Delay-Intolerant Covert Communications with Either Fixed or Random Transmit Power

Shihao Yan, *Member, IEEE,* Biao He, *Member, IEEE,* Xiangyun Zhou, *Senior Member, IEEE,* Yirui Cong, *Member, IEEE,* and A. Lee Swindlehurst, *Fellow, IEEE*

***Abstract*—In this work, we study delay-intolerant covert communications in additive white Gaussian noise (AWGN) channels with a finite blocklength, i.e., a finite number of channel uses. Considering the maximum allowable number of channel uses to be N, it is not immediately clear whether the actual number of channel uses, denoted by n, should be as large as N or smaller for covert communications. This is because a smaller n reduces a warden's chance to detect the communications due to fewer observations, but also reduces the chance to transmit information. We show that $n = N$ is indeed optimal to maximize the amount of information bits that can be transmitted, subject to any covert communication constraint in terms of the warden's detection error probability. To better make use of the warden's uncertainty due to the finite blocklength, we also propose to use uniformly distributed random transmit power to enhance covert communications. Our examination shows that the amount of information that can be covertly transmitted logarithmically increases with the number of random power levels, which indicates that most of the benefit of using random transmit power is achieved with just a few different power levels.**

***Index Terms*—Covert communications, delay-intolerant, finite blocklength, covertness, random transmit power.**

I. Introduction

As people become more dependent on wireless devices to share private information, crucial concerns about the security and privacy of wireless communications are emerging since a large amount of confidential information (e.g., email/bank account information and password, credit card details) is transferred over wireless networks [1, 2]. In addition to the secrecy and integrity of the transmitted information, in some scenarios a user may wish to transmit messages over wireless networks without being detected. This is due to the fact that the exposure of the transmission may disclose the user's location, which may violate the privacy of the user. Against this background, as the line of last defence, hiding wireless transmissions meets the ever-increasing desire of strong security and privacy, which is also explicitly desired by government and military bodies (e.g., a stealth fighter desires to hide itself from enemies while communicating with its own military bases).

In fact, hiding wireless communications was addressed by spread spectrum techniques in the early 20th century and a review of spread spectrum techniques can be found in [4]. However, the fundamental limits of hiding wireless communications have never been fully examined in terms of the amount of information that can be conveyed covertly without being detected. Consequently, the covertness achieved by spread spectrum techniques has never been proven theoretically and thus the ability of spread spectrum to hide wireless transmissions is unknown. As such, cutting-edge research on wireless communication security has called for a rethinking and generalisation of spread spectrum at a more fundamental level, which has inspired the emergence of a new security paradigm termed covert communications (e.g., [5–7]).

In covert communications, a transmitter (Alice) intends to communicate with a legitimate receiver (Bob) without being detected by a warden (Willie), who is observing this communication. Considering additive white Gaussian noise (AWGN) channels, a square root law has been derived in [8], which states that Alice can transmit no more than $\mathcal{O}(\sqrt{n})$ bits in n channel uses covertly and reliably to Bob. Following [8], the scaling constant of the amount of information with respect to the square root of n was characterized for a broad class of discrete memoryless channels (DMCs) and AWGN channels in [9]. We note that this square root law requires a pre-shared secret to be established between Alice and Bob prior to Alice's transmission. This pre-shared secret is proven to be unnecessary for the square root law when the channel quality from Alice to Bob is higher than that from Alice to Willie, for the binary symmetric channel (BSC) [10], DMC [11], and AWGN channel [11]. Specifically, it is shown that keyless covert communications without a pre-shared secret are achievable when the quality of the channel from the transmitter to the receiver is higher than that of the channel from the transmitter to the warden [12].

In the square root law we have $\mathcal{O}(\sqrt{n})/n \to 0$ as $n \to \infty$, which indicates that the covert rate is asymptotically zero. That is, the average number of bits that can be covertly and reliably transmitted per channel use asymptotically approaches zero. However, in some scenarios a positive rate has been shown to be achievable (e.g., [10, 13–16]). For example, it is proved that a positive rate can be obtained when Willie has uncertainty about the receiver noise variance in AWGN channels [14, 16], when Willie does not exactly know the receiver noise model in BSC channels [10], or when Willie

S. Yan is with the School of Engineering, Macquarie University, Sydney, NSW 2109, Australia (e-mail: shihao.yan@mq.edu.au).

B. He and A. L. Swindlehurst are with the Center for Pervasive Communications and Computing, University of California, Irvine, CA 92697 USA (e-mails: {biao.he, swindle}@uci.edu).

X. Zhou is with Research School of Engineering, Australian National University, Canberra, ACT 2601, Australia (e-mail: xiangyun.zhou@anu.edu.au).

Y. Cong is with the College of Intelligence Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China (e-mail: congyirui11@nudt.edu.cn).

This work was partially supported by the Australian Research Council's Discovery Projects (DP180104062). Part of this work has been presented in IEEE ICC 2017 [3].

lacks knowledge of his channel characteristics in AWGN and block fading channels [15, 16]. In addition to noise or channel uncertainty, the impact of unknown transmission times or uniformed jammers on covert communications was examined in [17] and [18], respectively. Furthermore, covert communications with channel uncertainty in fading channels was studied in [19], where Willie's detection performance limits and the achieved covert rates were analyzed. Besides the fundamental limit of covert communications, some work has focused on constructing practical encoding schemes and characterising the required key size in order to achieve the covert communication limits (e.g., [11, 12, 20]). The authors of [21] characterized the second order asymptotics of the number of bits in binary input DMCs that can be transmitted from Alice to Bob subject to some constraints on the probability of error and the divergence between the channel output distributions induced with and without covert communications. Their results provide useful guidelines on how to expurgate a random code while maintaining the channel resolvability properties. Covert communications over parallel Gaussian channels were studied in [22], where the author focused on deriving an optimal power allocation strategy to maximize the achievable number of transmitted information bits subject to some specific constraints on Bob's decoding error probability and Willie's detection error probability.

In the literature of covert communications, only [14] mentions the impact of a finite n (i.e., a limited delay) on the detection performance at Willie. It is numerically shown that with noise uncertainty at Willie there may exist an optimal n that maximizes the rate from Alice to Bob subject to the covert communication constraint $\xi \geq 1-\epsilon$, where ξ is Willie's detection error probability (i.e., the sum of false positive and missed detection rates) and ϵ is a predetermined small number. Besides the detection performance at Willie, the assumption of a finite n also has a significant impact on the achievable rate R of the transmission from Alice to Bob [23], which has not been considered in the context of covert communications. Thus, the impact of finite n on covert communications has never been fully examined. We note that this impact cannot be intuitively revealed since, as the blocklength n increases, although Alice has more time slots over which to spread the transmit power, Willie will also have more time slots to collect observations in order to detect the covert transmission. In particular, although the achievable channel coding rate from Alice to Bob increases with n [23], a fact which aids the covert communication from Alice to Bob, the sufficient test statistic, i.e., the average power of each received symbol, at Willie converges to a constant value (i.e., the uncertainty in the test statistic decreases), enabling him to make a more reliable detection. Thus, it is not immediately obvious that n should take the largest possible value N. In the limit where N approaches infinity, Willie will have exact knowledge of the received power and we cannot hide any transmitted signal whose power will increase the received power at Willie beyond the noise power value that he already knows. This is confirmed by our prior works (e.g., [19, 24–26]), which have explicitly indicated that the optimal detection performance at Willie is independent of the known noise power in the asymptotic scenario with $N \to \infty$. This limit intuitively shows that making n as large as possible is not necessarily the optimal strategy, and motivates studying the impact of a finite n on the performance of covert communications, which has not been previously clarified. We note that the square-root law states that $\mathcal{O}(\sqrt{n})$ bits can be transmitted covertly and reliably in n channel uses, which seems to suggest that n should be as large as possible in the context of covert communications. However, the proof of the square-root law in [8] is an asymptotic argument for large enough n, and it does not establish the behavior of the number of information bits that can be transferred when $n \leq N$, where N is finite and possibly small. This situation has not been addressed in the prior literature. Therefore, the established square-root law cannot be directly applied to conclude that n should be as large as possible with the constraint $n \leq N$. This leaves a significant gap in our understanding on the fundamental limit of covert communications in some practical application scenarios, since in some scenarios the length of a codeword is always finite. For example, to achieve transmission efficiency (e.g., short delay) we may require the codeword to be short, e.g., the order of 100 channel uses for vehicle-to-vehicle communication or real-time video processing [27].

In this work, we first study the impact of finite n on both the achievable rate from Alice to Bob and the detection performance at Willie for AWGN channels, which allows us to explicitly examine the effect of n on covert communications. We then propose to use uniformly distributed random transmit power to enhance the performance of covert communications with a finite blocklength. Our main contributions are summarized below.

