Information-theoretic limits of covert communication over additive-noise channels

Cécile Bouette

THÈSE pour obtenir le titre de Docteur en Sciences
spécialité Sciences et Technologies de l'Information et de la Communication

École doctorale EM2P
Économie, Management, Mathématiques et Physique

Information-theoretic limits of covert communication over additive-noise channels

---

CÉCILE BOUETTE

Laboratoire ETIS-UMR 8051, CY Cergy Paris Université, ENSEA, CNRS

JURY:

Michèle Wigger	Professeure à Télécom Paris	Rapporteur
Sidharth Jaggi	Professeur à l'University of Bristol	Rapporteur
Malcolm Egan	Chargé de recherche à l'INRIA	Examinateur
Elsa Dupraz	Maître de conférences à l'IMT Atlantique	Examinatrice
Olivier Rioul	Professeur à Télécom Paris	Président du jury
Inbar Fijalkow	Professeure des universités à l'ENSEA	Directrice de thèse
Laura Luzzi	Maître de conférences à l'ENSEA	Co-encadrante
Ligong Wang	Chercheur à ETH Zurich	Co-encadrant

Defended on the 23rd of January 2025

ACKNOWLEDGEMENTS

This thesis was achievable due to the steadfast support and scientific guidance of my supervisors, Laura Luzzi and Ligong Wang. I am profoundly grateful for their direction. I would like also to express my sincere gratitude to Matthieu Bloch for his assistance with my research.
I wish to express my deep appreciation to my advisor, Inbar Fijalkow, for being my guide at ETIS. I also extend my thanks to Malcolm Egan and Florentina Nicolau for being part of my thesis monitoring committee and their thoughtful feedback.
Additionally, I want to thank Michèle Wigger, Sidharth Jaggi, Malcolm Egan, Elsa Dupraz, and Olivier Rioul for agreeing to be members of my Ph.D. committee. I thank the reviewers for their feedback on the manuscript and the jury for their thought-provoking questions during the defense.

Finally, I thank my family for their love and support over these three years and in particular Pierre-Louis, whose caring attentiveness and encouragement made this time of my life all the more enjoyable.

Contents

1 Introduction

1.1 Covert communication in physical layer security

Communication confidentiality is classically ensured by encryption [65], however, cryptographic security can be compromised with sufficiently large computational power. Recently, there has been growing interest in securing the physical environment used for communication, known as the physical layer of the Open Systems Interconnection (OSI) model (Figure 1).

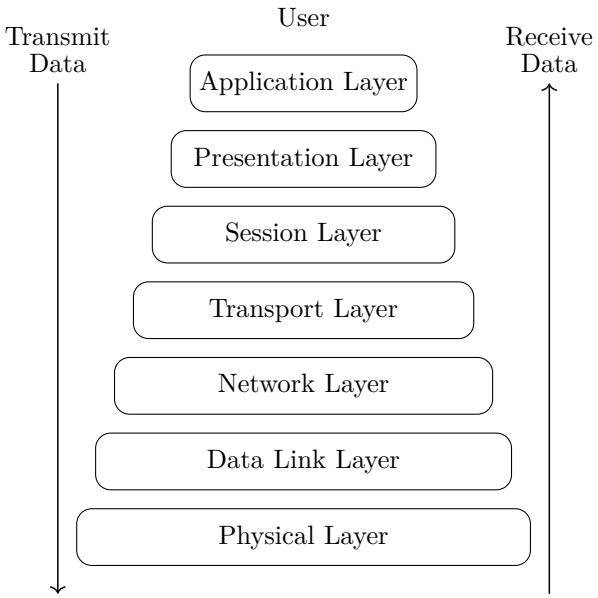

Fig. 1: OSI model.

Securing the physical layer could in principle replace cryptography or network encapsulation as everything ultimately depends on the physical layer. Thus physical-layer security aims to exploit the randomness inherent in wired or wireless communication channels to guarantee confidentiality even against computationally unlimited attackers.

Physical layer security was first investigated in the security paradigm of Wyner's wiretap channel model [81] where it is required that the adversary should not learn any information about the transmitted message while the intended receiver should be able to decode the message without any error. To characterize the quantity of information that could be sent securely over a noisy channel in the presence of an eavesdropper who observes the channel output, Wyner [81] introduced the metric known as secrecy capacity. It was shown that there exist wiretap codes [81, 19] that can achieve a positive rate of communication if the channel of the eavesdropper is more noisy than the channel of the legitimate receiver. Another possible scenario investigated within the field of physical layer security is the generation of secret keys at multiple terminals sharing a common random source [53] such as a wireless environment. These keys can then be used for encryption. Even though their exchange is vulnerable to jamming and man-in-the-middle attacks, appropriate countermeasures [3, 54] have been studied to mitigate these vulnerabilities. Physical layer security can be leveraged to provide many other services such as authentication using physical unclonable functions [56] or radio-frequency (RF) physical layer authentication (fingerprinting) [79].

Covert communication In this thesis, we focus on another problem in physical security, namely "covert communication", also known as "communication with a low probability of detection". In this scenario, the legitimate transmitter and receiver want to prevent a potential eavesdropper from making a good guess on whether a communication is ongoing or not. Covertness is desirable in many applications, since merely revealing *who* is communicating, *when* and *from where* can leak sensitive information, even if the content of the communication is not disclosed. We aim at ensuring covertness while ensuring reliable communication for the legitimate receiver: meaning ensuring a small error in decoding and theoretically, an error-free communication if the time of communication is sufficiently long.

Spread-Spectrum Communication Practical covert communication has been studied in the context of spread-spectrum technology [68, Pt. 5, Ch. 1], where the energy of the signal is spread over a large bandwidth. With this technique, the spectral density peaks of the signal approach the noise floor, which makes the signal harder to detect in the frequency domain. While spread-spectrum communication is widely used when the communicating parties have a hardware advantage over the eavesdropper, its security is often debated. In fact, there is little theoretical assurance about how difficult it is for an eavesdropper to detect the communication. Moreover, its usage is mostly confined to military contexts, for instance during the Cold War, its development and use were initiated by the MIT, Bell Labs, and USSR laboratories [82]. However, it is impractical for individuals where hardware resources may be limited and frequency hopping might be restricted by law.

Steganography Covert communication has also been studied in the context of steganography [48], which involves concealing information in digital fixed-size objects. The terminology for steganography was established at the 1996 Information Hiding Workshop [32]: an original, unaltered message (such as an image or software binary code) is called a *cover object*; the sender tries to hide an embedded message called a *stego object* by transforming the cover object using a secret key, and the resulting encrypted object is sent to the receiver. In addition, according to Kerckhoff's principle [30] which states that the security of a cryptographic system shouldn't rely on the secrecy of the algorithm, the eavesdropper is assumed to have full knowledge of the steganographic system and the distributions of the cover objects, except for the knowledge of the secret key. The communication-theoretic definition of steganographic capacity is reminiscent of the channel capacity definition: steganographic capacity is the maximal amount of hidden information in any cover object in the limit when its size goes to infinity [32] or in several finite-alphabet objects in the limit when the number of objects goes to infinity [48]. It has been shown that the amount of information that can be hidden in an object, such that the object appears unchanged, is proportional to the square root of the size of the object itself [48, 30], which is sometimes called the *square root law* of steganography.

Covert communication in information theory Our work addresses the problem of covert communication over noisy communication channels and adopts a theoretical approach in the spirit of Shannon. This study builds on the first studies on the limits of covert communication through the lens of information theory in [1, 77, 7]. Generally, the setup of covert communication involves two legitimate users and an eavesdropper who can either share the same channel as the legitimate receiver [77] or potentially face a different level of noise [7]. This framework is somewhat similar to that of spread-spectrum technology and steganography, but the approach to the problem is more fundamental.

Square root law In the framework of covert communication, Bash, Goeckel, and Towsley [1] first determined the order of magnitude of the message length which the legitimate users can reliably communicate over a noisy channel while ensuring a low probability of detection by an eavesdropper. This

work showed that the capacity (in nats per channel use) of the channel under a covertness constraint is zero because the maximum message length that can be transmitted reliably and covertly scales like the square root of the total number of channel uses. This phenomenon is sometimes called the *square root law* and is reminiscent of the square root law of steganography. While [1] considered additive white Gaussian noise (AWGN) channels, the square root law was also established for binary symmetric channels by Che, Bakshi, and Jaggi [16]. The exact asymptotic throughput and the corresponding scaling constant for the square root law—which we shall formally define later on—were characterized for discrete memoryless channels (DMCs) and AWGN channels in Wang, Wornell, Zheng [77] and Bloch [7]. The square root law does not hold if the transmitter is not required to be switched off when it is not transmitting a message. Hou and Kramer [41] considered a different scenario where the sender can transmit random signals to confuse the eavesdropper when it is not transmitting a message, and showed that in this case the square root law may be beaten. Other works have identified situations in which the square root law may be beaten when the channel statistics are not known e.g. when there is noise uncertainty for the eavesdropper channel [15, 50] or when the number of transmit antennas scales up with the blocklength [4].

