# Covert Communication Gains from Adversary's Ignorance of Transmission Time

Boulat A. Bash, *Member, IEEE,* Dennis Goeckel, *Fellow, IEEE,*
and Don Towsley, *Fellow, IEEE*

**Abstract**

The recent *square root law* (SRL) for covert communication demonstrates that Alice can reliably transmit $\mathcal{O}(\sqrt{n})$ bits to Bob in $n$ uses of an additive white Gaussian noise (AWGN) channel while keeping ineffective any detector employed by the adversary; conversely, exceeding this limit either results in detection by the adversary with high probability or non-zero decoding error probability at Bob. This SRL is under the assumption that the adversary knows *when* Alice transmits (if she transmits); however, in many operational scenarios he does not know this. Hence, here we study the impact of the adversary's ignorance of the time of the communication attempt. We employ a slotted AWGN channel model with $T(n)$ slots each containing $n$ symbol periods, where Alice may use a single slot out of $T(n)$. Provided that Alice's slot selection is secret, the adversary needs to monitor all $T(n)$ slots for possible transmission. We show that this allows Alice to reliably transmit $\mathcal{O}(\min\{\sqrt{n \log T(n)}, n\})$

B. A. Bash is with Raytheon BBN Technologies, Cambridge, MA (email: bbash@bbn.com).

D. Goeckel is with the Electrical and Computer Engineering Department, University of Massachusetts, Amherst, MA (email: goeckel@ecs.umass.edu).

D. Towsley is with the College of Information and Computer Sciences, University of Massachusetts, Amherst, MA (email: towsley@cs.umass.edu).

This research was sponsored by the National Science Foundation under grants CNS-1018464, CNS-0964094, and ECCS-1309573. B. A. Bash acknowledges financial support from Raytheon BBN Technologies and DARPA under contract number HR0011-16-C-0111.

An abridged version of this manuscript was presented at the International Symposium on Information Theory (ISIT) 2014 [1].

bits to Bob (but no more) while keeping the adversary's detector ineffective. To achieve this gain over SRL, Bob does not have to know the time of transmission provided $T(n) < 2^{c_{\mathrm{T}} n}$, $c_{\mathrm{T}} = \mathcal{O}(1)$.

## I. Introduction

Recent revelations of massive surveillance programs [2] have emphasized the need for secure communication systems that do not just protect the content of the user's message from being decoded, but prevent the detection of its transmission in the first place. Indeed, encrypted data or even just the transmission of a signal can arouse suspicion, and even the most theoretically robust cryptographic security scheme can be defeated by a determined adversary using non-computational methods such as side-channel analysis. Covert, or *low probability of detection/intercept* (LPD/LPI) communication capability is thus very important in extremely sensitive situations (e.g., during military operations or for organization of civil unrest).

In the covert communication scenario, Alice transmits a message to Bob over a noisy channel while the adversary, warden Willie, attempts to detect her transmission. The channel from Alice to Willie is also subject to noise. Thus, while Alice transmits low-power covert signals to Bob, Willie attempts to classify these signals as either noise on his channel or signals from Alice. We recently showed that the *square root law* (SRL) governs covert communication: provided that Alice and Bob pre-share a secret of sufficient length, she can reliably transmit $\mathcal{O}(\sqrt{n})$ bits to Bob in $n$ uses of an additive white Gaussian noise (AWGN) channel while keeping ineffective any detector employed by Willie. Conversely, attempting to transmit more that $\mathcal{O}(\sqrt{n})$ bits either results in detection by Willie with probability one or non-zero decoding error probability at Bob as $n \to \infty$ [3], [4]. Follow-on work has addressed the size of the pre-shared secret [5], [6], the optimal constant hidden by the asymptotic (big-$\mathcal{O}$) notation [6], [7], network aspects of covert communication [8], [9], extensions of the SRL to quantum channels where the adversary is only limited by the laws of quantum mechanics [10]–[12], and, finally, practical limitations on the adversary that improve the covert throughput to beyond the SRL limit [13]–[17].

Studies of covert communication up to now assume that Willie knows when Alice may start to transmit (if she does). However, in many practical scenarios (e.g., delay-constrained communication), Alice's message may have to be short relative to the total time available to transmit it (e.g., a few seconds out of the day when both Alice and Bob are available). Additionally, Willie might not know when the communication starts (e.g., a possible transmission time can be secretly pre-arranged in advance). This forces Willie to monitor a much longer time period than the duration of Alice's transmission, limiting his capabilities. Here we show how Alice can leverage Willie's ignorance of her transmission time to transmit significant additional information covertly to Bob.

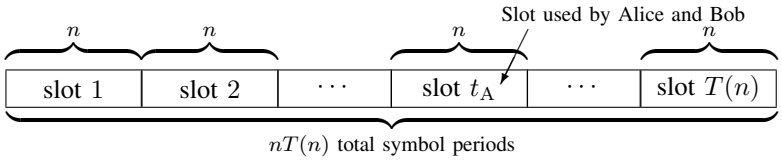


Fig. 1: Slotted channel: each of the $T(n)$ slots contains $n$ symbol periods. Alice and Bob use slot $t_{\mathrm{A}}$ to communicate.

In our scenario, Alice communicates to Bob over an AWGN channel. Willie also has an AWGN channel from Alice. Unlike the setting in [3], [4], the channel is slotted, as shown in Figure 1. Each of $T(n)$ slots contains $n$ symbol periods, where $T(n)$ is an increasing function of $n$. If Alice used all $nT(n)$ symbol periods for transmission, then, by the SRL in [3], [4], she could reliably transmit $\mathcal{O}(\sqrt{nT(n)})$ covert bits to Bob. However, to model a practical scenario where Alice is constrained to a short message relative to the total available transmission time, her communication is restricted to a single slot $t_{\mathrm{A}}$ which is kept secret from Willie. While this certainly reduces the amount of transmissible data, a natural question is whether an improvement can be made over a naïve application of the SRL [3], [4], which allows Alice to reliably transmit $\mathcal{O}(\sqrt{n})$ covert bits in this scenario. Here we demonstrate that Alice can transmit $\mathcal{O}\left(\min\{\sqrt{n\log T(n)}, n\}\right)$ bits reliably on this channel while maintaining arbitrarily low probability of detection by Willie. Conversely, we show that the transmission of

$\omega(\sqrt{n \log T(n)})$ bits[1] either results in Alice being detected with high probability or unreliable communication.

The improvement stems from Willie not knowing $t_{\mathrm{A}}$ and being forced to monitor all $T(n)$ slots. When Willie knows which slot Alice might use, by Theorem 2 in [3], [4] she can be detected if she transmits more than $\mathcal{O}(\sqrt{n})$ bits, since it is improbable that Willie's observations will look like AWGN noise that he expects. His optimal detector is a threshold on the power observed in slot $t_{\mathrm{A}}$ [17]. However, if Willie does not know $t_{\mathrm{A}}$, he has to examine all $T(n)$ slots. Effectively, Willie's test statistic is the maximum slot power (see the remark following the proof of Theorem 1.1). When only noise is observed, the average maximum slot power is substantially higher than average power in any single slot. Hence, to avoid false alarms, Willie's threshold when he does not know $t_{\mathrm{A}}$ must also be greater than when he knows it. This allows Alice to transmit additional covert information.

Our main result is stated formally as follows:

**Theorem.** *Suppose the channel between Alice and each of Bob and Willie experiences independent additive white Gaussian noise (AWGN) with constant power $\sigma_{\mathrm{b}}^2 > 0$ and $\sigma_{\mathrm{w}}^2 > 0$, respectively, and that Alice's transmitter is subject to the average power constraint $P_{\max} \in (0, \infty)$. Also suppose that, if Alice chooses to transmit, she uses one of the $T(n)$ slots chosen randomly. Each slot contains $n$ symbol periods, where $T(n) = \omega(1)$. Then, for any $\epsilon > 0$, Alice can reliably transmit $\mathcal{O}\left(\min\{\sqrt{n \log T(n)}, n\}\right)$ bits to Bob in a selected slot while maintaining a probability of detection error by Willie greater than $\frac{1}{2} - \epsilon$. Conversely, if Alice tries to transmit $\omega(\sqrt{n \log T(n)})$ bits using $n$ consecutive symbol periods, either Willie detects her*

[1]Throughout this paper we employ asymptotic notation [18, Ch. 3.1] where $f(n) = \mathcal{O}(g(n))$ denotes an asymptotic upper bound on $f(n)$ (i.e., there exist constants $m, n_0 > 0$ such that $0 \leq f(n) \leq mg(n)$ for all $n \geq n_0$), $f(n) = o(g(n))$ denotes an upper bound on $f(n)$ that is not asymptotically tight (i.e., for any constant $m > 0$, there exists constant $n_0 > 0$ such that $0 \leq f(n) < mg(n)$ for all $n \geq n_0$), and $f(n) = \omega(g(n))$ denotes a lower bound on $f(n)$ that is not asymptotically tight (i.e., for any constant $m > 0$, there exists constant $n_0 > 0$ such that $0 \leq mg(n) < f(n)$ for all $n \geq n_0$).

*with arbitrarily low probability of error or Bob cannot decode her message with arbitrary low probability of decoding error as* $n \to \infty$.

As in [3], [4], covert communication requires that Alice and Bob possess a common randomness resource. This corresponds to a secret codebook[2] needed in our proofs that is shared between Alice and Bob prior to communication, analogous to the one-time pad in the information-theoretic analysis of encryption [20]. This follows "best practices" in security system design as the security of the covert communication system depends only on the shared secret [21]. Remarkably, we demonstrate that the *multiplicative* increase (by a factor of $\sqrt{\log T(n)}$) in the number of covert bits that Alice can transmit reliably to Bob does not require Bob to know the timing of the transmission if $T(n) < 2^{c_{\mathrm{T}} n}$, where $c_{\mathrm{T}} > 0$ is a constant; to realize the $\sqrt{\log T(n)}$ gain when $T(n) \geq 2^{c_{\mathrm{T}} n}$ only an additive expense of an extra $\log T(n)$ secret bits is needed to indicate to Bob the slot employed by Alice. Thus, at most $\log T(n)$ secret bits are required in excess of those needed to enable the SRL on a single $n$-symbol slot. Timing is therefore a very useful resource for covert communication. It also necessitates a vastly different analysis than that in [3], [4]. Specifically, the relative entropy based bounds on the probability of detection error used in [3], [4] are too loose to yield our achievability results, and we thus have to apply other techniques from mathematical statistics.

After introducing our slotted channel model in Section II, we prove the achievability and the converse in Sections III and IV, respectively. We discuss the relationship of our paper to other work in covert communication in Section V and conclude in Section VI.

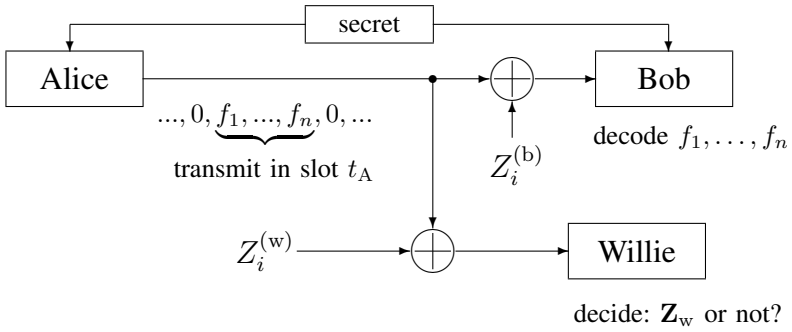


Fig. 2: System framework: Alice and Bob share a secret before transmission. If Alice chooses to transmit, she encodes information into a vector of real symbols $\mathbf{f} = \{f_i\}_{i=1}^{n}$ and uses random slot $t_A$ to send it on an AWGN channel to Bob (to ensure reliable decoding $t_A$ is secretly shared with Bob before the transmission if $T(n) \geq 2^{c_T n}$, where $c_T$ is a constant). Upon observing the channel from Alice, Willie has to classify his vector of readings $\mathbf{Y}_w$ as either an AWGN vector $\mathbf{Z}_w = \{Z_i^{(w)}\}_{i=1}^{nT(n)}$ or a vector that contains a slot with transmissions corrupted by AWGN.

