Covert Communication With Gaussian Noise: From Random Access Channel to Point-to-Point Channel

Masahito Hayashi, *Fellow, IEEE*, and Ángeles Vázquez-Castro, *Senior Member, IEEE*

Abstract— We propose a covert communication protocol for the spread-spectrum multiple random access with additive white Gaussian noise (AWGN) channel. No existing paper has studied covert communication for the random access channel. Our protocol assumes binary discrete phase-shift keying (BPSK) modulation, and it works well under imperfect channel state information (I-CSI) for both the legitimate and adversary receivers, which is a realistic assumption in the low power regime. Also, our method assumes that the legitimate users share secret variables in a similar way as the preceding studies. Although several studies investigated the covert communication for the point-to-point communication, no existing paper considers the covert communication under the above uncertainty assumption even for point-to-point communication. Our protocol under the above uncertainty assumption allows $O(n)$ legitimate senders and $O(n/\log n)$ active legitimate senders. Furthermore, our protocol can be converted to a protocol for point-to-point communication that works under the above uncertainty assumption.

Index Terms— Covert communication, information hiding, additive white Gaussian noise, random access channel, central limit theorem, universal code.

I. INTRODUCTION

A. Background: Point-to-Point Covert Communication

COVERT communication is a technology to hide the existence of communication, and has been actively studied. This type of communication is often called communication with low probability of detection. In this technology, the legitimate sender intends to transmit an information message to the legitimate receiver while making such communication undetectable by the adversary. This task can be achieved when adversary's observation with the silent case is imitated by adversary's observation under the existence of a communication between the legitimate sender and the legitimate

receiver. In fact, when the output of the silent case is written as a convex combination of other outputs in the channel to the adversary, the above task can be easily achieved. Here, we call this condition the redundant condition, and its rigorous definition is given in Section III-C. Under the above condition, using the method of wire-tap channel [1], [2], [3], the papers [4], [5] consider this problem for the point-to-point channel under the discrete memoryless condition. They showed that the covert transmission length $O(n)$ is possible with n uses of the channel. The idea of this method is that the transmitter transmits an independent and identically distributed (i.i.d.) sequence for the no-communication mode, which makes the problem more similar to covert communication in the presence of a jammer as discussed in [6, Remark 2].

However, the redundant condition does not hold in general. For example, when the form of the channel to the adversary is known, the additive white Gaussian noise (AWGN) channel and the binary symmetric channel (BSC) do not satisfy this condition. The papers [7], [8], and [9] discussed this problem for the cases of AWGN and BSC, respectively. Then, the papers [5] and [10] studied the covert communication problem under general discrete memoryless channel for the point-to-point channel when the redundant condition does not hold. They showed that the optimal covert transmission length is $O(\sqrt{n})$ in the non-redundant case with n uses of the channel. Following these studies, the papers [11], [12], and [13] studied this problem for AGWN as considering continuous time models. Also, the paper [14] extended this discussion to multiple-input multipl-output (MIMO) AWGN channels.

The fundamental assumption of the non-redundant case is using preshared secure keys between the legitimate sender and the legitimate receiver. Due to the preshared keys, the legitimate users can realize covert communication even when the channel to the adversary has smaller noise than the channel to the legitimate receiver. That is, to preshare keys is a mandatory resource for covert communication whenever Willie's channel is not worse than Bob's channel. Later, many subsequent studies [15], [16], [17], [18], [19], [20], [21], [22] analyzed the point-to-point covert communication over the AWGN channel under different additional assumptions.

B. Background: Random Access Channel

Our focus in this work is the random access channel where the point-to-point is a special case. A number of studies [23], [24], [25], [26], [27] have already discussed the random access channel from the viewpoint of information theory, however,
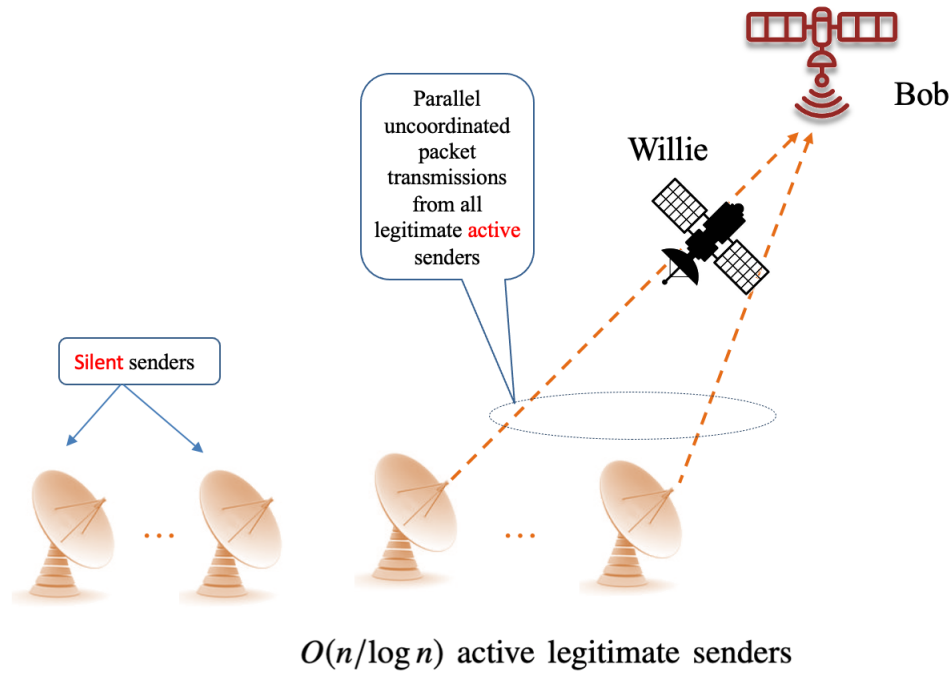
Fig. 1. This figure illustrates a random access channel with one legitimate receiver (Bob), one adversary (Willie), and multiple legitimate senders. Our protocol allows $O(n/\log n)$ active legitimate senders. Illegitimate receivers are shown in black.

TABLE I
COMPARISON WITH EXISTING RESULTS FOR ACCESS CHANNEL

	Covertness	Random access channel	Type of channel
[29]	Yes	No	Discrete
[23], [24], [27]	No	Yes	Discrete
[25], [26], [28]	No	Yes	AWGN
This paper	Yes	Yes	AWGN

these works do not address the covertness property of this channel. Recently, the paper [28] addressed the anti-jamming secrecy in the random access channel, but it did not discuss covertness. Another recent paper [29] discussed the covert communication for the access channel assuming discrete time, but it did not address random access channel nor AWGN channel.

C. Problem Statement and Novelty

In this paper, we present an information theoretical study of the covertness of (direct sequence) random access and one time pad encryption. We assume BPSK modulation and novel assumptions on the channel knowledge by the legitimate users and the adversary, Willie. Furthermore, to realize covertness, depending on the legitimate sender, our method uses preshared secret binary symbols, which is often called secret chips. Fig. 1 shows its illustrative scenario and the relation with existing results for access channel are summarized as Table I clarifying the novelty of our work. Throughout this paper, we assume logarithms with base e.

In our setting, n expresses the number of uses of the channel during one coding block-length. Our method allows $O(n)$

legitimate senders and $O(n/\log n)$ active legitimate senders under the assumption of preshared secure keys between these legitimate senders and the legitimate receiver.[1] To achieve this performance, we pose a novel and realistic channel condition (in the low power regime) as follows because it is impossible to achieve covertness of $O(n/\log n)$ bits in the non-redundant case when the channel parameter is completely known to the adversary, i.e., the channel is identified by the adversary. In a realistic scenario, it is difficult for both legitimate and adversary parties to obtain a complete knowledge of the channel parameters. Therefore, it is natural to assume that all parties (legitimate and non-legitimate) do not have a complete knowledge of the channel parameters while we assume that the channel parameters are fixed during one coding block-length. The latter assumption is justified when our signal model and asymptotic results hold within the channel coherence time. Under the above realistic conditions for the uncertainty, our protocol guarantees that the legitimate receiver retrieves the message.

Another novelty in our formulation is the universality of our proposed codes. In information theory a code is called a universal code when the code does not depend on the channel parameter in the above way, i.e., our code construction does not require full knowledge of the channel [30], [31]. In our case, this means that our method allows the dispersion of signal intensities from the senders in the detection of both receivers due to the effect of the fading fluctuation. As a consequence, we make the realistic assumption that for a

---

[1]In contrast, the recent paper [29] considers the case when the number of senders is fixed to K and the size of transmitted bits behaves as $o(\sqrt{n})$.

known scenario of interest, an upper and lower bound of the channel coefficients can be estimated.

In addition, our method assumes preshared secure keys between the legitimate sender and the legitimate receiver in the same way as [5], [10], [11], [12], and [13] and it works even when the channel to the adversary has smaller noise than the channel to the legitimate receiver. That is, for every coding block, each legitimate sender shares secret binary symbols as pre-shared secrets with the legitimate receiver while the legitimate sender can send only one bit. Hence, when the senders need to transmit ℓ bits, it is sufficient that they repeat this protocol ℓ times. In practice, this is realistically achieved using well known and widely available spreading codes. Each legitimate sender has n different preshared secret binary symbols. That is, the number of secret binary symbols at the legitimate receiver is n times the number of legitimate senders.

Our encoder is very simple for random access channel. That is, when a legitimate sender is active and intends to transmit one bit X, the sender encodes the intended bit X into n channel inputs by using one time pad encryption with n preshared secret binary symbols. Then, the legitimate receiver recovers the transmitted bits by using the preshared secret bits.

The most novel point of our work is the covertness analysis for the adversary. In our analysis, the covertness evaluation is reduced to the difference between the Gaussian distribution and the distribution of the weighted sample mean of n independent random variables subject to the average output distribution of BPSK modulation. Although a variant of the central limit theorem [32] guarantees that the distribution of the weighted sample mean of n independent random variables approaches to a Gaussian distribution, our covertness analysis needs the evaluation of the variational distance between the above two distributions. When the fading coefficients from a sender in Willie's detection does not depend on the sender, it is sufficient to discuss the variational distance between the distribution of the sample mean and a Gaussian distribution. Such a case was discussed in [33, (1.3)]. However, the general case requires more difficult analysis. Fortunately, the recent papers [34] and [35] have studied this mathematical problem by using Poincaré constant [34], [35], [36], [37]. Applying this result, we derive our covertness analysis.

D. From Random Access Channel Protocol to Point-to-Point Channel

Another novelty of our work is that we consider the fact that our protocol can be converted to a protocol for point-to-point communication that works under the above assumptions. Under this conversion, we obtain a covert communication protocol for the point-to-point channel that has $\lceil (l_n + 1)/2 \rceil$ different values as the channel input power, and achieves the covert transmission of $O(n/\log n)$ bits, where l_n is the number of bits the legitimate sender wants to transmit to the legitimate receiver. Although existing studies assuming the redundant case [4], [5] achieve covert transmission length $O(n)$ with n uses of the channel, existing studies assuming the non-redundant case [5], [10], [11], [12], [13] achieve

TABLE II
COMPARISON WITH EXISTING RESULTS FOR POINT-TO-POINT CHANNEL

	Transmission length	Bob knows channel parameters	Willie knows channel parameters	Type of channel
[10], [5]	$O(\sqrt{n})$	Yes	Yes	Discrete
[11–13]	$O(\sqrt{n})$	Yes	Yes	AWGN
[15–22]	$O(n)$	Yes	No	AWGN
This paper	$O(n/\log n)$	No	No	AWGN

n is the number of uses of channel.

covert transmission length $O(\sqrt{n})$, which is much smaller than the transmission length of the conventional communication. To resolve this problem, many subsequent researchers [15], [16], [17], [18], [19], [20], [21], [22] introduced the uncertainty of the channel parameters only of the channel to the adversary under the AWGN channel. In fact, due to the uncertainty, the adversary cannot distinguish the output of the Gaussian mixture input distribution from the output of zero input. That is, this modification enables the channel model to satisfy the redundant condition, which leads the covert transmission length $O(n)$. However, these studies assume that the legitimate receiver knows the channel parameters of his/her own channel, which is an unequal assumption, i.e., an unrealistic assumption. To make a fair assumption, we pose the novel and realistic channel condition introduced above that the channel parameters of the AWGN channels to both the legitimate and adversary receivers present some uncertainty, i.e. are not completely known to them. Fortunately, our protocol on the point-to-point channel achieves the covert transmission of $O(n/\log n)$ bits when both receivers (the legitimate receiver and the adversary) have uncertainty in their detection. In this sense, our method has an advantage over existing methods even under the Gaussian point-to-point channel.