- We consider covert communications with a maximum of N channel uses, and thus the actual number of channel uses n for the covert transmission from Alice to Bob is constrained by $n \leq N$. Although a larger n offers more observations to Willie for detecting the transmission, we analytically prove that the optimal value of n that maximizes the amount of information achieved in one block (denoted by η), subject to a specific covert communication constraint, is N. This result is consistent with the conclusion suggested by the square-root law in the limit of large n [8]. As such, this contribution can be regarded as a theoretical proof of the conclusion drawn from the square-root law under the practical constraint $n \leq N$ for finite and possibly small N, which is an extension of the square-root law.
- We characterize the relation between the maximum allowable transmit power per channel use, P^*, and the number of channel uses, n, subject to the covert communication constraint. Our examination shows that P^* decreases as n increases, which is due to the fact that increasing n forces Alice to allocate less power for each channel use to meet the covert communication constraint. Nevertheless, we show that the maximum allowable total transmit power (i.e., nP^*) increases as n increases, which indicates that the achievable η should increase with n. The results in this work, for the first time, provide important insights

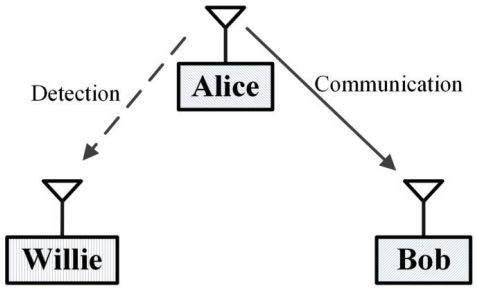

Fig. 1. Illustration of the system model for covert communications.

on the design of covert communications with a finite blocklength.

- Finally, we propose to use uniformly distributed random transmit power to enhance the performance of covert communications. Specifically, we study the impact of Alice using continuous uniform transmit power (CUTP) and discrete uniform transmit power (DUTP) on covert communications. Our results indicate that the achievable η can be significantly improved by CUTP and DUTP, especially when η achieved by using a fixed transmit power (FTP) is low. This result is different from the case with an infinite n, where adopting random transmit power does not facilitate covert communications. We also find that the η achievable by DUTP increases logarithmically with the number of transmit power levels M, which indicates that using a relatively small value of M can provide most of the achievable benefit.

The rest of this paper is organized as follows. Section II details the system model and performance metrics. Section III focuses on FTP and examines the impact of the finite blocklength n on covert communications. Section IV analyzes the performance of covert communications when CUTP or DUTP is adopted at Alice. Section V provides numerical results to confirm our analysis and compare the performance of FTP, CUTP, and DUTP. Finally, Section VI draws concluding remarks.

Notation: Scalar variables are denoted by italic symbols. Vectors and matrices are denoted by lower-case and upper-case boldface symbols, respectively. Given a vector $\mathbf{x}$, $x[i]$ denotes the i-th element of $\mathbf{x}$. The expectation operator is denoted by $\mathbb{E}\{\cdot\}$ and $\mathcal{CN}(0,\sigma^2)$ denotes the circularly-symmetric complex normal distribution with zero mean and variance σ^2.

II. System Model

A. Channel Model

The system model for covert communications is illustrated in Fig. 1, where each of Alice, Bob, and Willie is equipped with a single antenna. We assume the channel from Alice to Bob and the channel from Alice to Willie are only subject to AWGN. In covert communications, Alice transmits n complex-valued symbols $x[i]$ $(i = 1, 2, \ldots, n)$ to Bob, while Willie is passively collecting n observations on Alice's transmission to detect whether or not Alice has transmitted signals to Bob. In this work, we assume that the signals are constrained by a maximum blocklength denoted by N, which implies that $n \leq N$. We denote the AWGN at Bob and Willie as $r_b[i]$ and $r_w[i]$, respectively, where $r_b[i] \sim \mathcal{CN}(0,\sigma_b^2)$, $r_w[i] \sim \mathcal{CN}(0,\sigma_w^2)$, and σ_b^2 and σ_w^2 are the noise variances at Bob and Willie, respectively. In addition, we assume that $x[i]$, $r_b[i]$, and $r_w[i]$ are mutually independent. The transmit power of Alice for each block (n channel uses) is denoted as P, i.e., we have $\mathbb{E}\{|x[i]|^2\} = P$. We consider that P is fixed for different blocks in Section II and P is uniformly distributed over different blocks in Section III. We assume that Willie is aware of the value of P when it is fixed. When the value of P for each block is chosen randomly from either a CUTP or DUTP distribution, Willie is not aware of the actual value of P, but knows its distribution. We assume that Bob is always aware of the value of P by means of a shared secret key. The consequence of this assumption for both the CUTP and DUTP will be addressed later. Furthermore, we assume that Alice adopts Gaussian signaling, i.e., $x[i] \sim \mathcal{CN}(0,P)$.

B. Binary Hypothesis Testing at Willie

In order to detect the presence of covert communications, Willie must distinguish between the following two hypotheses:

$$\begin{cases} \mathcal{H}_0: \ y_w[i] = r_w[i], \\ \mathcal{H}_1: \ y_w[i] = x[i] + r_w[i], \end{cases} \tag{1}$$

where $\mathcal{H}_0$ denotes the null hypothesis where Alice has not transmitted signals, $\mathcal{H}_1$ denotes the alternative hypothesis where Alice has transmitted, and $y_w[i]$ is the received signal at Willie.

In this work, we adopt the total detection error probability to measure Willie's performance, which is defined as

$$\xi = \alpha + \beta, \tag{2}$$

where $\alpha \triangleq \Pr(\mathcal{D}_1|\mathcal{H}_0)$ is the false positive rate, $\beta \triangleq \Pr(\mathcal{D}_0|\mathcal{H}_1)$ is the missed detection rate, and $\mathcal{D}_1$ and $\mathcal{D}_0$ are the binary decisions that infer whether Alice's transmission is present or not, respectively. In covert communications, Willie's ultimate goal is to detect the presence of Alice's transmission with the minimum detection error probability ξ^*, which is achieved by using the optimal detector that minimizes ξ. Then, the covert communication constraint can be expressed as $\xi^* \geq 1-\epsilon$ for a given ϵ. Normally, the value of ϵ is small in order to provide good covertness.

C. Effective Throughput for a Finite Blocklength

The received signal at Bob for each symbol is

$$y_b[i] = x[i] + r_b[i]. \tag{3}$$

As pointed out by [23], the decoding error probability at Bob is not negligible when n is finite. As such, for a given decoding error probability δ the channel coding rate from Alice to Bob can be approximated by [23, 28]

$$R \approx \log_2(1+\gamma_b) - \sqrt{\frac{\gamma_b(\gamma_b+2)}{n(\gamma_b+1)^2}}\frac{Q^{-1}(\delta)}{\ln(2)}, \tag{4}$$

where $\gamma_b = P/\sigma_b^2$ is the signal-to-noise ratio (SNR) at Bob, and $Q^{-1}(\cdot)$ is the inverse Q-function. In this work we do not consider a specific code that can achieve the channel coding rate given in (4). Instead, we consider Gaussian signaling (i.e., $x[i] \sim \mathcal{CN}(0, P)$) as mentioned in Section II-A. We note that while the channel coding rate expression in (4) was derived using a second-order asymptotic analysis, it is based on the Normal approximation in [23], which has been shown to be near-optimal for reasonably long block lengths (e.g., $n \geq 100$) and for the Gaussian signaling assumption with average power constraints that we make in this work.

The square root law states that Alice can covertly and reliably transmit no more than $\mathcal{O}(\sqrt{n})$ bits in n channel uses to Bob. Such scaling-law results are obtained when $n \to \infty$. As such, these results cannot be applied in covert communications with a finite n. In this work, we examine the amount of information that can be transmitted reliably from Alice to Bob for a given ϵ. Noting that the decoding error probability of a channel with a finite n is not negligible, we adopt the amount of information bits η that can be transmitted from Alice to Bob as the main performance metric for covert communications with a finite blocklength, while utilizing $\xi^* \geq 1 - \epsilon$ as the constraint. Mathematically, η is given by

$$\eta = nR(1 - \delta). \tag{5}$$

We note that η quantifies the expected number of information bits that can be reliably transmitted from Alice to Bob, excluding information bits that suffer from decoding errors. In this work, we assume that the channel coding rate is determined for a given decoding error probability as in (4), i.e., we do not consider the optimization of R or δ to maximize η, although they both are functions of n. The reason will be discussed later.

The ultimate goal of our covert communication design is to achieve the maximum η while guaranteeing the covert communication constraint $\xi^* \geq 1 - \epsilon$. However, due to the complicated expression of R given in (4) this maximum η is mathematically intractable. As such, in this work we first focus on the design of the number of channel uses and Alice's transmit power in order to maximize the total transmit power nP subject to $\xi^* \geq 1-\epsilon$ in Section III and then examine adopting uniformly distributed random transmit power to facilitate covert communications in terms of improving η in Section IV.