## II. Prerequisites

### A. Channel Model

We use the discrete-time slotted AWGN channel model with real-valued symbols depicted in Figures 1 and 2. The channel has $T(n)$ slots, each containing $n$ symbol periods. Alice selects slot $t_A$ uniformly at random prior to transmission. If Alice chooses to transmit, she sends a vector of $n$ real-valued symbols $\mathbf{f} = \{f_i\}_{i=1}^{n}$ during slot $t_A$. The AWGN on Bob's channel is described by an independent and identically distributed (i.i.d.) sequence $\{Z_i^{(b)}\}_{i=1}^{nT(n)}$ of $nT(n)$ zero-mean Gaussian random variables with variance $\sigma_b^2$ (i.e., $Z_i^{(b)} \sim \mathcal{N}(0, \sigma_b^2)$). Bob receives $\mathbf{Y}_b = \{\mathbf{Y}_b(t)\}_{t=1}^{T(n)}$ where $\mathbf{Y}_b(t) = [Y_{(t-1)n+1}^{(b)}, \ldots, Y_{tn}^{(b)}]$ is a vector of observations collected during slot $t$. If Alice transmits during slot $t_A$, $Y_{(t_A-1)n+i}^{(b)} = f_i + Z_{(t_A-1)n+i}^{(b)}$. For any slot that is

[2]The requirement of a pre-shared secret was shown to be unnecessary [5], [6] for the standard SRL in [3], [4] if Bob has a better channel than Willie; [6] was extended [19] to the asynchronous scenario described in this paper while it was in review.

not used for transmission, $Y^{(\mathrm{b})}_{(t-1)n+i} = Z^{(\mathrm{b})}_{(t-1)n+i}$ (this includes all slots $\{t : t \neq t_{\mathrm{A}}\}$, and slot $t_{\mathrm{A}}$ when Alice does not transmit).

Similarly, Willie observes $\mathbf{Y}_{\mathrm{w}} = \{\mathbf{Y}_{\mathrm{w}}(t)\}_{t=1}^{T(n)}$ where $\mathbf{Y}_{\mathrm{w}}(t) = [Y^{(\mathrm{w})}_{(t-1)n+1}, \ldots, Y^{(\mathrm{w})}_{tn}]$ is a vector of observations collected during slot $t$. The AWGN on Willie's channel is described by an i.i.d. sequence $\{Z^{(\mathrm{w})}_i\}_{i=1}^{nT(n)}$ of $nT(n)$ zero-mean Gaussian random variables with variance $\sigma^2_{\mathrm{w}}$ (i.e., $Z^{(\mathrm{w})}_i \sim \mathcal{N}(0, \sigma^2_{\mathrm{w}})$). If Alice transmits during slot $t_{\mathrm{A}}$, $Y^{(\mathrm{w})}_{(t_{\mathrm{A}}-1)n+i} = f_i + Z^{(\mathrm{w})}_{(t_{\mathrm{A}}-1)n+i}$. For any slot that is not used for transmission, $Y^{(\mathrm{w})}_{(t-1)n+i} = Z^{(\mathrm{w})}_{(t-1)n+i}$ (again, this includes all slots $\{t : t \neq t_{\mathrm{A}}\}$, and slot $t_{\mathrm{A}}$ when Alice does not transmit). Willie does not know $t_{\mathrm{A}}$ and has to perform a statistical hypothesis test on his entire set of observations $\mathbf{Y}_{\mathrm{w}}$ to determine whether Alice is transmitting.

### *B. Hypothesis Testing*

Willie performs a statistical hypothesis test [22] on $\mathbf{Y}_{\mathrm{w}}$, where the null hypothesis $H_0$ is that Alice does not transmit and each sample is an i.i.d. realization of AWGN $Y^{(\mathrm{w})}_i \sim \mathcal{N}(0, \sigma^2_{\mathrm{w}})$. The alternate hypothesis $H_1$ is that Alice transmits, and the samples from one of the slots come from a different distribution. The rejection of $H_0$ when it is true is a false alarm (FA) and the acceptance of $H_0$ when it is false is a missed detection (MD). The lower bound on the sum of the probabilities of these events $\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}}$ characterizes the necessary trade-off between the false alarms and missed detections in a hypothesis test. As in [3], [4], Alice desires $\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}} \geq 1 - \epsilon$ for an arbitrary choice of $\epsilon > 0$, ensuring that any hypothesis test Willie constructs is ineffective.[3]

[3]Willie's probability of error is $\mathbb{P}^{(\mathrm{w})}_{\mathrm{e}} = \pi_0 \mathbb{P}_{\mathrm{FA}} + \pi_1 \mathbb{P}_{\mathrm{MD}}$, where $\pi_0$ and $\pi_1$ are the prior probabilities of hypotheses $H_0$ and $H_1$. A random guess of Alice's transmission state and "uninformative" priors $\pi_0 = \pi_1 = \frac{1}{2}$ yield $\mathbb{P}^{(\mathrm{w})}_{\mathrm{e}} = \frac{1}{2}$. By lower-bounding $\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}} \geq 1 - \epsilon$, Alice ensures that, for uninformative priors, any detector Willie employs can only be slightly better than a random guess. Our scaling results apply when $\pi_0 \neq \pi_1$ as well, since $\mathbb{P}^{(\mathrm{w})}_{\mathrm{e}} \geq \min(\pi_0, \pi_1)\,(\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}})$.

## III. Achievability

### A. Proof Idea and Preliminaries

We proved the achievability theorems in [3], [4] by upper-bounding the relative entropy [23, Ch. 10] between the distributions of Willie's sequence of observations $\mathbf{Y}_{\mathrm{w}}$ corresponding to hypotheses $H_0$ and $H_1$. Here we take a different approach by explicitly analyzing Willie's optimal detector assuming that his only unknowns are:

a) slot $t_{\mathrm{A}}$ chosen uniformly at random from $\{1, \ldots, T(n)\}$ by Alice for transmission (if she transmits), as depicted in Figures 1 and 2; and

b) a secret shared between Alice and Bob prior to the potential transmission.

Thus, Willie is given Alice's channel input distribution, the distribution of the AWGN on his channel from Alice, and the slot boundaries depicted in Figure 1. This effectively provides Willie with a complete statistical model of observations $\mathbf{Y}_{\mathrm{w}}$, allowing him to construct the likelihood functions $f_0(\mathbf{Y}_{\mathrm{w}})$ and $f_1(\mathbf{Y}_{\mathrm{w}})$ under hypotheses $H_0$ and $H_1$, respectively. Alice's transmission state is binary (either she transmits or she does not) and the optimal detector for Willie is the likelihood ratio test (LRT) by the Neyman–Pearson lemma [22, Ch. 3.2 and 13.1]. Here we determine what it takes for Alice to ensure that Willie's optimal detector performs only slightly better than a random guess of her transmission state, and how much data she can reliably transmit to Bob in this manner.

The LRT compares the likelihood ratio $\Lambda(\mathbf{Y}_{\mathrm{w}}) = \frac{f_1(\mathbf{Y}_{\mathrm{w}})}{f_0(\mathbf{Y}_{\mathrm{w}})}$ to a threshold $\tau(n)$. $H_0$ or $H_1$ is chosen based on whether $\Lambda(\mathbf{Y}_{\mathrm{w}})$ is smaller or larger than $\tau(n)$ (if it equals the threshold, a random decision is made):

$$\Lambda(\mathbf{Y}_{\mathrm{w}}) \underset{H_1}{\overset{H_0}{\lessgtr}} \tau(n). \tag{1}$$

The LRT statistic $\Lambda(\mathbf{Y}_{\mathrm{w}})$ is a function of the sequence of observations $\mathbf{Y}_{\mathrm{w}}$, and, as such, is a random variable. Per its definition in Section II, $\mathbf{Y}_{\mathrm{w}}$ is parameterized by the slot length $n$ and which hypothesis is true (that is, Alice's transmission state). Let $\Lambda_s^{(n)} \equiv \Lambda(\mathbf{Y}_{\mathrm{w}})$ where $s \in \{0, 1\}$ indicates the true hypothesis ($H_0$ or $H_1$). Since one-to-one transformations of both sides in (1)

do not affect the performance of the test, we analyze a detector that is equivalent to the one defined in (1) but employs the test statistic $L_s^{(n)} \equiv g_n\left(\Lambda_s^{(n)}\right)$, where $g_n(x)$ is a one-to-one function defined later. Denote by $K^{(n)} \xrightarrow{\mathcal{P}} Q$ and $K^{(n)} \xrightarrow{\mathcal{D}} Q$ convergence of random variable $K^{(n)}$ to random variable $Q$ in probability and in distribution, respectively. The following lemma establishes sufficient conditions for the covertness of Alice's transmission:

**Lemma 1.** *If the LRT statistic is described by random variables:*

$$L_0^{(n)} = S^{(n)} + V_0^{(n)} \text{ when } H_0 \text{ is true} \tag{2}$$

$$L_1^{(n)} = S^{(n)} + V_1^{(n)} \text{ when } H_1 \text{ is true}, \tag{3}$$

*where* $V_0^{(n)} \xrightarrow{\mathcal{P}} 0$ *and* $V_1^{(n)} \xrightarrow{\mathcal{P}} 0$*, as well as* $S^{(n)} \xrightarrow{\mathcal{D}} Z$ *with* $Z \sim \mathcal{N}(0,1)$*, then* $\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}} \geq 1 - \epsilon$ *for any* $\epsilon > 0$ *and a sufficiently large* $n$*.*

*Proof:* See Appendix A. ■

In order to employ Lemma 1, we require a one-to-one function $g_n(x)$ that re-scales $\Lambda_s^{(n)}$ so that the convergence conditions hold. In the proofs that follow, we show that $\Lambda_s^{(n)} = \frac{1}{T(n)} \sum_{t=1}^{T(n)} U_t^{(n)}$, where $\{U_t^{(n)}\}_{t=1}^{T(n)}$ is a sequence of $T(n)$ independent random variables, each corresponding to a slot of length $n$. Since Alice is limited to slot $t_{\mathrm{A}}$ for a potential transmission, the only random variable in $\{U_t^{(n)}\}_{t=1}^{T(n)}$ that is distributed differently under each hypothesis is $U_{t_{\mathrm{A}}}^{(n)}$. We thus denote by $U_{t_{\mathrm{A}}}^{(n,0)}$ and $U_{t_{\mathrm{A}}}^{(n,1)}$ the random variable corresponding to slot $t_{\mathrm{A}}$ under $H_0$ and $H_1$, respectively. Regardless of Alice's transmission state, $\{U_t^{(n)}\}_{\substack{t=1 \\ t \neq t_{\mathrm{A}}}}^{T(n)}$ is an i.i.d. sequence, with $U_{t_{\mathrm{A}}}^{(n,0)}$ distributed identically to the elements of $\{U_t^{(n)}\}_{\substack{t=1 \\ t \neq t_{\mathrm{A}}}}^{T(n)}$. Therefore,

$$\Lambda_s^{(n)} = \frac{1}{T(n)} \sum_{\substack{t=1 \\ t \neq t_{\mathrm{A}}}}^{T(n)} U_t^{(n)} + \frac{U_{t_{\mathrm{A}}}^{(n,s)}}{T(n)}, \quad s \in \{0,1\}. \tag{4}$$

Let's denote by $\mu_{\mathrm{U}}(n)$ and $\sigma_{\mathrm{U}}^2(n)$ respectively the mean and variance of $U_{t_{\mathrm{A}}}^{(n,0)}$ and $U_t^{(n)}$, $t \neq t_{\mathrm{A}}$.

We define $g_n(x)$ as:

$$g_n(x) = \frac{(x - (T(n)-1)\mu_{\mathrm{U}}(n))T(n)}{\sigma_{\mathrm{U}}(n)\sqrt{T(n)-1}}. \tag{5}$$

Thus, the re-scaled test statistic $L_s^{(n)}$ is expressed as follows:

$$L_0^{(n)} = \frac{1}{\sqrt{T(n)-1}} \sum_{\substack{t=1 \\ t \neq t_{\mathrm{A}}}}^{T(n)} \frac{U_t^{(n)} - \mu_{\mathrm{U}}(n)}{\sigma_{\mathrm{U}}(n)} + \frac{U_{t_{\mathrm{A}}}^{(n,0)}}{\sigma_{\mathrm{U}}(n)\sqrt{T(n)-1}} \tag{6}$$

when $H_0$ is true, and

$$L_1^{(n)} = \frac{1}{\sqrt{T(n)-1}} \sum_{\substack{t=1 \\ t \neq t_{\mathrm{A}}}}^{T(n)} \frac{U_t^{(n)} - \mu_{\mathrm{U}}(n)}{\sigma_{\mathrm{U}}(n)} + \frac{U_{t_{\mathrm{A}}}^{(n,1)}}{\sigma_{\mathrm{U}}(n)\sqrt{T(n)-1}} \tag{7}$$

when $H_1$ is true. Provided $U_t^{(n)}$ satisfies the regularity conditions required by the central limit theorem (CLT) for triangular arrays [24, Theorem 27.2], $\frac{1}{\sqrt{T(n)-1}} \sum_{\substack{t=1 \\ t \neq t_{\mathrm{A}}}}^{T(n)} \frac{U_t^{(n)} - \mu_{\mathrm{U}}(n)}{\sigma_{\mathrm{U}}(n)} \xrightarrow{\mathcal{D}} Z$, where $Z \sim \mathcal{N}(0,1)$. Thus, the weighted sum in (6) and (7) corresponds to $S^{(n)}$ in Lemma 1. Now consider the term corresponding to slot $t_{\mathrm{A}}$, $\frac{U_{t_{\mathrm{A}}}^{(n,s)}}{\sigma_{\mathrm{U}}(n)\sqrt{T(n)-1}}$. It effectively offsets $Z$'s mean away from zero and its distribution depends on Alice's transmission state $s$. Thus, depending on which hypothesis is true, it maps to either $V_0^{(n)}$ or $V_1^{(n)}$ in Lemma 1. To prove achievability of covert communication, we show that there exists a coding scheme for Alice such that the random variable describing the LRT statistic has the form given in (6) and (7), with the terms in the sums satisfying the regularity conditions required by the CLT and the term corresponding to slot $t_{\mathrm{A}}$ converging to zero in probability. This allows us to establish the covertness of Alice's transmission by applying Lemma 1. We prove reliability by extending the random coding arguments from [3], [4].