Finally, we remark that our main focus is the analysis of the asymptotic performance of our code to guarantee covertness and therefore practical issues such as BPSK symbol acquisition or outage probability (e.g for concrete statistics assumptions on the channel dynamics) and practical channel estimation are out of the scope of this work and is left for future work. The relation with existing results for point-to-point channel is summarized in Table II, clarifying the novelty of our work and results.

This paper is organized as follows. Section II describes our formulation of random access channel model, and states our result in this case. Section III explains what protocol is obtained for the particular case of point-to-point channel model. Then, Section III compares our obtained code for the point-to-point channel model with simple applications of the methods [4], [5]. Section IV shows that Bob correctly recovers the message with almost probability one under both models in the asymptotic case. Section V formulates covertness with respect to Willie, and states our covertness result. In addition, Section V shows its proof for the case with equal fading including the point-to-point channel model while its proof with

the general case with unequal fading is shown in Appendix. Section VI presents a discussion of our results.

II. RANDOM ACCESS CHANNEL

A. Random Access Channel Model

Our random access channel model has m senders $\mathcal{A}_1, \ldots, \mathcal{A}_m$, one adversary, Willie, and the legitimate receiver, Bob. The task of our protocol is formulated as follows. Each sender $\mathcal{A}_i$ intends to send one bit M_i to Bob within the channel coherence time when $\mathcal{A}_i$ is active. If $\mathcal{A}_i$ is silent, he/she does not need to send it to Bob. Also, the senders want to hide the existence of their communication to the adversary, Willie. For the practical implementation, we assume that the channel is AWGN and each sender can use only BPSK modulation.

To realize the hidden communication, the senders and Bob share secret random variables that are not known to Willie, in the same way as [5] and [10]. That is, the sender $\mathcal{A}_i$ has binary random variables $S_{i,1}, \ldots, S_{i,n}$ that are subject to the uniform distribution independently. The legitimate receiver, Bob also knows all the binary random $S_{i,j}$. However the adversary, Willie, does not know $S_{i,j}$.

We consider a random access channel with Gaussian channel as follows. Assume that only l senders $\mathcal{A}_{i_1}, \ldots, \mathcal{A}_{i_l}$ are active and other senders are silent. When $\mathcal{A}_{i_k}$ inputs n variable $X_{i_k,1}, \ldots, X_{i_k,n}$, Bob receives

$$Y_j = N_{B,j} + \sum_{k=1}^{l} a_{i_k} X_{i_k,j} \qquad (1)$$

for $j = 1, \ldots, n$. Similarly, Willie receives

$$Z_j = N_{W,j} + \sum_{k=1}^{l} b_{i_k} X_{i_k,j} \qquad (2)$$

for $j = 1, \ldots, n$. Here, a_{i_k} and b_{i_k} are the fading coefficients within the channel coherence time in Bob's and Willie's detection. Hence, a_{i_k} and b_{i_k} are positive constants during a coherent time. That is, we treat a_{i_k} and b_{i_k} as constants in the following discussion. In the following, we denote the maximum $\max_k a_{i_k}$ ($\max_k b_{i_k}$) and the minimum $\min_k a_{i_k}$ ($\min_k b_{i_k}$) by $\bar{a}$ ($\bar{b}$) and $\underline{a}$ ($\underline{b}$), respectively. Hence, we make the realistic assumption that for a known scenario of interest, upper and lower bounds of the channel coefficients can be estimated. That is, Bob and Willie know $\bar{a}$, $\bar{b}$, $\underline{a}$, and $\underline{b}$. Also, $N_{B,1}, \ldots, N_{B,n}$ ($N_{W,1}, \ldots, N_{W,n}$) are independent Gaussian variable with average 0 and variance v_B (v_W).

In a realistic setting, Bob and Willie know rough values of the variance of noise power for Bob and Willie, denoted as v_B and v_W, but, they do not know their exact values due to the following reason. Willie and Bob know their own receiving device well. Hence, they know the noise generated in their receiving device. In this way, Bob and Willie have a similar performance for receiving the signal from the senders. However, a part of the noise is generated out of Willie's device, which can be considered as a background noise. To discuss Willie's detection of the existence of the communication, we focus on Willie's knowledge on the value of the variance v_W of Willie's observation. That is, we denote the set of possible

variance v_W of Willie's observation by $\mathcal{V}$. In the following, we assume that $\mathcal{V}$ is an open set of $\mathbb{R}$. This background noise uncertainty also affects to Bob, however, he has the pre-shared keys so that he can recover the message nevertheless such uncertainty. In contrast, since Willie doesn't have it, his ability is affected by such background noise uncertainty.

In fact, it is a common assumption that the channel is characterized by unknown parameters. In this case, the channel model is usually denoted as compound, and codes for such channels are called universal codes in information theory [30] (see e.g. the paper [31] for a Gaussian channel). Even in the above existing universal setting, we need to assume that the channel parameters belong to a certain subset. Otherwise, it is impossible to guarantee secure communication. Estimating channel coefficients is well known in signal processing. The transmission inserts "pilots" i.e. known symbols to measure the channel effect at reception. Hence, to assume some underlying "roughly" estimation of the channel dynamics is reasonable [38]. In the following, we propose our code that does not depend on these channel parameters except for $\underline{a}$ and $\frac{v_B}{\underline{a}^2}$.

B. Random Access Covert Protocol

Here, we present our protocol for random access covert communication. When the sender $\mathcal{A}_i$ is silent, the input signal $X_{i,j}$ is set to zero for $j = 1, \ldots, n$. When the sender $\mathcal{A}_i$ is active and the sender $\mathcal{A}_i$ sends the binary message M_i, $\mathcal{A}_i$ encodes the message as

$$X_{i,j} = \sqrt{\frac{t_n}{n}}(-1)^{M_i + S_{i,j}} \qquad (3)$$

for $j = 1, \ldots, n$. Here, the above code uses average power $\frac{t_n}{n}$ for each channel use, which is sufficiently small.

Next, we consider Bob's decoder. In order to recover M_i, using the secret variables $S_{i,j}$, Bob calculates the decision statistic $\bar{Y}_i := \sum_{j=1}^{n}(-1)^{S_{i,j}}Y_j$ from his receiving variables $Y_1, \ldots, Y_n$ as Fig. 2. For each i, Bob outputs one of three outcomes, silent, 1, and 0 as follows. When $|\bar{Y}_i| < \sqrt{n}\frac{\underline{a}t_n}{2}$, Bob considers that $\mathcal{A}_i$ is silent. When $\bar{Y}_i \leq -\sqrt{n}\frac{\underline{a}t_n}{2}$, Bob considers that M_i is 1. When $\bar{Y}_i \geq \sqrt{n}\frac{\underline{a}t_n}{2}$, Bob considers that M_i is 0.

Then, we have the following theorem for the analysis on the asymptotic performance of our code.

Theorem 1: Assume that the number of senders, m, and of active senders, l, and m, l are given as m_n, $l_n = \frac{n}{t_n}$ to satisfy

$$\frac{n}{l_n^2} \to 0, \qquad (4)$$

$$\frac{l_n}{n} \to 0, \qquad (5)$$

$$\frac{n\underline{a}^2}{8l_n(v_B + \bar{a}^2)} - \log m_n \to +\infty. \qquad (6)$$

Also, we assume that v_W belongs to an open set $\mathcal{V}$. Then, under the above presented protocol, Bob can recover the message with asymptotically zero error, and Willie cannot detect the existence of the communication regardless of the values of the channel parameter to Willie.

H_s : Bob considers that $\mathcal{A}_i$ is silent
H_0 : M_i is decoded to 0
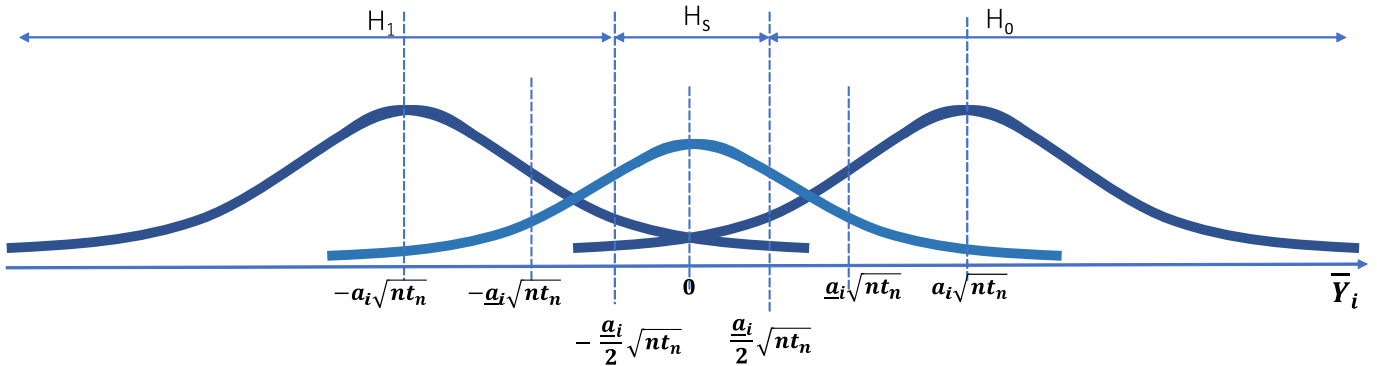H_1 : M_i is decoded to 1

Fig. 2. This figure shows how Bob makes his decision from $\bar{Y}_i$ for the random access case when they use the channel n times. Here, Bob needs to output one of three outcomes, silent, 1, and 0. The right graph shows the distribution of $\bar{Y}_i$ when $M_i = 0$, and the left graph shows the distribution of $\bar{Y}_i$ when $M_i = 1$. The central graph shows the distribution of $\bar{Y}_i$ when $\mathcal{A}_i$ is silent.

Here, we have not formulated Willie's detection. In Section V-A, we state the impossibility of Willie's detection after presenting its formal definition.

Interestingly, our encoder and our decoder do not depend on the values a_i, b_i, v_B, v_W of the channel parameters, and our decoder do not depend on the number l_n of active senders. But, the probability of correct decoding depends on the number l_n, and is close to 1 as long as the conditions (4) and (6) hold. Hence, in order that Bob knows whether his decoding is correct, he needs to know whether the number l_n is smaller than a certain threshold, which can deduced by Bob from the estimated received power.

For example, when $l_n = cn \log n$ with $c < \frac{8(v_B + \bar{a}^2)}{\underline{a}^2}$ and $m_n = O(n)$, the conditions (4) and (6) hold. Also, when $l_n = cn \log n$ with $c \leq \frac{8(v_B + \bar{a}^2)}{\underline{a}^2}$ and $m_n = l_n$, these conditions hold. Therefore, covert communication with random access code is asymptotically possible with $O(n)$ senders and $O(n/\log n)$ active senders when all active senders transmit only one bit.

Now, we consider the case when each active sender $\mathcal{A}_i$ wants transmit $u_n = g(n)$ bits, $M_{i,1}, \ldots, M_{i,u_n}$. In this case, the active sender $\mathcal{A}_i$ shares random binary symbols $S_{i,1,j}, \ldots, S_{i,u_n,j}$ with Bob for $j = 1, \ldots, n$, and the active sender $\mathcal{A}_i$ sets $X_{i,j}$ as

$$X_{i,j} = \sqrt{\frac{t_n}{n}} \sum_{j'=1}^{u_n} (-1)^{M_{i,j'} + S_{i,j',j}} \tag{7}$$

as the encoding. Here, we denote the numbers of senders and active senders by m_n' and l_n'. This situation can be considered as a special case of Theorem 1 with $m_n = g(n)m_n'$ and $l_n = g(n)l_n'$. Hence, when $m_n = g(n)m_n'$ and $l_n = g(n)l_n'$ satisfy the conditions (4) and (6), Bob can recover the message with asymptotically zero error, and Willie cannot detect the existence of the communication regardless of the values of the channel parameter to Willie.

Also, the condition (6) implies that t_n goes to zero. Since t_n is the power per user, the power per user needs to be

TABLE III
RELATION BETWEEN RANDOM ACCESS CHANNEL AND
POINT-TO-POINT CHANNEL

	random access channel	point-to-point channel
n	number of uses of channel	block length
l_n	transmission length (number of active senders)	transmission length
m_n	number of senders	this number is set to be l_n

zero asymptotically in our protocol. This agrees with the intuition that covertness requires as low power as possible. Further, since the values of the channel parameter to Willie are not contained in the assumption of this theorem, the covert communication is possible even if Willie's channel is better than Bob's channel.

