



# On Gaussian covert communication in continuous time

Ligong Wang

► **To cite this version:**

Ligong Wang. On Gaussian covert communication in continuous time. EURASIP Journal on Wireless Communications and Networking, 2019, 2019, 10.1186/s13638-019-1609-8 . hal-02467493

**HAL Id: hal-02467493**

**<https://hal.archives-ouvertes.fr/hal-02467493>**

Submitted on 5 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH

Open Access



# On Gaussian covert communication in continuous time

Ligong Wang

## Abstract

The paper studies covert communication over a continuous-time Gaussian channel. The covertness condition requires that the channel output must statistically resemble pure noise. When the additive Gaussian noise is “white” over the bandwidth of interest, a formal coding theorem is proven, extending earlier results on covert Gaussian communication in discrete time. This coding theorem is applied to study scenarios where the input bandwidth can be infinite and where positive or even infinite per-second rates may be achievable.

**Keywords:** Covert communication, Low probability of detection, Gaussian channel, Continuous time, Waveform channel, Prolate spheroidal wave functions

## 1 Introduction

Covert communication, or communication with low probability of detection [1–4], refers to scenarios where the transmitter and the receiver must keep a warden from discovering the fact that they are using the channel to communicate. On an additive white Gaussian noise (AWGN) channel, this means that the warden’s observation should be statistically close to pure noise. It was first shown in [1] that the AWGN channel obeys the so-called square-root law for covert communication: the number of information nats that can be communicated covertly over the channel can only grow proportionally to the square root of the total number of channel uses. The exact scaling law, when covertness is measured in terms of relative entropy, was determined in [3]. Similar results have been obtained for the binary symmetric channel [2] and general discrete memoryless channels [3, 4]. A number of further works have extended these results in several directions. Among them, some also consider total variation distance in place of relative entropy as the measure for covertness [5, 6].

A discrete-time AWGN channel is usually used to model a continuous-time communication channel with a bandwidth constraint on its input waveform, corrupted by Gaussian noise that is “white” with respect to that bandwidth. Using the sampling theorem, such a

continuous-time channel with bandwidth  $W$  Hz over the time interval  $[0, T]$  is approximately equivalent to  $2WT$  uses of a discrete-time AWGN channel [7]. Hence, one can roughly say (as in, e.g., a brief remark in [1]) that the number of nats that can be covertly communicated over this continuous-time channel is proportional to  $\sqrt{WT}$ .

In this paper, we first provide a rigorous mathematical framework to study covert communication over the continuous-time channel with AWGN over the bandwidth of interest. Formal treatment of continuous-time Gaussian channels is nontrivial because, in short, no nonzero signal can be both band-limited and time-limited (see, e.g., [8], Theorem 6.8.2). Indeed, Shannon’s capacity formula for the band-limited Gaussian channel [7] called for several follow-up works to acquire a clear physical meaning; see [9, 10] and references therein. In this paper, we adopt a model proposed by Wyner [9] where the input is required to be “strictly band-limited and approximately time-limited” and introduce a covertness constraint to that model.

There are some important technicalities in formulating the continuous-time model. In particular, we find it important to let the warden observe the output waveform over the entire real line. Indeed, even when the transmitted signal is strictly time-limited, one should *not* assume that the warden can only observe the channel output within the time period that the input occupies. We demonstrate this by constructing a slightly different model and proving a pathological result under that model.

Correspondence: [ligong.wang@ensea.fr](mailto:ligong.wang@ensea.fr)  
ETIS—Université Paris Seine, Université de Cergy-Pontoise, ENSEA, CNRS, Cergy-Pontoise, France

Under the proposed framework, we prove that the maximum number of nats that can be covertly communicated over  $T$  seconds and bandwidth  $W$  Hz is indeed proportional to  $\sqrt{WT}$ . Additionally, we show that binary phase shift keying (BPSK) is optimal in the sense that it achieves the dominant term in this maximum number of nats. The latter is related to the fact that BPSK can achieve the capacity per unit cost on the AWGN channel [11], i.e., BPSK is asymptotically optimal in the limit where the signal-to-noise ratio goes to zero.

Using the continuous-time result, we then investigate the regime where  $W$  is infinity or grows to infinity together with  $T$ . Our intention is to capture engineering insights to scenarios where information is transmitted over a large bandwidth and a relatively short time, such as in “spread-spectrum” communication [12]. We prove that, if  $W$  is infinity or grows large fast enough compared to  $T$ , then covert communication can have positive rates in nats per second. This is not surprising, since we already argued that information throughput grows like  $\sqrt{WT}$ . Additionally, we show that, if available bandwidth grows large fast enough, then, under the same average-power constraint on the input, the covert communication capacity is the same as the capacity without covertness constraint. Note, however, that traditionally, the capacity of the “infinite-bandwidth” AWGN channel is computed by bringing  $W$  to infinity *after* letting  $T \rightarrow \infty$ .

Our framework only applies to the case where the Gaussian noise is white over the bandwidth of interest. When the noise is colored, we are not able to prove a rigorous coding theorem. Instead, we use known formulas to calculate relative entropy and mutual information [13], which lead us to some conjectures.

The infinite-bandwidth results are related to our recent work [14], which shows that the continuous-time Poisson channel with neither bandwidth nor peak-power constraint permits transmission of infinitely many information nats per second. However, a close look at the two works reveals fundamental differences between the two channels. For example, in the Poisson case, a constraint on the average input power has no effect on covert communication, whereas in the Gaussian case, it generally does.

The recent work [15] also treats covert communication over a continuous-time Gaussian channel. It adopts a different model from the current paper: the transmitted signal is required to be strictly time-limited while satisfying some “spectral mask” constraints in the frequency domain. Under these constraints, the authors of [15] show that  $\sqrt{WT}$  growth of covert information is achievable using raised-cosine modulation. However, they do not prove any continuous-time converse result under that model.

The rest of this paper is arranged as follows. After introducing some notation, we formulate and solve the continuous-time covert communication problem in Section 2. We then study the infinite-bandwidth scenario in Section 3. We propose a slightly different model and show its deficiency in Section 4. The paper is concluded with some discussion in Section 5.

### 1.1 Some notation

We use uppercase letters like  $X$  to denote random variables, and corresponding lowercase letters like  $x$  to denote their realizations. We write a random real function on the interval  $[a, b]$  as  $X(t)$ ,  $t \in [a, b]$ ; sometimes we also use the shorthand notation  $X_a^b$ , where  $a$  and  $b$  might be replaced with  $-\infty$  or  $\infty$ , respectively. When the domain of the function is clear from context, we may further shorten it as  $X(\cdot)$ . To denote the realization of a random function, we replace  $X$  in the above by  $x$ . A vector  $(X_i, X_{i+1}, \dots, X_j)$  is written as  $X_i^j$ , where  $j$  may be replaced by  $\infty$ .