### *B. Average Power Constraint*

We first show achievability under an average power constraint $P_{\max} \in (0, \infty)$.

**Theorem 1.1** (Achievability)**.** *Suppose Alice has a slotted AWGN channel to Bob with $T(n) = \omega(1)$ slots, each containing $n$ symbol periods, and that her transmitter is subject to the average*

*power constraint $P_{\max} \in (0, \infty)$. Then, provided that Alice and Bob share a sufficiently long secret, if Alice chooses to, she can transmit $\mathcal{O}\left(\min\{\sqrt{n \log T(n)}, n\}\right)$ bits in a single slot while $\lim_{n\to\infty} \mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}} \geq 1 - \epsilon$ and $\lim_{n\to\infty} \mathbb{P}_{\mathrm{e}}^{(\mathrm{b})} \leq \delta$ for arbitrary $\epsilon > 0$ and $\delta > 0$.*

*Proof:* **Construction:** Alice secretly selects slot $t_{\mathrm{A}}$ uniformly at random out of the $T(n)$ slots. Alice's channel encoder takes as input blocks of length $M$ bits and encodes them into codewords of length $n$ symbols. We employ a random coding argument and independently generate $2^M$ codewords $\{\mathbf{c}(W_k), k = 1, 2, \ldots, 2^M\}$ from $\mathbb{R}^n$ for messages $\{W_k\}_{k=1}^{2^M}$, each according to $p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} p_X(x_i)$, where $X \sim \mathcal{N}(0, P_{\mathrm{f}})$ and symbol power $P_{\mathrm{f}} < \frac{\sigma_{\mathrm{w}}^2}{2}$ is defined later. The codebook[4] is used only to send a single message and, along with $t_{\mathrm{A}}$, is the secret not revealed to Willie, though he knows how it is constructed, including the value of $P_{\mathrm{f}}$.

**Analysis (Willie):** Denote by $\Upsilon_t = \sum_{Y_i \in \mathbf{Y}_{\mathrm{w}}(t)} Y_i^2$ the power in slot $t$. Since Willie's channel from Alice is corrupted by AWGN with power $\sigma_{\mathrm{w}}^2$, the likelihood function of the observations $\mathbf{Y}_{\mathrm{w}}$ under $H_0$ is:

$$f_0(\mathbf{Y}_{\mathrm{w}}) = \left(\frac{1}{2\pi\sigma_{\mathrm{w}}^2}\right)^{\frac{nT(n)}{2}} \exp\left[-\frac{1}{2\sigma_{\mathrm{w}}^2} \sum_{t=1}^{T(n)} \Upsilon_t\right]. \tag{8}$$

Since Willie does not know which of the $T(n)$ slots Alice randomly selects for communication, or the codebook Alice and Bob use, but knows that Alice's signal is Gaussian, the likelihood function of the observations $\mathbf{Y}_{\mathrm{w}}$ under $H_1$ is:

$$f_1(\mathbf{Y}_{\mathrm{w}}) = \frac{1}{(2\pi\sigma_{\mathrm{w}}^2)^{\frac{(T(n)-1)n}{2}} (2\pi(\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}}))^{\frac{n}{2}} T(n)} \times \sum_{t=1}^{T(n)} \exp\left[-\frac{\Upsilon_t}{2(\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}})} - \frac{\sum_{\substack{r=1 \\ r\neq t}}^{T(n)} \Upsilon_r}{2\sigma_{\mathrm{w}}^2}\right]. \tag{9}$$

[4]Another way of viewing the construction is as a choice of one of $T(n)$ codebooks, where the $i^{\text{th}}$ codebook has a block of non-zero symbols in the $i^{\text{th}}$ slot. Selection of the $t_{\mathrm{A}}$-th slot is equivalent to selection of the $t_{\mathrm{A}}$-th codebook and the message is encoded by choosing a codeword from the selected codebook.

The LRT statistic $\Lambda_s^{(n)}$ is the ratio between (8) and (9). Re-arranging terms yields:

$$\Lambda_s^{(n)} = \frac{1}{T(n)} \sum_{t=1}^{T(n)} \left( \frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}}} \right)^{\frac{n}{2}} \exp \left[ \frac{P_{\mathrm{f}} \Upsilon_t}{2\sigma_{\mathrm{w}}^2(\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}})} \right]. \tag{10}$$

When Alice does not transmit in the $i^{\text{th}}$ symbol period, $Y_i \sim \mathcal{N}(0, \sigma_{\mathrm{w}}^2)$ since Willie observes AWGN; when Alice transmits, $Y_i \sim \mathcal{N}(0, \sigma_{\mathrm{w}}^2 + P_{\mathrm{f}})$ by construction. Let $\{X_t\}$, $X_t \sim \chi_n^2$, $t = 1, \ldots, T(n)$ be a sequence of i.i.d. chi-squared random variables with $n$ degrees of freedom. Then $\Upsilon_t = \sigma_{\mathrm{w}}^2 X_t$ for all $t \in \{1, \ldots, T(n)\}$ under $H_0$ and $t \in \{1, \ldots, T(n)\} \setminus \{t_{\mathrm{A}}\}$ under $H_1$, while $\Upsilon_{t_{\mathrm{A}}} = (\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}}) X_{t_{\mathrm{A}}}$ under $H_1$. Let $U_t^{(n)} = \left( \frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}}} \right)^{\frac{n}{2}} \exp \left[ \frac{P_{\mathrm{f}} X_t}{2(\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}})} \right]$ for all $t \in \{1, \ldots, T(n)\} \setminus \{t_{\mathrm{A}}\}$, $U_{t_{\mathrm{A}}}^{(n,0)} = \left( \frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}}} \right)^{\frac{n}{2}} \exp \left[ \frac{P_{\mathrm{f}} X_t}{2(\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}})} \right]$, and $U_{t_{\mathrm{A}}}^{(n,1)} = \left( \frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}}} \right)^{\frac{n}{2}} \exp \left[ \frac{P_{\mathrm{f}} X_t}{2\sigma_{\mathrm{w}}^2} \right]$. Application of $g_n(x)$ in (5) to (10) yields the expression for $L_s^{(n)}$ in the form defined in (6) and (7).

Using the moment generating function (MGF) $\mathcal{M}_{\chi_n^2}(x) = (1 - 2x)^{-n/2}$ of a chi-squared random variable, we have:

$$\mu_{\mathrm{U}}(n) = \left( \frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}}} \right)^{\frac{n}{2}} \mathbb{E} \left[ \exp \left( \frac{P_{\mathrm{f}} X_t}{2(\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}})} \right) \right] = 1 \tag{11}$$

$$\sigma_{\mathrm{U}}^2(n) = \left( \frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}}} \right)^{n} \mathbb{E} \left[ \exp \left( \frac{P_{\mathrm{f}} X_t}{\sigma_{\mathrm{w}}^2 + P_{\mathrm{f}}} \right) \right] - 1 = \left( \frac{\sigma_{\mathrm{w}}^4}{\sigma_{\mathrm{w}}^4 - P_{\mathrm{f}}^2} \right)^{\frac{n}{2}} - 1. \tag{12}$$

Since $P_{\mathrm{f}} < \frac{\sigma_{\mathrm{w}}^2}{2}$, $\mu_{\mathrm{U}}(n)$ and $\sigma_{\mathrm{U}}^2(n)$ satisfy conditions for the CLT [24, Theorem 27.2], $\frac{1}{\sqrt{T(n)-1}} \sum_{\substack{t=1 \\ t \neq t_{\mathrm{A}}}}^{T(n)} \frac{U_t^{(n)} - \mu_{\mathrm{U}}(n)}{\sigma_{\mathrm{U}}(n)} \xrightarrow{\mathcal{D}} Z$, where $Z \sim \mathcal{N}(0, 1)$. Also, when Alice does not transmit, by Chebyshev's inequality:

$$\mathbb{P} \left( \left| \frac{U_{t_{\mathrm{A}}}^{(n,0)}}{\sigma_{\mathrm{U}}(n) \sqrt{T(n) - 1}} \right| > \delta \right) \leq \frac{\sigma_{\mathrm{U}}^2(n)}{(\delta \sigma_{\mathrm{U}}(n) \sqrt{T(n) - 1} - 1)^2}.$$

Since $T(n) = \omega(1)$ and $P_{\mathrm{f}} < \frac{\sigma_{\mathrm{w}}^2}{2}$, $\frac{U_{t_{\mathrm{A}}}^{(n,0)}}{\sigma_{\mathrm{U}}(n) \sqrt{T(n)-1}} \xrightarrow{\mathcal{P}} 0$ as $n \to \infty$.

When Alice transmits,

$$\mathbb{E}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right] = \left(\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2+P_{\mathrm{f}}}\right)^{\frac{n}{2}} \mathbb{E}\left[\exp\left(\frac{P_{\mathrm{f}}X_t}{2\sigma_{\mathrm{w}}^2}\right)\right] = \left(\frac{\sigma_{\mathrm{w}}^4}{\sigma_{\mathrm{w}}^4-P_{\mathrm{f}}^2}\right)^{\frac{n}{2}} \tag{13}$$

$$\mathbb{E}\left[\left(U_{t_{\mathrm{A}}}^{(n,1)}\right)^2\right] = \left(\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2+P_{\mathrm{f}}}\right)^{n} \mathbb{E}\left[\exp\left(\frac{P_{\mathrm{f}}X_t}{\sigma_{\mathrm{w}}^2}\right)\right] = \left(\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2+P_{\mathrm{f}}}\right)^{n} \left(\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2-2P_{\mathrm{f}}}\right)^{\frac{n}{2}}. \tag{14}$$

To complete the analysis of Willie's detector, we must show $\frac{U_{t_{\mathrm{A}}}^{(n,1)}}{\sigma_{\mathrm{U}}(n)\sqrt{T(n)-1}} \xrightarrow{\mathcal{P}} 0$ as $n\to\infty$. By Chebyshev's inequality:

$$\begin{aligned}\mathbb{P}\left(\left|\frac{U_{t_{\mathrm{A}}}^{(n,1)}}{\sigma_{\mathrm{U}}(n)\sqrt{T(n)-1}}\right| > \delta\right) &\leq \frac{\mathrm{Var}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right]}{\left(\delta\sigma_{\mathrm{U}}(n)\sqrt{T(n)-1} - \mathbb{E}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right]\right)^2} \\ &= \left(\frac{\delta\sigma_{\mathrm{U}}(n)\sqrt{T(n)-1}}{\sqrt{\mathrm{Var}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right]}} - \frac{\mathbb{E}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right]}{\sqrt{\mathrm{Var}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right]}}\right)^{-2}. \end{aligned} \tag{15}$$

Thus, it is sufficient to show that, as long as $n\to\infty$ and $T(n)=\omega(1)$,

$$\mathbb{E}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right] \Bigg/ \sqrt{\mathrm{Var}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right]} \to 0, \text{ and} \tag{16}$$

$$\sigma_{\mathrm{U}}^2(n)(T(n)-1) \Bigg/ \mathrm{Var}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right] \to \infty. \tag{17}$$

We use (12), (13) and (14) to obtain:

$$\frac{\mathbb{E}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right]}{\sqrt{\mathrm{Var}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right]}} = \frac{1}{\sqrt{\left(1+\frac{P_{\mathrm{f}}^2}{\sigma_{\mathrm{w}}^4(1-2P_{\mathrm{f}}/\sigma_{\mathrm{w}}^2)}\right)^{\frac{n}{2}}-1}} \tag{18}$$

$$\frac{\sigma_{\mathrm{U}}^2(n)}{\mathrm{Var}\left[U_{t_{\mathrm{A}}}^{(n,1)}\right]} \geq \frac{\sigma_{\mathrm{U}}^2(n)}{\mathbb{E}\left[\left(U_{t_{\mathrm{A}}}^{(n,1)}\right)^2\right]} \tag{19}$$

$$= \left(1-\frac{2P_{\mathrm{f}}^2}{\sigma_{\mathrm{w}}^4(1-P_{\mathrm{f}}/\sigma_{\mathrm{w}}^2)}\right)^{\frac{n}{2}} - \left(\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2+P_{\mathrm{f}}}\right)^{-n}\left(\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2-2P_{\mathrm{f}}}\right)^{-\frac{n}{2}} \tag{20}$$

$$= \left(1-\frac{2P_{\mathrm{f}}^2}{\sigma_{\mathrm{w}}^4(1-P_{\mathrm{f}}/\sigma_{\mathrm{w}}^2)}\right)^{\frac{n}{2}} - o(1), \tag{21}$$

where (21) follows since $\sigma_{\mathrm{w}}^2 > 0$ and $P_{\mathrm{f}}$ satisfies $0 \leq P_{\mathrm{f}} < \sigma_{\mathrm{w}}^2/2$ by construction.