Information-theoretic metrics for covertness Choosing covertness conditions in information theory [77, 7] is similar to choosing conditions in steganography [13] and should allow to control the error probability of the best possible test the eavesdropper could design to detect a hidden message. More precisely, the chosen condition should guarantee a lower bound on the sum of the probability of false alarm and missed detection. Leveraging hypothesis testing theory [18, Section 11.7], we want to make sure that the eavesdropper cannot do much better than random guessing which ensures a low probability of detection. For this purpose, the difference between the output statistics of the channel when there is communication and pure noise should be controlled in terms of total variation distance [7] or Kullback-Leibler divergence [77] (see Section 2.3).

Key length Some studies on covert communication assume that a secret key is shared between the legitimate users, which gives the legitimate receiver the advantage needed in order to detect the communication and decode the message reliably [77, 7]. In particular, Bloch [7] investigated the conditions for needing a key and its size when needed over DMCs and AWGN channels. This work showed that the key length is also proportional to the square root of the number of channel uses when communicating at maximum message size. The square root scaling of the key also applies with any message size in the context of DMCs and was characterized in [12, Corollary 3]. The requirement for a key arises from the necessity of approximating the output of the channel when no communication is taking place with the actual output statistics of the code. Code design criteria for covert communication were revisited in [7] through the concept of channel resolvability [37]: the full set of codewords (corresponding to different values of keys and confidential messages) should be a resolvability code for the eavesdropper's channel, which approximates the covert output, while the subcodebook corresponding to a fixed key should be a good channel code for the legitimate receiver's channel.

Motivation of this thesis While it is widely used, the Gaussian noise model does not capture all practical scenarios of interest for wireless communications. For instance, experimental evidence suggests that dense wireless networks with interference are actually characterized by heavy-tailed noise [17]. The Laplace distribution is useful for modeling noise spikes due to rare events [46, Chapter 10]; generalized Gaussian distributions [55, 25] have been considered to model multiple-user interference in ultrawideband systems [2] and atmospheric noise [44]; α-stable distributions [22, 14] and in particular the Cauchy distribution [35] are used for modeling interference in wireless ad hoc networks. In the non-covert setting,

a general formula for the capacity of continuous channels is still an open problem but the capacity has been characterized or bounded in several special cases including exponential noise [72], Cauchy noise [26], generalized Gaussian noise [25], and α-stable noise [22]. In contrast, we will show that surprisingly covert communications are easier to study and we are able to derive a simple expression for the scaling constant of the square root law over rather general additive-noise channels.

1.2 Main contributions

The present work fits into the classic framework where the sender is turned off when there is no communication and the square root law holds. Furthermore, we assume that the legitimate receiver and the eavesdropper face exactly the same noise distribution. In this scenario, the legitimate receiver shares a key with the sender to provide the necessary advantage for detecting and decoding the communication. We extend the computation of the scaling constant for the square root law from AWGN channels to general memoryless additive noise channels including non-Gaussian additive noise, as well as Gaussian channels with memory.

We show that the scaling constant of covert communication remains the same between a Gaussian channel with memory and an AWGN channel. This contrasts with the classical channel coding scenario, where the presence of memory may allow an increase in transmission rates.

In the case of general additive noise, under mild integrability assumptions, we show that the square root scaling constant is upper-bounded by a simple expression that depends solely on the class of the noise distribution, namely on the probability density function (PDF) of the noise. For example when the noise is generalized Gaussian [55, 25], we show that the scaling constant depends only on one parameter p: where the smaller p is, the heavier the tail of the distribution is, and the more information can be sent covertly and reliably. We then show that, under some additional assumptions on the noise PDF, the said upper bound is tight, i.e., there exists a covert code that can asymptotically achieve it. We find that the optimal asymptotic throughput has a similar form to the one found in [77, Theorem 3] for discrete memoryless channels and remains unchanged if the noise varies in scale. We further provide (sometimes loose) upper bounds on the key length that is needed to achieve the optimal scaling constant.

Furthermore, we study covert communication in the finite blocklength regime. Under a *maximal* error probability criterion, we upper bound the second-order asymptotics of covert communication over an AWGN channel. Then we show that allowing a positive *average* probability of decoding error enhances the asymptotic amount of covert information that can be shared over an AWGN channel. Additionally, we provide upper and lower bounds on the first-order asymptotics.

1.3 Organization of this thesis

The remainder of the document is organized as follows:

Chapter 2 introduces the notation and the information-theoretic framework of covert communication. We define the scaling constant of the square root law of covert communication and review known results on its characterization in the case of discrete memoryless channels and the AWGN channel. Furthermore, we present the first original contribution of this thesis, which is to characterize the scaling constant for Gaussian channels with memory.

Chapter 3 considers the general case of memoryless additive noise, computes a general formula for an upper bound of the amount of information that can be sent reliably and covertly, and states sufficient conditions under which this upper bound is achievable. We explicitly compute this square root law bound in several special cases including generalized Gaussian, exponential, gamma noise, and Cauchy noise.

Chapter 4 considers covert communication over AWGN channels in the finite blocklength regime under two different constraints on the probability of decoding error. For a fixed maximal probability of error, we prove an upper bound on the first and second-order asymptotics. Under a fixed average probability of error, we show that the first-order asymptotics of covert communication depend on the probability of error and establish upper and lower bounds.
Chapter 5 examines the implications of our results and proposes some ideas for further development.

Some complementary material is provided in the Appendices, including definitions, and classical information theory results. More precisely:
Appendix A details mathematical definitions and theorems useful throughout the reading of this thesis.
Appendix B outlines the Neyman-Pearson lemma.
Appendix C consists of a summary of useful information-theoretic definitions, properties, and proofs of preliminary results.
Appendix D elaborates on covert communication in the more general case where the eavesdropper and the legitimate receiver do not face the same noise.

1.4 My publications

This work was presented in part in the following publications:

Journal paper

[J1] **C. Bouette**, L. Luzzi, and L. Wang, "Covert Communication Over Additive-Noise Channels," to appear in *IEEE Transactions on Information Theory*, 2025. [IEEE early access] [arXiv]

International conferences

[C1] **C. Bouette**, L. Luzzi, and L. Wang, "Covert communication over two types of additive noise channels," in *IEEE Information Theory Workshop* (ITW), 2023. [arXiv] [IEEExplore]

[C2] **C. Bouette**, L. Luzzi, and M. Bloch, "Covert Capacity of AWGN Channels under Average Probability of Error," submitted to *IEEE International Symposium on Information Theory* (ISIT), 2025.

Poster

C. Bouette, L. Luzzi, and L. Wang, "Covert Communication Over Additive-Noise Channels," recent results poster session, in *IEEE International Symposium on Information Theory*, 2024. [ETIS]

2 Fundamental limits of covert communication

2.1 Notations and preliminaries

We usually use upper-case letters like X to denote (real) random variables and lower-case letters like x to denote their realizations. A length-n random vector $(X_1, \ldots, X_n)$ is denoted X^n. We use P_X to denote the law (also called distribution) of the random variable X and P_{X^n} that of the random vector X^n. We denote the Lebesgue-Stieltjes associated measure of P_X by $\mathrm{d}P_X$ [80, Section 3.11]. When it exists, the probability density function (PDF) corresponding to P_X is denoted p_X. We denote the product of measures by $\otimes$.
The entropy of a discrete random variable X is denoted $H(X)$ [18], the differential entropy of a continuous random variable X is denoted $h(X)$ [18], and the mutual information between X and Y is denoted $I(X;Y)$; all of these are measured in nats.
We denote a deterministic code by $\mathcal{C}$ and a random code by C.
We denote $\|\cdot\|_2$ and $\|\cdot\|_1$ respectively the Euclidean norm and the absolute-value norm.
Let P_1 and P_2 be two real distributions on the same measurable space $(\Omega, \mathcal{T})$. The total variation distance [75] between P_1 and P_2 is

$$d_{TV}(P_1, P_2) = 2 \sup_{\omega \in \Omega} |P_1(\omega) - P_2(\omega)|. \tag{2.1}$$