We slightly deviate from standard notation to express the relative entropy in terms of two random variables (as opposed to in terms of their distributions). For example,

$$\mathcal{D}(Y_{-\infty}^{\infty} \parallel Z_{-\infty}^{\infty})$$

denotes the relative entropy between the distributions of  $Y(t)$ ,  $t \in (-\infty, \infty)$ , and  $Z(t)$ ,  $t \in (-\infty, \infty)$ , respectively. Mutual information between two continuous-time random processes or continuous-valued random variables is written, for example, like

$$I(X_{-\infty}^{\infty}; Y_{-\infty}^{\infty}).$$

For definition of relative entropy and mutual information for general probability distributions, we refer to [13].

We use  $W$  to denote the bandwidth that the input signal can employ, and  $T$  to denote the total time of communication. We shall often study the limit where the product  $WT \rightarrow \infty$ . Here,  $W$  and  $T$  can be functions of each other, or be such that  $W$  is fixed while  $T \rightarrow \infty$ , or vice versa. Further, we use  $\delta$  to denote the covertness parameter, which can be a positive constant, or a positive function of  $WT$  satisfying

$$\lim_{WT \rightarrow \infty} WT\delta = \infty, \quad (1)$$

$$\lim_{WT \rightarrow \infty} \frac{\delta}{WT} = 0. \quad (2)$$

We use  $o(a)$  to denote a “higher-order term” than  $a$ , in the sense that the ratio  $o(a)/a$  approaches zero in the limit where  $a \downarrow 0$ ; this limit usually coincides with  $WT \rightarrow \infty$ .

Finally, information is measured in nats, and  $\log$  denotes the natural logarithm.

## 2 A formal continuous-time treatment

Consider the continuous-time channel described by

$$Y(t) = X(t) + Z(t), \quad t \in \mathbb{R}, \quad (3)$$

where  $X(\cdot)$  is the (potentially random) channel input sent by the transmitter, which we require to be square-integrable: with probability one,

$$\int_{-\infty}^{\infty} X(t)^2 dt < \infty; \quad (4)$$

$Y(\cdot)$  is the channel output observed by both the intended receiver and the warden; and  $Z(\cdot)$  is the additive noise to the channel, which we assume to be generated randomly according to a zero-mean stationary Gaussian process and independent of  $X(\cdot)$ .

A codebook is specified by an encoder, which is a mapping from a message  $m$  taken from the message set  $\mathcal{M}$  to an input waveform  $x(\cdot)$ , and a decoder, which is a mapping from an output waveform  $y(\cdot)$  to the decoded message  $\hat{m} \in \mathcal{M}$ . We allow the transmitter and the receiver to use a random codebook: the distribution according to which the codebook is drawn is known to the warden, but the realization of the codebook is not.

We require that the input waveform be “strictly band-limited to  $[-W, W]$  and approximately time-limited to  $[0, T]$ ” in the sense of Wyner [9]. Formally, for every message  $m \in \mathcal{M}$ :

1. The Fourier Transform of  $x(\cdot)$ , which is given by

$$f \mapsto \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft} dt, \quad f \in \mathbb{R}, \quad (5)$$

must equal zero for all  $f \notin [-W, W]$ ;

2. The ratio

$$\frac{\int_0^T |x(t)|^2 dt}{\int_{-\infty}^{\infty} |x(t)|^2 dt} \geq 1 - \eta \quad (6)$$

for some  $\eta \in (0, 1)$ ;

3. The receiver maps  $y(t)$ ,  $t \in (-\infty, \infty)$ , to a decoded message; and
4. For some given  $\delta > 0$ , the following covertness condition must be satisfied:

$$\mathcal{D}(Y_{-\infty}^{\infty} \| Z_{-\infty}^{\infty}) \leq \delta. \quad (7)$$

Let  $M(W, T, \epsilon, \eta, \delta)$  be the largest possible value of  $|\mathcal{M}|$  such that the above conditions are satisfied and that the average probability of a decoding error for a uniformly chosen message is less than or equal to  $\epsilon$ .

In the rest of this section, we assume that  $Z(\cdot)$  has power spectral density (PSD)  $N(\cdot)$  that is constant within the bandwidth of interest:

$$N(f) = \frac{N_0}{2}, \quad |f| \leq W. \quad (8)$$

**Remark 1** In some parts of this paper, e.g., the next theorem,  $W$  is allowed to grow to infinity. There is no single noise process to satisfy (8) for every finite  $W$ , because a random process having a PSD that is constant over the entire real line does not exist. In cases where  $W \rightarrow \infty$ , our setting should be understood in such a way that the noise process “adapts” itself with our choice of input bandwidth  $W$ . Although this formulation (that the noise process adapts to the input signal) has no physical meaning, it serves as a mathematically valid route to study the limit where  $W \rightarrow \infty$ , providing engineering insights to wideband scenarios.

The next theorem extends Theorem 5 of [3] to the continuous-time setting. Furthermore, it shows that BPSK is optimal up to the dominant term in total throughput. Here, by BPSK, we mean a signaling scheme where the symbols take values in the set  $\{-a, a\}$  for some constant  $a$ . It is clear from the proof in Section 2.4 that BPSK is also optimal in achieving the discrete-time result ([3], Theorem 5)<sup>1</sup>.

**Theorem 1** If, for every finite  $W$  that we consider, the additive Gaussian noise has PSD satisfying (8), then, irrespectively of the value of  $N_0$ ,

$$\lim_{\eta \downarrow 0} \lim_{\epsilon \downarrow 0} \lim_{W T \rightarrow \infty} \frac{\log M(W, T, \epsilon, \eta, \delta)}{\sqrt{2WT\delta}} = 1. \quad (9)$$

Furthermore, the limit (9) can be achieved by modulating a set of waveforms with BPSK.

*Proof* See Sections 2.3 and 2.4.  $\square$

The next corollary follows immediately from Theorem 1.

**Corollary 1** If  $W$  is a constant that does not depend on  $T$ , then

$$\lim_{\eta \downarrow 0} \lim_{\epsilon \downarrow 0} \lim_{T \rightarrow \infty} \frac{\log M(W, T, \epsilon, \eta, \delta)}{\sqrt{T\delta}} = \sqrt{2W}. \quad (10)$$

Hence, not surprisingly, when available bandwidth is fixed, the amount of information that can be covertly communicated over the continuous-time Gaussian channel is approximately proportional to  $\sqrt{T\delta}$ .

In the rest of this section, we first review some mathematical tools (Section 2.1), which will allow us to reduce the continuous-time channel to a discrete-time one (Section 2.2). We then prove the converse (Section 2.3) and direct (Section 2.4) parts of Theorem 1. Some intermediate steps in the proof will be borrowed from [3].

<sup>1</sup>Optimality of BPSK in the discrete-time case was first observed by the author in 2015. That result has circulated among some researchers as private communication.

## 2.1 Prolate spheroidal wave functions

We provide a very brief introduction to the prolate spheroidal wave functions (PSWFs), which are powerful tools in formal treatment of continuous-time Gaussian channels; for more details, we refer the reader to [9, 10, 16, 17] and references therein.