Now consider two scaling regimes for $T(n)$:

- $T(n) = o(e^n)$: set $P_{\mathrm{f}} = \frac{c_{\mathrm{P}}^{(\mathrm{S})}\sigma_{\mathrm{w}}^2\sqrt{\log T(n)}}{\sqrt{n}}$ with $c_{\mathrm{P}}^{(\mathrm{S})} > 0$ a constant determined later. Taylor series expansion of $\log(1+x)$ at $x = 0$ yields:
$$\left(1 + \frac{P_{\mathrm{f}}^2}{\sigma_{\mathrm{w}}^4(1 - 2P_{\mathrm{f}}/\sigma_{\mathrm{w}}^2)}\right)^{\frac{n}{2}} = e^{\frac{n}{2}\log\left(1+\frac{P_{\mathrm{f}}^2}{\sigma_{\mathrm{w}}^4(1-2P_{\mathrm{f}}/\sigma_{\mathrm{w}}^2)}\right)} = e^{\frac{c_{\mathrm{P}}^{(\mathrm{S})}}{2}\log T(n) - o(\log T(n))},$$
implying that (18) converges to zero since $T(n) = \omega(1)$. Thus, (16) holds. Furthermore, Taylor series expansion of $\log(1-x)$ at $x = 0$ shows:
$$\left(1 - \frac{2P_{\mathrm{f}}^2}{\sigma_{\mathrm{w}}^4(1 - P_{\mathrm{f}}/\sigma_{\mathrm{w}}^2)}\right)^{\frac{n}{2}} = e^{\frac{n}{2}\log\left(1-\frac{2P_{\mathrm{f}}^2}{\sigma_{\mathrm{w}}^4(1-P_{\mathrm{f}}/\sigma_{\mathrm{w}}^2)}\right)} = e^{-c_{\mathrm{P}}^{(\mathrm{S})}\log T(n) - o(\log T(n))},$$
Thus, there exists $c_{\mathrm{P}}^{(\mathrm{S})} \in (0,1)$ such that (17) holds.
- $T(n) = \Omega(e^n)$: set $P_{\mathrm{f}} = c_{\mathrm{P}}^{(\mathrm{L})}\sigma_{\mathrm{w}}^2/2$ with $c_{\mathrm{P}}^{(\mathrm{L})} \in (0,1)$ a constant. Then, clearly, (16) holds since (18) converges to zero, and (17) holds for an appropriately chosen $c_{\mathrm{P}}^{(\mathrm{L})}$.

Therefore, setting $P_{\mathrm{f}} = \sigma_{\mathrm{w}}^2 \min\{\frac{c_{\mathrm{P}}^{(\mathrm{S})}\sqrt{\log T(n)}}{\sqrt{n}}, \frac{c_{\mathrm{P}}^{(\mathrm{L})}}{2}\}$ ensures convergence of the RHS of (15) to zero for any scaling of $T(n)$, and, by Lemma 1, ensures $\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}} \geq 1 - \epsilon$ for any $\epsilon > 0$.

**Analysis (Bob):** Let Bob employ the maximum likelihood (ML) decoder (i.e., minimum distance decoder). If Bob knows the value of $t_{\mathrm{A}}$, Alice can reliably (i.e., with Bob's decoding error probability, averaged over all the codebooks, decaying to zero as $n \to \infty$) transmit $M = \frac{n\gamma}{2}\log_2\left(1 + \frac{\sigma_{\mathrm{w}}^2}{2\sigma_{\mathrm{b}}^2}\min\left\{\frac{c_{\mathrm{P}}^{(\mathrm{S})}\sqrt{\log T(n)}}{\sqrt{n}}, \frac{c_{\mathrm{P}}^{(\mathrm{L})}}{2}\right\}\right)$ covert bits, where $\gamma \in (0,1)$ is a constant [3], [4]. However, knowledge of $t_{\mathrm{A}}$ is unnecessary for Bob if $T(n) < 2^{c_{\mathrm{T}}n}$, where $c_{\mathrm{T}} > 0$ is a constant, as we show next. Let's augment Alice and Bob's Gaussian codebook with the origin $\mathbf{c}(0) = \{0, \ldots, 0\}$ (indicating "no transmission") and have Bob attempt to decode each of the $T(n)$ slots. The squared distance between a codeword $\mathbf{c}(W_k)$ and $\mathbf{c}(0)$ is $P_{\mathrm{f}}X$, where $X \sim \chi_n^2$. Repeating the analysis of Bob's detection error probability from [3], [4] using the distance between $\mathbf{c}(W_k)$ and $\mathbf{c}(0)$ instead of $\mathbf{c}(W_i)$ yields a looser upper bound on the probability of the decoding error in each slot. By the union bound over all $T(n)$ slots, the overall probability of error is $\mathbb{P}_{\mathrm{e}}^{(\mathrm{b})} \leq T(n)2^{M - \frac{n}{2}\log_2\left(1+\frac{P_{\mathrm{f}}}{4\sigma_{\mathrm{b}}^2}\right)}$. If $T(n) = o(e^n)$, then clearly Bob's decoding error

probability decays to zero if Alice attempts to transmit $M = \frac{n\gamma}{2}\log_2\left(1 + \frac{c_{\mathrm{P}}^{(\mathrm{S})}\sigma_{\mathrm{w}}^2\sqrt{\log T(n)}}{4\sigma_{\mathrm{b}}^2\sqrt{n}}\right)$ bits in a randomly selected $n$-symbol slot $t_{\mathrm{A}}$. If $T(n) = \Omega(e^n)$, then, $P_{\mathrm{f}} = \frac{c_{\mathrm{P}}^{(\mathrm{L})}\sigma_{\mathrm{w}}^2}{2}$, and $T(n) < 2^{c_{\mathrm{T}}n}$ where $c_{\mathrm{T}} = \frac{1-\gamma}{2}\log_2\left(1 + \frac{c_{\mathrm{P}}^{(\mathrm{L})}\sigma_{\mathrm{w}}^2}{8\sigma_{\mathrm{b}}^2}\right)$ ensures that Bob's decoding error probability decays to zero if Alice attempts to transmit $M = \frac{n\gamma}{2}\log_2\left(1 + \frac{c_{\mathrm{P}}^{(\mathrm{L})}\sigma_{\mathrm{w}}^2}{8\sigma_{\mathrm{b}}^2}\right)$ bits in a randomly selected $n$-symbol slot $t_{\mathrm{A}}$. Therefore, $\mathcal{O}(\min\{\sqrt{n\log T(n)}, n\})$ covert bits can be transmitted reliably using slot $t_{\mathrm{A}}$. ■

*Remark:* The logarithm of (10) is the log-likelihood ratio:

$$\begin{aligned}\log\Lambda_s^{(n)} &= -\log T(n) + u(n)\log\sum_{t=1}^{T(n)}\exp\left[v(n)\Upsilon_t\right] \\ &= -\log T(n) + u(n)\,\mathrm{LogSumExp}(\{v(n)\Upsilon_t\}_{t=1}^{T(n)}),\end{aligned}$$

where $u(n) \equiv \left(\frac{\sigma_{\mathrm{w}}^2}{\sigma_{\mathrm{w}}^2+P_{\mathrm{f}}}\right)^{\frac{n}{2}}$ and $v(n) \equiv \frac{P_{\mathrm{f}}}{2\sigma_{\mathrm{w}}^2(\sigma_{\mathrm{w}}^2+P_{\mathrm{f}})}$. Since $\mathrm{LogSumExp}(\{x_i\})$ is an analytic approximation of $\max(\{x_i\})$ [25, Ch. 3.1.5], Willie's (approximate) sufficient statistic is the maximum slot power $\Upsilon_{\max} = \max_{t\in\{1,\ldots,T(n)\}}\Upsilon_t$. While this motivates the design of Willie's detector in the converse proof, in the achievability proofs we analyze the exact LRT.

### *C. Peak Power Constraint*

Unfortunately, representing real-valued codewords requires unbounded storage, which means that the length of the secret pre-shared by Alice and Bob is infinite. To address this, we consider a finite alphabet, which also satisfies a peak power constraint $P_{\max} \in (0, \infty)$ on the transmitter. The remark in [4, Sec. III] allows both improvement of Bob's decoding performance and reduction of the size of the pre-shared secret to $\mathcal{O}(n)$ bits (provided $T(n) < 2^{c_{\mathrm{T}}n}$). Approaches reported in [4], [6], [7] may reduce the pre-shared secret to $\mathcal{O}(\sqrt{n}\log n)$ bits, and even possibly to $\mathcal{O}(\sqrt{n})$ bits.

**Theorem 1.2** (Achievability under peak power constraint)**.** *Suppose Alice has a slotted AWGN channel to Bob with $T(n) = \omega(1)$ slots, each containing $n$ symbol periods, and that her trans-*

*mitter is subject to the peak power constraint* $P_{\max} \in (0, \infty)$*. Then, provided that Alice and Bob share a sufficiently long secret, if Alice chooses to, she can transmit* $\mathcal{O}\left(\min\{\sqrt{n \log T(n)}, n\}\right)$ *bits in a single slot while* $\lim_{n\to\infty} \mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}} \geq 1 - \epsilon$ *and* $\lim_{n\to\infty} \mathbb{P}_{\mathrm{e}}^{(\mathrm{b})} \leq \delta$ *for any* $\epsilon > 0$ *and* $\delta > 0$*.*

*Proof:* **Construction:** Alice secretly selects slot $t_{\mathrm{A}}$ uniformly at random out of the $T(n)$ slots in which to communicate. She encodes the input in blocks of length $M$ bits into codewords of length $n$ symbols with the symbols drawn from alphabet $\{-a, a\}$, where $a$ satisfies the peak power constraint $a^2 < P_{\max}$ and is defined later. Alice independently generates $2^M$ codewords $\{\mathbf{c}(W_k), k = 1, 2, \ldots, 2^M\}$ for messages $\{W_k\}$ from $\{-a, a\}^n$ according to $p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} p_X(x_i)$, where $p_X(-a) = p_X(a) = \frac{1}{2}$. As in the proof of Theorem 1.1, this single-use codebook is not revealed to Willie, though he knows how it is constructed, including the value of $a$. While in this proof the entire codebook is secretly shared between Alice and Bob, the amount of shared secret information can be reduced using the remark in [4, Sec. III].

**Analysis (Willie):** Since the model for the AWGN channel from Alice to Willie is the same as in Theorem 1.1, the likelihood function of the observations $\mathbf{Y}_{\mathrm{w}}$ under $H_0$ is given by (8). Since Willie does not know which of the $T(n)$ slots Alice randomly selects for communication, or the codebook Alice and Bob use, but knows how the codebook is constructed, the likelihood function of the observations $\mathbf{Y}_{\mathrm{w}}$ under $H_1$ is:

$$f_1(\mathbf{Y}_{\mathrm{w}}) = \frac{1}{(2\pi\sigma_{\mathrm{w}}^2)^{\frac{nT(n)}{2}} T(n)} \sum_{t=1}^{T(n)} \frac{1}{2^n} \sum_{\mathbf{b}\in\{-1,1\}^n} e^{-\frac{\sum_{\substack{r=1 \\ r\neq t}}^{T(n)} \Upsilon_r + \sum_{Y_i \in \mathbf{Y}_{\mathrm{w}}(t)} (Y_i - a b_i)^2}{2\sigma_{\mathrm{w}}^2}}, \tag{22}$$

where $\Upsilon_t = \sum_{Y_i \in \mathbf{Y}_{\mathrm{w}}(t)} Y_i^2$ denotes the power in slot $t$. The LRT statistic $\Lambda_s^{(n)}$ is the ratio between (8) and (22). Re-arranging terms yields:

$$\Lambda_s^{(n)} = \frac{1}{T(n)} \sum_{t=1}^{T(n)} \frac{\exp\left[-\frac{na^2}{2\sigma_{\mathrm{w}}^2}\right]}{2^n} \sum_{\mathbf{b}\in\{-1,1\}^n} e^{\frac{a}{\sigma_{\mathrm{w}}^2} \sum_{Y_i \in \mathbf{Y}_{\mathrm{w}}(t)} Y_i b_i}. \tag{23}$$

Let $U_t^{(n)} = \frac{\exp\left[-\frac{na^2}{2\sigma_{\rm w}^2}\right]}{2^n} \sum_{\mathbf{b}\in\{-1,1\}^n} \exp\left[\frac{a}{\sigma_{\rm w}^2}\sum_{Y_i\in\mathbf{Y}_{\rm w}(t)} Y_i b_i\right]$. Application of $g_n(x)$ in (5) to (23) yields the expression for $L_s^{(n)}$ in the form defined in (6) and (7).