Suppose that P_1 is absolutely continuous with respect to P_2 (denoted $P_1 \ll P_2$) which means that for any $\omega \in \mathcal{T}$, $P_2(\omega) = 0 \implies P_1(\omega) = 0$, then the Kullback-Leibler divergence [49, 75] between P_1 and P_2 is

$$D(P_1 \| P_2) = \int_\Omega \ln\left(\frac{\mathrm{d}P_1}{\mathrm{d}P_2}(w)\right) \mathrm{d}P_1, \tag{2.2}$$

where $\frac{\mathrm{d}P_1}{\mathrm{d}P_2}$ is the Radon–Nikodym derivative [36, p. 128] of P_1 with respect to P_2.
We denote by Q the Gaussian tail function:

$$\begin{aligned} Q : \mathbb{R} &\to \mathbb{R}^+ \\ x &\mapsto \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{d}t. \end{aligned} \tag{2.3}$$

The Lebesgue measure is denoted by λ. Let δ denote the Dirac distribution

$$\delta(w) = \begin{cases} 0, & 0 \notin w \\ 1, & 0 \in w \end{cases} \tag{2.4}$$

with φ_0 denoting its characteristic function, so

$$\varphi_0(t) = 1, \qquad t \in \mathbb{R}. \tag{2.5}$$

Throughout the manuscript, vectors of length n are denoted with a superscript n. A random variable Z following a Gaussian distribution with mean μ and standard deviation $\sigma > 0$ is denoted $Z \sim \mathcal{N}(0, \sigma)$. A multivariate Gaussian random vector Z^n of length n with any mean vector denoted μ^n of length n and the $n \times n$ symmetric positive definite covariance matrix Σ is denoted $Z^n \sim \mathcal{N}(\mu^n, \Sigma)$. I_n denotes the

$n \times n$ identity matrix.
We denote an exponential distribution of mean $\Lambda > 0$ by $\mathcal{E}(\Lambda)$.
We denote $\Gamma(\cdot)$ the gamma function and $\psi(\cdot)$ the digamma function [21]. We denote the exponential function indiscriminately as $x \mapsto e^x$ or $x \mapsto \exp(x)$.

2.2 System model and problem statement

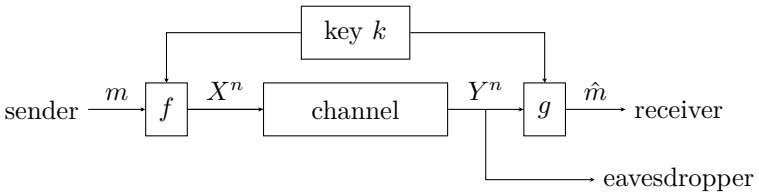

Fig. 2: General setup for covert communication.

We now define the general setup for covert communication, illustrated in Figure 2. The sender communicates with the legitimate receiver through a noisy channel and uses the channel n times. The input and output random variables X^n and Y^n take values in the alphabets $\mathcal{X}^n$ and $\mathcal{Y}^n$ respectively. We assume that the input alphabet $\mathcal{X}$ includes an "off" symbol denoted by 0, i.e. when the transmitter is not sending a message, it always transmits 0. $P_{Y^n|X^n}$ denotes the transition probability of the channel, modeling the noise added to a transmission.
The sender and the receiver are assumed to share a sufficiently long secret key $k \in \mathcal{K}$. A code $\mathcal{C} = (f, g)$ of length n for message set $\mathcal{M}$ and key set $\mathcal{K}$ consists of an encoder $f\colon \mathcal{M} \times \mathcal{K} \to \mathcal{X}^n, (m, k) \mapsto x^n$ and a decoder $g\colon \mathcal{Y}^n \times \mathcal{K} \to \mathcal{M}, (y^n, k) \mapsto \hat{m}$. We denote the input distribution random vector $X^n_{\mathcal{C}}$ and the codewords are the different possible input vectors $x^n_{\mathcal{C}} = (x_{1,\mathcal{C}}, \dots, x_{n,\mathcal{C}})$. The key and the message are assumed to be uniformly distributed and independent of each other. Unless explicitly stated, we will assume that the eavesdropper observes the same output as the legitimate receiver. We will assume that the eavesdropper knows the encoding and decoding functions f and g, but not the value of the secret key k.
Covertness requires that the eavesdropper should not be able to detect whether transmission is ongoing or not. Specifically, we consider the following covertness condition: for a chosen code $\mathcal{C}$, for some given $\Delta > 0$, the output distribution induced by the code must satisfy

$$D(P_{Y^n_{\mathcal{C}}} \| P_{Y^n|X^n=0^n}) \leq \Delta, \tag{2.6}$$

where $P_{Y^n_{\mathcal{C}}}$ denotes the distribution of the output sequence averaged over the messages and over the key:

$$P_{Y^n_{\mathcal{C}}}(\cdot) = \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} P_{Y^n|X^n}(\cdot | f(m,k)) \tag{2.7}$$

and $P_{Y^n|X^n=0^n}$ the output distribution corresponding to no-input i.e. the distribution of the noise induced by the channel.
Observing the output Y^n, the eavesdropper attempts to decide between two hypotheses: either no communication is happening, or there is an ongoing communication using the code $\mathcal{C}$.
For a chosen key k, the average probability of decoding error (or average probability of error) ε_k of the

legitimate receiver is

$$\varepsilon_k = \frac{1}{|\mathcal{M}|} \sum_{m=1}^{|\mathcal{M}|} \mathbb{P}\left[g(y^n, k) \neq m \mid x^n = f(m, k)\right] \tag{2.8}$$

and the average probability of decoding error ε of the code is

$$\varepsilon = \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} \mathbb{P}\left[g(y^n, k) \neq m \mid x^n = f(m, k)\right]. \tag{2.9}$$

2.3 Relation to hypothesis testing

The eavesdropper should not be able to determine whether a communication is happening or not. To this end, the eavesdropper's error is ensured to be very likely by requiring the covertness condition (2.6). Indeed hypothesis testing theory ensures that a small Kullback-Leibler divergence between the output distribution when there is communication and the noise distribution indicates that the eavesdropper can not do much better than random guessing when trying to identify whether the output he observes comes from the code output distribution or noise. We denote the two hypotheses that the eavesdropper aims to infer as

$$\begin{aligned} H_0 &: Y^n \sim P_{Y^n|X^n=0^n} \\ H_1 &: Y^n \sim P_{Y_{\mathcal{C}}^n}. \end{aligned} \tag{2.10}$$

We suppose further that $P_{Y_{\mathcal{C}}^n} \ll P_{Y^n|X^n=0^n}$. This hypothesis makes sense because if the sender generates an output that could not have been inherently produced by pure noise, then the communication will be automatically detected. From the Neyman-Pearson lemma (see Appendix B.1), we know that the optimal test to distinguish between the two hypotheses is a maximum likelihood test, involving the threshold of the Radon-Nikodym derivative of $P_{Y_{\mathcal{C}}^n}$ with respect to $P_{Y^n|X^n=0^n}$. Accordingly, we consider the following family of maximum likelihood tests indexed by $\gamma > 0$:

$$\begin{aligned} T_\gamma : \mathcal{Y}^n &\to \{0, 1\} \\ y^n &\mapsto \mathbb{1}\left\{\frac{\mathrm{d}P_{Y_{\mathcal{C}}^n}}{\mathrm{d}P_{Y^n|X^n=0^n}}(y^n) \geq \gamma\right\} \end{aligned} \tag{2.11}$$

where the output 0 indicates that the test chooses H_0 and the output 1, the test chooses H_1. The probability of false positive error (false alarm), meaning the probability of deciding for $P_{Y_{\mathcal{C}}^n}$ i.e. H_1 when in fact $Y^n \sim P_{Y^n|X^n=0^n}$, is

$$\beta = \int_{\mathcal{Y}^n} T_\gamma(y^n)\ \mathrm{d}P_{Y^n|X^n=0^n} \tag{2.12}$$

and the probability of false negative error (missed detection), meaning the probability of deciding for H_0 when actually $Y^n \sim P_{Y_{\mathcal{C}}^n}$ is

$$\kappa = 1 - \int_{\mathcal{Y}^n} T_\gamma(y^n)\ \mathrm{d}P_{Y_{\mathcal{C}}^n}. \tag{2.13}$$

Furthermore, the minimum error probability of the optimal hypothesis test conducted by an eavesdropper is in fact determined by the total variation distance $d_{TV}(\cdot, \cdot)$ between the two distributions; see Lemma B.2:

$$\min_\gamma(\beta + \kappa) = 1 - \frac{1}{2} d_{TV}\left(P_{Y_{\mathcal{C}}^n}, P_{Y^n|X^n=0^n}\right). \tag{2.14}$$

Recall Pinsker's inequality, historically first written in the form [57, p. 16] and well-known in the form:

Lemma 2.1 (Pinsker's inequality [47, Theorem 6.1]) *Let P_1 and P_2 be probability measures on the same measurable space $(\Omega, \mathcal{T})$ such that $P_1 \ll P_2$, then*

$$d_{TV}\left(P_1, P_2\right) \leq \sqrt{2}\sqrt{D\left(P_1 \| P_2\right)}. \tag{2.15}$$

Equations (2.14) and (2.15) imply that the sum of the probabilities of false alarm and missed detection can also be bounded in terms of the Kullback-Leibler divergence:

$$\min_{\gamma}(\beta + \kappa) \geq 1 - \frac{1}{\sqrt{2}}\sqrt{D\left(P_{Y_{\mathcal{C}}^n} \| P_{Y^n|X^n=0^n}\right)}. \tag{2.16}$$

In this light, the covertness condition (2.6) finally ensures that for any test chosen by the eavesdropper, its probability of error will be greater than $1 - \frac{1}{\sqrt{2}}\sqrt{\Delta}$. We note that because of the relation (2.14), an alternative (and less strict) notion of covertness can be defined by directly requiring the total variation distance $d_{TV}\left(P_{Y_{\mathcal{C}}^n}, P_{Y^n|X^n=0^n}\right)$ to be bounded by a constant. In this thesis, we prefer to use Kullback-Leibler divergence because it is easier to handle for theoretical analysis.