III. Fixed Transmit Power and Optimization of the Number of Channel Uses

In this section, we consider FTP where Alice's transmit power P is fixed and known by Willie. Specifically, we determine the optimal n and P that maximize nP subject to the covert communication constraint $\xi^* \geq 1 - \epsilon$.

A. Detection Performance at Willie

For FTP, Alice's transmit power P is fixed and is known by Willie. As such, the optimal test that minimizes ξ is the likelihood ratio test with $\lambda^* = 1$ as the threshold, which is given by

$$\frac{\mathbb{P}_1 \triangleq \prod_{i=1}^{n} f\left(y_w[i]|\mathcal{H}_1\right)}{\mathbb{P}_0 \triangleq \prod_{i=1}^{n} f\left(y_w[i]|\mathcal{H}_0\right)} \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} 1, \tag{6}$$

where $f(y_w[i]|\mathcal{H}_0) = \mathcal{CN}(0, \sigma_w^2)$ and $f(y_w[i]|\mathcal{H}_1) = \mathcal{CN}(0, P + \sigma_w^2)$ are the likelihood functions of $y_w[i]$ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. We note that $\lambda^* = 1$ is due to the assumption of unknown or equal *a priori* probabilities for $\mathcal{H}_0$ and $\mathcal{H}_1$, denoted respectively by P_0 and P_1. If both P_0 and P_1 are known, Willie's detection error probability is reformulated as $\xi = P_0\alpha + P_1\beta$ and the optimal test that minimizes this reformulated ξ is the likelihood ratio test with $\lambda^* = P_1/P_0$ as the threshold. We also note that the assumption of equal *a priori* probabilities is commonly adopted in the literature of covert communications (e.g., [8, 14, 17]).

After performing some algebraic manipulations, (6) can be reformulated as

$$T \triangleq \frac{1}{n}\sum_{i=1}^{n} |y_w[i]|^2 \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \Gamma^*, \tag{7}$$

where the test statistic T is the average power of each received symbol at Willie and Γ^* is the optimal threshold for T, which is given by

$$\Gamma^* = \frac{(P + \sigma_w^2)\sigma_w^2}{P} \ln\left(\frac{P + \sigma_w^2}{\sigma_w^2}\right). \tag{8}$$

As per (6) and (7), we note that the radiometer with Γ^* is indeed the optimal detector when Willie knows the likelihood functions exactly (i.e., when there are no nuisance parameters embedded in the likelihood functions). Following (7) and noting that T is a chi-squared random variable with $2n$ degrees of freedom, the likelihood functions of T under $\mathcal{H}_0$ and $\mathcal{H}_1$ are given by

$$f(T|\mathcal{H}_0) = \frac{T^{n-1}}{\Gamma(n)}\left(\frac{n}{\sigma_w^2}\right)^n e^{-\frac{nT}{\sigma_w^2}}, \tag{9}$$

$$f(T|\mathcal{H}_1) = \frac{T^{n-1}}{\Gamma(n)}\left(\frac{n}{P + \sigma_w^2}\right)^n e^{-\frac{nT}{P+\sigma_w^2}}, \tag{10}$$

where $\Gamma(n) = (n-1)!$ is the Gamma function. Then, following (9) and (10), for an arbitrary threshold Γ the false positive rate and missed detection rate are given by [13, 14]

$$\alpha = \Pr(T > \Gamma|\mathcal{H}_0) = 1 - \frac{\gamma\left(n, \frac{n\Gamma}{\sigma_w^2}\right)}{\Gamma(n)}, \tag{11}$$

$$\beta = \Pr(T < \Gamma|\mathcal{H}_1) = \frac{\gamma\left(n, \frac{n\Gamma}{P+\sigma_w^2}\right)}{\Gamma(n)}, \tag{12}$$

where $\gamma(\cdot, \cdot)$ is the lower incomplete Gamma function given by $\gamma(n, x) = \int_0^x e^{-t}t^{n-1}dt$. We note that Willie's minimum detection error probability ξ^* can be obtained by substituting Γ^* as given in (8) into (11) and (12).

B. Optimal Blocklength n and Alice's Transmit Power P

When FTP is adopted at Alice, the total transmit power over n channel uses is nP, which is flexible and under the control of Alice (i.e., both n and P are design parameters to determine). Thus, covert communications for any positive ϵ are feasible and thus a positive nP is achievable.

Although ξ^* can be obtained via an analytical expression as shown in the previous subsection, it is hard to use for further analysis due to its use of lower incomplete Gamma functions. As such, following Pinsker's inequality we adopt a lower bound on ξ^* in this section in order to analytically determine the optimal n and P. This lower bound is given by [8, 29, 30]

$$\xi^* \geq 1 - \sqrt{\frac{1}{2}\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1)}, \tag{13}$$

where $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1)$ is the Kullback-Leibler (KL) divergence from $\mathbb{P}_0$ to $\mathbb{P}_1$, which can be expressed as

$$\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) = n\left[\ln\left(\frac{P+\sigma_w^2}{\sigma_w^2}\right) - \frac{P}{P+\sigma_w^2}\right]. \tag{14}$$

As per (13), we can ensure $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ in order to guarantee $\xi^* \geq 1-\epsilon$. We also note that $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ is a stricter constraint relative to $\xi^* \geq 1-\epsilon$. From a conservative point of view, we adopt $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ as the covert communication constraint for FTP in this section to obtain analytical insights and we will numerically confirm in Section V that using $\xi^* \geq 1-\epsilon$ as the constraint leads to similar results. As such, the optimization of n and P in covert communications with FTP at Alice can be written as

$$\underset{n,P}{\text{argmax}} \quad nP, \tag{15a}$$

$$\text{s.t.} \quad \mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2, \tag{15b}$$

$$n \leq N. \tag{15c}$$

We present the solution to (15) in the following theorem.

***Theorem** 1:* The optimal values of n and P that maximize the total power nP subject to $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ and $n \leq N$, are, respectively, given by

$$n^* = N, \tag{16}$$

$$P^* = (\sigma_w^2 + P^*)\left[\ln\left(\frac{P^*}{\sigma_w^2}+1\right) - 2\epsilon^2 N\right], \tag{17}$$

where P^* is the solution to the fixed-point equation (17).

Proof: The detailed proof is provided in Appendix A. ■

Based on Theorem 1, we see that it is best for Alice to transmit over all the available channel uses for covert communications, provided that the transmit power is optimized to maintain the same level of covertness despite the fact that Willie has more observations when n is larger. The same level of covertness is achieved by reducing the transmit power when n becomes larger. It is interesting to observe that neither n^* nor P^* is related to R or δ. This demonstrates that the obtained n^* and P^* are globally optimal, regardless of the value of the channel coding rate R or decoding error probability δ. This is the main reason why we do not consider the optimization of R or δ in this work. We denote the maximum value of η as η^*, which is achieved by substituting P^* and n^* into (5).

IV. Continuous and Discrete Uniform Transmit Power with All Channel Uses

In this section, we focus on the case with CUTP or DUTP at Alice, where P is drawn from a continuous or discrete uniform distribution, respectively. We adopt uniform distributions since they provide a natural way of bounding the random transmit power between zero and the maximum power, but it is possible that other distributions may improve the performance of covert communications. We leave this interesting issue for future research. We still assume the average transmit power constraint (i.e., $\mathbb{E}\{|x[i]|^2\} = P$) for each block, i.e., P is fixed for each entire block with n channel uses. Following the previous section, we set $n = N$ in this section, which is optimal in terms of maximizing η for both CUTP and DUTP, since it is optimal for any realization of P. Specifically, we examine the detection performance at Willie, based on which we determine the maximum η achieved by CUTP and DUTP subject to the covert communication constraint $\xi^* \geq 1-\epsilon$.

A. Covert Communication Constraint and Optimal Detector at Willie

In this section, we use $\xi^* \geq 1-\epsilon$ directly instead of $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ as the covert communication constraint. This is due to the fact that Willie does not know each realization of P although he knows the distribution. As a result, Willie has to fix his detection threshold for all the realizations of P and thus the averaged KL divergence $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1)$ over all realizations of P is no longer a tight bound on Willie's detection error probability (it is only tight when Willie knows each realization of P and varies his detection threshold accordingly).