When Alice does not transmit, $\exp\left[\frac{a}{\sigma_{\rm w}^2}\sum_{Y_i\in\mathbf{Y}_{\rm w}(t)} Y_i b_i\right] \sim \log\mathcal{N}\left(0, \frac{na^2}{\sigma_{\rm w}^2}\right)$, where $\log\mathcal{N}(\mu,\sigma^2)$ denotes the log-normal distribution with location $\mu$ and scale $\sigma^2$. Thus,

$$\mu_{\rm U}(n) = \frac{\exp\left[-\frac{na^2}{2\sigma_{\rm w}^2}\right]}{2^n} \sum_{\mathbf{b}\in\{-1,1\}^n} \mathbb{E}\left[e^{\frac{a}{\sigma_{\rm w}^2}\sum_{Y_i\in\mathbf{Y}_{\rm w}(t)} Y_i b_i}\right] = 1.$$

To obtain $\sigma_{\rm U}^2(n)$, we calculate the second moment of $U_{t_{\rm A}}^{(n,0)}$ and $U_t^{(n)}$, $t \neq t_{\rm A}$:

$$\mathbb{E}\left[\left(U_t^{(n)}\right)^2\right] = \frac{e^{-\frac{na^2}{\sigma_{\rm w}^2}}}{2^{2n}} \sum_{\mathbf{b},\mathbf{d}\in\{-1,1\}^n} \mathbb{E}\left[e^{\frac{a}{\sigma_{\rm w}^2}\sum_{Y_i\in\mathbf{Y}_{\rm w}(t)} Y_i(b_i+d_i)}\right] = \frac{1}{2^{2n}} \sum_{\mathbf{b},\mathbf{d}\in\{-1,1\}^n} e^{\frac{a^2}{\sigma_{\rm w}^2}\sum_{i=1}^n b_i d_i} \tag{24}$$

$$= \frac{1}{2^{2n}} \sum_{\mathbf{b},\mathbf{z}\in\{-1,1\}^n} e^{\frac{a^2}{\sigma_{\rm w}^2}\sum_{i=1}^n z_i} \tag{25}$$

$$= \cosh^n\left(\frac{a^2}{\sigma_{\rm w}^2}\right), \tag{26}$$

where (24) follows from $\exp\left[\frac{a}{\sigma_{\rm w}^2}\sum_{Y_i\in\mathbf{Y}_{\rm w}(t)} Y_i(b_i+d_i)\right] \sim \log\mathcal{N}\left(0, \frac{a^2}{\sigma_{\rm w}^2}\sum_{i=1}^n (b_i+d_i)^2\right)$; (25) is since for a given $\mathbf{b}\in\{-1,1\}^n$ and any $\mathbf{d}\in\{-1,1\}^n$, $\mathbf{z} = [b_1d_1, b_2d_2, \ldots, b_nd_n] \in \{-1,1\}^n$ is unique; and (26) follows from Appendix B. Thus, $\frac{1}{\sqrt{T(n)-1}} \sum_{\substack{t=1\\ t\neq t_{\rm A}}}^{T(n)} \frac{U_t^{(n)}-\mu_{\rm U}(n)}{\sigma_{\rm U}(n)} \xrightarrow{\mathcal{D}} Z$, $Z\sim\mathcal{N}(0,1)$. Also, when Alice does not transmit, by Chebyshev's inequality:

$$\mathbb{P}\left(\left|\frac{U_{t_{\rm A}}^{(n,0)}}{\sigma_{\rm U}(n)\sqrt{T(n)-1}}\right| > \delta\right) \leq \frac{\sigma_{\rm U}^2(n)}{(\delta\sigma_{\rm U}(n)\sqrt{T(n)-1}-1)^2}.$$

Since $T(n) = \omega(1)$, $\frac{U_{t_{\rm A}}^{(n,0)}}{\sigma_{\rm U}(n)\sqrt{T(n)-1}} \xrightarrow{\mathcal{P}} 0$ as $n\to\infty$.

When Alice transmits, by construction, Willie observes $y_{(t_{\rm A}-1)n+i} \sim \mathcal{N}(ac_i, \sigma_{\rm w}^2)$, $i = 1,\ldots,n$, where $\mathbf{c} = [c_1,\ldots,c_n]$ is drawn equiprobably from $\{-1,1\}^n$. Thus,

$$\mathbb{E}\left[U_{t_{\rm A}}^{(n,1)}\right] = \frac{e^{-\frac{na^2}{2\sigma_{\rm w}^2}}}{2^{2n}} \sum_{\mathbf{c},\mathbf{b}\in\{-1,1\}^n} \mathbb{E}\left[e^{\frac{a}{\sigma_{\rm w}^2}\sum_{Y_i\in\mathbf{Y}_{\rm w}(t)} Y_i b_i}\right]$$

$$= \frac{1}{2^{2n}} \sum_{\mathbf{c},\mathbf{b}\in\{-1,1\}^n} e^{\frac{a^2}{\sigma_{\rm w}^2}\sum_{i=1}^n b_i c_i} = \cosh^n\left(\frac{a^2}{\sigma_{\rm w}^2}\right), \tag{27}$$

where (27) follows from $\exp\left[\frac{a}{\sigma_{\rm w}^2}\sum_{Y_i\in\mathbf{Y}_{\rm w}(t)}Y_i b_i\right] \sim \log\mathcal{N}\left(\frac{a^2}{\sigma_{\rm w}^2}\sum_{i=1}^n c_i b_i, \frac{na^2}{\sigma_{\rm w}^2}\right)$, the argument for (25) above, and Appendix B. By the definition of variance and the law of total expectation, $\mathrm{Var}[U_{t_{\rm A}}^{(n,1)}] \leq \frac{1}{2^n}\sum_{\mathbf{c}\in\{-1,1\}^n}\mathbb{E}\left[\left(U_{t_{\rm A}}^{(n,1)}\middle|\mathbf{c}\text{ sent}\right)^2\right]$, where

$$\begin{aligned}\mathbb{E}\left[\left(U_{t_{\rm A}}^{(n,1)}\middle|\mathbf{c}\text{ sent}\right)^2\right] &= \frac{e^{-\frac{na^2}{\sigma_{\rm w}^2}}}{2^{2n}}\sum_{\mathbf{b},\mathbf{d}\in\{-1,1\}^n}\mathbb{E}\left[e^{\frac{a}{\sigma_{\rm w}^2}\sum_{Y_i\in\mathbf{Y}_{\rm w}(t)}Y_i(b_i+d_i)}\right]\\ &= \frac{e^{-\frac{na^2}{\sigma_{\rm w}^2}}}{2^{2n}}\sum_{\mathbf{b},\mathbf{d}\in\{-1,1\}^n} e^{\frac{a^2}{\sigma_{\rm w}^2}\sum_{i=1}^n c_i(b_i+d_i)+\frac{(b_i+d_i)^2}{2}}\\ &= \frac{1}{2^{2n}}\sum_{\mathbf{b},\mathbf{d}\in\{-1,1\}^n} e^{\frac{a^2}{\sigma_{\rm w}^2}\sum_{i=1}^n c_i b_i+(c_i+b_i)d_i} \leq \cosh^n\left(\frac{a^2}{\sigma_{\rm w}^2}\right)\cosh^n\left(\frac{2a^2}{\sigma_{\rm w}^2}\right),\end{aligned} \tag{28}$$

with (28) following from $c_i + b_i \leq 2$ and arguments for (27) above. By Chebyshev's inequality:

$$\mathbb{P}\left(\left|\frac{U_{t_{\rm A}}^{(n,1)}}{\sigma_{\rm U}(n)\sqrt{T(n)-1}}\right| > \delta\right) \leq \frac{\cosh^n\left(\frac{a^2}{\sigma_{\rm w}^2}\right)\cosh^n\left(\frac{2a^2}{\sigma_{\rm w}^2}\right)}{\left(\delta\sigma_{\rm U}(n)\sqrt{T(n)-1} - \mathbb{E}\left[U_{t_{\rm A}}^{(n,1)}\right]\right)^2}. \tag{29}$$

Dividing both numerator and denominator of (29) by $\cosh^n\left(\frac{a^2}{\sigma_{\rm w}^2}\right)\cosh^n\left(\frac{2a^2}{\sigma_{\rm w}^2}\right)$, we note that $\frac{\mathbb{E}\left[U_{t_{\rm A}}^{(n,1)}\right]}{\cosh^{\frac{n}{2}}\left(\frac{a^2}{\sigma_{\rm w}^2}\right)\cosh^{\frac{n}{2}}\left(\frac{2a^2}{\sigma_{\rm w}^2}\right)} = \left(\frac{\cosh(a^2/\sigma_{\rm w}^2)}{\cosh(2a^2/\sigma_{\rm w}^2)}\right)^{\frac{n}{2}} \leq 1$, as $\frac{\cosh(x)}{\cosh(2x)} \leq 1$ for $x\in\mathbb{R}$. Also, $\frac{\sigma_{\rm U}^2(n)}{\cosh^n\left(\frac{a^2}{\sigma_{\rm w}^2}\right)\cosh^n\left(\frac{2a^2}{\sigma_{\rm w}^2}\right)} = \cosh^{-n}\left(\frac{2a^2}{\sigma_{\rm w}^2}\right)$. When $T(n) = o(e^n)$, setting $a^2 = \frac{c_{\rm P}^{(\rm S)}\sigma_{\rm w}^2\sqrt{\log T(n)}}{\sqrt{2n}}$ for a constant $c_{\rm P}^{(\rm S)}\in(0,1)$ ensures $\frac{U_{t_{\rm A}}^{(n,1)}}{\sigma_{\rm U}(n)\sqrt{T(n)-1}} \xrightarrow{\mathcal{P}} 0$ as $n\to\infty$. Convergence follows from noting that $\cosh^{-n}\left(\frac{2a^2}{\sigma_{\rm w}^2}\right) = \exp\left[-n\log\cosh\left(\frac{2a^2}{\sigma_{\rm w}^2}\right)\right] \geq \exp\left[-\frac{2na^4}{\sigma_{\rm w}^4}\right]$. When $T(n) = \Omega(e^n)$, convergence is obtained by setting $a^2 = \frac{c_{\rm P}^{(\rm L)}\sigma_{\rm w}^2}{2}$. Therefore, by Lemma 1, setting $a^2 = \sigma_{\rm w}^2\min\left\{\frac{c_{\rm P}^{(\rm S)}\sqrt{\log T(n)}}{\sqrt{2n}}, \frac{c_{\rm P}^{(\rm L)}}{2}\right\}$ ensures $\mathbb{P}_{\rm FA}+\mathbb{P}_{\rm MD}\geq 1-\epsilon$ for any $\epsilon>0$.

**Analysis (Bob):** Suppose Alice transmits $\mathbf{c}(W_k)$. As in the proof of Theorem 1.1, let Bob employ the ML decoder that suffers an error event $E_{k\to i}$ when the received vector $\mathbf{Y}_{\rm b}$ is closer in Euclidean distance to codeword $\mathbf{c}(W_i)$, $i\neq k$. Knowledge of $t_{\rm A}$ ensures reliable decoding by the application of Appendix C: Bob's error probability, averaged over all the codebooks, decays to zero

as $n \to \infty$ for a covert transmission containing $M = n\gamma \left(1 - \log_2 \left[1 + \exp\left(-\frac{\sigma_{\mathrm{w}}^2}{2\sigma_{\mathrm{b}}^2} \min\left\{\frac{c_{\mathrm{P}}^{(\mathrm{S})}\sqrt{\log T(n)}}{\sqrt{2n}}, \frac{c_{\mathrm{P}}^{(\mathrm{L})}}{2}\right\}\right)\right]\right)$ bits, where $\gamma \in (0, 1)$ is a constant.