Given any  $W, T > 0$ , there exists a countably infinite set of positive real numbers  $\{\lambda_i\}$  satisfying

$$1 > \lambda_1 > \lambda_2 > \dots \quad (11)$$

and a corresponding set of real functions  $\{\psi_i: \mathbb{R} \rightarrow \mathbb{R}\}_{i=1}^{\infty}$  such that the following properties are satisfied:

- Each  $\psi_i(\cdot)$  is band-limited to  $W$  Hz. Further, the functions  $\{\psi_i(\cdot)\}$  are orthonormal on  $\mathbb{R}$ , and complete in the space of square-integrable functions that are band-limited to  $W$  Hz;
- The restrictions of  $\{\psi_i(\cdot)\}$  to the time interval  $[0, T]$  are orthogonal:

$$\int_0^T \psi_i(t) \psi_j(t) dt = \begin{cases} \lambda_i, & i = j, \\ 0, & i \neq j. \end{cases} \quad (12)$$

Further, the restrictions of  $\{\psi_i\}$  to  $[0, T]$  are complete in the space of square-integrable functions on  $[0, T]$ .

It was shown by Slepian [17] that the coefficients  $\{\lambda_i\}$  above satisfy the following: for any  $\alpha \in (0, 1)$  (not dependent on  $W$  and  $T$ ), as  $WT \rightarrow \infty$ ,

$$\lambda_{2(1-\alpha)WT} \rightarrow 1 \quad (13)$$

$$\lambda_{2(1+\alpha)WT} \rightarrow 0. \quad (14)$$

Finally, for a zero-mean stationary Gaussian process  $Z(\cdot)$  with PSD

$$N(f) = \begin{cases} \frac{N_0}{2}, & |f| \leq W \\ 0, & |f| > W, \end{cases} \quad (15)$$

we have the following Karhunen-Loève expansion in terms of the above PSWFs:

$$Z(t) = \sum_{i=1}^{\infty} \hat{Z}_i \psi_i(t), \quad t \in [0, T], \quad (16)$$

where  $\{\hat{Z}_i\}$  are independent and identically distributed (IID) Gaussian random variables of mean zero and variance  $N_0/2$ .

## 2.2 Reduction to discrete time

Recall Condition 1 that the input signal  $x(t)$ ,  $t \in \mathbb{R}$ , must be band-limited to  $W$  Hz. Since the set  $\{\psi_i(\cdot)\}$  is complete in the space of square-integrable functions that are band-limited to  $W$  Hz, we can write  $X(\cdot)$  as

$$X(t) = \sum_{i=1}^{\infty} \hat{X}_i \psi_i(t), \quad t \in \mathbb{R} \quad (17)$$

for an infinite sequence of random variables  $\hat{X}_1, \hat{X}_2, \dots$ . Furthermore, the output  $Y(\cdot)$  can be passed through an ideal low-pass filter of cut-off frequency  $W$ . Doing so does not affect  $X(\cdot)$ , but will change the PSD of  $Z(\cdot)$  to the one given by (15), so that the resulting noise process will satisfy the expansion (16). We can then decompose the continuous-time channel into an infinite sequence of parallel discrete-time channels:

$$\hat{Y}_i = \hat{X}_i + \hat{Z}_i, \quad i = 1, 2, \dots, \quad (18)$$

where  $\hat{Z}_1, \hat{Z}_2, \dots$  are IID zero-mean Gaussian of variance  $N_0/2$ . One can see that the above reduction is optimal for both the receiver (i.e., decoding) and for the warden (i.e., detection of communication). We can hence base both the converse and the direct parts of the proof of Theorem 1 on the channel (18).

## 2.3 Converse part of Theorem 1

Consider a random code for the message set  $\{1, \dots, M\}$  that satisfies all aforementioned conditions. By a standard argument using Fano's inequality and the chain rule of mutual information, we have

$$(1 - \epsilon) \log M - 1 \leq \sum_{i=1}^{\infty} I(\hat{X}_i; \hat{Y}_i). \quad (19)$$

Further, noting that, under a second-moment constraint, the Gaussian input maximizes mutual information over a Gaussian channel [18], we have, for every  $i$ ,

$$I(\hat{X}_i; \hat{Y}_i) \leq \frac{1}{2} \log \left( 1 + \frac{2\rho_i}{N_0} \right) \leq \frac{\rho_i}{N_0}, \quad (20)$$

where we defined

$$\rho_i \triangleq \mathbb{E} \left[ |\hat{X}_i|^2 \right], \quad i = 1, 2, \dots \quad (21)$$

(The expectation is computed over the possibly random codebook and a uniformly chosen message.) Combining (19) and (20) yields

$$(1 - \epsilon) \log M - 1 \leq \frac{1}{N_0} \sum_{i=1}^{\infty} \rho_i. \quad (22)$$

To bound the right-hand side of (22), consider the covertness requirement. Recall that the processing on  $Y_{-\infty}^{\infty}$  described in Section 2.2 maps it to  $\hat{Y}_1^{\infty}$ . Clearly, it also maps  $Z_{-\infty}^{\infty}$  to  $\hat{Z}_1^{\infty}$ . Also recall that this reduction is optimal for the warden. Hence

$$\mathcal{D}(Y_{-\infty}^{\infty} \| Z_{-\infty}^{\infty}) = \mathcal{D}(\hat{Y}_1^{\infty} \| \hat{Z}_1^{\infty}). \quad (23)$$

It then follows by Condition 4 that

$$\delta \geq \mathcal{D}(\hat{Y}_1^{\infty} \| \hat{Z}_1^{\infty}) \geq \sum_{i=1}^{\infty} \mathcal{D}(\hat{Y}_i \| \hat{Z}_i), \quad (24)$$

where the second inequality follows by the same steps as Eq. (13) of [3]. For a fixed second moment, the relative

entropy on the right-hand side above is maximized by  $\hat{X}_i$  being zero-mean Gaussian ([3], Eq. (74)):

$$\mathcal{D}(\hat{Y}_i \| \hat{Z}_i) \geq \frac{\rho_i}{N_0} - \frac{1}{2} \log \left( 1 + \frac{2\rho_i}{N_0} \right). \quad (25)$$

Fix any  $\alpha \in (0, 1)$ . We combine (24) and (25), drop all summands with  $i > \lfloor 2(1+\alpha)WT \rfloor$  (note that they are non-negative), and use the convexity of the right-hand side of (25) in  $\rho_i$  to obtain

$$\begin{aligned} \delta &\geq \sum_{i=1}^{\lfloor 2(1+\alpha)WT \rfloor} \frac{\rho_i}{N_0} - \frac{1}{2} \log \left( 1 + \frac{2\rho_i}{N_0} \right) \\ &\geq \lfloor 2(1+\alpha)WT \rfloor \left( \frac{\bar{\rho}}{N_0} - \frac{1}{2} \log \left( 1 + \frac{2\bar{\rho}}{N_0} \right) \right), \end{aligned} \quad (26)$$

where we defined

$$\bar{\rho} \triangleq \frac{1}{\lfloor 2(1+\alpha)WT \rfloor} \sum_{i=1}^{\lfloor 2(1+\alpha)WT \rfloor} \rho_i. \quad (27)$$