2.4 Relation to channel resolvability

Recall that the covertness condition (2.6) requires that the output of the chosen code $\mathcal{C}$ should approximate the output of the all-zero symbol in Kullback-Leibler divergence.

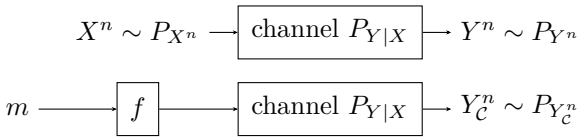

Fig. 3: Channel output approximation over a noisy channel.

Channel resolvability The problem of designing codes that approximate the output of a given source through a noisy channel was first studied by Han and Verdù [37], who defined the resolvability as the minimum code size per channel use required in order to generate an input that achieves an arbitrarily accurate approximation of the output statistics for any given input process. Such an approximation is sometimes also referred to as *soft covering* [20]. Although [37] considered the approximation of output statistics in terms of variational distance and normalized Kullback-Leibler divergence, the concept of resolvability can be extended to other metrics. In this thesis, we focus on the approximation in terms of unnormalized Kullback-Leibler divergence [38]. Although this notion can be defined for very general channel models, in this section we focus on memoryless channels and i.i.d. sources P_{X^n} for simplicity (see Figure 3).

Definition 2.1 (M-type [37, Definition 4]) *Let M be a positive integer. A probability distribution P on the measurable space $(\Omega, \mathcal{T})$ is said to be $M-$type if*

$$P(\omega) \in \left\{0, \frac{1}{M}, \frac{2}{M}, \ldots, 1\right\}, \qquad \text{for all } \omega \in \Omega. \tag{2.17}$$

Definition 2.2 (Resolvability code [38, Section III]) *Consider a code $\mathcal{C} = (f, g)$ for a message set $\mathcal{M}$ which has been generated i.i.d. from a distribution P_{X^n}. Let P_{Y^n} be the distribution induced by P_{X^n} through the channel $P_{Y^n|X^n}$; let $P_{Y_{\mathcal{C}}^n}$ the output distribution through the channel $P_{Y^n|X^n}$ for the code $\mathcal{C}$*

with equiprobable messages:

$$P_{Y_{\mathcal{C}}^n}(\cdot) = \frac{1}{|\mathcal{M}|} \sum_{m=1}^{|\mathcal{M}|} P_{Y^n|X^n}(\cdot|f(m)). \tag{2.18}$$

For a chosen $\eta > 0$, we call a resolvability code of size M, an $M-$type code such that

$$D(P_{Y_{\mathcal{C}}^n} \| P_{Y^n}) < \eta. \tag{2.19}$$

A general non-asymptotic bound for approximation of output statistics using random codes was given by Hayashi:

Theorem 2.1 (Random coding for approximation of output statistics [39, Theorem 14]**)** *Consider an input distribution P_{X^n} and a noisy channel $P_{Y^n|X^n}$. Let P_{Y^n} be the output distribution induced by P_{X^n} through the channel $P_{Y^n|X^n}$ and let C be a random code with codewords generated i.i.d. from the input distribution P_{X^n}. Then for all $\rho \in (0, 1]$,*

$$\mathbb{E}_{\mathsf{C}}\left[D\left(P_{Y_{\mathsf{C}}^n} \| P_{Y^n}\right)\right] \leq \frac{1}{\rho} \ln\left(1 + e^{-\rho \ln|\mathcal{M}| + \Psi(\rho|P_{Y^n|X^n}, P_{X^n})}\right) \tag{2.20}$$

where

$$\Psi(\rho|P_{Y^n|X^n}, P_{X^n}) = \ln\left(\mathbb{E}\left[\left(\frac{p_{Y^n|X^n}(Y^n|X^n)}{p_{Y^n}(Y^n)}\right)^{\rho}\right]\right), \tag{2.21}$$

$$P_{Y_{\mathsf{C}}^n}(\cdot) = \frac{1}{|\mathcal{M}|} \sum_{m=1}^{|\mathcal{M}|} P_{Y^n|X^n}(\cdot|X^n), \tag{2.22}$$

and the expectation in (2.20) is computed with respect to the random code C.

Remark 2.1 *Note that Theorem 2.1 was stated in [39] for discrete memoryless channels, but can be generalized to continuous outputs [39, Appendix D].*

Application to covert communication Note that covert communication fits within the previous framework, where the input X^n is constant and equal to the all-zero codeword, and the set of messages is the set $\mathcal{M} \times \mathcal{K}$ of message / key pairs. We should ensure that the whole code $\mathcal{C}$ of size $|\mathcal{M}| \times |\mathcal{K}|$ is a resolvability code for the eavesdropper i.e. the induced output distribution approximates the output of the all-zero input, while for a fixed key k, the corresponding subcode $\mathcal{C}(k)$ of size $|\mathcal{M}|$ is a good code for the legitimate receiver. In the case of random codes, for a given message size $|\mathcal{M}|$, Theorem 2.1 allows us to derive sufficient bounds for the key length to ensure covertness.

2.5 Square root law

Given $\varepsilon > 0$, we denote by $M^*(n, \varepsilon, \Delta)$ the maximum of $|\mathcal{M}|$ for which there exists a code $\mathcal{C}$ of length n that satisfies the covertness condition (2.6), and whose average probability of decoding error is at most ε. Except for special cases[1], covert communications are subject to the square root law, meaning that the amount of information that can be sent reliably and covertly scales like the square root of the number of channel uses. When the square root law holds, the capacity is therefore zero, and following [77], we can define a square root scaling constant as follows:

$$L \triangleq \lim_{\varepsilon \to 0} \liminf_{n \to \infty} \frac{\ln\left(M^*(n, \varepsilon, \Delta)\right)}{\sqrt{n\Delta}}. \tag{2.23}$$

[1] See Section 2.7.1.

In the remainder of this chapter, we will present known results about the characterization of L for some simple channel models, such as discrete memoryless channels and AWGN channels. In addition, we will provide new results on the characterization of L for the Gaussian channel with memory (Section 2.9). Furthermore, in Chapter 3, we will compute L for more general additive memoryless channels.

2.6 Information spectrum techniques

In the covert setup, the involved input and output distributions depend on the blocklength and the classical channel coding theorem [18, Section 7.7] does not hold. The basis to prove desired lower bounds for the scaling constant is a result by Shannon [66] who showed the existence of a code with a given average error probability as a function of its blocklength[2]. First, we introduce a definition:

Definition 2.3 (Information density and information spectrum [37]) *Given a joint distribution P_{X^n,Y^n} on $\mathcal{X}^n \times \mathcal{Y}^n$, the information density is given by:*

$$i_{X^n,Y^n}(x^n,y^n) = \ln\left(\frac{dP_{Y^n|X^n}(y^n|x^n)}{dP_{Y^n}(y^n)}\right) \tag{2.24}$$

The distribution of the random variable $\dfrac{i_{X^n,Y^n}(X^n,Y^n)}{n}$ where X^n and Y^n have joint distribution P_{X^n,Y^n} is referred to as the information spectrum.

A general proof for the existence of codes depending on the information spectrum derives from the following.