III. POINT-TO-POINT CHANNEL

A. Point-to-Point Channel Model

In the following, we discuss what protocol is obtained when the above protocol is applied to the point-to-point channel model. In this model, the legitimate sender, Alice, intends to transmit l_n bits to the legitimate receiver, Bob, with n uses of AWGN channel while the intensity of input can be fixed to a single value intended by Alice during one block length. This setting is often called point-to-point communication. Table III summarizes the relation between random access channel and point-to-point channel.

When Alice's j-th input variable is X, Bob receives

$$Y_j = N_{B,j} + aX_j \tag{8}$$

for $j = 1, \ldots, n$. The variables $N_{B,1}, \ldots, N_{B,n}$ are independent Gaussian random variables with mean 0 and variance v_B. Similarly, Willie receives

$$Z_j = N_{W,j} + bX_j \tag{9}$$

for $j = 1, \ldots, n$. The variables $N_{W,1}, \ldots, N_{W,n}$ are independent Gaussian random variables with mean 0 and variance v_W.
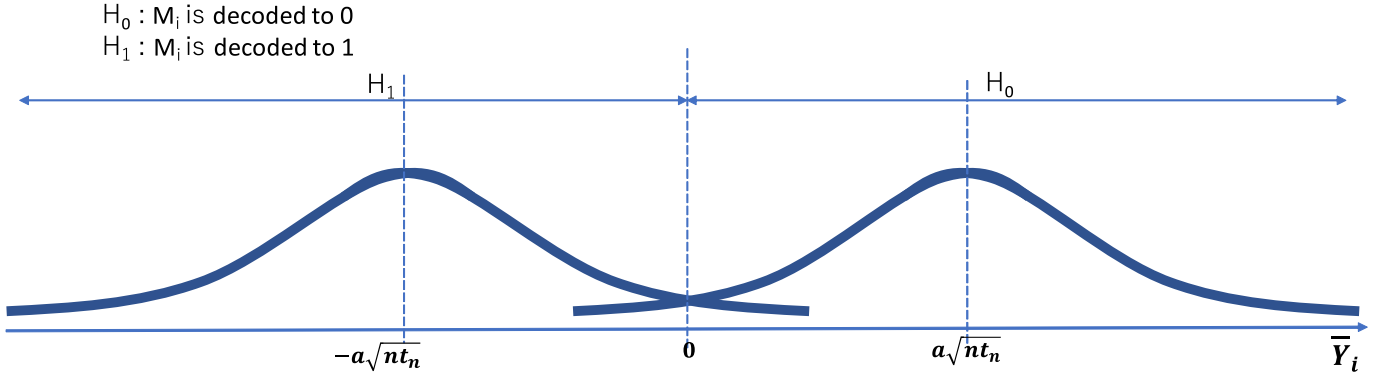
Fig. 3. This figure shows how Bob decodes the message M_i from $\bar{Y}_i$ for the point-to-point case. The right graph shows the distribution of $\bar{Y}_i$ when $M_i = 0$, and the left graph shows the distribution of $\bar{Y}_i$ when $M_i = 1$.

Then, we make the same assumption for v_B and v_W as the previous section. Here, when Alice is silent, all input variables $X_1, \ldots, X_n$ are zero. Since a and b are the fading coefficients in Bob's and Willie's detection, a and b are positive constants during a coherent time. That is, we treat a and b as constants in the following discussion. Therefore, our model is the model in the previous section with $\underline{a} = \bar{a} = a$, $\underline{b} = \bar{b} = b$, and $m = l$.

B. Covert Protocol

For covert transmission of l_n bits $M_1, \ldots, M_{l_n}$ within the coherence channel time, Alice and Bob share nl_n secret binary symbols $\{S_{i,j}\}_{1 \leq j \leq n, 1 \leq i \leq l_n}$. Then, Alice encodes the message $M_1, \ldots, M_{l_n}$ as

$$X_j = \sqrt{\frac{t_n}{n}} \sum_{i=1}^{l_n} (-1)^{M_i + S_{i,j}} \qquad (10)$$

for $j = 1, \ldots, n$. Here, the above code uses average power $\frac{t_n}{n}$ for each channel use, which is sufficiently small.

To see another form of X_j, we define A_j as $A_j := |\{i | M_i + S_{i,j} = 1 \mod 2\}|$, where $|A|$ expresses the number of elements of the set A. Then, A_j is independently subject to the binary distribution with l trials and with probability $1/2$. X_j has another form as

$$X_j = \sqrt{\frac{t_n}{n}} (l_n - 2A_j). \qquad (11)$$

The variable $\sqrt{\frac{t_n}{n}} (l_n - 2A_j)$ has average 0 and variance $\frac{t_n l_n}{n}$. Hence, the power for one channel use, i.e., the expectation of X_j^2 is $\frac{t_n l_n}{n}$, which converges to zero.

Next, we consider Bob's decoder. In order to recover M_i, using the secret variables $S_{i,j}$, Bob calculates the decision statistic $\bar{Y}_i := \sum_{j=1}^{n} (-1)^{S_{i,j}} Y_j$ from his receiving variables $Y_1, \ldots, Y_n$ as Fig. 3. When $\bar{Y}_i \leq 0$, Bob considers that M_i is 1. When $\bar{Y}_i > 0$, Bob considers that M_i is 0. Then, we have the following theorem for the analysis on the asymptotic performance of our code. Hence, as a special case of Theorem 1, we have the following theorem.

Theorem 2: When v_W belongs to an open set $\mathcal{V}$ and $l = l_n$ satisfies (4) and

$$\frac{na^2}{8l_n(v_B + a^2)} - \log l_n \to +\infty, \qquad (12)$$

Bob can recover the message with asymptotically zero error, and Willie cannot detect the existence of the communication.

Similar to Theorem 1, we have not formulated Willie's detection. In Section V-A, we state the impossibility of Willie's detection after presenting its formal definition.

This theorem shows that covert communication is asymptotically possible with transmission length $cn \log n$ with $c \leq \frac{8(v_B + \bar{a}^2)}{\underline{a}^2}$. Interestingly, our encoder and our decoder does not depend on the values a, b, v_B, v_W of the channel parameters as long as the condition (12) holds.

C. Relation With Redundant Condition

The existing studies [4], [5] showed the following. When the channel to Willie satisfies the redundant condition, a better transmission rate. However, our channel model does satisfy this condition even when the channel parameters of the channel to Willie are fixed. Therefore, these existing results cannot be applied to our case. To see this fact, we recall the definition of the redundant condition.

We denote the output sample space in the channel to Willie by $\mathcal{Z}$, which is potentially an infinite set. We also denote a measure on $\mathcal{Z}$ by $\nu(dz)$. We denote the set of Alice's input by $\mathcal{X}$, which is a finite set. Depending on Alice's input $x \in \mathcal{X}$, Willie's output is subject to the probability density function $p_{Z,x}$. When the following condition holds, the channel to Willie is called redundant. There exist two non-identical distributions P_X and P_X' such that

$$\sum_{x \in \mathcal{X}} P_X(x) p_{Z,x} = \sum_{x \in \mathcal{X}} P_X'(x) p_{Z,x}. \qquad (13)$$

In our model (9), the Willie's output is simplified as

$$Z = N_W + bX. \qquad (14)$$

Here, the variable N_W is a Gaussian random variables with mean 0 and variance v_W. Since our channel model of the channel to Willie does not satisfy the redundant condition with given v_W and b, we cannot directly apply the existing method by [4] and [5] to our model.

D. Comparison With Existing Methods

Many recent researchers [15], [16], [17], [18], [19], [20], [21], [22] have obtained achievability of covert transmission

of $O(n)$ bits when only Willie's channel has uncertainty and Bob knows a certain knowledge of his own channel unlike our assumption. When the fading coefficient a is known and only the noise power v_B is unknown, the decoder for the case with the maximum noise power works well. Hence, the existing method does work well in this case. For example, the paper [17] assumes that the fading coefficient a is 1 while v_B is unknown.[2] However, when the fading coefficient a is unknown in addition to v_B, the conventional decoder does not work because the decoder depends on the value of the fading coefficient a. To understand this difficulty, consider the following case with $n = 1$. The sender uses four points $\pm x_1$ and $\pm x_0$ in $\mathbb{R}$ with $x_1 > x_0 > 0$ for the encoding. When the receiver receives a value $y \in (x_0, x_1) \subset \mathbb{R}$, the maximum likelihood (ML) decoder depends on the value of the fading coefficient a. When $a < \frac{2y}{x_1 - x_0}$, the ML decoder estimates that the input is $+x_1$. Otherwise, it estimates that the input is $+x_0$. In this way, the decoder depends on the value of the fading coefficient a in general. In the case of conventional channel coding, people often employ universal coding to resolve this problem. The method of type [30] enables us to develop universal coding for the discrete memoryless channel case. The paper [31] proposed a universal coding that works for the continuous case including the AWGN channels. In this way, it is a crucial technology to develop a code that works independently of the channel parameters.

To see the difficulty to achieve the covert transmission under imperfect knowledge for both channels, the following part discusses what problem appears in simple applications of the original methods [4], [5] under our setting. Our channel model has a uncertainty for the variance v_W of the additive Gaussian noise in the channel to the adversary. This situation is formulated as follows. Users know that the variance of the additive Gaussian noise in the channel to the adversary, Willie, belongs to an open set $\mathcal{V}$, but they cannot identify which value in $\mathcal{V}$ is the true value v_W.

Since Alice is allowed to use various values as the channel input power in the Gaussian point-to-point channel, Alice can select the input X_j freely. Hence, we can apply the method [5] for the redundant case as follows. First we choose a sufficiently small positive number v' such that $v_W + b^2 v'$ belongs to $\mathcal{V}$. Here, v' expresses the possible error for Willie's knowledge about the variance of the additive Gaussian noise in the channel to Willie. That is, even when the true value is v_W, Willie cannot identify which of v_W and $v_W + b^2 v'$ is the true. Hence, Willie has to keep the possibility that $v_W + b^2 v'$ is the true as well as v_W.

Let $R_{v'}$ be a real number such that $R_{v'} < I(Y_j; X_j)$, where X_j is subject to the Gaussian distribution with variance v' and mean 0. Using the conventional random coding of rate $R_{v'}$ with respect to the above Gaussian distribution, we generate a code. Here, the choice of the code is a part of the preshared information between Alice and Bob. When Alice encodes the message M_n via the preshared code, using the preshared code, Bob decodes message M_n. However, as is discussed in the

paper [5], in this case, Willie cannot distinguish the received signals from the Gaussian distribution with variance $v_W + b^2 v'$ and mean 0, which is a special case of the silent case. In this way, this method achieves the covert transmission of $O(n)$ bits. Although this method has a higher covert transmission speed than our method, it has the following problem. When the decoder of this method is a maximum likelihood decoder, it depends on the value of a while it does not depend on the value of v_W. That is, the above method satisfies the condition (i), but does not satisfy the condition (ii);

(i)　　Willie's output of the code simulates the Gaussian distribution with variance $v_W + b^2 v'$ and mean 0.

(ii)　　Bob's decoder does not depend on the channel parameters a and v_B.

Although the paper [31] proposed a universal coding that works for the AWGN channel, the encoder of [31] is generated by a distribution with finite support. Hence, use of the method [31] satisfies the condition (ii), but does not satisfy condition (i).

As another idea, we apply the method [4] as follows. The method [4] employs the code for wire-tap channel. Since the paper [39, Appendix D] proposed a wire-tap code for the AWGN channels, the method [4] satisfies the condition (i) when Bob knows the channel parameters of the channel to Bob. That is, this alternative method does not satisfy the condition (ii). Fortunately, our code satisfies both conditions, i.e., our method is the first method to achieve both conditions (i) and (ii). In this sense, our code has an advantage over a simple application of the methods [4], [5].

IV. ANALYSIS OF BOB'S DECODING

A. Characterization of Correct Decoding

The aim of this section is the asymptotic evaluation of the probability that Bob correctly decodes all messages including the detection of the existence of communication from all senders. In this subsection, we derive a lower bound of this probability.

Bob's receiving signal is written as

$$Y_j = N_{B,j} + \sqrt{\frac{t_n}{n}} \sum_{k=1}^{l_n} a_{i_k} (-1)^{M_{i_k} + S_{i_k,j}}. \tag{15}$$

The variable $\bar{N}_{B,i} := \sum_{j=1}^{n} (-1)^{S_{i,j}} N_{B,j}$ is a Gaussian variable with average 0 and variance $n v_B$. The variables $M_{i_k} + S_{i_k,j} + S_{i,j}$ with $j = 1, \ldots, n$ and $k = 1, \ldots, l_n$ with $i_k \neq i$ are independent binary random variables subject to the uniform distribution.