However, as in Theorem 1.1, knowledge of $t_{\mathrm{A}}$ is unnecessary for Bob if $T(n) < 2^{c_{\mathrm{T}} n}$ where $c_{\mathrm{T}}$ is a constant. Again, let's augment Alice and Bob's codebook with the origin $\mathbf{c}(0) = \{0, \ldots, 0\}$ (indicating "no transmission") and have Bob attempt to decode each of the $T(n)$ slots. Denoting the decoding error probability in slot $t$ by $\mathbb{P}_{\mathrm{e}}^{(\mathrm{b})}(t)$, we employ the union bound over all slots to upper-bound the overall decoding error probability:

$$\mathbb{P}_{\mathrm{e}}^{(\mathrm{b})} \leq \sum_{t=1}^{T(n)} \mathbb{P}_{\mathrm{e}}^{(\mathrm{b})}(t). \tag{30}$$

The decoding error probability for one of the $T(n) - 1$ slots that Alice does not use is the probability that the received vector is closer to some codeword than the origin $\mathbf{c}(0)$:

$$\mathbb{P}_{\mathrm{e}}^{(\mathrm{b})}(t) = \mathbb{P}\left(\cup_{i=1}^{2^M} E_{0\to i}\right) \leq \sum_{i=1}^{2^M} \mathbb{P}(E_{0\to i}), t \neq t_{\mathrm{A}}, \tag{31}$$

where the inequality is the union bound. Since the Euclidean distance between each codeword and $\mathbf{c}(0)$ is $\sqrt{n}a$, by [26, Eq. (3.44)]:

$$\mathbb{P}(E_{0\to i}) = Q\left(\frac{\sqrt{n}a}{2\sigma_{\mathrm{b}}}\right) \leq \frac{1}{2}\exp\left(-\frac{na^2}{8\sigma_{\mathrm{b}}^2}\right), \tag{32}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}dt$ and the inequality is because of the upper bound $Q(x) \leq \frac{1}{2}e^{-x^2/2}$ [27, Eq. (5)]. Substituting (32) into (31) yields:

$$\mathbb{P}_{\mathrm{e}}^{(\mathrm{b})}(t) \leq 2^{M - \frac{na^2 \log_2 e}{8\sigma_{\mathrm{b}}^2}}, t \neq t_{\mathrm{A}}. \tag{33}$$

To upper-bound the decoding error probability for the slot that Alice uses to transmit, we combine the bounds in (33) and (51) as follows:

$$\mathbb{P}_{\mathrm{e}}^{(\mathrm{b})}(t_{\mathrm{A}}) \leq 2^{M - n\min\left\{1 - \log_2\left[1 + \exp\left(-\frac{a^2}{2\sigma_{\mathrm{b}}^2}\right)\right], \frac{a^2 \log_2 e}{8\sigma_{\mathrm{b}}^2}\right\}}. \tag{34}$$

Combining (30), (33), and (34) yields:

$$\mathbb{P}_{\mathrm{e}}^{(\mathrm{b})} \leq (T(n) - 1)2^{M - \frac{na^2 \log_2 e}{8\sigma_{\mathrm{b}}^2}} + 2^{M - n\min\left\{1 - \log_2\left[1 + \exp\left(-\frac{a^2}{2\sigma_{\mathrm{b}}^2}\right)\right], \frac{a^2 \log_2 e}{8\sigma_{\mathrm{b}}^2}\right\}}. \tag{35}$$

If $T(n) = o(e^n)$, then Alice sets $a^2 = \frac{c_{\mathrm{P}}^{(\mathrm{S})}\sigma_{\mathrm{w}}^2\sqrt{\log T(n)}}{\sqrt{2n}}$. Thus, for large enough $n$, $\frac{a^2\log_2 e}{8\sigma_{\mathrm{b}}^2} < 1-\log_2\left[1+\exp\left(-\frac{a^2}{2\sigma_{\mathrm{b}}^2}\right)\right]$ and Bob's decoding error probability decays to zero if Alice attempts to transmit $M = \frac{\gamma c_{\mathrm{P}}^{(\mathrm{S})}\sigma_{\mathrm{w}}^2\sqrt{n\log T(n)}\log_2 e}{8\sqrt{2}\sigma_{\mathrm{b}}^2}$ bits in a randomly selected $n$-symbol slot $t_{\mathrm{A}}$, where $\gamma \in (0,1)$. If $T(n) = \Omega(e^n)$, then, $a^2 = \frac{c_{\mathrm{P}}^{(\mathrm{L})}\sigma_{\mathrm{w}}^2}{2}$, and $T(n) < 2^{c_{\mathrm{T}}n}$ where $c_{\mathrm{T}} = \frac{(1-\gamma)c_{\mathrm{P}}^{(\mathrm{L})}\sigma_{\mathrm{w}}^2\log_2 e}{16\sigma_{\mathrm{b}}^2}$ ensures that Bob's decoding error probability decays to zero if Alice attempts to transmit $M = n\gamma\min\left\{1-\log_2\left[1+\exp\left(-\frac{\sigma_{\mathrm{w}}^2}{4\sigma_{\mathrm{b}}^2}\right)\right], \frac{c_{\mathrm{P}}^{(\mathrm{L})}\sigma_{\mathrm{w}}^2\log_2 e}{16\sigma_{\mathrm{b}}^2}\right\}$ bits in slot $t_{\mathrm{A}}$. ■

## IV. Converse

In this section we show that Alice cannot transmit $\omega(\sqrt{n\log T(n)})$ bits both reliably and covertly using one of the $T(n)$ $n$-symbol slots. Alice attempts to send one of $2^M$ (equally likely) $M$-bit messages reliably to Bob using a sequence of $n$ consecutive symbol periods out of $nT(n)$, where $M = \omega(\sqrt{n\log T(n)})$, and each message is encoded arbitrarily into $n$ symbols. Unlike in the previous section, here Willie is oblivious to the locations of the slot boundaries, Alice's codebook construction scheme and other properties of her signal. Nevertheless, by dividing his sequence of $nT(n)$ observations into a set of $T(n)$ non-overlapping subsequences each containing $n$ symbols, and employing a simple threshold detector on the maximum subsequence power, Willie can detect Alice if she attempts to transmit $\omega(\sqrt{n\log T(n)})$ covert bits reliably.

**Theorem 2.** *Suppose Alice's transmitter is subject to the average power constraint $P_{\max} \in (0,\infty)$. If Alice attempts to transmit $\omega(\sqrt{n\log T(n)})$ bits using a sequence of $n$ consecutive symbol periods that are arbitrarily located inside a sequence of $nT(n)$ symbol periods, then, as $n\to\infty$, either Willie detects her with high probability, or Bob cannot decode with arbitrarily low probability of error.*

*Proof:* First, consider $\log T(n) = \omega(n)$. By the standard arguments [23, Ch. 9], the average power constraint implies that Alice can reliably transmit at most $\mathcal{O}(n)$ bits in $n$ channel uses. Since $n$ is asymptotically smaller than $\sqrt{n\log T(n)}$, the claim holds trivially. Therefore, we focus on $\log T(n) = \mathcal{O}(n)$ for the remainder of the proof.

Willie divides the sequence $\mathbf{Y}_{\mathrm{w}}$ of $nT(n)$ observations into a set of $T(n)$ non-overlapping subsequences $\{\mathbf{Y}_{\mathrm{w}}(t)\}_{t=1}^{T(n)}$, with each $\mathbf{Y}_{\mathrm{w}}(t)$ containing $n$ consecutive observations. Denote by $\Upsilon_t = \sum_{Y_i \in \mathbf{Y}_{\mathrm{w}}(t)} Y_i^2$ the observed power in each subsequence and $\Upsilon_{\max} = \max_{t\in\{1,\ldots,T(n)\}} \Upsilon_t$. For a threshold $\tau$, Willie accuses Alice of transmitting if $\Upsilon_{\max} > \tau$, setting

$$\tau = \sigma_{\mathrm{w}}^2(n + c\sqrt{n \log T(n)}), \tag{36}$$

with constant $c > 0$ determined next.

Suppose Alice does not transmit. For an arbitrary $\mathbb{P}_{\mathrm{FA}}^* > 0$, we show that there exists $c > 0$ such that the probability of false alarm $\mathbb{P}(\Upsilon_{\max} > \tau) \leq \mathbb{P}_{\mathrm{FA}}^*$ as $n \to \infty$. Note that each $\Upsilon_t = \sigma_{\mathrm{w}}^2 X_t$ where $\{X_t\}$, $X_t \sim \chi_n^2$, $t = 1, \ldots, T(n)$ is a sequence of i.i.d. chi-squared random variables each with $n$ degrees of freedom. We have:

$$\mathbb{P}[\Upsilon_{\max} > \tau] = 1 - \mathbb{P}\left[X_{\max} \leq \tau/\sigma_{\mathrm{w}}^2\right] = 1 - \left[1 - \mathbb{P}\left[X_1 > n + c\sqrt{n \log T(n)}\right]\right]^{T(n)},$$

where $X_{\max} = \max_{t\in\{1,\ldots,T\}} X_t$. The Chernoff bound for the tail of a chi-squared distribution [28, Lemma 2.2] yields:

$$\mathbb{P}[\Upsilon_{\max} > \tau] \leq 1 - \left[1 - e^{\frac{n}{2}\log\left(1+\frac{c\log T(n)}{\sqrt{n}}\right) - \frac{c\sqrt{n\log T(n)}}{2}}\right]^{T(n)} \leq 1 - \left[1 - e^{-\frac{c^2 \log T(n)}{4\left(1+c\sqrt{\frac{\log T(n)}{n}}\right)}}\right]^{T(n)} \tag{37}$$

$$= 1 - \left[1 - \frac{1}{T(n)^{\frac{c^2}{4\left(1+c\sqrt{\frac{\log T(n)}{n}}\right)}}}\right]^{T(n)} \tag{38}$$

where (37) is the application of $\log(1+x) \leq \frac{x(2+x)}{2(1+x)}$ for $x \geq 0$ [29, Eq. (3)] and re-arrangement of terms. By the definition of the asymptotic notation [18, Ch. 3.1], when $\log T(n) = \mathcal{O}(n)$, there exist real constant $k > 0$ and integer $n_0 > 0$ such that $\log T(n) \leq kn$ for all $n \geq n_0$. Setting $c > 2(\sqrt{k} + \sqrt{1+k})$ ensures that (38) converges to zero as $n \to \infty$ (when $\log T(n) = o(n)$, setting $c > 2$ suffices). Therefore, for any desired upper bound on the false alarm probability,

there exists a constant $c$ such that setting the threshold as in (36) guarantees that upper bound for $n$ large enough.

Now suppose Alice uses an arbitrary codebook $\{\mathbf{c}(W_k), k = 1, \ldots, 2^{nR}\}$ and transmits codeword $\mathbf{c}(W_k)$ using $n$ consecutive symbol periods. Denote the average symbol power of $\mathbf{c}(W_k)$ by $P_{\mathrm{f}} = \frac{\|\mathbf{c}(W_k)\|^2}{n}$. Since Alice uses $n$ consecutive symbols, her transmission overlaps at most two of Willie's subsequences, which we denote $t_{\mathrm{A}}$ and $t_{\mathrm{B}}$. Denote by $P_{\mathrm{A}}$ and $P_{\mathrm{B}}$ the power from Alice's transmission in subsequences $t_{\mathrm{A}}$ and $t_{\mathrm{B}}$, respectively, with $P_{\mathrm{A}} + P_{\mathrm{B}} = nP_{\mathrm{f}}$. Willie's probability of missing Alice's transmission is:

$$\mathbb{P}_{\mathrm{MD}}^{(k)} = \mathbb{P}(\Upsilon_{\max} \leq \tau) = \mathbb{P}(\Upsilon_{t_{\mathrm{A}}} \leq \tau)\mathbb{P}(\Upsilon_{t_{\mathrm{B}}} \leq \tau) \prod_{\substack{t=1 \\ t \notin \{t_{\mathrm{A}}, t_{\mathrm{B}}\}}}^{T(n)} \mathbb{P}(\Upsilon_t \leq \tau), \tag{39}$$

where the factorization in (39) is because Alice's codeword and the noise in other subsequences are independent. $\prod_{t=1, t\notin\{t_{\mathrm{A}}, t_{\mathrm{B}}\}}^{T(n)} \mathbb{P}(\Upsilon_t \leq \tau) \leq 1$ does not depend on Alice's codeword. However, since the codeword is an unknown deterministic signal that is added to AWGN on Willie's channel to Alice, $\frac{\Upsilon_{t_{\mathrm{A}}}}{\sigma_{\mathrm{w}}^2} \sim \chi_n^2(P_{\mathrm{A}})$ and $\frac{\Upsilon_{t_{\mathrm{B}}}}{\sigma_{\mathrm{w}}^2} \sim \chi_n^2(P_{\mathrm{B}})$ are non-central chi-squared random variables with $n$ degrees of freedom and respective non-centrality parameters $\frac{P_{\mathrm{A}}}{\sigma_{\mathrm{w}}^2}$ and $\frac{P_{\mathrm{B}}}{\sigma_{\mathrm{w}}^2}$. Without loss of generality, assume that $P_{\mathrm{A}} \geq P_{\mathrm{B}}$. Thus, $P_{\mathrm{A}}$ satisfies $\frac{nP_{\mathrm{f}}}{2} \leq P_{\mathrm{A}} \leq nP_{\mathrm{f}}$ and the expected value and variance of $\Upsilon_{t_{\mathrm{A}}}$ are bounded as follows [30, App. D.1]:

$$\mathbb{E}\left[\Upsilon_{t_{\mathrm{A}}}\right] \geq \sigma_{\mathrm{w}}^2 n + \frac{nP_{\mathrm{f}}}{2} \tag{40}$$

$$\mathrm{Var}\left[\Upsilon_{t_{\mathrm{A}}}\right] \leq 2n\sigma_{\mathrm{w}}^4 + 4n\sigma_{\mathrm{w}}^2 P_{\mathrm{f}}. \tag{41}$$