Recall (2), which together with (26) implies that  $\bar{\rho}$  must tend to zero as  $WT \rightarrow \infty$ . Furthermore, by applying the Taylor expansion

$$\log \left( 1 + \frac{2\bar{\rho}}{N_0} \right) = \frac{2\bar{\rho}}{N_0} - \frac{2\bar{\rho}^2}{N_0^2} + o(\bar{\rho}^2) \quad (28)$$

to (26), we see that  $\bar{\rho}$  must satisfy

$$\bar{\rho} \leq \sqrt{\frac{\delta}{\lfloor 2(1+\alpha)WT \rfloor}} \cdot N_0 + o\left(\sqrt{\frac{\delta}{WT}}\right). \quad (29)$$

Next consider Condition 2. Since the functions  $\psi_i(\cdot)$  have unit energy and satisfy (12), Condition 2 requires

$$\frac{\sum_{i=1}^{\infty} \lambda_i \rho_i}{\sum_{i=1}^{\infty} \rho_i} \geq 1 - \eta. \quad (30)$$

Further using (14), we have the following requirement:

$$\begin{aligned} 1 - \eta &\leq \lim_{WT \rightarrow \infty} \frac{\sum_{i=1}^{\lfloor 2(1+\alpha)WT \rfloor} \rho_i}{\sum_{i=1}^{\infty} \rho_i} \\ &= \lim_{WT \rightarrow \infty} \frac{\lfloor 2(1+\alpha)WT \rfloor \bar{\rho}}{\sum_{i=1}^{\infty} \rho_i}. \end{aligned} \quad (31)$$

Together with (22), this implies

$$\begin{aligned} (1 - \epsilon) &\lim_{WT \rightarrow \infty} \frac{\log M}{\sqrt{2WT\delta}} \\ &\leq \frac{1}{N_0} \cdot \lim_{WT \rightarrow \infty} \frac{\sum_{i=1}^{\infty} \rho_i}{\sqrt{2WT\delta}} \\ &= \frac{1}{N_0} \cdot \lim_{WT \rightarrow \infty} \frac{\sum_{i=1}^{\infty} \rho_i}{\lfloor 2(1+\alpha)WT \rfloor \bar{\rho}} \cdot \frac{\lfloor 2(1+\alpha)WT \rfloor \bar{\rho}}{\sqrt{2WT\delta}} \\ &\leq \frac{1}{(1-\eta)N_0} \lim_{WT \rightarrow \infty} \frac{\lfloor 2(1+\alpha)WT \rfloor \bar{\rho}}{\sqrt{2WT\delta}} \\ &= \frac{1+\alpha}{(1-\eta)N_0} \lim_{WT \rightarrow \infty} \sqrt{\frac{2WT}{\delta}} \cdot \bar{\rho} \\ &\leq \frac{1+\alpha}{1-\eta} \lim_{WT \rightarrow \infty} \sqrt{\frac{2WT}{\delta}} \sqrt{\frac{\delta}{\lfloor 2(1+\alpha)WT \rfloor}} \\ &= \frac{\sqrt{1+\alpha}}{1-\eta}, \end{aligned} \quad (32)$$

where the second-to-last line follows by (29). Letting  $\epsilon$ ,  $\eta$ , and  $\alpha$  go to zero in the above yields the desired converse result.

#### 2.4 Direct part of Theorem 1

Fix some  $\alpha, \gamma \in (0, 1)$ , both of which will be chosen to be close to zero later. We use the first  $\lfloor 2(1-\alpha)WT \rfloor$  channels in (18) to communicate and discard the remaining channels. On these channels, we generate a random codebook by picking every entry in every codeword independently and equally likely to be  $a$  or  $-a$ , where

$$a \triangleq (1-\gamma)\sqrt{N_0} \cdot \left( \frac{\delta}{\lfloor 2(1-\alpha)WT \rfloor} \right)^{\frac{1}{4}}. \quad (33)$$

(Every  $\hat{X}_i$  with  $i > \lfloor 2(1-\alpha)WT \rfloor$  is chosen to be zero with probability one.) The bandwidth constraint is obviously satisfied. The time-limit constraint is also satisfied for any  $\eta > 0$  when  $WT$  is sufficiently large, by virtue of (13). For covertness, we have

$$\begin{aligned} \mathcal{D}(Y_{-\infty}^{\infty} \| Z_{-\infty}^{\infty}) &= \mathcal{D}(\hat{Y}_1^{\lfloor 2(1-\alpha)WT \rfloor} \| \hat{Z}_1^{\lfloor 2(1-\alpha)WT \rfloor}) \\ &= \sum_{i=1}^{\lfloor 2(1-\alpha)WT \rfloor} \mathcal{D}(\hat{Y}_i \| \hat{Z}_i) \\ &= \lfloor 2(1-\alpha)WT \rfloor \cdot \mathcal{D}(\hat{Y} \| \hat{Z}), \end{aligned} \quad (34)$$

where we dropped the subscript in the last line, as all pairs  $(\hat{X}_i, \hat{Y}_i)$ ,  $i = 1, \dots, \lfloor 2(1-\alpha)WT \rfloor$ , have the same joint distribution. Note that, by our choice of  $\hat{X}$ , the random variable  $\hat{Y}$  has the following probability density function (PDF):

$$f(y) = \frac{1}{\pi N_0} \left( \frac{1}{2} e^{-\frac{(y-a)^2}{N_0}} + \frac{1}{2} e^{-\frac{(y+a)^2}{N_0}} \right), \quad y \in \mathbb{R}. \quad (35)$$

The PDF of  $\hat{Z}$  is that of the Gaussian:

$$g(z) = \frac{1}{\pi N_0} e^{-\frac{z^2}{N_0}}, \quad z \in \mathbb{R}. \quad (36)$$

We hence have

$$\begin{aligned} \mathcal{D}(\hat{Y} \parallel \hat{Z}) &= \int_{-\infty}^{\infty} f(y) \log \frac{\frac{1}{2} e^{-\frac{(y-a)^2}{N_0}} + \frac{1}{2} e^{-\frac{(y+a)^2}{N_0}}}{e^{-\frac{y^2}{N_0}}} \\ &= -\frac{a^2}{N_0} + \int_{-\infty}^{\infty} f(y) \log \left( \frac{1}{2} e^{\frac{2ay}{N_0}} + \frac{1}{2} e^{-\frac{2ay}{N_0}} \right). \end{aligned} \quad (37)$$

For the integrand above, we have the following upper bound:

$$\log \left( \frac{1}{2} e^{\frac{2ay}{N_0}} + \frac{1}{2} e^{-\frac{2ay}{N_0}} \right) \leq \frac{2a^2 y^2}{N_0^2} - \frac{4a^4 y^4}{3N_0^4} + \frac{64a^6 y^6}{45N_0^6}. \quad (38)$$

Further note that, by (35), the second moment of  $\hat{Y}$  is  $\left(a^2 + \frac{N_0}{2}\right)$ , the fourth moment is  $\left(a^4 + 3a^2 N_0 + \frac{3N_0^2}{4}\right)$ , and the sixth moment is finite. We can thus continue (37) as

$$\begin{aligned} \mathcal{D}(\hat{Y} \parallel \hat{Z}) &\leq -\frac{a^2}{N_0} + \frac{2a^2}{N_0^2} \left(a^2 + \frac{N_0}{2}\right) \\ &\quad - \frac{4a^4}{3N_0^4} \left(a^4 + 3a^2 N_0 + \frac{3N_0^2}{4}\right) + o(a^4) \\ &= \frac{a^4}{N_0^2} + o(a^4) \\ &= (1-\gamma)^4 \cdot \frac{\delta}{[2(1-\alpha)WT]} + o\left(\frac{\delta}{WT}\right). \end{aligned} \quad (39)$$

Combining (34) and (39), we get

$$\mathcal{D}(Y_{-\infty}^{\infty} \parallel Z_{-\infty}^{\infty}) \leq (1-\gamma)^4 \delta + o(\delta), \quad (40)$$

which is smaller than  $\delta$  for large enough  $WT$ .