When Bob focuses on $\bar{Y}_i$ to recover M_i, only the term related to M_i is of his interest and the remaining terms can be considered as noises. Hence, by using $N(i) := \bar{N}_{B,i} + \sqrt{\frac{t_n}{n}} \sum_{k=1, i_k \neq i}^{l_n} a_{i_k} \sum_{j=1}^{n} (-1)^{M_{i_k} + S_{i_k,j} + S_{i,j}}$, the term $\bar{Y}_i$ can be rewritten as

$$\bar{Y}_i = \sum_{j=1}^{n} (-1)^{S_{i,j}} Y_j$$

$$= \sum_{j=1}^{n} (-1)^{S_{i,j}} \left(N_j + \sqrt{\frac{t_n}{n}} \sum_{k=1}^{l} a_{i_k} (-1)^{M_{i_k} + S_{i_k,j}} \right)$$

---

[2] The paper [9] considers the case when the channel parameter is unknown. But, it assumes the BSC channel, which is different from AWGN channel.

$$= \bar{N}_{B,i} + \sqrt{\frac{t_n}{n}} \sum_{k=1}^{l_n} a_{i_k} \sum_{j=1}^{n} (-1)^{M_{i_k}+S_{i_k,j}+S_{i,j}}$$

$$= \bar{N}_{B,i} + \sqrt{nt_n} a_i (-1)^{M_i}$$
$$+ \sqrt{\frac{t_n}{n}} \sum_{k=1, i_k \neq i}^{l_n} a_{i_k} \sum_{j=1}^{n} (-1)^{M_{i_k}+S_{i_k,j}+S_{i,j}}$$

$$= \sqrt{nt_n} a_i (-1)^{M_i} + N(i). \qquad (16)$$

That is, $N(i)$ is considered as a noise.

Assume that $\mathcal{A}_i$ is silent, i.e., $i \in (\{i_k\}_{k=1}^{l_n})^c$. Bob's decoding is correct when $|N(i)| < \sqrt{nt_n}\frac{a}{2}$. That is, the probability of Bob's correct decoding is $\Pr\left(|N(i)| < \sqrt{nt_n}\frac{a}{2}\right)$. See Fig. 2 to illustrate this process.

Assume that $\mathcal{A}_i$ is active. Bob's decoding is correct when $M_i = 0$ and $N(i) \leq \sqrt{nt_n}(a_i - \frac{a}{2})$. Also, Bob's decoding is correct when $M_i = 1$ and $N(i) \geq -\sqrt{nt_n}(a_i - \frac{a}{2})$. Since $N(i)$ is symmetric, i.e., the distribution of $N(i)$ is the same as the distribution of $-N(i)$, the probability of Bob's correct decoding is $\Pr\left(N(i) \leq \sqrt{nt_n}(a_i - \frac{a}{2})\right)$.

Then, we obtain the following lower bound for the probability that Bob correctly decodes all messages including the detection of the existence of communication from all senders.

$$P_n := \prod_{k=1}^{l_n} \Pr\left(N(i_k) \leq \sqrt{nt_n}(a_{i_k} - \frac{a}{2})\right)$$
$$\cdot \prod_{i \in (\{i_k\}_{k=1}^{l_n})^c} \Pr\left(|N(i)| < \sqrt{nt_n}\frac{a}{2}\right). \qquad (17)$$

Then, as shown in Appendix A, we have the following lemma.

Lemma 1: When l is given as $l_n = \frac{n}{t_n}$ and the conditions (4) and (6) hold, we have

$$P_n \to 1, \qquad (18)$$

Since the probability P_n is a non-negative value upper bounded by 1, its convergence speed to 1 is evaluated by the speed of the convergence of $\log P_n$ to 0. That is, this convergence is evaluated by the speed of the convergence of $\log(-\log P_n)$ to $-\infty$. The following expresses an upper bound of this convergence speed.

$$\log(-\log P_n) \leq -\frac{n\underline{a}^2}{8l_n(v_B + \bar{a}^2)} + \log m_n + \log 2 + o(1). \qquad (19)$$

The condition (6) guarantees that this lower bound goes to $-\infty$.

B. Single Sender Case

The single sender case can be evaluated by putting $\underline{a} = \bar{a} = a$ and $m_n = l_n$. In the same way as (16), we have

$$\bar{Y}_i = \sqrt{nt_n} a (-1)^{M_i} + N(i), \qquad (20)$$

where

$$N(i) := \bar{N}_i + a\sqrt{\frac{t_n}{n}} \sum_{i'=1, i \neq i'}^{l_n} \sum_{j=1}^{n} (-1)^{M_{i'}+S_{i',j}+S_{i,j}}. \qquad (21)$$

Since $N(i)$ is symmetric, in the same way as (17), the probability of correct decoding is lower bounded as

$$\prod_{k=1}^{l_n} \Pr\left(N(i) < \sqrt{nt_n}a\right). \qquad (22)$$

Then, as a special case of Lemma 1, we have the following lemma.

Lemma 2: When l is given as $l_n = \frac{n}{t_n}$ and the conditions (4) and (12) hold, we have

$$\prod_{k=1}^{l_n} \Pr\left(N(i) < \sqrt{nt_n}a\right) \to 1. \qquad (23)$$

V. WILLIE'S DETECTION

A. Formulation

Since the random access channel model contains the point-to-point channel as a special case with $\underline{b} = \bar{b} = b$ and $m_n = l_n$, we discuss the random access channel model in the following. Under the encoder (3), Willie's receiving signal is written as

$$Z_j = N_{W,j} + \frac{1}{\sqrt{n}} \sum_{k=1}^{l_n} b_{i_k} (-1)^{M_{i_k}+S_{i_k,j}}. \qquad (24)$$

We denote the distribution for Z_j by P_{Z_j} and denote the joint distribution for $Z = (Z_1, \ldots, Z_n)$ by P_Z. $P_{Z|M=m}$ expresses the conditional distribution under the condition $M(:= (M_1, \ldots, M_{l_n})) = m(:= (m_1, \ldots, m_{l_n}))$.

We denote the Gaussian distribution with average 0 and variance v by G_v. When all senders are silent, Willie's observation $(Z_1, \ldots, Z_n)$ is subject to the distribution $G_{v_W}^n$. Also, we assume that Willie does not know the exact value of the variance v_W of his receiving device while Willie knows its rough value because a part of the noise $N_{W,j}$ is generated out of Willie's device, which can be considered as a background noise. Then, we denote the set of possible variances with the silent case by $\mathcal{V}$. In the following, we assume that $\mathcal{V}$ is an open set of $\mathbb{R}$.

In this case, to cover Willie's advantageous scenario, we assume that Willie knows the secret message M, but he does not know the binary symbols $\{S_{i,j}\}$. That is, we show that Willie cannot detect the existence of the communication even though he knows the secret message M. When no communication is made, the joint distribution of Willie' receiving signal Z and the secret message M belongs to the set $\{G_{v_W}^n \times P_M | v_W \in \mathcal{V}\}$.

It is known that the statistical distinguishablity is characterized by the variational distance as follows, where we denote the variational distance between P and Q by $d_V(P, Q)$. Given a method $\mathcal{T}$ to distinguish P and Q, we denote the probability being incorrectly deciding the distribution to be P while the true is Q, by $\epsilon_1(\mathcal{T})$, and the probability being incorrectly deciding the distribution to be Q while the true is P, by $\epsilon_2(\mathcal{T})$. The sum of $\epsilon_1(\mathcal{T})$ and $\epsilon_2(\mathcal{T})$ is evaluated as.[3]

$$\epsilon_1(\mathcal{T}) + \epsilon_2(\mathcal{T}) \geq 1 - d_V(P, Q). \qquad (25)$$

---

[3]This fact is well known in the community of quantum information. For example, the reader might see the reference [40, Section 3.2].

Hence, the distinguishability between the real distribution $P_{Z,M}$ of making the communication and the case with no communication is measured by the minimum $\min_{v \in \mathcal{V}} d_V(P_{Z,M}, G_v^n \times P_M)$, which shows the ability of distinguishing the following two hypotheses. One is the hypothesis that the true distribution is $P_{Z,M}$, which corresponds to the case of making the communication. The other one is the hypothesis that the true distribution belongs to the set $\{G_{v_W}^n \times P_M | v_W \in \mathcal{V}\}$, which corresponds to the case of no communication. Hence, when our covertness measure $\min_{v \in \mathcal{V}} d_V(P_{Z,M}, G_v^n \times P_M)$ is sufficiently small with any l active senders, the situation with any l active senders cannot be distinguished with the situation with no communication, i.e., Willie cannot detect any active user.

Theorem 3: Assume that l is given as $l_n = \frac{n}{t_n}$ to satisfy the condition (5). Also, we assume that v_W belongs to $\mathcal{V}$, Then, our covertness measure goes to zero as under the protocol presented in Section II-B

$$\min_{v \in \mathcal{V}} d_V(P_{Z,M}, G_v^n \times P_M) \to 0. \tag{26}$$

The combination of Lemma 1 and Theorem 3 leads Theorem 1.

Further, under the case with equal fading in Willie's detection, the condition in Theorem 3 can be relaxed as follows.

Theorem 4: Assume that $\underline{b} = \bar{b} = b$ and l is given as $l_n = \frac{n}{t_n}$ to satisfy the condition

$$\frac{n}{l_n^2}, \frac{l_n^2}{n^3} \to 0. \tag{27}$$

Also, we assume that v_W belongs to $\mathcal{V}$, Then, our covertness measure goes to zero under the protocol presented in Section II-B as

$$\min_{v \in \mathcal{V}} d_V(P_{Z,M}, G_v^n \times P_M) \to 0. \tag{28}$$

When l_n is linear with n, the condition (5) does not hold, but the condition (27) does hold. In this sense, Theorem 4 has a weaker condition for l_n than Theorem 3. The case with equal fading in Willie's detection, i.e., the case with $\underline{b} = \bar{b} = b$, covers the case with the point-to-point channel. Thus, due to Theorem 4, one sender, Alice, can send l_n bits to Bob in one channel coherence time securely with covertness to Willie. the conditions (4) and (6) imply the condition (27), the combination of Lemma 2 and Theorem 4 leads Theorem 2.

B. Useful Formula for Our Covertness Measure

As a preparation of our proofs of Theorems 3 and 4, we prepare a useful formula for our covertness measure, i.e., the minimum variational distance as follows.

$$\min_{v \in \mathcal{V}} d_V(P_{Z,M}, G_v^n \times P_M)$$
$$\leq \min_{v \in \mathcal{V}} d_V(D(P_{Z,M}, P_Z \times P_M) + d_V(P_Z \times P_M, G_v^n \times P_M)$$
$$= d_V(P_{Z,M}, P_Z \times P_M) + \min_{v \in \mathcal{V}} d_V(P_Z, G_v^n). \tag{29}$$

The first term $d_V(P_{Z,M}, P_Z \times P_M)$ expresses the secrecy of the message, and the second term $\min_{v \in \mathcal{V}} d_V(P_Z, G_v^n)$

expresses the possibility that Willie detects the existence of the communication.

The variables $(M_i + S_{i_k,j})_{k,j}$ are independently subject to the binary uniform distribution under the condition $M = m$. The variables $Z = (Z_1, \ldots, Z_n)$ do not depend on m. That is, we have $P_{Z|M=m} = P_Z$. Hence, Willie has no information for the message $M = m$, i.e.,

$$d_V(P_{Z,M}, P_Z \times P_M) = 0. \tag{30}$$

In the following, we discuss $\min_{v \in \mathcal{V}} d_V(P_Z, G_v^n)$. We choose $v' := \frac{1}{n} \sum_{k=1}^{l} b_{i_k}^2$. When $c = \frac{l}{n}$ is sufficiently small, v' is small so that $v_W + v'$ belongs to $\mathcal{V}$ because $v_W \in \mathcal{V}$ and $\mathcal{V}$ is an open set. Hence, we have

$$\min_{v \in \mathcal{V}} d_V(P_Z, G_v^n) \leq d_V(P_Z, G_{v_W+v'}^n). \tag{31}$$

Combining (29), (30), and (31), we obtain

$$\min_{v \in \mathcal{V}} d_V(P_{Z,M}, G_v^n \times P_M) \leq d_V(P_Z, G_{v_W+v'}^n). \tag{32}$$

Applying the Pinsker inequality $d_V(P, Q)^2 \leq \frac{1}{2} D(P\|Q)$, where $D(P\|Q) := \int (\log P(x) - \log Q(x))P(x)dx$, we have $d_V(P_Z, G_{v_W+v'}^n)^2 \leq \frac{1}{2} D(P_{Z_1,\ldots,Z_n} \| G_{v_E+v'}^n)$. Hence, it is sufficient to show that the relative entropy $D(P_{Z_1,\ldots,Z_n} \| G_{v_E+v'}^n)$ between the joint distribution $P_{Z_1,\ldots,Z_n}$ of the random variables $Z_1, \ldots, Z_n$ and the n-fold Gaussian distribution $G_{v_E+v'}^n$ is sufficiently small. Combining (29), (30), and the above application of the Pinsker inequality, we have

$$\min_{v \in \mathcal{V}} d_V(P_{Z,M}, G_v^n \times P_M) \leq \sqrt{\frac{1}{2} D(P_{Z_1,\ldots,Z_n} \| G_{v_E+v'}^n)}. \tag{33}$$

Therefore, the remaining part evaluates $D(P_{Z_1,\ldots,Z_n} \| G_{v_E+v'}^n)$.

