+|H_{\mathrm{e}}|^2\rho^*(H_{\max})}\right)\right\}^+\right]\\
&\quad - E\left[\log\left(\frac{1+|H_{\max}|^2\rho^*(H_{\max})}{1+|H_{\mathrm{e}}|^2\rho^*(H_{\max})}\right)\right] \qquad (89)\\
&= \Pr(|H_{\mathrm{e}}|^2 \ge |H_{\max}|^2)\\
&\quad \cdot E\left[\log\frac{1+|H_{\mathrm{e}}|^2\rho^*(H_{\max})}{1+|H_{\max}|^2\rho^*(H_{\max})}\,\middle|\,|H_{\mathrm{e}}|^2 \ge |H_{\max}|^2\right]\\
&\le \Pr(|H_{\mathrm{e}}|^2 \ge |H_{\max}|^2)\\
&\quad \cdot E\left[\log\frac{|H_{\mathrm{e}}|^2}{|H_{\max}|^2}\,\middle|\,|H_{\mathrm{e}}|^2 \ge |H_{\max}|^2\right] \qquad (90)\\
&\le \frac{2\log 2}{K+1} \qquad (91)
\end{aligned}
$$

where (89) follows from substituting the bounds in Proposition 3, where (90) follows from the fact that $\log((1+|H_{\mathrm{e}}|^2a)/(1+|H_{\max}|^2a))$ is increasing in a for $|H_{\mathrm{e}}|^2 \ge |H_{\max}|^2$, and where (91) follows from the fact that $\Pr(|H_{\mathrm{e}}|^2 \ge |H_{\max}|^2) = 1/(1+K)$, since we assumed the channel coefficients to be i.i.d., and from the following "helper" lemma.

Lemma 2: If $H_1, H_2, \ldots, H_K, H_{\mathrm{e}}$ are i.i.d. unit-mean exponentials, then for $K \ge 2$ we have

$$
E\left[\log\frac{|H_{\mathrm{e}}|^2}{|H_{\max}|^2}\,\middle|\,|H_{\mathrm{e}}|^2 \ge |H_{\max}|^2\right] \le 2\log 2. \qquad (92)
$$

Proof of Lemma 2: First, we use the following fact.

Fact 1 ([6]): Let $V_1, V_2, \ldots, V_K, V_{K+1}$ be i.i.d. exponentially distributed random variables with mean λ, and let

$V_{\max}(K+1)$ and $V_{\max}(K)$, respectively, denote the largest and second-largest of these random variables. Then the joint distribution of $(V_{\max}(K), V_{\max}(K+1))$ satisfies

$$V_{\max}(K+1) = V_{\max}(K) + Y \tag{93}$$

where Y is an exponentially distributed random variable with mean λ that is independent of $V_{\max}(K)$.

Proceeding, we have

$$\begin{aligned} & E\left[\log \frac{|H_{\mathrm{e}}|^2}{|H_{\max}|^2} \,\middle|\, |H_{\mathrm{e}}|^2 \geq |H_{\max}|^2\right] && (94) \\ &= E\left[\log \frac{|H_{\max}|^2 + Y}{|H_{\max}|^2}\right] && (95) \\ &\leq E\left[\frac{Y}{|H_{\max}|^2}\right] && (96) \\ &= E[Y] E\left[\frac{1}{|H_{\max}|^2}\right] && (97) \\ &= E\left[\frac{1}{|H_{\max}|^2}\right] && (98) \end{aligned}$$

where (96) follows from the identity $\log(1+x) \leq x$ for $x > 0$, where (97) follows from the independence of Y and $H_{\max}$, and where (98) from the fact that $E[Y] = 1$. Since $|H_{\max}|^2 \geq \max(|H_1|^2, |H_2|^2)$, we obtain

$$E\left[\frac{1}{|H_{\max}|^2}\right] \leq E\left[\frac{1}{\max(|H_1|^2, |H_2|^2)}\right] = 2\log 2$$

which establishes (92). □

IX. Concluding Remarks

In this paper, a generalization of the wiretap channel to the case of parallel and fading channels with multiple receivers was considered. For parallel channels, we established the common-message secrecy capacity for the reversely degraded product broadcast channel and provided upper and lower bounds for general product broadcast channels. For independent messages over parallel channels, we determined the secrecy sum-capacity, again for the reversely degraded case. We also extended both results to Gaussian parallel channels.