We note that CUTP or DUTP does not affect the likelihood function under $\mathcal{H}_0$ (i.e., $\mathbb{P}_0$). Since the transmit power P in CUTP or DUTP is a random variable with a specific *a priori* probability $f_P(p)$, following (6) the likelihood function under $\mathcal{H}_1$ can be written as

$$\begin{aligned}\mathbb{P}_1 &= \int \prod_{i=1}^{N} f\left(y_w[i]|p,\mathcal{H}_1\right) f_P(p)dp \\ &= \int \prod_{i=1}^{N} \frac{\exp\left(-\frac{|y_w[i]|^2}{2(p+\sigma_w^2)}\right)}{\sqrt{2\pi(p+\sigma_w^2)}} f_P(p)dp, \\ &= \int \frac{\exp\left(-\frac{\sum_{i=1}^{N}|y_w[i]|^2}{2(p+\sigma_w^2)}\right)}{(2\pi(p+\sigma_w^2))^{N/2}} f_P(p)dp. \end{aligned} \tag{18}$$

As per (18), we note that $\mathbb{P}_1$ depends on $y_w[i]$ only through $\sum_{i=1}^{N}|y_w[i]|^2$, no matter the explicit expression of $f_P(p)$. As such, based on the Fisher-Neyman factorization theorem [31, Theorem 7.7.1], we can conclude that with CUTP or DUTP at Alice, the average power of each received symbol T given in (7) is still the sufficient test statistic. As such, the detector at Willie for an arbitrary threshold is given by

$$T = \frac{1}{N}\sum_{i=1}^{N}|y_w[i]|^2 \underset{\mathcal{D}_0}{\overset{\mathcal{D}_1}{\gtrless}} \Gamma. \tag{19}$$

Then, Willie has to find the optimal value of Γ that minimizes the detection error probability, since this optimal value is not

the same as that given in (8). To this end, we next derive the false positive and missed detection rates for CUTP and DUTP in the following subsections.

B. Continuous Uniform Transmit Power

In this subsection, we consider CUTP, in which Alice's transmit power P follows a continuous uniform distribution with probability density function (pdf)

$$f_P(x) = \begin{cases} \dfrac{1}{P_m} \text{ when } 0 < x \leq P_m \\ 0 \text{ when } x \geq P_m, \end{cases} \tag{20}$$

where P_m is Alice's maximum transmit power. With CUTP Alice first has to determine the value of P_m that maximizes η subject to $\xi^* \geq 1-\epsilon$. To this end, we first derive the false positive rate α and missed detection rate β in the following theorem.

***Theorem** 2:* For fixed maximum transmit power P_m and Willie's detection threshold Γ, the false positive and missed detection rates for CUTP are, respectively, given by

$$\alpha = 1 - \frac{\gamma\left(N, \frac{N\Gamma}{\sigma_w^2}\right)}{\Gamma(N)}, \tag{21}$$

$$\beta = \frac{1}{(N-1)P_m} \sum_{i=0}^{N-2} \left[\frac{P_m + \sigma_w^2}{i!} \gamma\left(i+1, \frac{N\Gamma}{P_m+\sigma_w^2}\right) - \frac{\sigma_w^2}{i!}\gamma\left(i+1, \frac{N\Gamma}{\sigma_w^2}\right)\right]. \tag{22}$$

Proof: The detailed proof is provided in Appendix B. ■

The optimal threshold that minimizes $\xi = \alpha + \beta$ can be solved by numerically using (21) and (22). Accordingly, the minimum detection error probability ξ^* can be determined as well. We next derive the expected channel coding rate achieved using CUTP for a fixed decoding error probability.

***Theorem** 3:* Taking the approximation in (4) as an equality, the expected channel coding rate achieved by Alice with CUTP for a fixed decoding error probability δ is derived as

$$\overline{R} = \frac{Q^{-1}(\delta)}{\gamma_{bm}\ln(2)\sqrt{N}} \left[2\tan^{-1}\sqrt{\frac{\gamma_{bm}}{2+\gamma_{bm}}} - \sqrt{\gamma_{bm}(2+\gamma_{bm})}\right] + \frac{1+\gamma_{bm}}{\gamma_{bm}}\log_2(1+\gamma_{bm}) - \frac{1}{\ln(2)}, \tag{23}$$

where $\gamma_{bm} = P_m/\sigma_b^2$.

Proof: The detailed proof is proved in Appendix C. ■

With the achieved ξ^* and $\overline{R}$, the optimal value of P_m for CUTP that maximizes η subject to the covert communication constraint can be obtained through

$$P_m^* = \underset{P_m}{\text{argmax}} \quad N\overline{R}(1-\delta), \tag{24a}$$

$$\text{s.t.} \quad \xi^* \geq 1-\epsilon. \tag{24b}$$

We denote the maximum η as η^*, which is achieved by substituting P_m^* into $\eta = N\overline{R}(1-\delta)$. We note that P_m^* is the solution to $\xi^* \geq 1-\epsilon$ since ξ^* is a monotonically decreasing function of P_m while $\overline{R}$ is a monotonically increasing function of P_m. This is confirmed by the following fact. The expected transmit power for CUTP over different realizations of P at Alice is given by

$$\overline{P}_t = \int_0^{P_m} \frac{x}{P_m} dx = \frac{P_m}{2}, \tag{25}$$

which monotonically increases with P_m. As such, Alice's expected total transmit power $N\overline{P}_t$ monotonically increases with P_m, which indicates that it becomes easier for Willie to detect Alice's transmission (i.e., ξ^* decreases) as P_m increases since the radiometer is the optimal detector at Willie. This also demonstrates that the expected η should monotonically increase with P_m as discussed in the proof of Theorem 1.

Following Theorem 3, we note that in order to achieve the expected channel coding rate $\overline{R}$ given in (23) for a fixed δ, Alice has to vary R together with her transmit power, so that the value of R corresponding to each realization of P can be obtained as in (4). This means that not only the number of transmit power levels but also the number of potential codebooks required for covert communications from Alice to Bob should approach infinity for CUTP. We note that Bob has to know each R and each realization of Alice's transmit power, which should be kept secret from Willie. As such, the size of the shared secrets between Alice and Bob for CUTP also approaches infinity. As a result, CUTP is hard to implement in practice and thus in the following subsection we focus on the more practical DUTP strategy for Alice to use random transmit power, where CUTP serves as a benchmark.

C. Discrete Uniform Transmit Power

In this subsection, we consider DUTP at Alice, where the number of transmit power levels together with the number of codebooks and the shared secret size are quantized. Specifically, P follows a discrete uniform distribution, i.e., we have $P = P_i$ with probability $1/M$, where

$$P_i = \frac{iP_m}{M}, \quad i = 1, 2, \ldots, M, \tag{26}$$

P_m is Alice's maximum transmit power and M is the number of transmit power levels that Alice can set. While Willie is unaware of the chosen value of P, he does know the parameters P_m and M of its distribution. We also note that $P_i > 0$ since $P_i = 0$ corresponds to the null hypothesis $\mathcal{H}_0$ in which Alice does not transmit signals to Bob. Then, for DUTP with a fixed M Alice first has to determine the value of P_m that maximizes η subject to $\xi^* \geq 1-\epsilon$. To this end, we next present the false positive rate α and the missed detection rate β associated with DUTP.

***Theorem** 4:* For fixed P_m, M, and Γ, the false positive and missed detection rates for DUTP are, respectively, given by

$$\alpha = 1 - \frac{\gamma\left(N, \frac{N\Gamma}{\sigma_w^2}\right)}{\Gamma(N)}, \tag{27}$$

$$\beta = \frac{1}{M}\sum_{i=1}^{M} \frac{\gamma\left(N, \frac{NM\Gamma}{iP_m + M\sigma_w^2}\right)}{\Gamma(N)}. \tag{28}$$

Proof: We note that DUTP does not affect the likelihood function under $\mathcal{H}_0$ (i.e., $\mathbb{P}_0$). As such, following (11) the false

positive rate α for DUTP is given by (27). Following (10) and (26), the likelihood function of T under $\mathcal{H}_1$ for DUTP is given by

$$f(T|\mathcal{H}_1)=\frac{1}{M}\sum_{i=1}^{M}\frac{T^{N-1}}{\Gamma(N)}\left(\frac{NM}{iP_m+M\sigma_w^2}\right)^N e^{-\frac{NMT}{iP_m+M\sigma_w^2}}. \quad (29)$$

Thus, following (12), the missed detection rate for DUTP is given by (28). ■

The optimal threshold at Willie that minimizes ξ for DUTP can be found numerically using (27) and (28). Accordingly, ξ^* can be determined as well. We next present the expected channel coding rate achieved using DUTP for a fixed decoding error probability.

***Theorem** 5:* Taking the approximation in (4) as an equality, the expected channel coding rate achieved by DUTP for a fixed decoding error probability δ is derived as

$$\overline{R}=\frac{1}{M}\sum_{i=1}^{M}\left[\log_2(1+\gamma_{bi})-\sqrt{\frac{\gamma_{bi}(\gamma_{bi}+2)}{N(\gamma_{bi}+1)^2}}\frac{Q^{-1}(\delta)}{\ln(2)}\right], \quad (30)$$

where $\gamma_{bi}=iP_m/M\sigma_b^2$.