C. Analysis of Case With Equal Fading

Now, we show Theorem 4 by using the result in [33]. We assume that $\underline{b} = \bar{b} = b$. To state the result by [33], we define χ^2 distance $\chi^2(P, Q)$ as

$$\chi^2(P, Q) := \int \frac{(p(x) - q(x))^2}{q(x)} dx, \tag{34}$$

where p and q are probability density functions of the distributions P and Q. We define Renyi divergence of order 2, $D_2(P\|Q)$ as

$$D_2(P\|Q) := \log \int \frac{p(x)^2}{q(x)} dx. \tag{35}$$

We have

$$D(P\|Q) \leq D_2(P\|Q) \leq \log(1 + \chi^2(P, Q)) \leq \chi^2(P, Q). \tag{36}$$

Proposition 1 [33, (1.3)]: $U_1, \ldots, U_n$ are n independent and identical distributed variables with average 0 and variance v_W. The distribution of U_i is absolutely continuous and has

the probability density function $P_U(u)$. $P_U(u)$ is symmetric, i.e., $P_U(u) = P_U(-u)$. Define

$$\mathcal{U}_n := \sum_{i=1}^n \frac{1}{\sqrt{n}} U_i. \tag{37}$$

Then, we have

$$\chi^2(P_{\mathcal{U}_n}, G_1) = \frac{(\alpha_4 - 3)^2}{24n^2} + O(\frac{1}{n^3}), \tag{38}$$

where $\alpha_4 := \mathbb{E}[G_1^4]$.

Since v' is simplified as $v' = \frac{l_n b^2}{n} = \frac{b^2}{t_n}$, we have

$$\alpha_4(j) := \mathbb{E}\Big[\Big(\frac{1}{\sqrt{v + v'}} Z_j\Big)^4\Big] = \frac{v'^2 + 6 v_W v' + 3 v^2}{(v_W + v')^2}$$

$$= \frac{v'^2 + 6 v_W v' + 3 v^2 - 3(v_W + v')^2}{(v_W + v')^2} + 3$$

$$= \frac{-2(\frac{b^2}{t_n})^2}{(v_W + \frac{b^2}{t_n})^2} + 3. \tag{39}$$

Since $\underline{b} = \overline{b} = b$, we have $b_{i_k} = b$ for all k, Proposition 1 guarantees that

$$\chi^2(P_{Z_j}, G_{v_W + v'}) = \chi^2(P_{\frac{1}{\sqrt{v+v'}} Z_j}, G_1)$$

$$= \frac{(\alpha_4(j) - 3)^2}{24 l^2} + O(\frac{1}{l_n^3})$$

$$\overset{(a)}{=} \frac{4(\frac{b^2}{t_n})^4}{24 l_n^2 (v_W + \frac{b^2}{t_n})^4} + O(\frac{1}{l_n^3})$$

$$= \frac{b^8}{6 n^2 t_n^2 (v_W + b^2 t_n^{-1})^4} + O(\frac{t_n^3}{n^3}), \tag{40}$$

where (a) follows from (39). Hence, we have

$$D(P_{Z_j} \| G_{v_W + v'}) \le \frac{b^8}{6 n^2 t_n^2 (v_W + b^2 t_n^{-1})^4} + O(\frac{t_n^3}{n^3}). \tag{41}$$

Thus,

$$D(P_{Z_1,\dots,Z_n} \| G_{v_W + v'}^n)$$

$$\le n\Big(\frac{b^8}{6 n^2 t_n^2 (v_W + b^2 t_n^{-1})^4} + O(\frac{t_n^3}{n^3})\Big)$$

$$= \frac{b^8}{6 n t_n^2 (v_W + b^2 t_n^{-1})^4} + O(\frac{t_n^3}{n^2}). \tag{42}$$

The condition (27) guarantees that $\frac{b^8}{6 n t_n^2 (v_W + b^2 t_n^{-1})^4}$ and $\frac{t_n^3}{n^2}$ go to zero, which implies that the term $D(P_{Z_1,\dots,Z_n} \| G_{v_W + v'}^n)$ goes to zero. Hence, combining (33), we obtain Theorem 4.

D. Analysis of Case With Unequal Fading

Now, we show Theorem 3. We consider the case with unequal fading in Willie's detection. For the analysis of this case, we have the following theorem.

Theorem 5: The relative entropy $D(P_{Z_1,\dots,Z_n} \| G_{v_W + v'}^n)$ is evaluated as

$$D(P_{Z_1,\dots,Z_n} \| G_{v_W + v'}^n)$$

$$\le \frac{\frac{1}{n v'^2} \sum_{k=1}^{l_n} b_{i_k}^4 \Big(\frac{(v_W + v')(e^{\frac{v'}{v_W + 2v'}} + e^{-\frac{v'}{v_W}})}{2\sqrt{v_W(v_W + 2v')}} - 1\Big)}{\frac{e^{-\frac{v' + v_W}{2 v_W}}}{2} + (1 - \frac{e^{-\frac{v' + v_W}{2 v_W}}}{2}) \frac{1}{n^2 v'^2} \sum_{k=1}^{l_n} b_{i_k}^4}. \tag{43}$$

Since the proof of this theorem is very long, Appendix C proves it by using the result in [35], which employs the Poincaré constant.

We choose $c_n := \frac{l_n}{n} = \frac{1}{t_n}$. We have $\sum_{k=1}^{l_n} b_{i_k}^4 \le l_n \overline{b}^4 = c_n \overline{b}^4 n$ and $v' \le \frac{l_n \overline{b}^2}{n} = c_n \overline{b}^2$. Since the condition (5) by using (66), the RHS of (43) is evaluated as

(RHS of (43))

$$\le \frac{\sum_{k=1}^{l_n} b_{i_k}^4}{n v'^2} \frac{1}{\frac{e^{-\frac{v' + v_W}{2 v_W}}}{2}} \Big(\frac{1}{6}\Big(\frac{v'}{v_W}\Big)^3 + O\Big(\Big(\frac{v'}{v_W}\Big)^4\Big)\Big)$$

$$= \Big(\sum_{k=1}^{l_n} b_{i_k}^4\Big) \frac{e^{1/2}}{3}\Big(\frac{v'}{n v_W^3}\Big)\Big(1 + O\Big(\frac{v'}{v_W}\Big)\Big) + O\Big(\Big(\frac{v'}{v_W}\Big)^4\Big)$$

$$\le c_n \overline{b}^4 \frac{e^{1/2}}{3}\Big(\frac{c \overline{b}^2}{v_W^3}\Big)(1 + O(c_n)) + O(c_n^4)$$

$$= \frac{c_n^2 e^{1/2}}{3}\Big(\frac{\overline{b}^2}{v_W}\Big)^3 + O(c_n^3). \tag{44}$$

Since the condition (5) guarantee that $c_n \to 0$, we have

$$D(P_{Z_1,\dots,Z_n} \| G_{v_W + v'}^n) \to 0. \tag{45}$$

Hence, combining (33), we complete the proof of Theorem 3.

Here, we compare the evaluation of (43) with the evaluation of (42). Eq. (43) has the order $O(\frac{1}{t_n^2}) = O(\frac{1}{\log n^2})$ for n, and Eq. (42) has the order $O(\frac{1}{n t_n^2}) = O(\frac{1}{n \log n^2})$ for n. Hence, Eq. (42) results in a much better covertness evaluation than Eq. (43). Therefore, combining this discussion and Lemma 1, we obtain Theorem 1.

VI. DISCUSSION AND OPEN PROBLEMS

We have discussed covert communication assuming BPSK over AWGN channels. Our results are composed of two contributions: covert communications for the random access channel and point-to-point communication. In the former case, we assume that the sender can choose the power to a certain value that is fixed during a coding block length. Our encoder and our decoder in both settings are quite simple. Thus, the proposed methods are easily implementable being the main complexity the problem of key sharing, as in traditional spread spectrum. However, we have not described the detail of the implementation of our method based on spread-spectrum principle, which is needed for practical application. Its detailed description is an important future study.

In our method, we assume that Willie does not have perfect knowledge on the channel parameter. That is, using this lack of Willie's knowledge, our method improves the transmission speed over existing methods [11], [12], [13] in the case of AWGN channel. The key idea for our method is convergence of the distribution of the weighted sample mean of n independent random variables to a Gaussian distribution, which is related to a kind of central limit theorem [32]. This method depends on the fact that our output distribution is a Gaussian

distribution. That is, it is impossible to extend this method to another channel model. Due to this reason, we also assume that Bob does not have perfect knowledge on the channel parameter. Our code is shown to guarantee that Bob reliably recovers the message under this condition. That is, our code is a universal code in this sense unlike the recent papers [15], [16], [17], [18], [19], [20], [21], and [22].

In addition, since our covertness analysis needs the evaluation of the variational distance, we need more precise evaluation than the variant of central limit theorem [32]. To resolve this problem, we have employed the results from the papers [33], [35], and [34]. As discussed in Section V-C, the result in [33] has been used for the analysis on the case with equal fading. For the general case with unequal fading, as discussed in Section V-D, the analysis on the papers [34], [35] employs Poincaré constant [34], [35], [36], [37]. Although Section V-D analyzes the case with unequal fading, the evaluation in Section V-C is tighter than in Section V-D. Hence, the above analysis has better evaluation for the ability of Willie's detection of the existence of the communication than the use of evaluation in Section V-D. Hence, Eq. (42) suggests a possibility to improve the evaluation (43). For this improvement, it is needed to extend Proposition 1 to the case (62), i.e., the case with unequal weighted sum. It is an interesting a future study.

Further, we have not proved the converse part for the transmission length for our covert communication model. Since our code construction is based on an elementary idea, there is a possibility to improve the transmission speed of our method. It is another future problem to derive the asymptotically tight transmission speed under both settings. In the relation to this topic, to state the advantage of our method for the point-to-point communication, we have pointed that existing methods [4], [5] need perfect knowledge for channel parameters. That is, our model requires a code to satisfy the two conditions (i) and (ii) defined in Section III-D simultaneously. Although our code satisfies both conditions, we have not shown that no code with transmission length $O(n)$ satisfies both conditions in the point-to-point communication. In fact, in our protocol for the point-to-point communication, our code is essentially constructed with bit-by-bit communication. Clearly, this method is not efficient when covertness is not discussed. Therefore, it is an interesting point whether or not our protocol can be improved by a more efficient method in the point-to-point scenario. It is another open problem to clarify this issue.

Also, our method needs pre-shared secret of $O(n^2)$ bits. Since this is larger than the size of transmitted message, it is better to reduce this size while keeping our transmission rate. One possibility for this solution is choosing the matrix $S_{i,j}$ to be a Toeplitz matrix. While this choice reduces the size of pre-shared secret to $O(n)$, it changes the stochastic behaviors of Y_j and Z_j. It is another future study to evaluate the performance of this modified code. Finally, it is an interesting future study to apply the theoretical limits obtained in this work to a practical engineering setting with realistic fading channel parameters and evaluate the performance with practical BPSK symbol acquisition and realistic channel dynamics using metrics such as covert security outage probability.