Theorem 2.2 (Shannon's achievability bound [66, Theorem 1][61, Theorem 18.5]**)** *For a given channel with transition law $P_{Y^n|X^n}$, for any input distribution P_{X^n} and induced output distribution P_{Y^n}, for any $\tau > 0$, the expectation of the average probability of error for a random code C with independent codewords generated i.i.d. from P_{X^n} is bounded as*

$$\mathbb{E}_{\mathsf{C}}[\varepsilon(\mathsf{C})] \leq \mathbb{P}\left[\frac{i_{X^n,Y^n}(X^n,Y^n)}{n} \leq \frac{\ln|\mathcal{M}|}{n} + \tau\right] + \exp(-n\tau). \tag{2.25}$$

In particular, there exists a code of blocklength n and size $|\mathcal{M}|$ whose codewords has been sampled from P_{X^n} with average probability of error ε such that:

$$\varepsilon \leq \mathbb{P}\left[\frac{i_{X^n,Y^n}(X^n,Y^n)}{n} \leq \frac{\ln|\mathcal{M}|}{n} + \tau\right] + \exp(-n\tau). \tag{2.26}$$

Later Verdù and Han [73] derived the corresponding converse:

Theorem 2.3 [73, Theorem 4] *For any code with average probability of error ε, size $|\mathcal{M}|$ and blocklength n*

$$\varepsilon \geq \mathbb{P}\left[\frac{i_{X^n,Y^n}(X^n,Y^n)}{n} \leq \frac{\ln|\mathcal{M}|}{n} - \tau\right] - \exp(-n\tau), \tag{2.27}$$

where P_{X^n} places probability mass $\frac{1}{M}$ on each codeword (i.e. P_{X^n} is the uniform distribution on the set of codewords) and P_{Y^n} is the output distribution induced by P_{X^n} through $P_{Y^n|X^n}$.

Remark 2.2 *Feinstein's lemma [28] gives a similar achievability bound to Theorem 2.2, this time for maximal error probability. In this work, we prefer to use Shannon's achievability bound (2.26) because Feinstein uses a greedy construction for his code whereas Shannon's idea is to use a random coding argument allowing us to follow the resolvability approach of Section 2.4.*

[2] As we will see in Chapter 4, information spectrum techniques also allow to characterize the first and second order asymptotics of covert communication.

2.7 Discrete memoryless channel

In this section, we consider the problem of covert communication over a discrete memoryless channel with finite input alphabet $\mathcal{X}$ and finite output alphabet $\mathcal{Y}$. We assume that the input alphabet contains an "off" symbol 0, which corresponds to the absence of transmission. It was shown in [77, 7] that the square root law holds on discrete memoryless channels at the exception of the case where the input 0 is redundant [77], namely where the off symbol output can be written as a mixture of other output distributions resulting from the transmission of meaningful symbols. This special case is considered in the following section.

2.7.1 Special case where the input 0 is redundant

Suppose that the output distribution of the channel, when there is no transmission, is a mixture of the output distributions of meaningful symbols. In that case, the square root law does not hold and the maximum amount of information that can be transmitted reliably under the covertness condition (2.6) scales like the blocklength n. Thus the capacity of the channel is positive:

Proposition 2.1 [77, Proposition 1] *We consider a channel with transition probability $P_{Y|X}$ where there exists an input distribution P_X on $\mathcal{X}$ such that*

$$P_X(0) = 0 \tag{2.28}$$

and

$$\sum_{x\in\mathcal{X}} P_X(x)P_{Y|X}(\cdot|x) = P_{Y|X}(\cdot|0). \tag{2.29}$$

Then, for all $\Delta > 0$:

$$\lim_{\varepsilon\to 0}\liminf_{n\to\infty} \frac{\ln\left(M^*(n,\varepsilon,\Delta)\right)}{n} = \max_{P_X} I(X;Y), \tag{2.30}$$

where the maximum is taken over distributions satisfying (2.28) and (2.29).

Proof: The fact that we can achieve this rate follows from the standard typicality argument of Shannon's channel coding theorem [18]: when the rate of the code is below $I(X;Y)$, the average probability of decoding error goes to 0 as n goes to infinity.
Furthermore, a random code generated by choosing the components of each codeword i.i.d. according to the distribution P_X, independently of the other codewords satisfies

$$\mathbb{E}\left[D(P_{Y_{\mathcal{C}}^n}\|P_{Y^n|X^n=0^n})\right] = \mathbb{E}\left[D(P_{Y_{\mathcal{C}}^n}\|P_{Y^n|X^n=0^n})\right] + D(P_{Y^n}\|P_{Y^n|X^n=0^n}), \tag{2.31}$$

where the expectation is computed with respect to P_X. We notice that a sufficiently long key ensures that the first term on the right-hand side of (2.31) is close to zero. The fact that P_X satisfies (2.29) implies that the second term on the right-hand side vanishes:

$$\begin{aligned} D(P_{Y^n}\|P_{Y^n|X^n=0^n}) &= n\ D(P_Y\|P_Z) \\ &= n\sum_{y\in\mathcal{Y}} P_Y(y)\ln\left(\frac{P_Y(y)}{P_{Y|X=0}(y|0)}\right) && (2.32)\\ &= n\sum_{y\in\mathcal{Y}} P_Y(y)\ln\left(\frac{P_Y(y)}{\sum_{x\in\mathcal{X}}P(x)P_{Y|X}(y|x)}\right) && (2.33)\\ &= 0, && (2.34) \end{aligned}$$

satisfying automatically the covertness condition (2.6).

We now show that it is not possible to achieve a rate greater than (2.30). We denote $X_{\mathcal{C},i}$ the i^{th} random component of the input distribution random vector $X_{\mathcal{C}}^n$. Suppose that the codebook's average input distribution

$$P_{\bar{X}}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} P_{X_{\mathcal{C},i}}(\cdot) \tag{2.35}$$

does not asymptotically satisfy (2.29): then

$$\begin{aligned}\lim_{n\to+\infty} D\left(P_{\bar{Y}} \| P_{Y|X=0}\right) &= \lim_{n\to+\infty} D\left(\sum_{x\in\mathcal{X}} P_{\bar{X}}(x) P_{Y|X}(\cdot|x) \| P_{Y|X=0}\right) \\ &> 0,\end{aligned} \tag{2.36}$$

where $P_{\bar{Y}}$ is the output induced by $P_{\bar{X}}$ through $P_{Y|X}$. Consequently, the code does not satisfy the covertness constraint (2.6) since

$$D\left(P_{Y_{\mathcal{C}}^n} \| P_{Y^n|X^n=0^n}\right) \geq n\ D\left(P_{\bar{Y}} \| P_{Y|X=0}\right) \tag{2.37}$$

where (2.37) holds because of the convexity of the Kullback-Leibler divergence and (2.37) is hence unbounded as n goes to infinity. Moreover, the concavity of the mutual information ensures that

$$\frac{1}{n} I(X^n; Y^n) \leq I(\bar{X}; \bar{Y}) \tag{2.38}$$

hence asymptotically the channel coding theorem [18] ensures that the maximum in the right hand-side of (2.30) is taken from input distribution that satisfies (2.29). Finally, any distribution P_X that satisfies (2.29) but not (2.28) is sub-optimal because

$$\begin{aligned}
I(X;Y) &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) P_X(x) \ln\left(\frac{P_{Y|X}(y|x)}{P_Y(y)}\right) \\
&= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) P_X(x) \ln\left(\frac{P_{Y|X}(y|x)}{P_{Y|X=0}(y|0)}\right) \\
&= \sum_{x\in\mathcal{X}\setminus 0}\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) P_X(x) \ln\left(\frac{P_{Y|X}(y|x)}{P_{Y|X=0}(y|0)}\right) \\
&= \sum_{x\in\mathcal{X}\setminus 0}\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) P_X(x) \left(\ln\left(\frac{P_{Y|X}(y|x)}{P_{Y'}(y)}\right) + \ln\left(\frac{P_{Y'}(y)}{P_Y(y)}\right)\right) \\
&= \sum_{x\in\mathcal{X}\setminus 0}\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) P_X(x) \ln\left(\frac{P_{Y|X}(y|x)}{P_{Y'}(y)}\right) - D\left(P_Y \| P_{Y'}\right) \\
&\quad - \sum_{y\in\mathcal{Y}} P_{Y|X}(y|0) P_X(0) \ln\left(\frac{P_{Y'}(y)}{P_Y(y)}\right) \\
&= \sum_{x\in\mathcal{X}\setminus 0}\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) P_X(x) \ln\left(\frac{P_{Y|X}(y|x)}{P_{Y'}(y)}\right) - D\left(P_Y \| P_{Y'}\right) \\
&\quad - P_X(0) \sum_{y\in\mathcal{Y}} P_Y(y) \ln\left(\frac{P_{Y'}(y)}{P_Y(y)}\right) \\
&= \sum_{x\in\mathcal{X}\setminus 0}\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) P_X(x) \ln\left(\frac{P_{Y|X}(y|x)}{P_{Y'}(y)}\right) - \left(1 - P_X(0)\right) D\left(P_Y \| P_{Y'}\right) \\
&\leq \sum_{x\in\mathcal{X}\setminus 0}\sum_{y\in\mathcal{Y}} P_{Y|X}(y|x) P_{X'}(x) \ln\left(\frac{P_{Y|X}(y|x)}{P_{Y'}(y)}\right) \\
&= I(X';Y'),
\end{aligned} \tag{2.39}$$

where $P_{X'}$ is P_X conditioned to $X \neq 0$ and Y' the output induced by X' through $P_{Y|X}$. □

2.7.2 Case when the input 0 is not redundant

If the output distribution of the DMC, when there is no transmission, is not a mixture of the output distributions of meaningful symbols and there is at least one input symbol other than 0, then the square root law holds. The corresponding scaling constant L defined in (2.23) was computed in [77].