For fading channels, we analyzed a fast-fading model in which the transmitter knows the instantaneous channels of all the legitimate receivers but not of the eavesdropper, but the eavesdropper has full information about all channels of all receivers. Interestingly, the common-message secrecy capacity does not decay to zero as the number of legitimate receiver grows. For the case of independent messages, we showed that an opportunistic architecture achieves the secrecy sum-capacity in the limit of large number of receivers.

In terms of future work, there are a number of interesting directions to pursue.

As one example, our formulation for the fading channel assumes that the fading coefficients of the legitimate receivers are revealed to the sender in a causal fashion. Implicitly, we are assuming the availability of an insecure, but authenticated feedback link between the receiver(s) and the sender that is used to provide CSI to the transmitter. The availability of this (digital) feedback link is reminiscent of the secret key generation protocols pioneered by Maurer [18]. Indeed, this feedback link can be used in a variety of ways rather than just providing CSI as is assumed here and exploring connections to the key-generation approach of Maurer may be fruitful.

Throughout this paper, we focused on Wyner's notion of perfect secrecy, which corresponds to requiring the block-length-normalized mutual information between the message and the output of the eavesdropper's channel to approach zero with increasing block length. As we mentioned at the outset, this is a significantly weaker notion of security than Shannon's, which requires that the mutual information be zero regardless of the block length. In work lying between these extremes, Maurer and Wolf [19] have observed that for the discrete memoryless wiretap channel, the secrecy notion of Wyner can be strengthened in the following sense—the *unnormalized* mutual information between the message and the output of the eavesdropper's channel can be driven to zero with the block length without sacrificing further rate. It remains to be seen if analogous results can be obtained for the Gaussian wiretap channel and the fading channels considered in this work.

The protocols investigated in this paper relied on time diversity (for the common message) and multiuser diversity (for independent messages) to enable secure communication. In situations where such forms of diversity is not available, it is of interest to develop a formulation for secure transmission, analogous to the outage formulation for slow-fading channels. Secondly, the impact of multiple antennas on secure transmission is far from being clear at this stage. While multiple antennas can theoretically provide significant gains in throughput in the conventional systems, a theoretical analysis for the case of confidential messages is naturally of great interest.

Finally, this paper has focused on architectural questions and associated separation theorems, using random coding arguments. As such, many of our constructions rely implicitly or explicitly on the existence of good "standard" scalar wiretap codes for discrete and Gaussian channels. The development of practical and flexible families of secure, capacity-achieving, low-complexity scalar wiretap codes has only begun, and remains a rich area for further research [22].

Appendix A
Concavity of Conditional Mutual Information

We establish the following.

Fact 2: For any random variables X, Y, and Z the quantity $I(X;Y|Z)$ is concave in $p(X)$.

Proof: Let T be a binary-valued random variable that determines the induced distribution on X, i.e.,

$$p(Y,Z,X|T=t) = \begin{cases} p(Y,Z|X)\, p_0(X), & t=0 \\ p(Y,Z|X)\, p_1(X), & t=1. \end{cases}$$

Hence, we have the Markov chain

$$T \rightarrow X \rightarrow (Y,Z). \tag{99}$$

To establish the concavity of $I(X;Y|Z)$ in $p(X)$ it suffices to show that

$$I(X;Y|Z,T) \leq I(X;Y|Z)$$

which follows from the following chain of inequalities:

$$\begin{aligned} & I(X;Y|Z,T) - I(X;Y|Z) \\ &= I(X;Y,Z|T) - I(X;Z|T) - I(X;Y,Z) - I(X;Z) \qquad (100) \\ &= I(X;Y,Z|T) - I(X;Z|T) - I(TX;Y,Z) - I(TX;Z) \\ &= I(X;Y,Z|T) - I(TX;Y,Z) - I(X;Z|T) - I(TX;Z) \\ &= I(T;Z) - I(T;Y,Z) = -I(T;Y|Z) \leq 0 \qquad (101) \end{aligned}$$

where (100) is a consequence of the chain rule for mutual information, and (101) follows from (99), whence $I(T;Z|X) = I(T;Y,Z|X) = 0$. □

Appendix B
Alternative Derivation of Lower Bound in Theorem 3

Following [10], we discretize the continuous-valued coefficients and thus create parallel subchannels, one for each quantized state. The number of parallel subchannels increases as the quantization becomes finer. In what follows, we only quantize the magnitude of the fading coefficients. The receiver can always rotate the phase, so it plays no part.