APPENDIX A
PROOF OF LEMMA 1

To evaluate the above quantity, we introduce two kinds of cummulant generating functions ϕ_b and $\phi_{g,v}$ as

$$\phi_b(s) := \log \frac{e^s + e^{-s}}{2}, \quad \phi_{g,v} := \frac{vs^2}{2}. \tag{46}$$

That is, ϕ_b is the cummulant generating function of the variable $(-1)^X$ when X obeys the binary distribution with probability $\frac{1}{2}$, and $\phi_{g,v}$ is the cummulant generating function of Gaussian distribution with average 0 and variance v. Therefore, the cummulant generating function of $N(i)$ is $\phi_{n,i}(s) := \phi_{g,nv_B}(s) + \sum_{k=1,i_k \neq i}^{l_n} n\phi_b(a_{i_k} s \sqrt{\frac{t_n}{n}})$. We have

$$\phi_{n,i}(s) \leq \phi_n(s) := \phi_{g,nv_B}(s) + l_n n\phi_b\left(\bar{a}s\sqrt{\frac{t_n}{n}}\right). \tag{47}$$

When n is large,

$$\phi_n(s) = \frac{s^2}{2}(nv_B + l_n\bar{a}^2 t_n) + O\left(nl\left(\frac{t_n}{n}\right)^{\frac{3}{2}} s^3\right)$$
$$= \frac{s^2}{2}(nv_B + l_n\bar{a}^2 t_n) + O\left(l\sqrt{\frac{t_n^3}{n}}s^3\right). \tag{48}$$

We assume that $\mathcal{A}_i$ is silent. Since the condition (4) guarantees $\frac{n}{l_n^2} \to 0$, using (48), we have

$$\beta_n := \max_s \left(s\sqrt{nt_n}\frac{a}{2} - \phi_n(s)\right)$$
$$= \frac{(\sqrt{nt_n}\frac{a}{2})^2}{2(nv_B + l_n\bar{a}^2 t_n)} + O\left(l_n\sqrt{\frac{t_n^3}{n}}\left(\frac{\sqrt{t_n}}{\sqrt{n}v_B}\right)^3\right)$$
$$= \frac{nt_n\bar{a}^2}{8n(v_B + \bar{a}^2)} + O\left(\frac{l_n t_n^3}{n^2 v_B^3}\right) = \frac{t_n\bar{a}^2}{8(v_B + \bar{a}^2)} + o(1). \tag{49}$$

Thus, under the condition (6), we have

$$m_n \exp(-\beta_n) = \exp(\log m_n - \beta_n) \to 0. \tag{50}$$

Also, the condition (6) guarantees

$$\left(1 - 2\exp(-\beta_n)\right)^{\frac{1}{2}\exp(\beta_n)} \to \frac{1}{e}. \tag{51}$$

Markov inequality implies

$$1 - \Pr\left(|N(i)| < \sqrt{nt_n}\frac{a}{2}\right)$$
$$\leq 2\exp\left(-\max_s\left(s\sqrt{nt_n}\frac{a}{2} - \phi_{n,i}(s)\right)\right)$$
$$\leq 2\exp(-\beta_n). \tag{52}$$

Therefore,

$$\prod_{i\in(\{i_k\}_{k=1}^{l_n})^c} \Pr\left(|N(i)| < \sqrt{nt_n}\frac{a}{2}\right) \geq \left(1 - 2\exp(-\beta_n)\right)^{m_n-l_n}. \tag{53}$$

We assume that $\mathcal{A}_i$ is active. Since we have

$$s\sqrt{nt_n}\left(a_i - \frac{a}{2}\right) - \phi_{n,i}(s) \geq s\sqrt{nt_n}\frac{a}{2} - \phi_n(s) \tag{54}$$

for $s \geq 0$, we have

$$\max_s \left(s\sqrt{nt_n}(a_i - \frac{a}{2}) - \phi_{n,i}(s) \right) \geq \beta_n. \quad (55)$$

Markov inequality implies

$$1 - \Pr\left(N(i) < \sqrt{nt_n}(a_i - \frac{a}{2}) \right)$$

$$\leq \exp\left(-\max_s \left(s\sqrt{nt_n}(a_i - \frac{a}{2}) - \phi_{n,i}(s) \right) \right) \leq \exp(-\beta_n). \quad (56)$$

Therefore, we have

$$\prod_{k=1}^{l_n} \Pr\left(N(i_k) \leq \sqrt{nt_n}(a_{i_k} - \frac{a}{2}) \right) \geq (1 - \exp(-\beta_n))^{l_n}. \quad (57)$$

Combining (53) and (57), we have

$$P_n = \prod_{k=1}^{l_n} \Pr\left(N(i_k) \leq \sqrt{nt_n}(a_{i_k} - \frac{a}{2}) \right)$$

$$\cdot \prod_{i \in \{i_k\}_{k=1}^{l_n}} \Pr\left(|N(i)| < \sqrt{nt_n}\frac{a}{2} \right)$$

$$\geq (1 - \exp(-\beta_n))^{l_n} \cdot \left(1 - 2\exp(-\beta_n) \right)^{m_n - l_n}$$

$$\geq \left(1 - 2\exp(-\beta_n) \right)^{m_n}$$

$$= \left(\left(1 - 2\exp(-\beta_n) \right)^{\frac{1}{2}\exp(\beta_n)} \right)^{2m_n \exp(-\beta_n)} \xrightarrow{(a)} 1, \quad (58)$$

where (a) follows from (50) and (51). Hence, we obtain (18).

In addition, using (49) and (51), we have

$$\log(-\log P_n)$$

$$\overset{(a)}{\leq} \log\left(-\log\left(\left(1 - 2\exp(-\beta_n) \right)^{\frac{1}{2}\exp(\beta_n)} \right)^{2m_n \exp(-\beta_n)} \right)$$

$$= \log\left(2m_n \exp(-\beta_n) \left(-\log\left(\left(1 - 2\exp(-\beta_n) \right)^{\frac{1}{2}\exp(\beta_n)} \right) \right) \right)$$

$$= \log(2m_n) - \beta_n + \log\left(-\log\left(\left(1 - 2\exp(-\beta_n) \right)^{\frac{1}{2}\exp(\beta_n)} \right) \right)$$

$$\overset{(b)}{=} -\frac{n\underline{a}^2}{8l_n(v_B + \overline{a}^2)} + \log m_n + \log 2 + o(1) + \log(1 + o(1))$$

$$= -\frac{n\underline{a}^2}{8l_n(v_B + \overline{a}^2)} + \log m_n + \log 2 + o(1), \quad (59)$$

where (a) follows from (58) and (b) follows from (49) and (51). This relation implies (19).

APPENDIX B
PREPARATION FOR PROOF OF THEOREM 5

For our proof of Theorem 5, we make several preparations in this subsection. First, we introduce the Poincaré constant. Assume that the random variable H is subject to a distribution

P on $\mathbb{R}$. We define the Poincaré constant $C(P)$ for the distribution P as

$$C(P) := \inf_{f:\text{smooth on } \mathbb{R}} \frac{E[(f'(H))^2]}{V[f(H)]}. \quad (60)$$

The value $C(P)$ will be used in (63). For example, it is known that the Poincaré constant $C(G_v)$ is calculated as [41], [42]

$$C(G_v) = \frac{1}{v}. \quad (61)$$

Using the Poincaré constant, we have the following proposition.

Proposition 2 [35, Theorem 1]: $U_1, \ldots, U_n$ are n independent and identical distributed variables with average 0 and variance 1. The distribution of U_i is absolutely continuous and has the probability density function $P_U(u)$. Define

$$\mathcal{U}_n := \sum_{i=1}^n \alpha_i U_i, \quad L(\alpha) := \sum_{i=1}^n \alpha_i^4, \quad (62)$$

where $\sum_{i=1}^n \alpha_i^2 = 1$. Then, we have

$$D(P_{\mathcal{U}_n} \| G_1) \leq \frac{L(a)}{\frac{C(p)}{2} + (1 - \frac{C(p)}{2})L(a)} D(P_U \| G_1). \quad (63)$$

As another preparation, we introduce the probability density function $P_{v,v'}$ of the random variable $\sqrt{\frac{v'}{v'+v}}\left((-1)^X + \sqrt{\frac{v}{v'}}N \right)$, where X is the binary variable subject to the uniform distribution and N is the standard Gaussian variable.

The Poincaré constant of $P_{v,v'}$ is evaluated as follows.
Lemma 3: We have

$$C(P_{v,v'}) \geq e^{-\frac{v'+v}{2v}}. \quad (64)$$

Also, we evaluate the relative entropy between $P_{v,v'}$ and the Gaussian distribution in the following lemma.
Lemma 4:

$$D(P_{v,v'} \| G_1) \leq \frac{(v+v')(e^{\frac{v'}{v+2v'}} + e^{-\frac{v'}{v}})}{2\sqrt{v(v+2v')}} - 1. \quad (65)$$

In addition, when $\frac{v'}{v}$ is small, we have

$$\frac{(v+v')(e^{\frac{v'}{v+2v'}} + e^{-\frac{v'}{v}})}{2\sqrt{v(v+2v')}} - 1 = \frac{1}{6}\left(\frac{v'}{v} \right)^3 + O\left(\left(\frac{v'}{v} \right)^4 \right). \quad (66)$$

The proofs of Lemmas 3 and 4 will be given in Appendices E and D, respectively.

APPENDIX C
PROOF OF THEOREM 5

In this subsection, we show Theorem 5 by using the above preparation. We prepare l_n independent standard Gaussian variables $N_1, \ldots, N_{l_n}$. We have

$$Z_j = N_{W,j} + \frac{1}{\sqrt{n}} \sum_{k=1}^{l_n} b_{i_k}(-1)^{M_i + S_{i_k,j}}$$

$$= \sum_{k=1}^{l_n} \frac{b_{i_k}}{\sqrt{n}}\left((-1)^{M_i + S_{i_k,j}} + \sqrt{\frac{v_W}{v'}}N_k \right). \quad (67)$$

We apply Proposition 2 to the case $n = l_n$, $i = k$, $\alpha_k = \frac{b_{i_k}}{\sqrt{nv'}}$, $U_k = \sqrt{\frac{v'}{v'+v_W}}\left((-1)^{M_i+S_{i_k,j}} + \sqrt{\frac{v_W}{v'}}N_k\right)$. Then, we have

$$\sum_{k=1}^{l_n} \alpha_k^4 = \frac{1}{n^2 v'^2} \sum_{k=1}^{l_n} b_{i_k}^4. \tag{68}$$

Since $Z_j = \sqrt{v_W + v'}\left(\sum_{k=1}^{l_n} \alpha_k S_k\right)$, applying Proposition 2, we have

$$D(P_{Z_j}\|G_{v_W+v'})$$

$$= D(P_{\sum_{k=1}^{l_n} \alpha_k S_k}\|G_1)$$

$$\leq \frac{\frac{1}{n^2 v'^2}\sum_{k=1}^{l_n} b_{i_k}^4}{\frac{C(P_{v_W,v'})}{2} + \left(1 - \frac{C(P_{v_W,v'})}{2}\right)\frac{1}{n^2 v'^2}\sum_{k=1}^{l_n} b_{i_k}^4} D(P_{v_W,v'}\|G_1)$$

$$\overset{(a)}{\leq} \frac{\frac{1}{n^2 v'^2}\sum_{k=1}^{l_n} b_{i_k}^4\left(\frac{(v_W+v')(e^{\frac{v'}{v_W+2v'}} + e^{-\frac{v'}{v_W}})}{2\sqrt{v_W(v_W+2v')}} - 1\right)}{\frac{C(P_{v_W,v'})}{2} + \left(1 - \frac{C(P_{v_W,v'})}{2}\right)\frac{1}{n^2 v'^2}\sum_{k=1}^{l_n} b_{i_k}^4}$$

$$\overset{(b)}{\leq} \frac{\frac{1}{n^2 v'^2}\sum_{k=1}^{l_n} b_{i_k}^4\left(\frac{(v_W+v')(e^{\frac{v'}{v_W+2v'}} + e^{-\frac{v'}{v_W}})}{2\sqrt{v_W(v_W+2v')}} - 1\right)}{\frac{e^{-\frac{v'+v_W}{2v_W}}}{2} + \left(1 - \frac{e^{-\frac{v'+v_W}{2v_W}}}{2}\right)\frac{1}{n^2 v'^2}\sum_{k=1}^{l_n} b_{i_k}^4}, \tag{69}$$

where (a) and (b) follow from Lemma 4 and Lemma 3, respectively. Therefore,

$$D(P_{Z_1,\ldots,Z_n}\|G_{v_W+v'}^n)$$

$$= \sum_{j=1}^{n} D(P_{Z_j}\|G_{v_W+v'})$$

$$\leq n \frac{\frac{1}{n^2 v'^2}\sum_{k=1}^{l_n} b_{i_k}^4\left(\frac{(v_W+v')(e^{\frac{v'}{v_W+2v'}} + e^{-\frac{v'}{v_W}})}{2\sqrt{v_W(v_W+2v')}} - 1\right)}{\frac{e^{-\frac{v'+v_W}{2v_W}}}{2} + \left(1 - \frac{e^{-\frac{v'+v_W}{2v_W}}}{2}\right)\frac{1}{n^2 v'^2}\sum_{k=1}^{l_n} b_{i_k}^4}$$

$$= \frac{\frac{1}{nv'^2}\sum_{k=1}^{l_n} b_{i_k}^4\left(\frac{(v_W+v')(e^{\frac{v'}{v_W+2v'}} + e^{-\frac{v'}{v_W}})}{2\sqrt{v_W(v_W+2v')}} - 1\right)}{\frac{e^{-\frac{v'+v_W}{2v_W}}}{2} + \left(1 - \frac{e^{-\frac{v'+v_W}{2v_W}}}{2}\right)\frac{1}{n^2 v'^2}\sum_{k=1}^{l_n} b_{i_k}^4}. \tag{71}$$