Proof: The result in (30) follows from (4) and (26). ■

With the obtained ξ^* and $\overline{R}$, the optimal value of P_m for DUTP can be written as

$$P_m^* = \underset{P_m}{\text{argmax}} \quad N\overline{R}(1-\delta), \quad (31a)$$

$$\text{s.t.} \quad \xi^* \geq 1-\epsilon. \quad (31b)$$

We denote the maximum η as η^*, which is achieved by substituting P_m^* into $\eta = N\overline{R}(1-\delta)$. We note again that P_m^* is the solution to $\xi^* = 1-\epsilon$, since ξ^* is a monotonically decreasing function of P_m. This can be observed directly from (28), which shows that for an arbitrary threshold Γ the missed detection rate is a monotonically decreasing function of P_m. As per (27), the false positive rate α is not a function of P_m. We also note that $\overline{R}$ is a monotonically increasing function of P_m. This can be confirmed by Alice's expected transmit power for DUTP over different realizations of P, given by

$$\overline{P}_t = \frac{1}{M}\sum_{i=1}^{M}P_i = \frac{P_m}{2}\left(1+\frac{1}{M}\right), \quad (32)$$

which monotonically increases with P_m.

We note that the missed detection rate for FTP given in (12) is a special case of that for DUTP given in (28) with $M=1$ and $P_m=P$. Also, the channel coding rate for FTP given in (4) is a special case of the expected channel coding rate for DUTP given in (30) with $M=1$ and $P_m=P$. As such, FTP can be interpreted as a special case of DUTP with $M=1$ and $P_m=P$. Not surprisingly, the results for CUTP can be obtained by letting $M\to\infty$ in DUTP, i.e., the missed detection rate given in (22) can be obtained from (28) with $M\to\infty$ and the channel coding rate given in (23) can be achieved from (30) with $M\to\infty$. Intuitively, ξ^* achieved by DUTP should increase with M since the system complexity increases with M. This is due to the fact that M is not only the number of transmit power levels that Alice has to support, but also the number of codebooks shared between Alice and Bob. In particular, Alice's R should vary with her transmit power in order to achieve the expected channel coding rate $\overline{R}$ for a fixed decoding error probability δ as discussed in the previous subsection. In addition, a larger M requires more secrets shared between Alice and Bob, since the index of the adopted transmit power level and the associated R should be agreed between Alice and Bob, which should be kept secret from Willie. As such, in the following section we numerically examine the tradeoff between the achievable covertness and the system complexity and shared secret size.

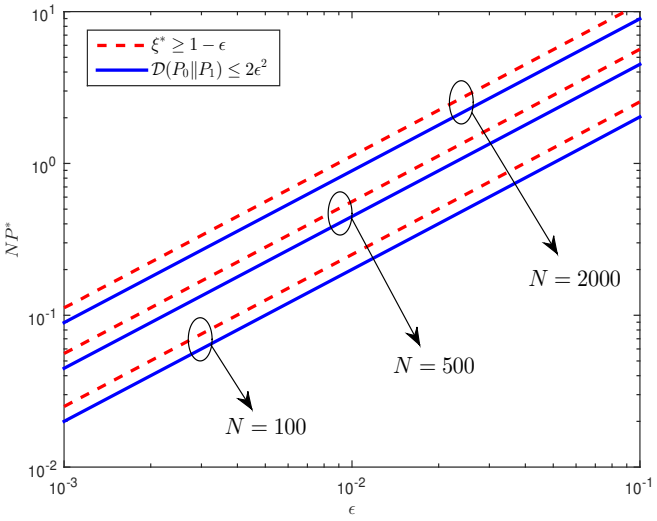

Fig. 2. Maximum allowable total transmit power NP^* versus ϵ for different values of N, where $\sigma_b^2=\sigma_w^2=0$ dB.

V. Numerical Results

In this section, we first provide numerical results on FTP subject to $\xi \geq 1-\epsilon$ to verify our analysis with $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ as the covert communication constraint. We then examine the covert performance of DUTP with FTP and CUTP as the benchmarks, which leads to many useful insights on the tradeoff between the achievable covertness and system complexity issues, such as the shared secret size between Alice and Bob.

A. Optimization of the Number of Channel Uses for FTP

In Fig. 2, we plot the maximum allowable total transmit power NP^* over the entire block versus ϵ. In this figure and the following figures, the curves for $\xi \geq 1-\epsilon$ are achieved by numerically evaluating the false positive and detection rates as per (11) and (12), respectively. In this figure, we observe that NP^* is larger with the constraint $\xi \geq 1-\epsilon$ than with $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ as the constraint. This is due to the fact that equality in (13) cannot be achieved, and hence the constraint $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) \leq 2\epsilon^2$ is stricter than $\xi \geq 1-\epsilon$. We also observe that NP^* increases (hence η increases) as N increases, which can be explained by Theorem 1. Finally, we observe that NP^* decreases (hence η decreases) as ϵ decreases, which demonstrates the tradeoff between the covert requirement and

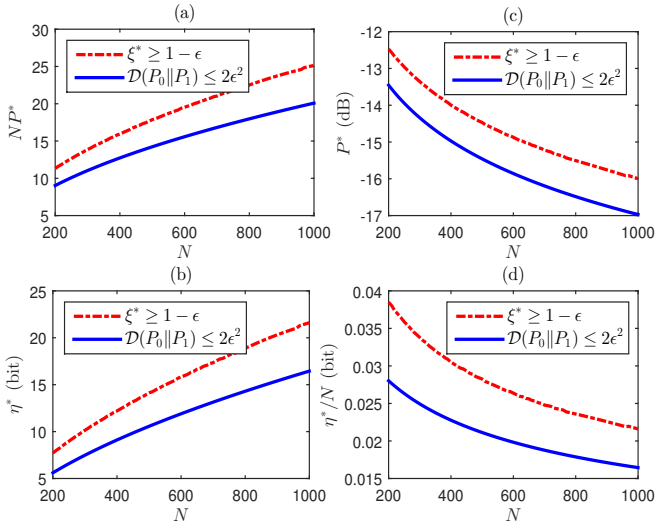

Fig. 3. NP^*, η, P^*, and η/N versus N, where $\sigma_b^2 = -5$ dB, $\sigma_w^2 = 0$ dB, $\delta = 0.15$, and $\epsilon = 0.1$.

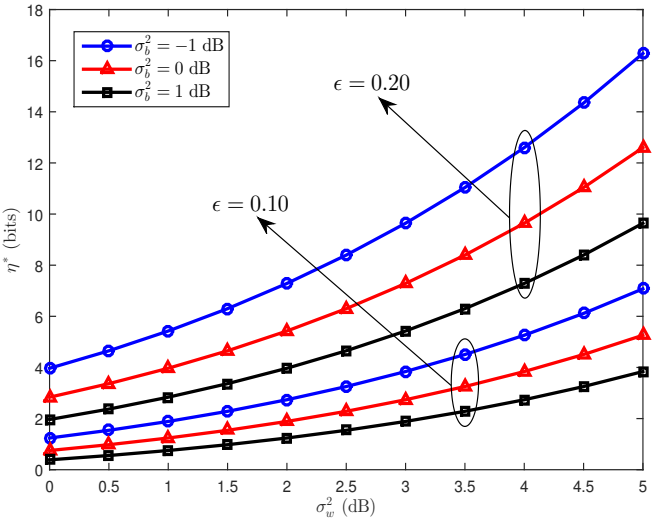

Fig. 4. Maximum η versus the noise variance at Willie σ_w^2 for different values of σ_b^2 and ϵ, where $N = 100$ and $\delta = 0.20$.

the achievable η (e.g., a stricter covert requirement leads to a smaller η).

In Fig. 3, we plot NP^*, η^*, P^*, and η/N versus N. As expected, we first observe that NP^* and η^* monotonically increase as N increases in Fig. 3(a) and Fig. 3(b), respectively. This confirms the correctness of adopting nP as an indicator of η in Section III. Although NP^* increases, it is interesting to observe that the maximum allowable transmit power P^* monotonically decreases as N increases in Fig. 3(c). This is due to the fact that as the number of observations at Willie increases, Alice has to reduce her transmit power for each channel use in order to meet the same covert communication constraint. In Fig. 3(d), we observe that η per channel use, i.e., η^*/N, monotonically decreases as N increases. This is mainly due to the fact that the maximum allowable transmit power P^* monotonically decreases as N increases as shown in Fig. 3(c). Although the channel coding rate slightly increases as N increases for a fixed P^*, this minor increase cannot counteract the decrease caused by the decrease in P^*. These observations from Fig. 3 demonstrate that increasing N helps Alice to allocate less transmit power to each channel use in order to maintain the same level of covertness, but increases the total transmit power over all the channel uses, which in turn improves η subject to the same covert communication constraint.

In Fig. 4, we plot η^* subject to $\xi^* \geq 1-\epsilon$ versus the noise variance at Willie σ_w^2. In this figure, we first observe that η^* increases as σ_w^2 increases since the uncertainty in Willie's received power increases with σ_w^2. This is different from the case with an infinite blocklength $n \to \infty$, in which the maximum covert rate is not a function of σ_w^2. This is due to the fact that with $n \to \infty$ the AWGN power at Willie is deterministic and thus increasing σ_w^2 does not decrease Willie's detection performance, since the change of σ_w^2 can be counteracted by Willie varying the detection threshold. As expected, we also observe that η^* decreases as the noise variance at Bob σ_b^2 increases. Finally, we observe that as ϵ increases slightly from 0.10 to 0.20, η^* significantly increases. This observation demonstrates that η^* is very sensitive to ϵ.