Since $\mathbb{P}(\Upsilon_{t_{\mathrm{B}}} \leq \tau) \leq 1$, Chebyshev's inequality with (40) and (41) yields:

$$\mathbb{P}_{\mathrm{MD}}^{(k)} \leq \mathbb{P}\left[|\Upsilon_{t_{\mathrm{A}}} - \mathbb{E}[\Upsilon_{t_{\mathrm{A}}}]| > \mathbb{E}[\Upsilon_{t_{\mathrm{A}}}] - \sigma_{\mathrm{w}}^2[n + c\sqrt{n \log T(n)}]\right] \leq \frac{2\sigma_{\mathrm{w}}^4 + 4\sigma_{\mathrm{w}}^2 P_{\mathrm{f}}}{\left(\frac{\sqrt{n}P_{\mathrm{f}}}{2} - c\sigma_{\mathrm{w}}^2\sqrt{\log T(n)}\right)^2}. \tag{42}$$

Therefore, if $P_{\mathrm{f}} = \omega\left(\sqrt{\frac{\log T(n)}{n}}\right)$, as $n \to \infty$, $\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}}$ can be made arbitrarily small. The proof of the non-zero lower bound on the decoding error probability $\mathbb{P}_{\mathrm{e}}^{(\mathrm{b})}$ if Alice tries to transmit

$\omega(\sqrt{n \log T(n)})$ bits in a single slot using average symbol power $P_{\mathrm{f}} = \mathcal{O}\left(\sqrt{\frac{\log T(n)}{n}}\right)$ follows from the arguments in the proof of Theorem 2 in [3], [4]. ■

## V. Discussion

Here we relate our work to other studies of covert communication. An overview of the area can be found in [31].

### *A. Relationship with Steganography*

Steganography is an ancient discipline [32] of hiding messages in innocuous objects. Modern steganographic systems [33] hide information by altering the properties of fixed-size, finite-alphabet covertext objects (e.g., images), and are subject to a similar SRL as covert communication: $\mathcal{O}(\sqrt{n})$ symbols in covertext of size $n$ may safely be modified to hide an $\mathcal{O}(\sqrt{n} \log n)$-bit message in the resulting stegotext [34]. The similarity between the SRLs in these disciplines comes from the mathematics of statistical hypothesis testing, as discussed in [4]. The extra $\log n$ factor is because of the lack of noise in the steganography context. However, arguably the earliest work on SRL [35] shows its achievability without the $\log n$ factor in the presence of an "active" adversary that corrupts stegotext using AWGN. This was re-discovered independently and published with the converse in [3], [4].

Batch steganography uses multiple covertext objects to hide a message and is subject to the steganographic SRL described above [36], [37]. The batch steganography interpretation of covert communication using the timing-based degree-of-freedom that is described here is equivalent to using only one of $T(n)$ covertext objects of size $n$ to embed a message. Willie, who knows that one covertext object is used but not which one, has to examine all of them. We are not aware of any work on this particular problem, but it is likely that one could extend our result to it. We also note that more recent work on steganography shows that an empirical model of the covertext suffices to break the steganographic SRL, allowing the embedding of $\mathcal{O}(n)$ bits in an $n$-symbol covertext [38]. However, this technique relies on embedding messages in covertext by

*replacing* part of it—something that cannot be done in standard communication systems unless Alice controls Willie's noise source.

### *B. Related Work in Physical Layer Covert Communication*

The emergence of radio-frequency (RF) communication systems necessitated the development of means to protect them from jamming, detection, and eavesdropping. Spread-spectrum techniques [39] address these issues by transmitting a signal that requires bandwidth $W_{\mathrm{M}}$ on a much wider bandwidth $W_{\mathrm{s}} \gg W_{\mathrm{M}}$, thus, effectively suppressing the power spectral density of the signal below the noise floor. This provides both covertness as well as the resistance to jamming, fading, and other interference.

However, while the spread-spectrum architectures are well-developed, the fundamental SRL for covert communication has been derived only recently [3], [4]. This resulted in the revival of the field, with follow-on work focusing on reducing the size of the pre-shared secret [5], [6], fully characterizing the optimal constant hidden by the big-$\mathcal{O}$ notation of the SRL [6], [7], and extending the SRL to quantum channels with Willie limited only by the laws of quantum mechanics [10]–[12]. Finally, while here we improve on the SRL by exploiting Willie's ignorance of transmission timing, other studies explore even stronger assumptions on his limitations. In particular, authors in [13], [16], [17] examine the impact of the errors in Willie's estimate of noise variance $\sigma_{\mathrm{w}}^2$ at his receiver, which allows $\mathcal{O}(n)$ covert bits to be transmitted in $n$ uses of the channel even when Willie has upper and lower bounds on $\sigma_{\mathrm{w}}^2$. Thus, positive-rate rather than SRL-governed covert communication is possible when Willie's knowledge of the channel is incomplete. However, successive work has demonstrated that the converse of Section IV still holds even if Willie does not know $\sigma_{\mathrm{w}}^2$, but under the restrictive assumption that Willie knows the slot boundaries [40]. The main result of [40] can then be combined with the approach of Theorem 2 to remove the requirement of Willie knowing the slot boundaries, as described in Corollary 2 of [40].

## VI. Conclusion and Future Work

We have shown that secretly pre-arranging a choice of a single $n$-symbol period slot out of $T(n)$ allows Alice to reliably transmit $\mathcal{O}(\min\{n, \sqrt{n \log T(n)}\})$ bits on an AWGN channel to Bob while rendering Willie's detector arbitrarily close to ineffective. Surprisingly, the multiplicative increase in transmitted information over the result in [3], [4] is obtained without the need for Bob to know which slot holds the transmission if $T(n) < 2^{c_{\mathrm{T}} n}$, where $c_{\mathrm{T}}$ is a constant, and, when $T(n) \geq 2^{c_{\mathrm{T}} n}$ only an additive expense of an extra $\log T(n)$ pre-shared secret bits is needed. In the future we plan on combining this work with our recent results on jammer-assisted covert communication [9] to enable covert networks.

## Appendix A
## Proof of Lemma 1

Consider any $\epsilon > 0$. Suppose Willie chooses threshold $\tau(n)$ arbitrarily. The false alarm probability is lower-bounded using the fact that $S^{(n)}$ is independent of which hypothesis is true:

$$\mathbb{P}[L_0^{(n)} > \tau(n) | H_0 \text{ is true}] \geq \mathbb{P}\left[S^{(n)} \geq \tau(n) + \delta \middle| |V_0^{(n)}| < \delta\right].$$

Similarly the probability of missed detection is lower-bounded as follows:

$$\mathbb{P}[L_1^{(n)} \leq \tau(n) | H_1 \text{ is true}] \geq \mathbb{P}\left[S^{(n)} \leq \tau(n) - \delta \middle| |V_1^{(n)}| < \delta\right].$$

Denoting by $E_{\mathrm{C}}(\tau(n), \delta)$ the event that either $S^{(n)} \geq \tau(n) + \delta$ or $S^{(n)} \leq \tau(n) - \delta$,

$$\mathbb{P}(E_{\mathrm{C}}(\tau(n), \delta)) = 1 - F_{S^{(n)}}(\tau(n) + \delta) + F_{S^{(n)}}(\tau(n) - \delta),$$

where $F_{S^{(n)}}(\cdot)$ is the distribution function for $S^{(n)}$. Denote the standard Gaussian distribution function by $\Phi(z) = \int_{-\infty}^{z} \phi(t) \mathrm{d}t$ where $\phi(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}}$ is the standard Gaussian density function. The convergence of $F_{S^{(n)}}(z)$ to $\Phi(z)$ is pointwise in $z$, and, since $\tau(n)$ is the $n^{\text{th}}$ value in an arbitrary sequence, we cannot use this fact directly. However, let's choose finite constants $G < 0$ and $H > 0$, and partition the real number line into three regions as shown in Figure 3. Clearly,

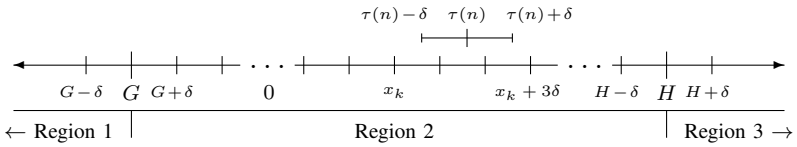


Fig. 3: The real number line partitioned into three regions for the analysis of $\mathbb{P}(E_{\mathrm{C}}(\tau(n),\delta))$. $G$, $H$ and $\delta$ are the constants that we select. $\tau(n)$ satisfying $G \leq \tau(n) \leq H$ is illustrated.

for any $n$, $\tau(n)$ is in one of these regions. Next we demonstrate that (47) holds for an arbitrary $\tau(n)$ by appropriately selecting $G$, $H$, and $\delta$.

Consider $\tau(n) < G$, or region 1 in Figure 3: $\mathbb{P}(E_{\mathrm{C}}(\tau(n),\delta)) \geq 1 - F_{S^{(n)}}(\tau(n)+\delta) \geq 1 - F_{S^{(n)}}(G+\delta)$. Because the convergence of $F_{S^{(n)}}(z)$ to $\Phi(z)$ is pointwise, given $\delta$, $\epsilon$, and $G = \Phi^{-1}(\epsilon/6) - \delta$, there exists $n_2$ such that, for all $n \geq n_2$, $\mathbb{P}(E_{\mathrm{C}}(\tau(n),\delta)) \geq 1 - \Phi(G+\delta) - \frac{\epsilon}{6} = 1 - \frac{\epsilon}{3}$ when $\tau(n) < G$. Similarly for $\tau(n) > H$, or region 3 in Figure 3: $\mathbb{P}(E_{\mathrm{C}}(\tau(n),\delta)) \geq F_{S^{(n)}}(\tau(n)-\delta) \geq F_{S^{(n)}}(H+\delta)$. Again, because the convergence of $F_{S^{(n)}}(z)$ to $\Phi(z)$ is pointwise, given $\delta$, $\epsilon$, and $H = \Phi^{-1}(1-\epsilon/6) + \delta$, there exists $n_3$ such that, for $\tau(n) > H$ and all $n \geq n_3$,

$$\mathbb{P}(E_{\mathrm{C}}(\tau(n),\delta)) \geq \Phi(H+\delta) - \frac{\epsilon}{3} = 1 - \frac{\epsilon}{3}. \tag{43}$$

Finally, consider $\tau(n)$ satisfying $G \leq \tau(n) \leq H$, or region 2 in Figure 3. Let's assume that $H$ and $G$ are selected so that $H - G$ is an integer multiple of $\delta$ (e.g., using larger $H$ than necessary, which results in the lower bound in (43) being smaller). Consider a sequence $\{x_k\}_{k=0}^{(H-G)/\delta+2}$ where $x_0 = G - \delta$, $x_1 = G$, $x_2 = G + \delta$, $x_3 = G + 2\delta, \ldots, x_{(H-G)/\delta} = H - \delta$, $x_{(H-G)/\delta+1} = H$, $x_{(H-G)/\delta+2} = H + \delta$. Sequence $\{x_k\}_{k=0}^{(H-G)/\delta+2}$ partitions region 2 into $\frac{H-G}{\delta} + 2$ subregions, and, for any $\tau(n)$ satisfying $G \leq \tau(n) \leq H$, there exists $k \in \left\{0, \ldots, \frac{H-G}{\delta} + 2\right\}$ such that $x_k \leq \tau(n) - \delta < \tau(n) + \delta \leq x_k + 3\delta$, as illustrated in Figure 3. Since $F_{S^{(n)}}(z)$ is monotonic,

$$\mathbb{P}(E_{\mathrm{C}}(\tau(n),\delta)) \geq 1 - F_{S^{(n)}}(x_k + 3\delta) + F_{Z(n)}(x_k). \tag{44}$$

Since the convergence of $F_{S^{(n)}}(z)$ to $\Phi(z)$ is pointwise, for a given $x_k$, $\delta$, and $\epsilon$, there exists $m_k$ such that for all $n \geq m_k$,

$$\mathbb{P}(E_{\mathrm{C}}(\tau(n),\delta)) \geq 1 - \left(\Phi(x_k+3\delta) + \frac{\epsilon}{12}\right) + \left(\Phi(x_k) - \frac{\epsilon}{12}\right) = 1 - \int_{x_k}^{x_k+3\delta} \phi(t)\mathrm{d}t - \frac{\epsilon}{6} \quad (45)$$

$$\geq 1 - \frac{3\delta}{\sqrt{2\pi}} - \frac{\epsilon}{6} \quad (46)$$

where (46) follows from $\phi(t) \leq \frac{1}{\sqrt{2\pi}}$. Setting $\delta = \frac{\epsilon\sqrt{2\pi}}{18}$ and $n_4 = \max_{\left\{0,\ldots,\frac{H-G}{\delta}\right\}}(m_k)$ yields the desired lower bound for all $n \geq n_4$ when $\tau(n)$ satisfies $G \leq \tau(n) \leq H$. Thus, for any $\tau(n)$, when $n \geq n_0$ where $n_0 = \max(n_2, n_3, n_4)$,

$$\mathbb{P}\left(E_{\mathrm{C}}\left(\tau(n), \epsilon\sqrt{2\pi}/18\right)\right) \geq 1 - \frac{\epsilon}{3}. \quad (47)$$

Since $V_0^{(n)} \xrightarrow{\mathcal{P}} 0$ and $V_1^{(n)} \xrightarrow{\mathcal{P}} 0$, there exists $n_1 > 0$ such that for all $n \geq n_1$, $\mathbb{P}\left(|V_0^{(n)}| > \frac{\epsilon\sqrt{2\pi}}{18}\right) < \frac{\epsilon}{3}$ and $\mathbb{P}\left(|V_0^{(n)}| > \frac{\epsilon\sqrt{2\pi}}{18}\right) < \frac{\epsilon}{3}$. The intersection of these events and the event $E_{\mathrm{C}}(S(n), \epsilon\sqrt{2\pi}/9)$ yields a detection error event. By combining their probabilities using DeMorgan's Law and the union bound, we lower-bound $\mathbb{P}_{\mathrm{FA}} + \mathbb{P}_{\mathrm{MD}} \geq 1 - \epsilon$ for all $n \geq \max\{n_0, n_1\}$.