APPENDIX D
PROOF OF LEMMA 4

The probability density function $p_{v,v'}$ of the distribution $P_{v,v'}$ for $\sqrt{\frac{v'}{v'+v}}\left((-1)^X + \sqrt{\frac{v}{v'}}N\right)$ is

$$p_{v,v'}(x) = \frac{1}{2}\sqrt{\frac{v'+v}{2\pi v}}e^{-\frac{(v'+v)\left(x-\sqrt{\frac{v'}{v'+v}}\right)^2}{2v}}$$

$$+ \frac{1}{2}\sqrt{\frac{v'+v}{2\pi v}}e^{-\frac{(v'+v)\left(x+\sqrt{\frac{v'}{v'+v}}\right)^2}{2v}}. \tag{72}$$

Then, we have (70), as shown at the bottom of the next page. Hence,

$$\chi^2(P_{v,v'}, G_1) + 1$$

$$= \int_{-\infty}^{\infty}\left(\frac{1}{2}\sqrt{\frac{v'+v}{v}}e^{-\frac{(v'+v)\left(x-\sqrt{\frac{v'}{v'+v}}\right)^2}{2v}+\frac{x^2}{2}}\right.$$

$$\left.+ \frac{1}{2}\sqrt{\frac{v'+v}{v}}e^{-\frac{(v'+v)\left(x+\sqrt{\frac{v'}{v'+v}}\right)^2}{2v}+\frac{x^2}{2}}\right)^2 \sqrt{\frac{1}{2\pi}}e^{-\frac{x^2}{2}}\,dx$$

$$= \int_{-\infty}^{\infty}\left(\frac{v'+v}{4v}e^{-\frac{(v'+v)\left(x-\sqrt{\frac{v'}{v'+v}}\right)^2}{v}+x^2}\right.$$

$$+ \frac{v'+v}{4v}e^{-\frac{(v'+v)\left(x+\sqrt{\frac{v'}{v'+v}}\right)^2}{v}+x^2}$$

$$\left.+ \frac{v'+v}{2v}e^{-\frac{(v'+v)\left(x-\sqrt{\frac{v'}{v'+v}}\right)^2}{2v}-\frac{(v'+v)\left(x+\sqrt{\frac{v'}{v'+v}}\right)^2}{2v}+x^2}\right)$$

$$\times \sqrt{\frac{1}{2\pi}}e^{-\frac{x^2}{2}}\,dx$$

$$= \sqrt{\frac{1}{2\pi}}\int_{-\infty}^{\infty}\left(\frac{v'+v}{4v}e^{-\frac{v+2v'}{2v}\left(x-\frac{2\sqrt{v'(v'+v)}}{v+2v'}\right)^2+\frac{v'}{v+2v'}}\right.$$

$$+ \frac{v'+v}{4v}e^{-\frac{v+2v'}{2v}\left(x+\frac{2\sqrt{v'(v'+v)}}{v+2v'}\right)^2+\frac{v'}{v+2v'}}$$

$$\left.+ \frac{v'+v}{2v}e^{-\frac{v+2v'}{2v}x^2-\frac{v'}{v}}\right)dx$$

$$= 2\cdot\frac{v'+v}{4v}\sqrt{\frac{v}{v+2v'}}e^{\frac{v'}{v+2v'}} + \frac{v'+v}{2v}\sqrt{\frac{v}{v+2v'}}e^{-\frac{v'}{v}}$$

$$= \frac{v+v'}{2\sqrt{v(v+2v')}}\left(e^{\frac{v'}{v+2v'}} + e^{-\frac{v'}{v}}\right), \tag{73}$$

which implies that

$$\chi^2(P_{v,v'}, G_1) = \frac{(v+v')(e^{\frac{v'}{v+2v'}} + e^{-\frac{v'}{v}})}{2\sqrt{v(v+2v')}} - 1. \tag{74}$$

Combination of (36) and (74) yields (65).

When v' is small, $\frac{1}{2\sqrt{v(v+2v')}} = \frac{1}{2v}(1 - \frac{1}{2}\frac{2v'}{v} + \frac{3}{8}(\frac{2v'}{v})^2 - \frac{5}{16}(\frac{2v'}{v})^3 + O(\frac{1}{v}(\frac{v'}{v})^4)) = \frac{1}{2v}\left(1 - \frac{v'}{v} + \frac{3}{2}(\frac{v'}{v})^2 - \frac{5}{2}(\frac{v'}{v})^3 + O(\frac{1}{v}(\frac{v'}{v})^4)\right)$. Hence, using (70) in the top of this page, we obtain (66).

APPENDIX E
PROOF OF LEMMA 3

To show Lemma 3, we prepare the following proposition.

Proposition 3 ([34, Proposition 5]): Two absolutely continuous distributions μ and ν satisfy the inequality

$$\frac{\min_x \frac{d\nu}{d\mu}(x)}{\max_x \frac{d\nu}{d\mu}(x)}C_P(\mu) \leq C_P(\nu). \tag{75}$$

We define the function $f(x)$ as

$$f(x) := \frac{p_{v,v'}(x)}{p_{v,0}(x)} = \frac{1}{2}\sqrt{\frac{v'+v}{v}}e^{-\frac{(v'+v)}{2v}\left(x-\sqrt{\frac{v'}{v'+v}}\right)^2+\frac{x^2}{2}}$$

$$+ \frac{1}{2}\sqrt{\frac{v'+v}{v}}e^{-\frac{(v'+v)}{2v}\left(x+\sqrt{\frac{v'}{v'+v}}\right)^2+\frac{x^2}{2}}$$

$$= \frac{1}{2}\sqrt{\frac{v'+v}{v}}e^{-\frac{v'}{2v}x^2+\frac{(v'+v)}{v}\sqrt{\frac{v'}{v'+v}}x-\frac{v'}{2v}}$$

$$+ \frac{1}{2}\sqrt{\frac{v'+v}{v}}e^{-\frac{v'}{2v}x^2-\frac{(v'+v)}{v}\sqrt{\frac{v'}{v'+v}}x-\frac{v'}{2v}}. \tag{76}$$

For $x \geq 0$, we have

$$f(x) \leq \sqrt{\frac{v'+v}{v}}e^{-\frac{v'}{2v}x^2+\frac{(v'+v)}{v}\sqrt{\frac{v'}{v'+v}}x-\frac{v'}{2v}}$$

$$= \sqrt{\frac{v'+v}{v}}e^{-\frac{v'}{2v}\left(x^2-2\sqrt{\frac{v'+v}{v'}}x\right)-\frac{v'}{2v}}$$

$$= \sqrt{\frac{v'+v}{v}}e^{-\frac{v'}{2v}\left(x-\sqrt{\frac{v'+v}{v'}}\right)^2+\frac{v+v'}{2v}-\frac{v'}{2v}}$$

$$= \sqrt{\frac{v'+v}{v}}e^{-\frac{v'}{2v}\left(x-\sqrt{\frac{v'+v}{v'}}\right)^2+\frac{1}{2}} \leq \sqrt{\frac{v'+v}{v}}e^{\frac{1}{2}}. \tag{77}$$

The derivative of f is calculated as

$$\frac{df}{dx}(x) = \frac{1}{2}\sqrt{\frac{v'+v}{v}}\left(-\frac{v'}{v}x+\frac{v'+v}{v}\sqrt{\frac{v'}{v'+v}}\right)$$

$$\times e^{-\frac{v'}{2v}x^2+\frac{v'+v}{v}\sqrt{\frac{v'}{v'+v}}x-\frac{v'}{2v}}$$

$$+ \frac{1}{2}\sqrt{\frac{v'+v}{v}}\left(-\frac{v'}{v}x-\frac{v'+v}{v}\sqrt{\frac{v'}{v'+v}}\right)$$

$$\times e^{-\frac{v'}{2v}x^2-\frac{v'+v}{v}\sqrt{\frac{v'}{v'+v}}x-\frac{v'}{2v}}. \tag{78}$$

Assume that $x \geq 0$. The relation $\frac{df}{dx}(x) = 0$ holds if and only if

$$\left(-\frac{v'}{v}x+\frac{v'+v}{v}\sqrt{\frac{v'}{v'+v}}\right)e^{-\frac{v'}{2v}x^2+\frac{v'+v}{v}\sqrt{\frac{v'}{v'+v}}x-\frac{v'}{2v}}$$

$$= \left(\frac{v'}{v}x+\frac{v'+v}{v}\sqrt{\frac{v'}{v'+v}}\right)e^{-\frac{v'}{2v}x^2-\frac{v'+v}{v}\sqrt{\frac{v'}{v'+v}}x-\frac{v'}{2v}}. \tag{79}$$

The above relation is equivalent to

$$e^{2\frac{\sqrt{(v'+v)v'}}{v}x} = e^{2\frac{v'+v}{v}\sqrt{\frac{v'}{v'+v}}x}$$

$$= \frac{\frac{v'}{v}x+\frac{(v'+v)}{v}\sqrt{\frac{v'}{v'+v}}}{-\frac{v'}{v}x+\frac{v'+v}{v}\sqrt{\frac{v'}{v'+v}}} = \frac{x+\sqrt{\frac{v'+v}{v'}}}{-x+\sqrt{\frac{v'+v}{v'}}}. \tag{80}$$

That is,

$$2\frac{\sqrt{(v'+v)v'}}{v}x$$

$$= \log\left(x+\sqrt{\frac{v'+v}{v'}}\right) - \log\left(-x+\sqrt{\frac{v'+v}{v'}}\right). \tag{81}$$

We denote the RHS and the LHS by $g(x)$ and $h(x)$, respectively. Their derivatives are calculated as

$$\frac{dg}{dx}(x) = 2\frac{\sqrt{(v'+v)v'}}{v} \tag{82}$$

$$\frac{dh}{dx}(x) = \frac{1}{x+\sqrt{\frac{v'+v}{v'}}} + \frac{1}{-x+\sqrt{\frac{v'+v}{v'}}} = \frac{2\sqrt{\frac{v'+v}{v'}}}{-x^2+\frac{v'+v}{v'}}. \tag{83}$$

Hence, we have

$$\frac{\frac{dg}{dx}(0)}{\frac{dh}{dx}(0)} = \frac{v+v'}{v} > 1. \tag{84}$$

Since $\frac{\frac{dg}{dx}(x)}{\frac{dh}{dx}(x)}$ is monotonically decreasing for x, the solution of $\frac{dg}{dx}(x) - \frac{dh}{dx}(x) = 0$ in $(0, \frac{v'+v}{v'})$ is only one element