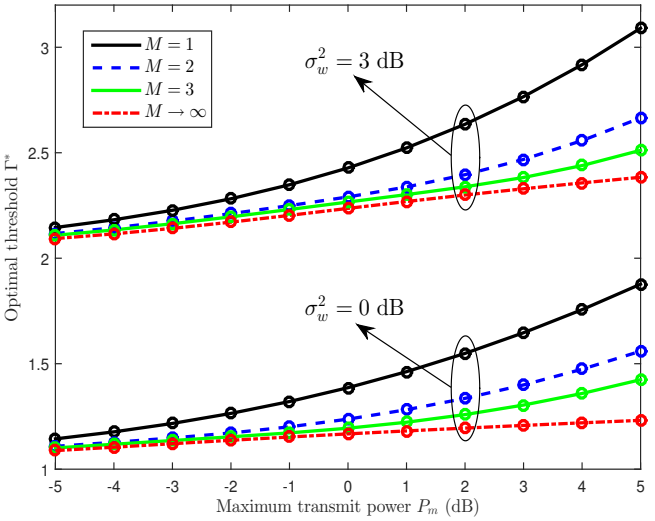

Fig. 5. Willie's optimal detection threshold Γ^* versus the maximum transmit power P_m for different values of M and σ_w^2, where $N = 100$.

B. Comparison among FTP, CUTP, and DUTP

In Fig. 5, we plot Willie's optimal detection threshold Γ^* versus the maximum transmit power P_m. In this figure, $M = 1$ and $M \to \infty$ represent FTP and CUTP, respectively. The curves represent theoretical results and the circles represent numerical results. We first observe that Γ^* decreases as M increases, which is due to the fact that for a fixed P_m the expected transmit power $\overline{P}_t$ decreases with M. Furthermore, we observe that Γ^* is always greater than σ_w^2. As such, we can solve $\partial\xi/\partial\Gamma = 0$ for CUTP and DUTP with Γ being greater than σ_w^2 and less than Γ^* for FTP (i.e., $M = 1$) as given in (8), which can significantly facilitate achieving Γ^* theoretically or numerically.

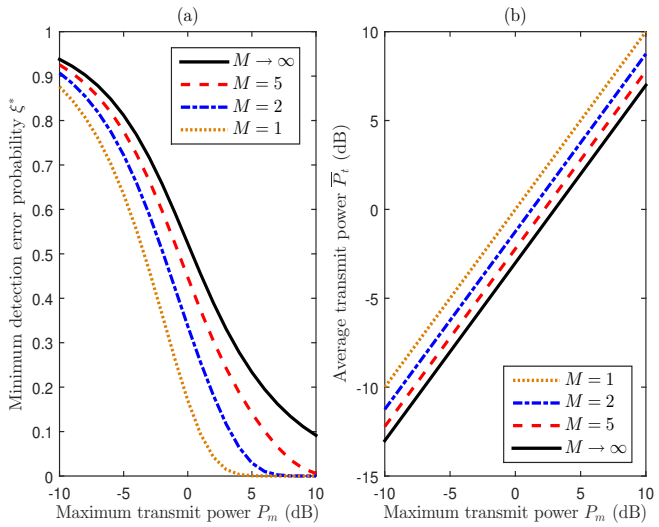

Fig. 6. Minimum detection error probability ξ^* and expected transmit power $\overline{P}_t$ versus the maximum transmit power P_m for different values of M, where $\sigma_w^2 = 5$ dB and $N = 100$.

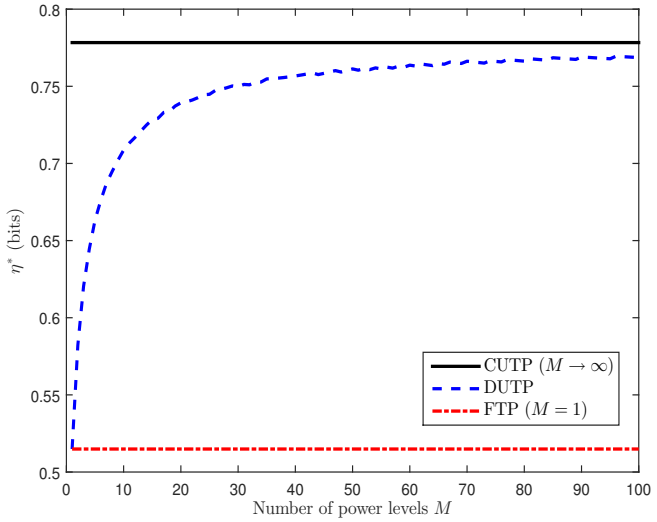

Fig. 7. Maximum η versus the number of power levels at Alice (M), where $\sigma_w^2 = 5$ dB, $\sigma_b^2 = 0$ dB, $N = 100$, $\epsilon = 0.05$, and $\delta = 0.1$.

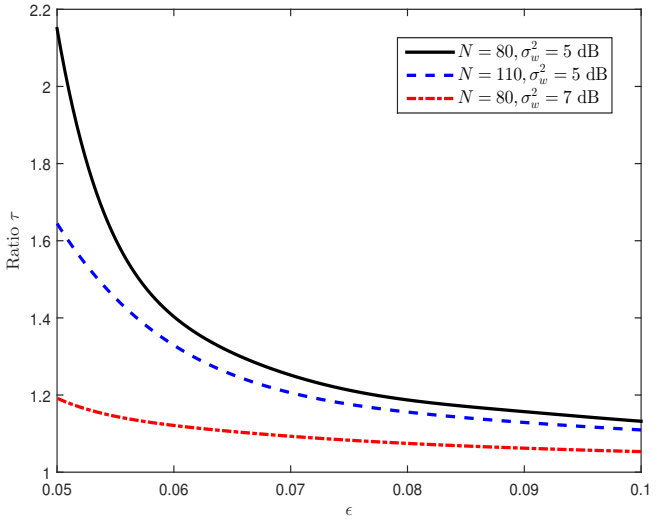

Fig. 8. The ratio τ versus ϵ, where $\sigma_b^2 = 0$ dB and $\delta = 0.1$.

In Fig. 6, we plot the minimum total error ξ^* and the expected transmit power $\overline{P}_t$ versus P_m for different values of M. In Fig. 6 (a), we first observe that ξ^* increases with M for a fixed P_m, which demonstrates that adopting random transmit power can indeed make it harder for Willie to detect Alice's transmission. This can be explained by Fig. 6 (b), which shows that as M increases, $\overline{P}_t$ decreases for a fixed P_m. Intuitively, this is due to the fact that as M increases the uncertainty in Alice's transmit power increases and thus it becomes harder for Willie to make correct decisions. However, we note that for a predetermined ξ^* we have different corresponding values of P_m for different values of M in Fig. 6 (a), which lead to different values of $\overline{P}_t$ corresponding to different values of M for a fixed ξ^* in Fig. 6 (b). As such, in this figure it is hard to examine whether a larger M leads to a larger $\overline{P}_t$ (and thus a larger η^*), an issue that will be further examined in the following figure. Finally, we note that a larger ξ^* corresponds to a larger $\overline{P}_t$, which indicates that we could use $\overline{P}_t$ as an alternative metric of the detection performance at Willie when ξ^* is not achievable in a closed-form expression.

In Fig. 7, we plot η^* subject to $\xi^* \geq 1 - \epsilon$ versus different values of M. In this figure, we first observe that CUTP can significantly outperform FTP in terms of achieving a much (approximately 5 times) higher η^* than FTP. This indicates that with a finite blocklength, adopting random transmit power at Alice can indeed facilitate achieving covertness, which is due to the fact that varying Alice's transmit power can boost the non-zero uncertainty (caused by AWGN due to the finite blocklength) on the received power at Willie. Then, as expected we also observe that for DUTP, η^* increases as M increases, while when $M = 1$ DUTP is identical to FTP and as $M \to \infty$ DUTP approaches CUTP. This demonstrates that increasing M has a larger impact on increasing the uncertainty at Willie compared with decreasing η^* at Bob. In addition to the number of transmit power levels that Alice has to support, M also affects the size of the shared secrets between Alice and Bob. Thus, in addition to the tradeoff between the number of transmit power levels that Alice has to support and the achieved covertness, Fig. 7 also demonstrates the tradeoff between the shared secret size and the achieved covertness. Surprisingly, we observe that η^* increases approximately logarithmically with M, which indicates that small values of M can achieve most of the benefit of using variable power. For example, we see that the value of η^* achieved by DUTP with $M = 5$ is within 90% of that for CUTP. This demonstrates the practical usefulness of adopting random transmit power to facilitate achieving covert communications. In particular, we see that it provides a method for improving the achievable covertness without significantly increasing the system complexity or the shared secret size.