We quantize the channel gains into one of the q values

$$A_1 = 0 < A_2 < \cdots < A_q < A_{q+1} = J.$$

(Any slot where the channel gain of any receiver exceeds J is simply skipped). Receiver k is in state $l \in \{1, 2, \ldots, q\}$ at time t if $A_l \leq |H_k(t)|^2 < A_{l+1}$. When in state l, the receiver's channel gain is pessimistically discretized to $\sqrt{A_l}$. Since there are K independent receivers, there are a total of $M = q^K$ possible super-states, which we number as $S_1, S_2, \ldots, S_M$. Denote the quantized gain of receiver k in S_m by the double subscript S_{km}. Let $p(S_m)$ denote the probability of state S_m. Also let $p_k(A_l)$ be the probability that a receiver k is in state l, i.e., $p_k(A_l) = \sum_{k=1:S_{kk}=A_l}^{M} p(S_k)$. In super-state S_m, the channel of receiver k and the eavesdropper are

$$\begin{aligned} y_{km}(t) &= \sqrt{S_{km}}\, x(t) + z_k(t) \\ y_{\mathrm{e}m}(t) &= H_{\mathrm{e}}(t)\, x(t) + z_{\mathrm{e}}(t). \end{aligned}$$

By selecting $U_m \sim \mathcal{CN}(0,P)$ and $X_m = U_m$, the argument in the summation in (56) (with the eavesdropper output $(Y_{\mathrm{e}m}, H_{\mathrm{e}})$) is

$$\begin{aligned} & \{I(U_m;Y_{km}) - I(U_m;Y_{\mathrm{e}m},H_{\mathrm{e}})\}^+ \\ &= \{I(X_m;Y_{km}) - I(X_m;Y_{\mathrm{e}m},H_{\mathrm{e}})\}^+ \\ &= \{I(X_m;\sqrt{S_{km}}X_m + Z_k) - I(X_m;H_{\mathrm{e}}X_m + Z_{\mathrm{e}},H_{\mathrm{e}})\}^+ \\ &= \{\log(1 + S_{km}P) - E[\log(1 + |H_{\mathrm{e}}|^2 P]\}^+. \end{aligned}$$

Substituting into (56), we have that we can achieve rate

$$\begin{aligned} & \bar{R}_Q(P) \qquad (102) \\ &= \min_{k \in \{1,\ldots,K\}} \sum_{m=1}^{M} p(S_m) \left\{ \log\left(\frac{1 + S_{km}P}{\exp\{E[\log(1 + |H_{\mathrm{e}}|^2 P)]\}} \right) \right\}^+ \qquad (103) \\ &= \min_{k \in \{1,\ldots,K\}} \sum_{l=1}^{q} p_k(A_l) \left\{ \log\left(\frac{1 + A_l P}{\exp\{E[\log(1 + |H_{\mathrm{e}}|^2 P)]\}} \right) \right\}^+ \qquad (104) \end{aligned}$$

where the second equality follows from rewriting the summation over the states of each individual user. By taking $q \to \infty$ (with J fixed), and invoking the dominated convergence theorem (Appendix C), (104) converges to

$$\begin{aligned} & \min_{k \in \{1,\ldots,K\}} \int_0^J \left\{ \log\left(\frac{1 + xP}{\exp\{E[\log(1 + |H_{\mathrm{e}}|^2 P)]\}} \right) \right\}^+ p_k(x)\, dx \\ &= \min_{k \in \{1,\ldots,K\}} \int_0^\infty \left\{ \log\left(\frac{1 + xP}{\exp\{E[\log(1 + |H_{\mathrm{e}}|^2 P)]\}} \right) \right\}^+ p_k(x)\, dx \\ &\quad - \int_J^\infty \left\{ \log\left(\frac{1 + xP}{\exp\{E[\log(1 + |H_{\mathrm{e}}|^2 P)]\}} \right) \right\}^+ p_k(x)\, dx. \end{aligned} \tag{105}$$