Theorem 2.4 *[77, Theorem 3] For a discrete memoryless channel whose capacity-achieving input distribution support is equal to the set of all possible input symbols, if we denote its capacity-achieving output distribution by P_Y^*, then*

$$L \leq \sqrt{2}\sqrt{\mathrm{Var}_{P_{Y|X=0}}\left(\ln\left(\frac{P_{Y|X=0}(Y)}{P_Y^*(Y)}\right)\right)}. \tag{2.40}$$

Remark 2.3 *We notice that if the capacity-achieving output distribution is the uniform distribution over $\mathcal{Y}$, then*

$$L \leq \sqrt{2}\sqrt{\mathrm{Var}_{P_{Y|X=0}}\left(\ln\left(P_{Y|X=0}(Y)\right)\right)}. \tag{2.41}$$

2.8 Additive White Gaussian Noise channel

Covert communication on an additive white Gaussian noise (AWGN) channel satisfies the square root law [1, 77, 7]. The corresponding scaling constant L (2.23) has been characterized in [77]. We briefly present the key results for this setting. We consider an additive channel with i.i.d. Gaussian noise:

$$Y_i = X_i + Z_i, \qquad Z_i \sim \mathcal{N}(0, \sigma^2), \quad \sigma > 0, \qquad i = 1, 2, \ldots, n. \tag{2.42}$$

We assume that the eavesdropper and the legitimate receiver see the same outputs (Figure 4).

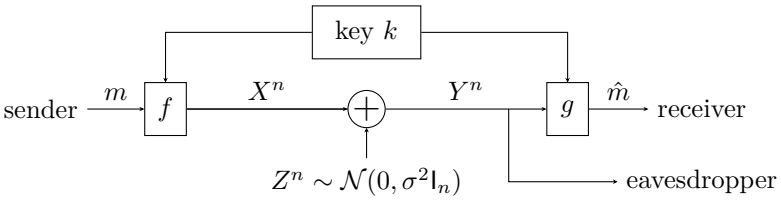

Fig. 4: Covert communication over an AWGN channel.

In the case of an additive channel, the covertness condition (2.6) can be rewritten as follows:

$$D\left(P_{Y_C^n} \| P_{Z^n}\right) \leq \Delta. \tag{2.43}$$

Theorem 2.5 *[77, Theorem 5] For the channel (2.42) under the constraint (2.43),*

$$L = 1, \tag{2.44}$$

irrespectively of the noise power σ^2.

In particular [77] showed that (2.44) is achievable with random coding using i.i.d. Gaussian inputs[3] with average power constraint $O\left(\frac{1}{\sqrt{n}}\right)$ and [76] showed that (2.44) is also achievable with random coding using i.i.d. binary phase-shift keying (BPSK) inputs with amplitude $\frac{\sqrt{2}\sigma\Delta^{\frac{1}{4}}}{n^{\frac{1}{4}}}$. In addition [7] considered again the setting in Figure 4 but under a total variation distance covertness constraint and recovered

[3] In Chapter 3, we will consider a more general class of continuous channels and will recover Theorem 2.5 as a special case.

the square root law for this setting [7, Theorem 6]. Moreover [7] showed that the minimum key length required to achieve covertness scales like $o(\sqrt{n})$.

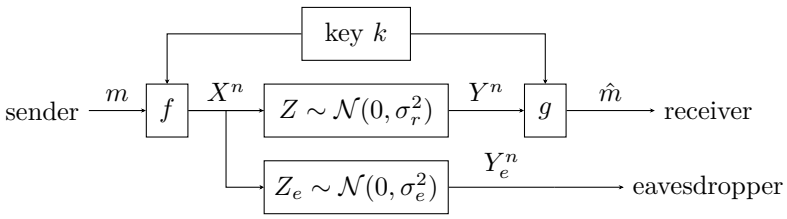

Fig. 5: Covert communication over the degraded AWGN channel.

Degraded AWGN channel We notice that in the special case where the eavesdropper and the legitimate receiver do not see the same outputs, the value for L changes. Consider an i.i.d. additive Gaussian noise channel with noise power σ_e^2 for the eavesdropper and σ^2 for the legitimate receiver (see Figure 5):

$$\begin{aligned} Y_i &= X_i + Z_i, \qquad Z_i \sim \mathcal{N}(0, \sigma^2), \quad \sigma > 0, \qquad i = 1, 2, \ldots, n, \\ Y_{e,i} &= X_{e,i} + Z_{e,i}, \qquad Z_{e,i} \sim \mathcal{N}(0, \sigma_e^2), \quad \sigma_e > 0, \qquad i = 1, 2, \ldots, n. \end{aligned} \tag{2.45}$$

The covertness constraint (2.6) can now be written in the form

$$D(P_{Y_{e,\mathcal{C}}^n} \| P_{Z_e^n}) \leq \Delta \tag{2.46}$$

where $P_{Z_e^n}$ denotes the distribution of the noise vector Z_e^n, and $P_{Y_{e,\mathcal{C}}^n}$ that of the output sequence averaged over the messages and the key:

$$\begin{aligned} p_{Y_{e,\mathcal{C}}^n}(y^n) &= \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} p_{Y_e^n|X^n}(y^n | f(m,k)) \\ &= \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_e^n} e^{-\frac{\|y^n - f(m,k)\|_2^2}{2\sigma_e^2}}. \end{aligned} \tag{2.47}$$

Theorem 2.6 *[78, Theorem 2] The scaling constant for the maximum amount of information that can be sent covertly and reliably to the legitimate receiver on the channel (2.45) satisfying the covertness condition (2.46) is*

$$L = \frac{\sigma_e^2}{\sigma^2}. \tag{2.48}$$

For completeness, we include the proof of Theorem 2.6 in Appendix D.

Remark 2.4 *We notice that no key is needed when $\sigma_e > \sigma$ (see Appendix D.1) which is a similar result to [7, Theorem 6].*

Remark 2.5 *One can recover Theorem 2.5 as a special case of Theorem 2.6 where the eavesdropper and the legitimate receiver listen on the same AWGN channel.*

2.9 New result: Gaussian channel with memory

In this section, we present the first original contribution of this thesis, which was presented in [C1]. We consider an extension of the setting of Figure 4 where the noise is Gaussian but not i.i.d. We assume that the legitimate receiver and the eavesdropper observe the same outputs and consider the channel with colored noise depicted in Figure 6:

$$Y_i = X_i + Z_i, \qquad i = 1, 2, \ldots, n, \tag{2.49}$$

where the noise sequence is modeled by a Gaussian process

$$Z^n \sim \mathcal{N}(\mu_n, \Sigma_n), \tag{2.50}$$

with the mean vector μ_n and symmetric positive definite covariance matrix Σ_n. Note that Σ_n being positive definite implies that it is invertible (i.e., non-singular).

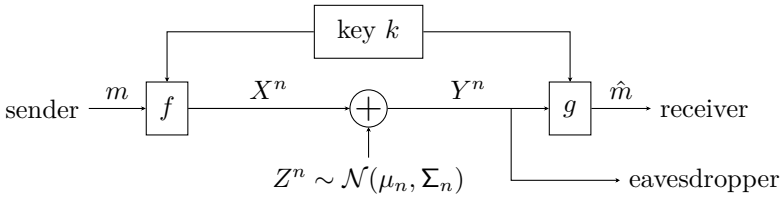

Fig. 6: Covert communication over the Gaussian channel with memory.

We shall show that the fundamental limit for covert communication over the additive Gaussian noise channel with memory in Figure 6 is the same as over the AWGN channel in Figure 4: not only does $\ln\left(M^*(n, \varepsilon, \Delta)\right)$ grow like $\sqrt{n}$, but we also have $L = 1$ as in Theorem 2.5. This result contrasts with the standard (non-covert) capacity of the Gaussian noise channel under an average power constraint, which in general will change if AWGN is replaced by colored Gaussian noise [18, Section 9.5].