*Remark:* Lemma 1 holds when $S^{(n)}$ converges to any distribution provided it has a continuous density, however, here we do not need such generality.

## APPENDIX B
## PROOF OF EQUALITY $\frac{1}{2^n}\sum_{\mathbf{x}\in\{-1,1\}^n} \exp\left[a\sum_{i=1}^n x_i\right] = \cosh^n(a)$

We argue by induction on $n$. The base case $n = 1$ is trivial. Assume the claim holds for $n$. Now,

$$\frac{1}{2^{n+1}} \sum_{\mathbf{x}\in\{-1,1\}^{n+1}} \exp\left[a\sum_{i=1}^{n+1} x_i\right] = \frac{1}{2\cdot 2^n} \sum_{\substack{\mathbf{x}\in\{-1,1\}^n \\ x_{n+1}\in\{-1,1\}}} \exp[ax_{n+1}] \exp\left[a\sum_{i=1}^{n} x_i\right] = \cosh(a)\cosh^n(a).$$

## APPENDIX C
## ANALYSIS OF $\mathbb{P}_{\mathrm{e}}^{(\mathrm{b})}$ UNDER PEAK POWER CONSTRAINT AND KNOWN $t_{\mathrm{A}}$

Here we analyze Bob's decoding error probability $\mathbb{P}_{\mathrm{e}}^{(\mathrm{b})}$ when the transmission time $t_{\mathrm{A}}$ is known and Alice uses binary modulation $\{-a, a\}$ that satisfies the peak power constraint and

ensures that $\mathbb{P}_{\rm FA}+\mathbb{P}_{\rm MD}\geq 1-\epsilon$. This appendix provides an alternative to the analysis of $\mathbb{P}_{\rm e}^{\rm (b)}$ in the proof of Theorem 1.2 in [3], [4] (the analysis of $\mathbb{P}_{\rm e}^{\rm (b)}$ in the proof of Theorem 1.2 in [3], [4] contains minor technical errors which do *not* change the main results). The construct presented here is adapted for the proof of Theorem 1.2 in Section III.

Suppose Alice transmits $\mathbf{c}(W_k)$. Recall that $\mathbf{c}(W_k)$ is drawn from a codebook $\{\mathbf{c}(W_m), m=1,2,\ldots,2^M\}$ containing codewords that are independently generated according to $p_{\mathbf{X}}(\mathbf{x})=\prod_{i=1}^n p_X(x_i)$, where $p_X(-a)=p_X(a)=\frac{1}{2}$. Bob uses an ML decoder which suffers an error event $E_{k\to i}$ when the received vector $\mathbf{Y}_b$ is closer in Euclidean distance to codeword $\mathbf{c}(W_i)$, $i\neq k$. The decoding error probability, averaged over all the codebooks, is then:

$$\mathbb{P}_e^{\rm (b)}=\mathbb{E}_{\mathbf{c}(W_k)}\left[\mathbb{P}\left(\cup_{\substack{i=0\\ i\neq k}}^{2^M}E_{k\to i}\right)\right]\leq\sum_{\substack{i=1\\ i\neq k}}^{2^M}\mathbb{E}_{\mathbf{c}(W_k)}\mathbb{P}(E_{k\to i}),\tag{48}$$

where $\mathbb{E}_X[\cdot]$ denotes the expectation over random variable $X$, and the inequality in (48) is the union bound. Let $\|\mathbf{d}\|_2=\|\mathbf{c}(W_k)-\mathbf{c}(W_i)\|_2$ denote the Euclidean ($\mathcal{L}_2$) distance between two codewords. Then, by [26, Eq. (3.44)]:

$$\mathbb{E}_{\mathbf{c}(W_k)}\mathbb{P}(E_{k\to i})=\mathbb{E}_{\mathbf{d}}\left[Q\left(\frac{\|\mathbf{d}\|_2}{2\sigma_{\rm b}}\right)\right]\leq\mathbb{E}_{\mathbf{d}}\left[\frac{1}{2}\exp\left(-\frac{\|\mathbf{d}\|_2^2}{8\sigma_{\rm b}^2}\right)\right],\tag{49}$$

where $Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-t^2/2}{\rm d}t$ and the bound in (49) is because $Q(x)\leq\frac{1}{2}e^{-x^2/2}$ [27, Eq. (5)]. The Euclidean distance $\|\mathbf{d}\|_2$ depends on the number of locations $j$ where $\mathbf{c}(W_k)$ and $\mathbf{c}(W_i)$ differ, which is binomially-distributed by construction, where each location is different with probability $\frac{1}{2}$. When the codewords are different in $j$ locations, we have $\|\mathbf{d}\|_2^2=4ja^2$, and thus:

$$\mathbb{E}_{\mathbf{d}}\left[\frac{1}{2}\exp\left(-\frac{\|\mathbf{d}\|_2^2}{8\sigma_{\rm b}^2}\right)\right]=\frac{1}{2}\sum_{j=0}^n\exp\left(-\frac{ja^2}{2\sigma_{\rm b}^2}\right)\binom{n}{j}\frac{1}{2^n}=\frac{1}{2^{n+1}}\left[1+\exp\left(-\frac{a^2}{2\sigma_{\rm b}^2}\right)\right]^n,\tag{50}$$

where (50) is an application of binomial theorem. Substitution of (50) into (48) yields

$$\mathbb{P}_e^{\rm (b)}\leq 2^{M-n\left(1-\log_2\left[1+\exp\left(-\frac{a^2}{2\sigma_{\rm b}^2}\right)\right]\right)}.\tag{51}$$

Thus, if Alice attempts to transmit $M=n\gamma\left(1-\log_2\left[1+\exp\left(-\frac{a^2}{2\sigma_{\rm b}^2}\right)\right]\right)$ bits, where $\gamma\in(0,1)$ is a constant, Bob's decoding error probability decays to zero as $n\to\infty$. If $a^2=\mathcal{O}(1)$, then

clearly $M = \mathcal{O}(n)$ bits. If $a^2 = o(1)$, then, application of the bounds $\log_2(1+x) \leq \frac{x}{\ln 2}$ and $1 - e^{-x} \geq x - \frac{x^2}{2}$ yields $M = \mathcal{O}(na^2)$. In particular, if $\sigma_{\rm w}^2$ is known to Alice and $a^2 = \frac{2\sqrt{2}\epsilon\sigma_{\rm w}^2}{\sqrt{n}}$ as prescribed by the analysis of Willie's detector in [4, Theorem 1.2], $M = \mathcal{O}(\epsilon\sigma_{\rm w}^2\sqrt{n}/\sigma_{\rm b}^2)$.

## REFERENCES


[1] B. A. Bash, D. Goeckel, and D. Towsley, "LPD Communication when the Warden Does Not Know When," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Honolulu, HI, Jul. 2014.

[2] BBC, "Edward Snowden: Leaks that exposed US spy programme," http://www.bbc.com/news/world-us-canada-23123964, Jan. 2014.

[3] B. A. Bash, D. Goeckel, and D. Towsley, "Square root law for communication with low probability of detection on AWGN channels," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Cambridge, MA, Jul. 2012.

[4] ——, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Select. Areas Commun.*, vol. 31, no. 9, pp. 1921–1930, 2013, arXiv:1202.6423.

[5] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, arXiv:1304.6693.

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

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

[8] S. Kadhe, S. Jaggi, M. Bakshi, and A. Sprintson, "Reliable, deniable, and hidable communication over multipath networks," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Honolulu, HI, Jul. 2014, arXiv:1401.4451.

[9] R. Soltani, B. A. Bash, D. Goeckel, S. Guha, and D. Towsley, "Covert single-hop communication in a wireless network with distributed artificial noise generation," in *Proc. Conf. Commun. Control Comp. (Allerton)*, Monticello, IL, 2014.

[10] B. A. Bash, S. Guha, D. Goeckel, and D. Towsley, "Quantum Noise Limited Communication with Low Probability of Detection," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013.

[11] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, "Quantum-secure covert communication on bosonic channels," *Nat Commun*, vol. 6, Oct 2015.

[12] A. Sheikholeslami, B. A. Bash, D. Towsley, D. Goeckel, and S. Guha, "Covert communication over classical-quantum channels," arXiv:1601.06826 [quant-ph], 2016.

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

[14] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *Proc. of IEEE Int. Symp. Inform. Theory (ISIT)*, Honolulu, HI, Jul. 2014, arXiv:1311.1411.

[15] T. S. Han, H. Endo, and M. Sasaki, "Reliability and secrecy functions of the wiretap channel under cost constraint," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6819–6843, Nov 2014.

[16] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *Proc. Inform. Theory Workshop (ITW)*, Hobart, Tasmania, Australia, Nov 2014, pp. 30–34.

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

[18] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, 2nd ed. Cambridge, Massachusetts: MIT Press, 2001.

[19] K. S. K. Arumugam and M. R. Bloch, "Keyless asynchronous covert communication," in *Proc. Inform. Theory Workshop (ITW)*, Sep. 2016.

[20] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, pp. 656–715, 1949.

[21] A. J. Menezes, S. A. Vanstone, and P. C. V. Oorschot, *Handbook of Applied Cryptography*. Boca Raton, FL, USA: CRC Press, Inc., 1996.

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

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

[24] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[26] U. Madhow, *Fundamentals of Digital Communication*. Cambridge, UK: Cambridge University Press, 2008.

[27] M. Chiani, D. Dardari, and M. K. Simon, "New exponential bounds and approximations for the computation of error probability in fading channels," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 840–845, Jul. 2003.

[28] S. Dasgupta and A. Gupta, "An Elementary Proof of a Theorem of Johnson and Lindenstrauss," *Random Struct. Algorithms*, vol. 22, no. 1, pp. 60–65, Jan. 2003.

[29] F. Topsøe, "Some bounds for the logarithmic function," *RGMIA Research Rep. Collection*, vol. 7, no. 2, 2004.

[30] D. Torrieri, *Principles of Spread-spectrum Communication Systems*. Boston, MA, USA: Springer, 2005.

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

[32] Herodotus, c. 440 BCE, 5.35 and 7.239.

[33] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*, 1st ed. New York: Cambridge University Press, 2009.

[34] A. D. Ker, "The square root law does not require a linear key," in *Proc. ACM Workshop Multimedia Security*, Roma, Italy, 2010, pp. 213–224.

[35] V. Korzhik, G. Morales-Luna, and M. H. Lee, "On the existence of perfect stegosystems," in *Proc. 4th Int. Workshop Digital Watermarking (IWDW)*, Siena, Italy, Sep. 2005, pp. 30–38.

[36] A. D. Ker, "A capacity result for batch steganography," *IEEE Signal Process. Lett.*, vol. 14, no. 8, pp. 525–528, Aug. 2007.

[37] ——, "Batch steganography and pooled steganalysis," in *Proc. Int. Inform. Hiding Workshop*, Alexandria, VA, 2006, pp. 265–281.

[38] S. Craver and J. Yu, "Subset selection circumvents the square root law," in *Proc. SPIE Media Forensics Security*, San Jose, CA, 2010, pp. 754 103–1–754 103–6.

[39] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook*. McGraw-Hill, 1994.

[40] 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.