---

$$\chi^2(P_{v,v'}, G_1) + 1 = \frac{(v+v')(e^{\frac{v'}{v+2v'}}+e^{-\frac{v'}{v}})}{2\sqrt{v(v+2v')}}$$

$$= \frac{v(1+\frac{v'}{v})(2+\frac{v'}{v+2v'}-\frac{v'}{v}+\frac{1}{2}(\frac{v'}{v+2v'})^2+\frac{1}{6}(\frac{v'}{v+2v'})^3+\frac{1}{2}(\frac{v'}{v})^2+\frac{1}{6}(\frac{v'}{v})^3+O((\frac{v'}{v})^4)))}{2v}$$

$$\cdot \left(1-\frac{v'}{v}+\frac{3}{2}(\frac{v'}{v})^2-\frac{5}{2}(\frac{v'}{v})^3+O\left(\frac{1}{v}\left(\frac{v'}{v}\right)^4\right)\right)$$

$$= \frac{(1+\frac{v'}{v})(2+\frac{v'}{v}(1-2\frac{v'}{v}+(2\frac{v'}{v})^2)-\frac{v'}{v}+\frac{1}{2}(\frac{v'}{v})^2(1-4\frac{v'}{v})+\frac{1}{6}(\frac{v'}{v})^3+\frac{1}{2}(\frac{v'}{v})^2+\frac{1}{6}(\frac{v'}{v})^3)}{2}$$

$$\cdot \left(1-\frac{v'}{v}+\frac{3}{2}(\frac{v'}{v})^2-\frac{5}{2}(\frac{v'}{v})^3\right) + O\left(\left(\frac{v'}{v}\right)^4\right)$$

$$= \frac{(1+\frac{v'}{v})(2-(\frac{v'}{v})^2+\frac{7}{3}(\frac{v'}{v})^3)}{2} \cdot \left(1-\frac{v'}{v}+\frac{3}{2}(\frac{v'}{v})^2-\frac{5}{2}(\frac{v'}{v})^3\right) + O\left(\left(\frac{v'}{v}\right)^4\right)$$

$$= (1+\frac{v'}{v})\left(1-\frac{1}{2}(\frac{v'}{v})^2+\frac{7}{6}(\frac{v'}{v})^3\right) \cdot \left(1-\frac{v'}{v}+\frac{3}{2}(\frac{v'}{v})^2-\frac{5}{2}(\frac{v'}{v})^3\right) + O\left(\left(\frac{v'}{v}\right)^4\right)$$

$$= \left(1-\frac{1}{2}(\frac{v'}{v})^2+\frac{7}{6}(\frac{v'}{v})^3\right) \cdot \left(1-\frac{v'}{v}+\frac{3}{2}(\frac{v'}{v})^2-\frac{5}{2}(\frac{v'}{v})^3+\frac{v'}{v}-(\frac{v'}{v})^2+\frac{3}{2}(\frac{v'}{v})^3\right) + O\left(\left(\frac{v'}{v}\right)^4\right)$$

$$= \left(1-\frac{1}{2}\left(\frac{v'}{v}\right)^2+\frac{7}{6}\left(\frac{v'}{v}\right)^3\right) \cdot \left(1+\frac{1}{2}\left(\frac{v'}{v}\right)^2-\left(\frac{v'}{v}\right)^3\right) + O\left(\left(\frac{v'}{v}\right)^4\right) = 1+\frac{1}{6}\left(\frac{v'}{v}\right)^3 + O\left(\left(\frac{v'}{v}\right)^4\right). \tag{70}$$

x_0. We have $\frac{dg}{dx}(x_0) = \frac{dh}{dx}(x_0)$. Hence, the solution of $g(x) - h(x) = 0$ in $(0, \frac{v'+v}{v'})$ is only one element x_1. We have $g(x_1) = h(x_1)$. We have $g(x) > h(x)$ for $x < x_1$, and $g(x) < h(x)$ for $x > x_1$. That is, $\frac{df}{dx}(x) > 0$ for $x < x_1$, and $\frac{df}{dx}(x) < 0$ for $\frac{v'+v}{v'} > x > x_1$. Also, $\frac{df}{dx}(x) < 0$ for $x \geq \frac{v'+v}{v'}$. Hence, we find that $\min_{x \geq 0} f(x)$ is realized with $x = 0$ and $\max_{x \geq 0} f(x)$ is realized with $x = x_1$. Thus, we have

$$\min_{x \geq 0} f(x) = f(0) = \sqrt{\frac{v'+v}{v}} e^{-\frac{v'}{2v}}. \qquad (85)$$

Combining (77) and (85), we have

$$\frac{\min_x \frac{d\nu}{d\mu}(x)}{\max_x \frac{d\nu}{d\mu}(x)} \geq e^{-\frac{v'+v}{2v}}. \qquad (86)$$

Therefore, (64) follows from the combination of (61), (86), and Proposition 3.

REFERENCES

[1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

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

[3] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, Apr. 2006.

[4] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun. 2014, pp. 601–605.

[5] 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, Jun. 2016.

[6] H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Covert communication in the presence of an uninformed, informed, and coordinated jammer," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Espoo, Finland, Jun./Jul. 2022, pp. 306–311.

[7] B. A. 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.

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

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

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

[11] S. Yan, Y. Cong, S. V. Hanly, and X. Zhou, "Gaussian signalling for covert communications," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3542–3553, Jul. 2019.

[12] L. Wang, "On Gaussian covert communication in continuous time," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 283, Dec. 2019.

[13] Q. E. Zhang, M. R. Bloch, M. Bakshi, and S. Jaggi, "Undetectable radios: Covert communication under spectral mask constraints," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 992–996.

[14] A. Abdelaziz and C. E. Koksal, "Fundamental limits of covert communication over MIMO AWGN channel," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Las Vegas, NV, USA, Oct. 2017, pp. 1–9.

[15] 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.

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

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

[18] R. Soltani, D. Goeckel, D. Towsley, B. A. Bash, and S. Guha, "Covert wireless communication with artificial noise generation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7252–7267, Nov. 2018.

[19] 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.

[20] J. Hu, S. Yan, X. Zhou, F. Shu, and J. Li, "Covert wireless communications with channel inversion power control in Rayleigh fading," Aug. 2018, *arXiv:1803.07812*.

[21] O. Shmuel, A. Cohen, O. Gurewitz, and A. Cohen, "Multi-antenna jamming in covert communication," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Paris, France, Jul. 2019, pp. 987–991.

[22] W. Xiong, Y. Yao, X. Fu, and S. Li, "Covert communication with cognitive jammer," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1753–1757, Oct. 2020.

[23] P. Minero, M. Franceschetti, and D. N. C. Tse, "Random access: An information-theoretic perspective," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 909–930, Feb. 2012.

[24] R. C. Yavas, V. Kostina, and M. Effros, "Random access channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 67, no. 4, pp. 2115–2140, Apr. 2021.

[25] R. C. Yavas, V. Kostina, and M. Effros, "Gaussian multiple and random access channels: Finite-blocklength analysis," *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 6983–7009, Nov. 2021.

[26] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access Gaussian channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2528–2532.

[27] M. Ebrahimi, F. Lahouti, and V. Kostina, "Coded random access design for constrained outage," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2732–2736.

[28] Á. Vázquez-Castro, "Asymptotically guaranteed anti-jamming spread spectrum random access without pre-shared secret," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 332–343, 2022.

[29] K. S. K. Arumugam and M. R. Bloch, "Covert communication over a K-user multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7020–7044, Nov. 2019.

[30] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[31] M. Hayashi, "Universal channel coding for general output alphabet," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 302–321, Jan. 2019.

[32] V. Bally, L. Caramellino, and G. Poly, "Convergence in distribution norms in the CLT for non identical distributed random variables," *Electron. J. Probab.*, vol. 23, pp. 1–51, Jan. 2018.

[33] S. G. Bobkov, G. P. Chistyakov, and F. Götze, "Rényi divergence and the central limit theorem," *Ann. Probab.* vol. 47, no. 1, pp. 270–323, Jan. 2019.

[34] F. Barthe and B. Klartag, "Spectral gaps, symmetries and log-concave perturbations," 2019, *arXiv:1907.01823*.

[35] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, "On the rate of convergence in the entropic central limit theorem," *Probab. Theory Rel. Fields*, vol. 129, no. 3, pp. 381–390, Jul. 2004.

[36] S. Bobkov and M. Ledoux, "Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution," *Probab. Theory Rel. Fields*, vol. 107, no. 3, pp. 383–400, 1997.

[37] S. G. Bobkov, "Isoperimetric and analytic Inequal. For log-concave probability measures," *Ann. Probab.*, vol. 27, no. 4, pp. 1903–1921, Oct. 1999.

[38] L. Sun, T. Xu, S. Yan, J. Hu, X. Yu, and F. Shu, "On resource allocation in covert wireless communication with channel estimation," *IEEE Trans. Commun.*, vol. 68, no. 10, pp. 6456–6469, Oct. 2020.

[39] M. Hayashi and R. Matsumoto, "Secure multiplex coding with dependent and non-uniform multiple messages," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2355–2409, May 2016.

[40] M. Hayashi, *Quantum Information Theory: Mathematical Foundation* (Graduate Texts in Physics). Berlin, Germany: Springer, 2017.

[41] M. Ledoux, "Concentration of measure and logarithmic Sobolev inequalities," in *Séminaire de Probabilités XXXIII* (Lecture Notes in Mathematics), vol. 1709, J. Azéma, M. Émery, M. Ledoux, and M. Yor, Eds. Berlin, Germany: Springer, 1999, pp. 120–216.

[42] L. Gross, "Logarithmic Sobolev inequalities and contractivity properties of semigroups," in *Dirichlet Forms* (Lecture Notes in Mathematics), vol. 1563, G. Dell'Antonio and U. Mosco, Eds. Berlin, Germany: Springer, 1993, pp. 54–88.

Masahito Hayashi (Fellow, IEEE) was born in Japan in 1971. He received the B.S. degree from the Faculty of Sciences, Kyoto University, Japan, in 1994, and the M.S. and Ph.D. degrees in mathematics from Kyoto University in 1996 and 1999, respectively.

He was with the Japan Society of the Promotion of Science (JSPS), Kyoto University, as a Research Fellow from 1998 to 2000 and the Laboratory for Mathematical Neuroscience, Brain Science Institute, RIKEN, as a Researcher from 2000 to 2003. He was with the ERATO Quantum Computation and Information Project, Japan Science and Technology Agency (JST), as the Research Head from 2003 to 2006 and the ERATO-SORST Quantum Computation and Information Project, JST, as the Group Leader from 2006 to 2007. He was with the Superrobust Computation Project Information Science and Technology Strategic Core (21st Century COE by MEXT) Graduate School of Information Science and Technology, The University of Tokyo, as an Adjunct Associate Professor from 2004 to 2007. He was with the Graduate School of Information Sciences, Tohoku University, as an Associate Professor from 2007 to 2012. In 2012, he joined the Graduate School of Mathematics, Nagoya University, as a Full Professor. He was with the Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology, Shenzhen, China, as a Chief Research Scientist from 2020 to 2023. In 2023, he joined the School of Data Science, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen) as a Full Professor and also joined the International Quantum Academy (SIQA) as a Chief Research Scientist. Also, he was with the Centre for Quantum Technologies, National University of Singapore, as a Visiting Research Associate Professor from 2009 to 2012 and has been a Visiting Research Professor since 2012. He was with the Center for Advanced Intelligence Project, RIKEN, as a Visiting Scientist from 2017 to 2020. He was with the Shenzhen Institute for Quantum Science and Engineering, Southern University of Science and Technology as a Visiting Professor from 2018 to 2020 and the Peng Cheng Laboratory, Center for Quantum Computing, Shenzhen, as a Visiting Professor from 2019 to 2020. In 2006, he has published the book *Quantum Information: An Introduction* (Springer), whose revised version was published as *Quantum Information Theory: Mathematical Foundation* from *Graduate Texts in Physics* (Springer) in 2016. In 2016, he has published other two books *Group Representation for Quantum Theory* and *A Group Theoretic Approach to Quantum Information* (Springer). His research interests include classical and quantum information theory and classical and quantum statistical inference. In 2022, he was elected as a fellow of IMS and an Asia–Pacific Artificial Intelligence Association (AAIA). In 2011, he received Information Theory Society Paper Award (2011) for "Information-Spectrum Approach to Second-Order Coding Rate in Channel Coding." In 2016, he received the Japan Academy Medal from the Japan Academy and the JSPS Prize from Japan Society for the Promotion of Science. He is on the editorial board of *New Journal of Physics* and *International Journal of Quantum Information*.

Ángeles Vázquez-Castro (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in telecommunications engineering from Vigo University, Spain, in 1994 and 1998, respectively, and the M.Sc. degree in analytical philosophy from the University of Barcelona, Spain, in 2014. Until 2002, she was with University Carlos III University, Madrid, and University of Southern California at Los Angeles as an Assistant Professor and a Visiting Professor, respectively. From 2002 to 2004, she held a research fellow position with the European Space Agency to work with the Space Research and Technology Centre, Noordwijk, The Netherlands. From 2008 to 2012, she was a part-time Visiting Professor with the University of Oslo, Norway. Since 2004, she has been an Associate Professor with Universitat Autònoma de Barcelona, Spain. Since 2011, she has been a Researcher with the Centre for Space Research (CERES), Institut d'Estudis Espacials de Catalunya (IEEC-UAB). Her results have been used for the development of satellite communication standards such as DVB-SH (channel models), DVB-RCS (coding), DVB-S2 (optimization of adaptive coding and modulation), or ITU-R (integrated terrestrial-satellite architectures). She has authored or coauthored more than 150 peer-reviewed scientific papers, 13 book chapters, one co-edited book on *Network Coding and Subspace Designs* (Springer, 2018), and is on the editorial board of the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY. She holds three patents, two for classical (multibeam) space communication, and one (licensed) for quantum (secure) space communication. Her research covers classical and quantum communication and information theory applied to space systems design. Her current research interests include quantum-enabled information security protocols design and their optimization for space systems. She is a Founder/member of the European Satellite Network Communication Experts (SatNEx) currently funded by the European Space Agency.