To establish $\bar{R}^-(P)$ in (16a), it remains to show that for J sufficiently large, the second integral above is arbitrarily small. This follows since

$$\int_0^\infty \{\log(1 + xP) - E[\log(1 + |H_{\mathrm{e}}|^2 P)]\}^+ p_k(x)\, dx < \infty$$

and hence we have that, for any $\epsilon > 0$ and for J sufficiently large

$$\int_J^\infty \left\{ \log\left(\frac{1 + xP}{\exp\{E[\log(1 + |H_{\mathrm{e}}|^2 P)]\}} \right) \right\}^+ p_k(x)\, dx < \epsilon. \tag{106}$$

Appendix C
Upper Bound on $f_P(x)$

Claim 1: Suppose that $H_{\mathrm{e}} \sim \mathcal{CN}(0,1)$. For all $P > 0$, the function

$$f_P(x) \triangleq \left\{\log(1 + xP) - E[\log(1 + |H_{\mathrm{e}}|^2 P)]\right\}^+ \tag{107}$$

defined in (63) is bounded according to $f_P(x) \leq g(x)$, where $g(x) = \log(1 + x) + \gamma/\log 2$.

Proof: First suppose that $P < 1$. In this case

$$\begin{aligned} f_P(x) &= \left\{\log(1 + xP) - E[\log(1 + |H_{\mathrm{e}}|^2 P)]\right\}^+ \\ &\leq \log(1 + xP) \\ &\leq \log(1 + x) < g(x). \end{aligned}$$

When $P \geq 1$, we have

$$\begin{aligned} f_P(x) &= \left\{\log(1 + xP) - E[\log(1 + |H_{\mathrm{e}}|^2 P)]\right\}^+ \\ &\leq \left\{\log(1 + xP) - E[\{\log(|H_{\mathrm{e}}|^2 P)\}^+]\right\}^+ \\ &\leq \left\{\log(1 + xP) - \{E[\log(|H_{\mathrm{e}}|^2 P)]\}^+\right\}^+ \\ &= \left\{ \log(1 + xP) - \left\{ -\frac{\gamma}{\log 2} + \log P \right\}^+ \right\}^+ \end{aligned}$$

$$\begin{aligned} &\leq \left\{\log(1+xP) + \frac{\gamma}{\log 2} - \log P\right\}^+ \\ &\leq \log(1+x) + \frac{\gamma}{\log 2} = g(x) \end{aligned} \tag{108}$$

where (108) follows from the fact that the function $\{\cdot\}^+$ is convex, so by Jensen's inequality $E[\{v\}^+] \geq \{E[v]\}^+$. □

Since $f_P(x) \geq 0$, and since $E[g(|H|^2)] < \infty$, the dominated convergence applies to $f_P(x)$.

Appendix D
Convergence Claim in Section VIII-A

For each J fixed, we need to show that

$$\begin{aligned} &R_K^-(P) \\ &= \lim_{q\to\infty}\sum_{l=1}^{q} \Pr(A_l)\left\{\log\left(\frac{1+A_l\rho(A_l)}{\exp\{E[\log(1+|H_\mathrm{e}|^2\rho(A_l))]\}}\right)\right\}^+ \\ &= \int_0^J \left\{\log\left(\frac{1+a\rho(a)}{\exp\{E[\log(1+|H_\mathrm{e}|^2\rho(a))]\}}\right)\right\}^+ p(a)\, da. \end{aligned} \tag{109}$$