In Fig. 8, we plot the ratio τ, which is defined as the ratio of η^* achieved by CUTP to that achieved by FTP, versus ϵ. In this figure, we first observe that τ monotonically decreases

with ϵ. We note that $\xi^* \geq 1-\epsilon$ becomes stricter as ϵ decreases. As such, this observation demonstrates that the benefit of adopting random transmit power in covert communications in terms of improving η^* becomes more dominant as the covert communication constraint becomes stricter. We also observe that τ increases as N decreases. This demonstrates the benefit of adopting random transmit power to improve the performance of covert communications with a finite and small blocklength; i.e., as the blocklength becomes smaller the improvement achieved by random transmit power becomes more significant. Finally, we observe that τ significantly increases as the noise power at Willie σ_w^2 decreases. Based on the above three observations, we can conclude that the benefits of using random transmit power in terms of increasing η^* become more dominant when η^* achieved by FTP decreases (e.g., when N, ϵ, or σ_w^2 is smaller).

VI. Conclusion

This work examined the impact of a finite number of channel uses $n \leq N$ on covert communications over AWGN channels. We showed that the amount of information that can be transmitted in covert communications is maximized when all available channel uses are utilized, i.e., $n^* = N$, although Willie will have more observations to detect the covert communications as n increases. Varying Alice's transmit power was proposed as a method for significantly enhancing the performance of covert communications, especially when the number η of information bits achieved by FTP is small. Our examinations showed that the η achieved by DUTP logarithmically increases with the number of transmit power levels M, which demonstrates the practical usefulness of using random transmit power to enhance covert communications. For practical implementations, our results suggest that the data transmission should extend across the maximum allowed time interval in order to enhance the covertness of the communications. In addition, randomly varying the transmit power can enhance covert communications when the allowed delay is short or the covert requirement is extremely strict.

Appendix A

We present our proof of Theorem 1 in the following 5 steps.

Step 1: We note that η and $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1)$ are both monotonically increasing functions of P and n. As such, we can conclude that the equality in the constraint (15b) is always met in order to maximize η. Thus, we have $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) = 2\epsilon^2$ and following (14) we have

$$n = \frac{2\epsilon^2}{f(\gamma_w)}, \tag{33}$$

where

$$f(\gamma_w) \triangleq \frac{\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1)}{n} = \ln(\gamma_w + 1) - \frac{\gamma_w}{\gamma_w + 1}, \tag{34}$$

and $\gamma_w = P/\sigma_w^2$ is the SNR at Willie.

Step 2: We note that $f(0) = 0$ and we derive the first derivative of $f(\gamma_w)$ with respect to γ_w as

$$\frac{\partial f(\gamma_w)}{\partial \gamma_w} = \frac{\gamma_w}{(\gamma_w + 1)^2} \geq 0, \tag{35}$$

which leads to the fact that $f(\gamma_w)$ is a monotonically increasing function of γ_w. With the constraint $\mathcal{D}(\mathbb{P}_0\|\mathbb{P}_1) = 2\epsilon^2$, n is a monotonically decreasing function of $f(\gamma_w)$ as per (33), which indicates that n is a monotonically decreasing function of γ_w and thus of P as well.

Step 3: We next prove that either $n = 1$ or $n = N$ maximizes $n\gamma_w$. To this end, in the following we first show that $n\gamma_w$ initially decreases and then increases with n. Following (33) and (34), we have

$$n\gamma_w = \frac{2\epsilon^2}{g(\gamma_w)}, \tag{36}$$

where $g(\gamma_w)$ is given by

$$g(\gamma_w) = \frac{\ln(1+\gamma_w)}{\gamma_w} - \frac{1}{1+\gamma_w}. \tag{37}$$

We then derive the first derivative of $g(\gamma_w)$ with respect to γ_w as

$$\frac{\partial g(\gamma_w)}{\partial \gamma_w} = \frac{h(\gamma_w)}{\gamma_w^2(1+\gamma_w)^2}, \tag{38}$$

where

$$h(\gamma_w) = 2\gamma_w^2 + \gamma_w - (1+\gamma_w)^2 \ln(1+\gamma_w). \tag{39}$$

We note that there are *only* two solutions to $h(\gamma_w) = 0$ for $\gamma_w \geq 0$, including $\gamma_w = 0$ and $\gamma_w = \gamma_w^\dagger$. We obtain $\gamma_w^\dagger \approx 2.1626$ by numerically solving $h(\gamma_w) = 0$. We also note that as $\gamma_w \to \infty$ we have $h(\gamma_w) \to -\infty$. Then, we can conclude that $h(\gamma_w) \geq 0$ for $\gamma_w < \gamma_w^\dagger$ and $h(\gamma_w) \leq 0$ for $\gamma_w \geq \gamma_w^\dagger$. As such, noting $\gamma_w^2(1+\gamma_w)^2 \geq 0$ and following (38), we have $\partial g(\gamma_w)/\partial \gamma_w \geq 0$ for $\gamma_w < \gamma_w^\dagger$ and $\partial g(\gamma_w)/\partial \gamma_w \leq 0$ for $\gamma_w \geq \gamma_w^\dagger$. This indicates that $g(\gamma_w)$ initially increases and then decreases with γ_w. As per (36), we know that $n\gamma_w$ monotonically decreases with $g(\gamma_w)$, which leads to the fact that $n\gamma_w$ first decreases and then increases as γ_w increases (i.e., $n\gamma_w$ has one minimum value but no maximum value). We recall that n is a monotonically decreasing function of γ_w under the constraint (33), which is proved following (35). Therefore, we conclude that $n\gamma_w$ first decreases and then increases as n increases, and thus the maximum value of $n\gamma_w$ is achieved either at $n = 1$ or $n = N$.

Step 4: We next prove that $n = N$ and not $n = 1$ maximizes $n\gamma_w$. Substituting $\gamma_w^\dagger$ into (33), we have $n^\dagger = 2\epsilon^2/f(\gamma_w^\dagger)$. For $0 < \epsilon < 0.4835$, we have $n^\dagger < 1$ since $f(\gamma_w^\dagger) > 0.4675$. When $n^\dagger < 1$, $n\gamma_w$ increases with n since $n \geq 1$. As such, for $0 < \epsilon < 0.4835$ the optimal value of n that maximizes $n\gamma_w$ is N (i.e., $n^* = N$). For $0.4835 \leq \epsilon \leq 0.5$, we have $n^\dagger < 2$ again since $f(\gamma_w^\dagger) > 0.4675$. We next confirm that even for $n^\dagger < 2$ we still have $n^* = N$. To this end, we only have to confirm $n\gamma_w$ for $n = 2$ is larger than that for $n = 1$. When $n = 1$, following (33) we have $f(\gamma_w) = 2\epsilon^2$. The maximum value of γ_w that guarantees $f(\gamma_w) = 2\epsilon^2$ (i.e., $n = 1$) is obtained when $\epsilon = 0.5$ since $f(\gamma_w)$ is a monotonically increasing function of γ_w as proved by (35). We obtain this maximum value by solving $f(\gamma_w) = 0.5$ as $\gamma_w^{n=1} < 2.3145$, which leads to $n\gamma_w < 2.3145$ when $n = 1$. When $n = 2$, following (33) we have $f(\gamma_w) = \epsilon^2$. The minimum value of γ_w that guarantees $f(\gamma_w) = \epsilon^2$ (i.e., $n = 2$) is obtained when $\epsilon =$

0.4835. We obtain this minimum value by solving $f(\gamma_w) = (0.4835)^2$ as $\gamma_w^{n=2} > 1.16$, which shows that $n\gamma_w > 2.32$ when $n = 2$. As such, we have $n\gamma_w < 2.3145$ when $n = 1$ and $n\gamma_w > 2.32$ when $n = 2$, which means that $n\gamma_w$ for $n = 2$ is larger than $n\gamma_w$ for $n = 1$. We recall that $n\gamma_w$ monotonically increases with n when $n \geq n^\dagger$. Therefore, for $0.4835 \leq \epsilon \leq 0.5$ the optimal value of n that maximizes $n\gamma_w$ is N.

Step 5: So far, we have proved $n^* = N$. Then, substituting $n^* = N$ into (33), we obtain the fixed-point equation in (17).

We note that there may be other methods to achieve this proof. For example, a geometrical argument based on the dimension of the observation vector space under $\mathcal{H}_0$ may provide an alternative approach, where one would have to prove that, for a fixed total transmit power, the KL divergence decreases with n. In our approach, we proved that, for a fixed KL divergence, the maximum allowable total transmit power increases with n. Since the KL divergence monotonically increases with the total transmit power for a fixed value of n, we can conclude that the alternative approach follows a similar philosophy.