We now analyze the maximum possible values for  $M$  for which the decoder can decode correctly. To this end, like [3], we use an information-spectrum result [19, 20], which guarantees that a sequence of codes can have vanishing error probability provided

$$\begin{aligned} &\lim_{WT \rightarrow \infty} \frac{\log M}{\sqrt{2WT\delta}} \\ &\leq p\text{-}\liminf_{WT \rightarrow \infty} \frac{1}{\sqrt{2WT\delta}} \log \frac{p(\hat{Y}_1^n | \hat{X}_1^n)}{p(\hat{Y}_1^n)} \\ &= p\text{-}\liminf_{WT \rightarrow \infty} \frac{[2(1-\alpha)WT]}{\sqrt{2WT\delta}} \log \frac{p(\hat{Y} | \hat{X})}{p(\hat{Y})}, \end{aligned} \quad (41)$$

where  $p\text{-}\liminf$  denotes the limit inferior in probability, namely, the largest number such that the probability that the random variable in consideration is greater than this number tends to one in the limit, and where we slightly

abuse notation to use  $p(\cdot)$  and  $p(\cdot|\cdot)$  to denote PDF and conditional PDF of the corresponding random variables or random vectors. Recall that  $p(\hat{Y}) = f(\hat{Y})$ , where  $f(\cdot)$  is given by (35), while given  $\hat{X} = x$ ,  $\hat{Y}$  is Gaussian with mean  $x$  and variance  $\frac{N_0}{2}$ . Also recall that  $x$  equals either  $a$  or  $-a$ . We hence have

$$\begin{aligned} \frac{p(\hat{Y} | \hat{X})}{p(\hat{Y})} &= \frac{e^{-\frac{(\hat{Y}-\hat{X})^2}{N_0}}}{\frac{1}{2} e^{-\frac{(\hat{Y}-a)^2}{N_0}} + \frac{1}{2} e^{-\frac{(\hat{Y}+a)^2}{N_0}}} \\ &= \frac{e^{\frac{2\hat{X}\hat{Y}}{N_0}}}{\frac{1}{2} e^{\frac{2a\hat{Y}}{N_0}} + \frac{1}{2} e^{-\frac{2a\hat{Y}}{N_0}}}. \end{aligned} \quad (42)$$

Using

$$\log \left( \frac{1}{2} e^{\frac{2a\hat{Y}}{N_0}} + \frac{1}{2} e^{-\frac{2a\hat{Y}}{N_0}} \right) \leq \frac{2a^2 \hat{Y}^2}{N_0^2}, \quad (43)$$

we obtain

$$\log \frac{p(\hat{Y} | \hat{X})}{p(\hat{Y})} \geq \frac{2}{N_0} \hat{X} \hat{Y} - \frac{2a^2}{N_0^2} \hat{Y}^2. \quad (44)$$

Noting that

$$\mathbb{E}[\hat{X} \hat{Y}] = a^2 \quad (45)$$

$$\mathbb{E}[\hat{Y}^2] = a^2 + \frac{N_0}{2}, \quad (46)$$

we can compute the expected value of the right-hand side of (44) to be

$$\mathbb{E} \left[ \frac{2}{N_0} \hat{X} \hat{Y} - \frac{2a^2}{N_0^2} \hat{Y}^2 \right] = \frac{a^2}{N_0} - \frac{2a^4}{N_0^2}. \quad (47)$$

Similarly, the variance can be shown to be

$$\text{Var} \left[ \frac{2}{N_0} \hat{X} \hat{Y} - \frac{2a^2}{N_0^2} \hat{Y}^2 \right] = \frac{2a^2}{N_0} + o(a^2). \quad (48)$$

Using (44), (47), (48), and Chebyshev's inequality, we obtain

$$\begin{aligned} &p\text{-}\liminf_{WT \rightarrow \infty} \frac{[2(1-\alpha)WT]}{\sqrt{2WT\delta}} \log \frac{p(\hat{Y} | \hat{X})}{p(\hat{Y})} \\ &\geq p\text{-}\liminf_{WT \rightarrow \infty} \frac{[2(1-\alpha)WT]}{\sqrt{2WT\delta}} \left( \frac{2}{N_0} \hat{X} \hat{Y} - \frac{2a^2}{N_0^2} \hat{Y}^2 \right) \\ &= \lim_{WT \rightarrow \infty} \frac{[2(1-\alpha)WT]}{\sqrt{2WT\delta}} \cdot \frac{a^2}{N_0} \\ &= (1-\gamma)^2 \sqrt{1-\alpha}. \end{aligned} \quad (49)$$

Recalling (41), the proof is completed when we bring both  $\gamma$  and  $\alpha$  to zero.

### 3 Exploring infinite bandwidth

In this section, we study cases where available bandwidth grows without bound. Section 3.1 considers the scenario

where the additive white Gaussian noise also has unlimited bandwidth. To this end, we assume the noise PSD is constant over a finite bandwidth  $W$  and then let  $W$  grow to infinity, either together with  $T$  or with  $T$  held fixed; recall Remark 1. Section 3.2 considers the case where the noise PSD is constant within a certain bandwidth and zero elsewhere (while no bandwidth limit is imposed on the input). Both these sections directly use Theorem 1 to obtain the desired results. Finally, Section 3.3 considers colored noise, where our analysis is less rigorous and does not lead to an explicit coding theorem; we state our findings there as “conjectures”.

### 3.1 White noise with unbounded bandwidth

**Corollary 2** Assume that the noise process satisfies (8) for every finite  $W$  that we consider. If the limit

$$\lim_{WT \rightarrow \infty} \frac{W\delta}{T} = c \tag{50}$$

exists, where  $c$  may equal infinity, then the per-second covert communication capacity is

$$\lim_{\eta \downarrow 0} \lim_{\epsilon \downarrow 0} \lim_{WT \rightarrow \infty} \frac{\log M(W, T, \epsilon, \eta, \delta)}{T} = \sqrt{2c}. \tag{51}$$

We note the following. If  $W$  grows large more slowly than  $T$ , then the per-second capacity for covert communication is zero for any finite  $\delta$ , as in the finite-bandwidth case. If  $W$  grows large linearly with  $T$ , then a positive covert communication rate is achievable as long as  $\delta$  is positive and bounded away from zero. If  $W$  grows large faster than  $T$ , i.e., if  $W/T \rightarrow \infty$  as  $WT \rightarrow \infty$ , then there exists some  $\delta$  that tends to zero slowly enough when  $WT \rightarrow \infty$ , for which the largest per-second covert communication rate is infinity.