In turn, defining, for $a \in [A_l, A_{l+1}]$

$$f_q(a) = \left\{\log\left(\frac{1+A_l\rho(A_l)}{\exp\{E[\log(1+|H_\mathrm{e}|^2\rho(A_l))]\}}\right)\right\}^+ \tag{110}$$

we have that

$$\begin{aligned} \sum_{l=1}^{q} \Pr(A_l) f_q(A_l) &= \sum_{l=1}^{q} \int_{A_l}^{A_{l+1}} p(a) f_q(a)\, da \\ &= \int_0^J p(a) f_q(a)\, da \end{aligned} \tag{111}$$

and the convergence claim (109) follows since

$$\begin{aligned} &\lim_{q\to\infty} \sum_{l=1}^{q} \Pr(A_l) f_q(A_l) \\ &= \lim_{q\to\infty} \int_0^J p(a) f_q(a)\, da, \\ &= \int_0^J p(a) \lim_{q\to\infty} f_q(a)\, da, \qquad (112) \\ &= \int_0^J p(a) \left\{\log\left(\frac{1+a\rho(a)}{\exp\{E[\log(1+|H_\mathrm{e}|^2\rho(a))]\}}\right)\right\}^+ da \qquad (113) \end{aligned}$$

where the order of limit and integration can be interchanged since $f_q(\cdot)$ in (110) satisfies the dominated convergence Theorem (cf. Appendix C).

Acknowledgment

The authors wish to thank the reviewers for helpful suggestions that enhanced and clarified the paper.

References

[1] M. Adams and V. Guillemin, *Measure Theory and Probability*. Boston, MA: Birkäuser, 1996.

[2] J. Barros and M. R. D. Rodrigues, "Secrecy capacity of wireless channels," in *Proc. Int. Symp. Information Theory*, Seattle, WA, Jul. 2006, pp. 356–360.

[3] G. Caire and S. Shamai (Shitz), "On the capacity of some channels with channel state information," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2007–2019, Sep. 1999.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[5] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.

[6] H. A. David, *Order Statistics*. New York: Wiley, 1981.

[7] W. Diffie and M. E. Hellman, "New directions in cryptography," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 6, pp. 644–654, Nov. 1976.

[8] A. A. El Gamal, "Capacity of the product and sum of two un-matched broadcast channels," *Probl. Inf. Transm.*, pp. 3–23, 1980.

[9] A. Fiat and M. Naor, "Broadcast encryption," in *Proc. 13th Annu. Int. Cryptology Conf. Advances in Cryptology*, Santa Barbara, CA, 1994, pp. 480–491.

[10] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[11] P. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inf. Theory*, submitted for publication.

[12] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting with multiuser diversity," in *Proc. 44th Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sep. 2006.

[13] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 4, pp. 451–456, Jul. 1978.

[14] L. Li and A. J. Goldsmith, "Optimal resource allocation for fading broadcast channels—Part I: Ergodic capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 1083–1102, Mar. 2001.

[15] Z. Li, R. Yates, and W. Trappe, "Secrecy capacity of independent parallel channels," in *Proc. 44th Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sep. 2006.

[16] Z. Li, R. Yates, and W. Trappe, "Secret communication with a fading eavesdropper channel," in *Proc. Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 1296–1300.

[17] Y. Liang and H. V. Poor, "Secure communication over fading channels," in *Proc. 44th Annu. Allerton Conf. Communication, Control and Computing*, Monticello, IL, Sep. 2006.

[18] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 2, pp. 733–742, Mar. 1993.

[19] U. M. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," in *Proc. EUROCRYPT (Lecture Notes in Computer Science)*. Berlin, Germany: Springer-Verlag, 2000, vol. 1807, pp. 351–368.

[20] R. Negi and S. Goel, "Secret communication using artificial noise," in *Proc. IEEE Vehicular Technology Conf.*, Sep. 2005, vol. 3, pp. 1906–1910.

[21] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, pp. 656–715, 1949.

[22] A. Thangaraj, S. Dihidar, A. R. Calderbank, S. W. McLaughlin, and J. M. Merolla, "Applications of ldpc codes to the wiretap channel," *IEEE Trans. Inf. Theory*, submitted for publication.

[23] D. N. C. Tse, "Optimal power allocation over parallel gaussian broadcast channels,", unpublished.

[24] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[25] A. D. Wyner, "The wiretap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, 1975.