Theorem 2.7 *For the channel* (2.49) *with Gaussian noise* (2.50)*, under the covertness requirement* (2.43)*,*

$$L = 1 \tag{2.51}$$

irrespectively of μ_n *and* Σ_n*.*

Proof: We prove the theorem operationally by showing a one-to-one correspondence between codes for the colored Gaussian channel (2.49) and for the i.i.d. Gaussian channel (2.42).
Since Σ_n is invertible, there exists an invertible $n \times n$ matrix A such that

$$Z^n = \mathsf{A}\tilde{Z}^n + \mu_n, \tag{2.52}$$

where $\tilde{Z}^n$ is a standard Gaussian vector, i.e., it consists of i.i.d. entries $\mathcal{N}(0, 1)$. Now consider the AWGN channel:

$$\tilde{Y}_i = \tilde{X}_i + \tilde{Z}_i, \qquad i = 1, \ldots, n. \tag{2.53}$$

with time-i input $\tilde{X}_i$ and output $\tilde{Y}_i$, respectively. Given any code $\mathcal{C} = (f, g)$ for the channel (2.49) with colored Gaussian noise (2.50), there is a corresponding code $\tilde{\mathcal{C}} = (\tilde{f}, \tilde{g})$ for the AWGN channel (2.53), and vice versa. Indeed, given $\mathcal{C}$, we construct $\tilde{\mathcal{C}}$ via the following mappings:

- for all $m \in \mathcal{M}$,

$$\tilde{f}(m) = \mathsf{A}^{-1} f(m); \tag{2.54}$$

- for all $\tilde{y}^n \in \mathbb{R}^n$,

$$\tilde{g}\left(\tilde{y}^n\right) = g\left(\mathsf{A}\tilde{y}^n + \mu_n\right). \tag{2.55}$$

Reversely, given $\tilde{\mathcal{C}}$, we construct $\mathcal{C}$ via:

- for all $m \in \mathcal{M}$,

$$f(m) = \mathsf{A}\tilde{f}(m); \tag{2.56}$$

- for all $y^n \in \mathbb{R}^n$,

$$g\left(y^n\right) = \tilde{g}\left(\mathsf{A}^{-1}(y^n - \mu_n)\right). \tag{2.57}$$

By this construction, a decoding error occurs with code $\mathcal{C}$ on the channel (2.49) with colored noise (2.50) if, and only if, a decoding error occurs with code $\tilde{\mathcal{C}}$ on the i.i.d. channel (2.53). Consequently, the error probabilities of the two codes (when used on their corresponding channels) are equal.

The one-to-one correspondence applies to any ensemble of random codes on the two channels as well. Furthermore, averaged over the random codes,

$$D(P_{Y^n} \| P_{Z^n}) = D(P_{\tilde{Y}^n} \| P_{\tilde{Z}^n}), \tag{2.58}$$

because the same invertible mapping—subtraction by μ_n and then multiplication by A^{-1}—maps Y^n to $\tilde{Y}^n$ and Z^n to $\tilde{Z}^n$, hence the data-processing inequality for the Kullback-Leibler divergence (Appendix C.1) holds in both directions.

We have now shown that the corresponding random codes on the two channels have exactly the same error probability and covertness property. The theorem then follows because, by Theorem 2.5, $L = 1$ for the AWGN channel. □

Remark 2.6 *Because of the one-to-one correspondence between the codes $\mathcal{C}$ and $\tilde{\mathcal{C}}$, the key length required for the Gaussian channel with memory is the same as for the i.i.d. Gaussian case. We will provide upper bounds for the key length requirements for general continuous channels in Section 3.5.*

3 Covert communication over general memoryless additive-noise channels

In this chapter, we present the second original contribution of this thesis, which was presented in part in [J1].
We consider covert communication over general additive channels with memoryless noise. As in Chapter 2, we focus on the asymptotic regime where the blocklength tends to infinity. First, we introduce technical assumptions about the noise probability density function (PDF) in Section 3.1. Under these assumptions, we establish a general upper bound for L in Section 3.2. Furthermore, in Section 3.3 we show that this bound is achievable if some additional conditions are satisfied. In Section 3.4 we derive an explicit formula for the scaling constant L (or an upper bound) for some particular channels. In Section 3.5, we present some upper bounds on the key length required for the achievability result.

3.1 Problem setup and technical assumptions

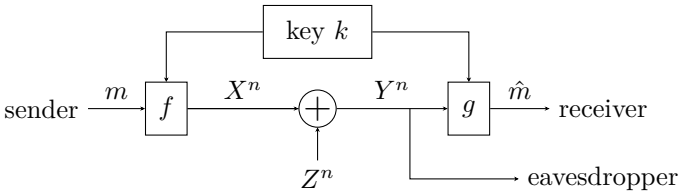

Fig. 7: General setup for covert communication over a memoryless additive channel.

We consider the setup illustrated in Fig. 7, where a transmitter and a receiver communicate in the presence of an eavesdropper over an additive noise channel described by

$$Y_i = X_i + Z_i, \qquad i = 1, 2, \ldots, n. \tag{3.1}$$

We assume that the noise stochastic process $\{Z_i\}_{i\in\{1,2,\ldots,n\}}$ is independent of the message and the secret key. We further assume that $\{Z_i\}_{i\in\{1,2,\ldots,n\}}$ is independent and identically distributed (i.i.d.) according to a PDF $p_Z(z)$, $z \in \mathbb{R}$. A fortiori, p_Z is Lebesgue-measurable and $\int_{\mathbb{R}} p_Z(z)\mathrm{d}z = 1$.
We make the following technical assumptions on p_Z: there exists some $\zeta \in (0,1)$ such that

$$\int_{\mathbb{R}} p_Z(z) \left(\ln(p_Z(z))\right)^4 \mathrm{d}z < \infty \tag{3.2}$$

$$\int_{\mathbb{R}} p_Z(z)^{\zeta} \mathrm{d}z < \infty \tag{3.3}$$

$$\int_{\mathbb{R}} p_Z(z)^{\zeta} \left(\ln(p_Z(z))\right)^4 \mathrm{d}z < \infty. \tag{3.4}$$

Remark 3.1 *An example of a distribution that fails to meet these integrability assumptions is given by probability density functions of the form* $p_Z(z) = 1/z \cdot 1/(\ln(z))^s$, $z \in (\exp((s-1)^{\frac{1}{1-s}}), +\infty)$, $s > 1$.

These imply some further integrability properties:

Lemma 3.1 *If p_Z satisfies (3.2)–(3.4), then, for every $k \in \{0,1,2,3,4\}$, the integrals*

$$\int_{\mathbb{R}} p_Z(z)^{\nu(z)} \left|\ln(p_Z(z))\right|^k \mathrm{d}z \tag{3.5}$$

are uniformly bounded over $\nu\colon \mathbb{R} \to [\zeta, 1]$, $z \mapsto \nu(z)$.

Proof: For any $k \in \{0,1,2,3,4\}$, $\nu\colon \mathbb{R} \to [\zeta,1]$, and $z \in \mathbb{R}$,

$$p_Z(z)^{\nu(z)} \left|\ln(p_Z(z))\right|^k \le \left(p_Z(z) + p_Z(z)^{\zeta}\right)\left(1 + (\ln(p_Z(z)))^4\right), \tag{3.6}$$

where the right-hand side does not depend on ν and is integrable due to (3.2)–(3.4). □

We recall that in the case of an additive channel, the covertness condition (2.6) can be rewritten as follows:

$$D\left(P_{Y_{\mathcal{C}}^n} \| P_{Z^n}\right) \le \Delta. \tag{3.7}$$

3.2 A general upper bound (Converse)

The following theorem provides a general upper bound on L. We shall later show that, under some additional assumptions on p_Z, this bound is tight.

Theorem 3.1 *For the memoryless additive-noise channel (3.1) with p_Z satisfying (3.2)–(3.4),*

$$L \le \sqrt{2}\sqrt{\mathrm{Var}\left[\ln(p_Z(Z))\right]}. \tag{3.8}$$

The following lemma will be used in the proof of Theorem 3.1.