Appendix B

We note that CUTP does not affect the likelihood function under $\mathcal{H}_0$ (i.e., $\mathbb{P}_0$). As such, (21) follows from (11) directly. We now derive the missed detection rate β for CUTP. Following (9) and (20), the likelihood function of T under $\mathcal{H}_1$ for CUTP can be written as

$$f(T|\mathcal{H}_1) = \frac{T^{N-1}}{\Gamma(N)} \int_0^{P_m} \left(\frac{N}{x+\sigma_w^2}\right)^N e^{-\frac{NT}{x+\sigma_w^2}} f_P(x)dx$$
$$= \frac{T^{N-1}}{\Gamma(N)P_m} \int_0^{P_m} \left(\frac{N}{x+\sigma_w^2}\right)^N e^{-\frac{NT}{x+\sigma_w^2}} dx$$
$$\stackrel{a}{=} \frac{N^N T^{N-1}}{\Gamma(N)P_m} \int_{\frac{1}{P_m+\sigma_w^2}}^{\frac{1}{\sigma_w^2}} y^{N-2} e^{-NTy} dy$$
$$\stackrel{b}{=} \frac{N}{\Gamma(N)P_m} \left[\gamma\left(N\text{–}1, \frac{NT}{\sigma_w^2}\right) - \gamma\left(N\text{–}1, \frac{NT}{P_m+\sigma_w^2}\right)\right], \quad (40)$$

where $\stackrel{a}{=}$ is achieved by setting $y = 1/x+\sigma_w^2$ and $\stackrel{b}{=}$ is achieved with the aid of the following identity [32, Eq. (3.351)]:

$$\int_0^u x^n e^{-\mu x} dx = \mu^{-n-1}\gamma(n+1, u\mu). \quad (41)$$

Then, following (40) for a fixed threshold Γ the missed detection rate for the CUTP is given by

$$\beta = \Pr(T < \Gamma|\mathcal{H}_1) = \int_0^\Gamma f(T|\mathcal{H}_1)dT$$
$$= \frac{N}{\Gamma(N)P_m} \int_0^\Gamma \left[\gamma\left(N\text{–}1, \frac{Nx}{\sigma_w^2}\right) - \gamma\left(N\text{–}1, \frac{Nx}{P_m+\sigma_w^2}\right)\right] dx$$
$$\stackrel{c}{=} \frac{N}{(N-1)P_m} \sum_{i=0}^{N-2} \frac{1}{i!} \left[\left(\frac{N}{P_m+\sigma_w^2}\right)^i \int_0^\Gamma x^i e^{-\frac{N}{P_m+\sigma_w^2}x} dx - \left(\frac{N}{\sigma_w^2}\right)^i \int_0^\Gamma x^i e^{-\frac{N}{\sigma_w^2}x} dx\right], \quad (42)$$

where $\stackrel{c}{=}$ is achieved with the aid of the following identity [32, Eq. (8.353.6)]:

$$\gamma(n, x) = (n-1)! \left[1 - e^{-x} \sum_{m=0}^{n-1} \frac{x^m}{m!}\right]. \quad (43)$$

We then solve the integrals in (42) as per (41) and obtain the desired result in (22).

Appendix C

Following (20), we have the pdf of P/σ_b^2 as

$$f_{\frac{P}{\sigma_b^2}}(x) = \begin{cases} \frac{\sigma_b^2}{P_m}, & 0 < x \leq \frac{P_m}{\sigma_b^2} \\ 0, & x \geq \frac{P_m}{\sigma_b^2}. \end{cases} \quad (44)$$

Then, as per (4) the expected channel coding rate achieved by CUTP for a fixed δ is approximated by

$$\overline{R} = \int_0^{\frac{P_m}{\sigma_b^2}} \left[\log_2(1+x) - \sqrt{\frac{x(x+2)}{N(x+1)^2}} \frac{Q^{-1}(\delta)}{\ln(2)}\right] \frac{\sigma_b^2}{P_m} dx$$
$$= \frac{\sigma_b^2}{P_m} \int_0^{\frac{P_m}{\sigma_b^2}} \log_2(1+x)dx - \frac{Q^{-1}(\delta)\sigma_b^2}{\sqrt{N}\ln(2)P_m} \int_0^{\frac{P_m}{\sigma_b^2}} \frac{\sqrt{x(x+2)}}{x+1} dx. \quad (45)$$

We finally solve the integrals in (45) with the aid of the following two identities

$$\int_0^u \ln(x+a)dx = (u+a)\ln(u+a) - u - a\ln(a), \quad (46)$$
$$\int_0^u \frac{\sqrt{x(x+2)}}{x+1} dx = \sqrt{u(u+2)} - 2\tan^{-1}\sqrt{\frac{u}{u+2}}, \quad (47)$$

which leads to the desired result in (23).

References

[1] M. Bloch and J. Barros, *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[2] X. Zhou, L. Song, and Y. Zhang, *Physical Layer Security in Wireless Communications*, CRC Press, 2013.

[3] S. Yan, B. He, Y. Cong, and X. Zhou, "Covert communication with finite blocklength in AWGN channels," in *Proc. IEEE ICC*, May 2017.

[4] M. K. Simon, Jim K. Omura, Robert A. Scholtz, and Barry K. Levitt, *Spread Spectrum Communications Handbook*, McGraw-Hill, 1994.

[5] P. H. Che, S. Kadhe, M. Bakshi, C. Chan, S. Jaggi, and A. Sprintson, "Reliable, deniable and hidable communication: A quick survey," in *Proc. IEEE Inf. Theory Workshop*, Nov. 2014, pp. 227–231.

[6] B. A. Bash, D. Goeckel, D. Towsley, and S. Guha, "Hiding information in noise: Fundamental limits of covert wireless communication," *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 26–31, Dec. 2015.

[7] B. He, S. Yan, X. Zhou, and V. Lau, "On covert communication with noise uncertainty," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 941–944, Apr. 2017.

[8] B. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1921–1930, Sep. 2013.

[9] L. Wang, G. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Infp. Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[10] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise, in *Proc. IEEE Int'l. Symp. Info. Theory*, Jul. 2013, pp. 2945–2949.

[11] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.

[12] K. Suria, K. Arumugam, and M. R. Bloch, "Keyless covert communication over multiple-access channels," in *Proc. IEEE Int'l. Symp. Info. Theory*, Jul. 2016, pp. 2229–2233.

[13] S. Lee and R. Baxley, "Achieving positive rate with undetectable communication over AWGN and Rayleigh channels," in *Proc. IEEE ICC, 2014*, Jun. 2014, pp. 780–785.

[14] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 7, pp. 1195–1205, Oct. 2015.

[15] T. Sobers, B. Bash, D. Goeckel, S. Guha, and D. Towsley, "Covert communication with the help of an uninformed jammer achieves positive rate," in *Proc. Asilomar Conf. on Signals, Systems, and Comput.*, Nov. 2015, pp. 625–629.

[16] D. Goeckel, B. Bash, S. Guha, and D. Towsley, "Covert communications when the warden does not know the background noise power," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 236–239, Feb. 2016.

[17] B. A. Bash, D. Goeckel, and D. Towsley, "LPD communication when the warden does not know when," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2014, pp. 606–610.

[18] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6193–6206, Sep. 2017.

[19] K. Shahzad, X. Zhou, and S. Yan, "Covert communication in fading channels under channel uncertainty," in *Proc. IEEE VTC Spring*, Jun. 2017.

[20] Q. Zhang, M. Bakshi, and S. Jaggi, "Computationally efficient deniable communication," in *Proc. IEEE Int'l. Symp. Info. Theory*, Jul. 2016, pp. 2234–2238.

[21] M. Tahmasbi and M. R. Bloch, "Second-order asymptotics of covert communications over noisy channels," in *Proc. IEEE Int'l. Symp. Info. Theory*, Jul. 2016, pp. 2224–2228.

[22] N. Letzepis, "Optimal power allocation for parallel Gaussian channels with LPD constraints," in *Proc. IEEE Milcom*, Nov. 2016.

[23] Y. Polyanskiy, H. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[24] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J.Wang, "Covert communications in wireless relay networks," in *Proc. IEEE GlobeCOM*, Dec. 2017.

[25] J. Hu, S. Yan, X. Zhou, F. Shu, J. Li, and J. Wang, "Covert communication achieved by a greedy relay in wireless networks," *IEEE Trans. Wireless Commun.*, DOI: 10.1109/TWC.2018.2831217, May 2018.

[26] J. Hu, K. Shahzad, S. Yan, X. Zhou, F. Shu, and J. Li, "Covert communications with a full-duplex receiver over wireless fading channels," in *Proc. IEEE ICC*, May 2018, pp. 1–6, arXiv:1711.03684.

[27] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of spectrum sharing networks using rate adaptation," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2823–2835, Aug. 2015.

[28] G. Ozcan and M. C. Gursoy, "Throughput of cognitive radio systems with finite blocklength codes," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2541–2554, Nov. 2013.

[29] E. Lehmann and J. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York: Springer, 2005.

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Hoboken, NJ, 2002.

[31] M. H. DeGroot, *Probability and Statistics*, 4th ed. Pearson, 2011.

[32] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 7th ed., Academic, San Diego, CA, 2007.