Next, consider an average-power constraint  $P$  that is imposed on the input: every input signal must satisfy

$$\frac{1}{T} \cdot \int_{-\infty}^{\infty} x(t)^2 dt \leq P. \tag{52}$$

As the next theorem shows, when  $W$  grows sufficiently fast, the covert communication capacity equals the capacity of the infinite-bandwidth Gaussian channel under the same power constraint. In other words, the covertness requirement has no effect on capacity in this case.

**Theorem 2** If

$$\frac{P}{N_0} \leq \sqrt{2c}, \tag{53}$$

where  $c$  is given in (50), then

$$\lim_{\eta \downarrow 0} \lim_{\epsilon \downarrow 0} \lim_{T \rightarrow \infty} \frac{\log M(W, T, \epsilon, \eta, \delta)}{T} = \frac{P}{N_0}. \tag{54}$$

*Proof* The converse holds because one cannot achieve a larger per-second capacity than the right-hand side of (54) even without a covertness constraint, and when the power constraint (52) is imposed on the average over all codewords instead of every codeword; see [9].

The achievability proof is similar to that of Theorem 1 given in Section 2.4, but, instead of (33), we choose

$$a \triangleq \sqrt{\frac{P}{2W}}. \tag{55}$$

The power constraint is satisfied: every  $x(\cdot)$  satisfies

$$\int_{-\infty}^{\infty} x(t)^2 dt = \lfloor 2(1 - \alpha)WT \rfloor \cdot a^2 \leq PT. \tag{56}$$

For covertness, instead of (39), we now have

$$\mathcal{D}(\hat{Y} \parallel \hat{Z}) \leq \frac{P^2}{4W^2 N_0^2} + o(W^{-2}), \tag{57}$$

so (40) becomes

$$\begin{aligned} \mathcal{D}(Y_{-\infty}^{\infty} \parallel Z_{-\infty}^{\infty}) &\leq \frac{\lfloor 2(1 - \alpha)WT \rfloor P^2}{4W^2 N_0^2} + o\left(\frac{T}{W}\right) \\ &\leq (1 - \alpha) \frac{T}{2W} \cdot \frac{P^2}{N_0^2} + o\left(\frac{T}{W}\right). \end{aligned} \tag{58}$$

The right-hand side of (58) is less than  $\delta$  when  $WT$  is large enough, by (50) and (53). Thus, we conclude that the covertness condition is satisfied for large enough  $WT$ .

Using the information-spectrum method, we know that a sequence of codes can have vanishing probability of decoding error if

$$\begin{aligned} \lim_{WT \rightarrow \infty} \frac{\log M}{T} &\leq p\text{-}\liminf_{WT \rightarrow \infty} \frac{1}{T} \log \frac{p(\hat{Y}_1^n | \hat{X}_1^n)}{p(\hat{Y}_1^n)} \\ &= p\text{-}\liminf_{WT \rightarrow \infty} \frac{\lfloor 2(1 - \alpha)WT \rfloor}{T} \log \frac{p(\hat{Y} | \hat{X})}{p(\hat{Y})}. \end{aligned} \tag{59}$$

In place of (49), we now have

$$p\text{-}\liminf_{WT \rightarrow \infty} \frac{\lfloor 2(1 - \alpha)WT \rfloor}{T} \log \frac{p(\hat{Y} | \hat{X})}{p(\hat{Y})} \geq (1 - \alpha) \frac{P}{N_0}. \tag{60}$$

The proof is completed by letting  $\alpha \downarrow 0$ .  $\square$

### 3.2 White noise with finite bandwidth

In this section, we assume that  $W = \infty$ , i.e., there is no bandwidth constraint at all on the input (as opposed to letting  $W \rightarrow \infty$  together with  $T$  as in Section 3.1). But we assume that the additive Gaussian noise is band-limited and has PSD

$$N(f) = \begin{cases} \frac{N_0}{2}, & |f| \leq W_0 \\ 0, & |f| > W_0, \end{cases} \tag{61}$$



where  $W_0$  is some constant that depends on neither  $T$  nor  $\delta$ . Not surprisingly, the input bandwidth will be effectively limited by the noise bandwidth  $W_0$ .

**Corollary 3** *If the additive Gaussian noise has PSD given by (61), then*

$$\lim_{\eta \downarrow 0} \lim_{\epsilon \downarrow 0} \lim_{T \rightarrow \infty} \frac{\log M(\infty, T, \epsilon, \eta, \delta)}{\sqrt{T\delta}} = \sqrt{2W_0}. \quad (62)$$

*Proof* If the input signal has nonzero energy outside the frequency range  $[-W_0, W_0]$ , then  $\mathcal{D}(Y_{-\infty}^{\infty} \| Z_{-\infty}^{\infty})$  will be infinity, violating the covertness constraint for any finite  $\delta$ ; this can be seen either using Theorem 10.5.1 of [13] or by noting that  $Z(\cdot)$  has an orthonormal expansion in the PSWFs for  $W_0$  and  $T$ , but  $X(\cdot)$  and  $Y(\cdot)$  do not. Hence, the input frequency must be restricted to  $[-W_0, W_0]$ . The claim then follows by replacing  $W$  with  $W_0$  in Theorem 1.  $\square$

**Remark 2** *In the context of Corollary 3, one could restrict the receiver to only observing  $y(t)$ ,  $t \in [0, T]$ , without affecting (62). This is because the receiver can perform the reduction to discrete time as in Section 2.2 merely using its observation on  $[0, T]$ , thanks to the orthogonality property (12). (This would not be possible if the noise were not band-limited, because the receiver would then need to first use an ideal low-pass filter, which cannot operate on the finite interval  $[0, T]$ .)*

*Note that, in this slightly different setting, we still allow the warden to observe the entire real line. As we shall see in Section 4, restricting the warden's observation to  $[0, T]$  will result in serious artifacts.*

### 3.3 Colored noise

We shall not study colored noise with the same rigor as we studied white noise, due to technical difficulties. Instead, we turn to formulas for mutual information and relative entropy of continuous-time Gaussian processes in terms of PSD [13]. These formulas provide useful engineering insights, but are not sufficient to prove rigorous coding theorems like Theorem 1. Our findings below therefore remain conjectures.

We first consider the scenario where the additive noise occupies infinite bandwidth, but has finite energy. (Unlike white noise as considered in Section 3.1, colored noise can have infinite bandwidth.) Formally, let the stationary Gaussian noise  $Z(t)$ ,  $t \in \mathbb{R}$ , have PSD  $N(f)$ ,  $f \in \mathbb{R}$ , that is positive for all  $f \in \mathbb{R}$  and symmetric around  $f = 0$ , and satisfies

$$\int_{-\infty}^{\infty} N(f) df < \infty. \quad (63)$$

We choose the input signal  $X(t)$ ,  $t \in \mathbb{R}$ , to be generated from a stationary Gaussian process with PSD

$$S(f) = \begin{cases} \beta N(f), & f \in [-W, W] \\ 0, & \text{otherwise,} \end{cases} \quad (64)$$

where  $\beta$  and  $W$  will be specified later. (Recall that there is no bandwidth constraint on the input, so  $W$  can be arbitrarily large.) We then have (Theorem 10.5.1 of [13])

$$\begin{aligned} \mathcal{D}(Y_{-\infty}^{\infty} \| Z_{-\infty}^{\infty}) &= T \cdot \frac{1}{2} \int_{-W}^W \left( \frac{S(f)}{N(f)} - \log \left( 1 + \frac{S(f)}{N(f)} \right) \right) df \\ &\leq T \cdot \frac{1}{4} \int_{-W}^W \left( \frac{S(f)}{N(f)} \right)^2 df \\ &= \beta^2 WT. \end{aligned} \quad (65)$$

Hence, the covertness condition is satisfied if we choose

$$\beta = \sqrt{\frac{\delta}{WT}}. \quad (66)$$

With this choice, we can compute the mutual information using ([13], Theorem 10.3.1):

$$\begin{aligned} I(X_{-\infty}^{\infty}; Y_{-\infty}^{\infty}) &= T \cdot \int_{-W}^W \frac{1}{2} \log \left( 1 + \frac{S(f)}{N(f)} \right) df \\ &= \frac{WT}{2} \log(1 + \beta) \\ &\approx \frac{\sqrt{WT\delta}}{2}. \end{aligned} \quad (67)$$

Since  $W$  can be chosen arbitrarily large, we are allowed to choose it as a function of  $T$  such that

$$\lim_{T \rightarrow \infty} \frac{W\delta}{T} = \infty, \quad (68)$$

in which case the per-second mutual information, given by (67) divided by  $T$ , will grow to infinity as  $T$  grows large. Also note that the power in the above-chosen input signal is given by

$$\beta \cdot \int_{-W}^W N(f) df \leq \beta \cdot \int_{-\infty}^{\infty} N(f) df. \quad (69)$$

For any  $\delta$  (which may be a function of  $T$ ), one can choose  $W$  to grow fast enough so that  $\beta$  as defined in (66) tends to zero as  $T \rightarrow \infty$ . This will ensure that (69) tends to zero as  $T$  grows large, namely, that the input power vanishes as  $T$  grows large.

To summarize, we make the following conjecture.

**Conjecture 1** *If the Gaussian noise process has PSD  $N(f)$  that is positive on the entire real line, then the per-second covert communication capacity without bandwidth constraint on the input is infinity. Furthermore, this holds irrespectively of whether an average-power constraint is imposed on the input or not.*

We next consider the case where the noise is band-limited:

$$N(f) > 0 \iff |f| \leq W_0, \quad (70)$$

where  $W_0$  is a constant that does not depend on  $T$ . Assume that the input follows some stationary Gaussian process. Note that, if the input PSD  $S(f)$  is positive on any interval where  $N(f) = 0$ , then  $\mathcal{D}(Y_{-\infty}^{\infty} \| Z_{-\infty}^{\infty})$  will be infinity. Hence, the input process must also be limited to the frequencies in  $[-W_0, W_0]$ . Let

$$\lambda(f) \triangleq S(f)/N(f), \quad f \in [-W_0, W_0]. \quad (71)$$

By the covertness condition (7) and by Theorem 10.5.1 of [13], we require

$$\begin{aligned} \delta &\geq \mathcal{D}(Y_{-\infty}^{\infty} \| Z_{-\infty}^{\infty}) \\ &= T \cdot \frac{1}{2} \int_{-W_0}^{W_0} (\lambda(f) - \log(1 + \lambda(f))) df. \end{aligned} \quad (72)$$

The integrand is convex in  $\lambda(f)$ , so we obtain

$$\bar{\lambda} - \log(1 + \bar{\lambda}) \leq \frac{\delta}{W_0 T} \quad (73)$$

where

$$\bar{\lambda} \triangleq \frac{1}{2W_0} \int_{-W_0}^{W_0} \lambda(f) df. \quad (74)$$

From (73) we obtain that, for large  $T$ ,

$$\bar{\lambda} \lesssim \sqrt{\frac{2\delta}{W_0 T}}. \quad (75)$$

This implies the following (approximate) upper bound on the mutual information:

$$\begin{aligned} I(X_{-\infty}^{\infty}; Y_{-\infty}^{\infty}) &= T \cdot \int_{-W_0}^{W_0} \frac{1}{2} \log(1 + \lambda(f)) df \\ &\leq W_0 T \log(1 + \bar{\lambda}) \\ &\lesssim \sqrt{2W_0 T \delta}, \end{aligned} \quad (76)$$

where the first inequality follows because the integrand is concave in  $\lambda(f)$ . This expression is the same as in the case where  $N(f)$  is constant for  $f \in [-W_0, W_0]$ . Also note that, if we choose

$$\lambda(f) = \bar{\lambda} \approx \sqrt{\frac{2\delta}{W_0 T}}, \quad f \in [-W_0, W_0], \quad (77)$$

then we obtain approximate equality in both (72) and (76). We hence make the following conjecture.

**Conjecture 2** *Corollary 3 holds whenever the Gaussian noise  $Z(t)$ ,  $t \in \mathbb{R}$ , has PSD that is positive within  $[-W_0, W_0]$  and zero elsewhere.*

#### 4 Deficiency of a time-limited model

If there is no bandwidth limit on the input signal  $x(\cdot)$ , as is the case in Corollary 3, then  $x(\cdot)$  can be made to be strictly time-limited to an interval  $[0, T]$ . One naturally wonders whether it is possible to formulate a time-limited covert communication model, where all parties have access only to the interval  $[0, T]$ . In this section, we show that, at least in the way we construct it below, such a model is invalid, because it leads to serious artifacts.

Our time-limited model is characterized as follows.

1. The transmitter maps a message to  $x(t)$ ,  $t \in [0, T]$ .
2. The receiver maps  $y(t)$ ,  $t \in [0, T]$ , to a decoded message.
3. The covertness constraint is

$$\mathcal{D}(X_0^T \| Y_0^T) \leq \delta. \quad (78)$$

Let  $\bar{M}(T, \epsilon, \delta)$  denote the largest possible cardinality of the message set such that the above conditions are satisfied and that the average probability of a decoding error is at most  $\epsilon$ .

Let us now assume that the additive noise  $Z(\cdot)$  has PSD given by (61). Under the model in Section 2, we have shown that the maximum amount of information that can be communicated in  $T$  seconds is proportional to  $\sqrt{T\delta}$ ; see Corollary 3. Under the new time-limited model, however, one can communicate an arbitrarily large amount of information over any fixed period of time.

**Theorem 3** *Let  $Z(t)$ ,  $t \in \mathbb{R}$ , have PSD given by (61). Under the above model, for any positive  $T$ ,  $\epsilon$ , and  $\delta$ ,*

$$\bar{M}(T, \epsilon, \delta) = \infty. \quad (79)$$

*Proof* Only the direct part needs proof. To this end, for a positive integer  $k$ , we generate  $k^3$  IID Gaussian random variables  $\{X_i\}$  of mean zero and variance  $k^{-2}$ . Let

$$X(t) = \begin{cases} \sum_{i=1}^{k^3} X_i \psi_i(t), & t \in [0, T], \\ 0, & \text{otherwise,} \end{cases} \quad (80)$$

where  $\{\psi_i(\cdot)\}$  are the PSWFs for  $W_0$  and  $T$ ; see Section 2.1. Clearly,  $X(t)$  is strictly time-limited to  $[0, T]$ . By (12), the channel can be reduced, for both the warden and the receiver, to  $k^3$  parallel Gaussian channels

$$Y_i = X_i + Z_i, \quad i \in \{1, \dots, k^3\}. \quad (81)$$

For covertness, we have the following bound:

$$\mathcal{D}(Y_0^T \| Z_0^T) = k^3 \left( \frac{2k^{-2}}{N_0} - \log \left( 1 + \frac{2k^{-2}}{N_0} \right) \right) \leq \frac{2}{kN_0^2}. \quad (82)$$

Hence, when we let  $k \rightarrow \infty$ , the covertness condition (78) will be met for any positive constant  $\delta$ . The input-output

mutual information can be calculated to be

$$I(X_0^T; Y_0^T) = \frac{k^3}{2} \log \left( 1 + \frac{2k^{-2}}{N_0} \right), \quad (83)$$

which grows to infinity when  $k \rightarrow \infty$ . As in Section 2.4, one can use information-spectrum methods to show that the amount of information that can be communicated with arbitrarily small probability of error indeed grows to infinity with  $k$ ; details are omitted.  $\square$

We provide some intuitive explanation to this artifact. A time-limited warden cannot employ an ideal low-pass filter, because the impulse response of such a filter occupies the entire real line. Hence, the time-limited warden cannot fully exploit the fact that the noise is band-limited. It is perhaps also interesting to understand this from the perspective of *memory*. Because the additive noise  $Z(\cdot)$  has memory (due to its finite bandwidth), its values on  $(-\infty, 0)$  and  $(T, \infty)$  can provide information about its values within  $[0, T]$ , helping the warden detect communication. For example, consider a communication scheme where  $X(0) \neq 0$  with a nonzero probability (which would be the case for the scheme used in the proof of Theorem 3). A time-unlimited warden will see a discontinuity in  $Y(\cdot)$  at  $t = 0$ , from which it can immediately determine that communication is taking place. This is not possible for a warden that is limited to  $[0, T]$ .

## 5 Discussion

We provided a rigorous mathematical framework for studying covert communication over the continuous-time Gaussian channel. We then used this framework to study the scenario where the input bandwidth can be infinitely large. We showed that, roughly speaking, over an AWGN channel where the transmitter can employ unbounded bandwidth, covert communication has the same per-second capacity as standard, non-covert communication.

We pointed out that one must be careful when formulating the continuous-time model. In particular, we believe that the model should allow the warden to observe not only the time window when communication might take place, but also before and after that time window. Essentially, this is because the channel has memory. The same issue would also arise in discrete-time channels with memory, unless one assumes, for example, that the channel behaves independently before, during, and after the communication window, as in [21, 22].

It remains to prove Conjectures 1 and 2 on colored noise in a fashion similar to our treatment of white noise. Doing so seems to require additional tools in functional analysis.

### Abbreviations

AWGN: Additive white Gaussian noise; BPSK: Binary phase shift keying; PDF: Probability density function; PSD: Power spectral density; PSWF: Prolate spheroidal wave function

### Authors' contributions

LW is responsible for the research and the writing of the manuscript. The author read and approved the final manuscript.

### Funding

The author declares no funding for this research.

### Competing interests

The author declares that he has no competing interests.

Received: 2 July 2019 Accepted: 28 November 2019

Published online: 27 December 2019

### References

1. B. A. Bash, D. Goekel, D. Towsley, Limits of reliable communication with low probability of detection on AWGN channels. *IEEE Sel. J. Areas Commun.* **31**(9), 1921–1930 (2013)
2. P. H. Che, M. Bakshi, S. Jaggi, *Reliable deniable communication: hiding messages in noise*, (2013). <https://doi.org/10.1109/isit.2013.6620765>
3. L. Wang, G. W. Wornell, L. Zheng, Fundamental limits of communication with low probability of detection. *IEEE Trans. Inform. Theory.* **62**(6), 3493–3503 (2016)
4. M. Bloch, Covert communication over noisy channels: a resolvability perspective. *IEEE Trans. Inform. Theory.* **62**(5), 2334–2354 (2016)
5. M. Tahmasbi, M. Bloch, First- and second-order asymptotics in covert communication. *IEEE Trans. Inform. Theory.* **65**(4), 2190–2212 (2019)
6. Q. Zhang, M. Bakshi, S. Jaggi, Computationally efficient covert communication. *Subm. to IEEE Trans. Inform. Theory* (2018). arXiv:1607.02014v2
7. C. E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 and 623–656 (1948)
8. A. Lapidoth, *A Foundation in Digital Communication*, Second ed. (Cambridge University Press, 2017). <https://doi.org/10.1017/9781316822708>
9. A. D. Wyner, Capacity of the band-limited Gaussian channel. *Bell Syst. Tech. J.* **45**, 359–395 (1966)
10. R. G. Gallager, *Information Theory and Reliable Communication*. (Wiley, 1968)
11. S. Verdú, On channel capacity per unit cost. *IEEE Trans. Inform. Theory.* **36**, 1019–1030 (1990)
12. M. Simon, J. Omura, R. Scholtz, B. Levitt, *Spread Spectrum Communications Handbook*. (McGraw-Hill, 1994)
13. M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. (San Francisco, Holden-Day, 1964)
14. L. Wang, *The continuous-time Poisson channel has infinite covert communication capacity*, (2018). <https://doi.org/10.1109/isit.2018.8437350>
15. Q. Zhang, M. R. Bloch, M. Bakshi, S. Jaggi, *Undetectable radios: covert communication under spectral mask constraints*, (2019). <https://doi.org/10.1109/isit.2019.8849818>
16. D. Slepian, H. Landau, H. Pollak, Prolate spheroidal wave functions, Fourier analysis and uncertainty – I & II. *Bell Syst. Tech. J.* **40**, 43–84 (1961)
17. D. Slepian, Some asymptotic expansions of prolate spheroidal wave functions. *Math. J. Phys.* **44**, 99–140 (1965)
18. T. M. Cover, J. A. Thomas, *Elements of Information Theory*, second ed. (Wiley, New York, 2006)
19. S. Verdú, T. S. Han, A general formula for channel capacity. *IEEE Trans. Inform. Theory.* **40**(4), 1147–1157 (1994)
20. T. S. Han, *Information Spectrum Methods in Information Theory*, (2003). [https://doi.org/10.1007/978-3-662-12066-8\\_7](https://doi.org/10.1007/978-3-662-12066-8_7)
21. P. H. Che, M. Bakshi, C. Chan, S. Jaggi, *Reliable deniable communication with channel uncertainty*, (2014). <https://doi.org/10.1109/itw.2014.6970786>
22. T. Sobers, B. Bash, S. Guha, D. Towsley, D. Goekel, Covert communication in the presence of an uninformed jammer. *IEEE Trans. Wirel. Comm.* **16**(9), 6193–6206 (2017)

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.