Lemma 3.2 *Consider p_Z satisfying (3.2)–(3.4). For any $\gamma \in [0, 1-\zeta)$, let the random variable $\tilde{Z}$ have PDF*

$$p_{\tilde{Z}}(\tilde{z}) = \alpha \cdot p_Z(\tilde{z})^{1-\gamma}, \quad \tilde{z} \in \mathbb{R}, \tag{3.9}$$

where

$$\alpha = \left(\int_{\mathbb{R}} p_Z(z)^{1-\gamma} \mathrm{d}z\right)^{-1}. \tag{3.10}$$

Then the following hold:

1) *For $\gamma \in [0, 1-\zeta)$,*

$$h(\tilde{Z}) = -\frac{\ln(\alpha)}{\gamma} + \frac{1-\gamma}{\gamma} D(P_{\tilde{Z}} \| P_Z). \tag{3.11}$$

2) *For any random variable Y satisfying*

$$D(P_Y \| P_Z) \le D(P_{\tilde{Z}} \| P_Z), \tag{3.12}$$

we have

$$h(Y) \le h(\tilde{Z}). \tag{3.13}$$

That is, $p_{\tilde{Z}}$ as in (3.9) maximizes the differential entropy for a given Kullback-Leibler divergence to P_Z.

3) *For $\gamma \downarrow 0$,*

$$D(P_{\tilde{Z}} \| P_Z) = \frac{\gamma^2}{2} \mathrm{Var}\left[\ln(p_Z(Z))\right] + O(\gamma^3) \tag{3.14}$$

$$h(\tilde{Z}) - h(Z) = \gamma\, \mathrm{Var}\left[\ln(p_Z(Z))\right] + O(\gamma^2). \tag{3.15}$$

4) *The function $\gamma \mapsto D(P_{\tilde{Z}} \| P_Z)$ is continuous on $[0, 1-\zeta)$.*

Proof: We first prove 1) as follows:

$$\begin{aligned} D(P_{\tilde{Z}}\|P_Z) &= \int_{\mathbb{R}} p_{\tilde{Z}}(z)\ln\left(\frac{p_{\tilde{Z}}(z)}{p_Z(z)}\right)\mathrm{d}z \\ &= -h(\tilde{Z}) - \int_{\mathbb{R}} p_{\tilde{Z}}(z)\ln(p_Z(z))\mathrm{d}z \\ &= -h(\tilde{Z}) - \frac{1}{1-\gamma}\int_{\mathbb{R}} p_{\tilde{Z}}(z)\ln(\alpha p_Z(z)^{1-\gamma})\mathrm{d}z + \frac{\ln(\alpha)}{1-\gamma} \\ &= -h(\tilde{Z}) + \frac{1}{1-\gamma}h(\tilde{Z}) + \frac{\ln(\alpha)}{1-\gamma} \\ &= \frac{\gamma}{1-\gamma}\,h(\tilde{Z}) + \frac{\ln(\alpha)}{1-\gamma} \end{aligned} \tag{3.16}$$

which implies (3.11).

We next show 2). For any random variable Y satisfying (3.12), we have

$$\begin{aligned} 0 &\le D(P_Y\|P_{\tilde{Z}}) \\ &= -h(Y) - \int_{\mathbb{R}} p_Y(y)\ln\left(p_{\tilde{Z}}(y)\right)\mathrm{d}y \\ &= -h(Y) - \int_{\mathbb{R}} p_Y(y)\ln\left(\alpha p_Z(y)^{1-\gamma}\right)\mathrm{d}y \\ &= -h(Y) - \ln(\alpha) - (1-\gamma)\int_{\mathbb{R}} p_Y(y)\ln\left(p_Z(y)\right)\mathrm{d}y \\ &= -h(Y) - \ln(\alpha) + (1-\gamma)D(P_Y\|P_Z) + (1-\gamma)h(Y) \\ &\le -\gamma\,h(Y) - \ln(\alpha) + (1-\gamma)D(P_{\tilde{Z}}\|P_Z) \qquad (3.17) \\ &= \gamma\big(h(\tilde{Z}) - h(Y)\big), \qquad (3.18) \end{aligned}$$

where (3.17) follows from (3.12); and (3.18) from (3.16). Inequality (3.18) implies (3.13).

We next show 3). There exists $\theta\colon \mathbb{R}\to(0,\gamma)$ such that the Taylor expansion of $p_Z(z)^{-\gamma}$ with the Lagrange form of the remainder is

$$p_Z(z)^{-\gamma} = 1 - \gamma\ln(p_Z(z)) + \frac{\gamma^2}{2}\left(\ln(p_Z(z))\right)^2 - \frac{\gamma^3}{6}\left(\ln(p_Z(z))\right)^3 p_Z(z)^{-\theta(z)} \qquad \forall z\in\mathbb{R}. \tag{3.19}$$

The normalization factor α is then

$$\begin{aligned} \alpha &= \left(\int_{\mathbb{R}} p_Z(z)\left(1 - \gamma\ln(p_Z(z)) + \frac{\gamma^2}{2}\left(\ln(p_Z(z))\right)^2 - \frac{\gamma^3}{6}\left(\ln(p_Z(z))\right)^3 p_Z(z)^{-\theta(z)}\right)\mathrm{d}z\right)^{-1} \\ &= \left(1 + \gamma h(Z) + \frac{\gamma^2}{2}\mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] + O(\gamma^3)\right)^{-1} \qquad (3.20) \\ &= 1 - \gamma h(Z) - \frac{\gamma^2}{2}\mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] + \gamma^2 h(Z)^2 + O(\gamma^3), \qquad (3.21) \end{aligned}$$

where (3.20) follows by Lemma 3.1 and (3.21) follows by the Taylor expansion of $x\mapsto(1+x)^{-1}$. Note that

$$\begin{aligned} \ln\alpha &= -\ln\left(1 + \gamma h(Z) + \frac{\gamma^2}{2}\mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] + O(\gamma^3)\right) \\ &= -\gamma h(Z) - \frac{\gamma^2}{2}\mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] + \frac{\gamma^2}{2}h(Z)^2 + O(\gamma^3) \\ &= -\gamma h(Z) - \frac{\gamma^2}{2}\mathrm{Var}\left[\ln(p_Z(Z))\right] + O(\gamma^3). \end{aligned} \tag{3.22}$$

We also have

$$\begin{aligned}
&\int_{\mathbb{R}} p_Z(z)^{1-\gamma} \ln(p_Z(z))\mathrm{d}z \\
&\quad = \int_{\mathbb{R}} p_Z(z) \left(1 - \gamma \ln(p_Z(z)) + \frac{\gamma^2}{2} \left(\ln(p_Z(z))\right)^2 - \frac{\gamma^3}{6} \left(\ln(p_Z(z))\right)^3 p_Z(z)^{-\theta(z)}\right) \ln(p_Z(z))\mathrm{d}z \\
&\quad = -h(Z) - \gamma \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] + \frac{\gamma^2}{2} \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^3\right] + O(\gamma^3). \qquad (3.23)
\end{aligned}$$

(The expectations above are finite by Lemma 3.1.) We can now compute the Taylor expansion of $D(P_{\tilde{Z}} \| P_Z)$ in γ:

$$\begin{aligned}
D(P_{\tilde{Z}} \| P_Z) &= \int_{\mathbb{R}} p_{\tilde{Z}}(z) \ln\left(\alpha \frac{p_Z(z)^{1-\gamma}}{p_Z(z)}\right) \mathrm{d}z \\
&= \ln(\alpha) - \gamma\alpha \int_{\mathbb{R}} p_Z(z)^{1-\gamma} \ln(p_Z(z))\mathrm{d}z \\
&= -\gamma h(Z) - \frac{\gamma^2}{2} \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] + \frac{\gamma^2}{2} h(Z)^2 + O(\gamma^3) \\
&\quad - \gamma \left(1 - \gamma h(Z) - \frac{\gamma^2}{2} \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] + \gamma^2 h(Z)^2 + O(\gamma^3)\right) \\
&\quad \times \left(-h(Z) - \gamma \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] + \frac{\gamma^2}{2} \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^3\right] + O(\gamma^3)\right) \qquad (3.24) \\
&= -\gamma h(Z) - \frac{\gamma^2}{2} \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] + \frac{\gamma^2}{2} h(Z)^2 + \gamma h(Z) + \gamma^2 \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] \\
&\quad - \gamma^2 h(Z)^2 + O(\gamma^3) \\
&= \frac{\gamma^2}{2} \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] - \frac{\gamma^2}{2} h(Z)^2 + O(\gamma^3) \\
&= \frac{\gamma^2}{2} \mathrm{Var}\left[\ln(p_Z(Z))\right] + O(\gamma^3), \qquad (3.25)
\end{aligned}$$

where (3.24) follows by (3.20), (3.21), (3.22) and (3.23). Similarly we compute the Taylor expansion of $h(\tilde{Z}) - h(Z)$: using (3.11) we find

$$\begin{aligned}
h(\tilde{Z}) - h(Z) &= -\frac{\ln(\alpha)}{\gamma} + \frac{1-\gamma}{\gamma} D(P_{\tilde{Z}} \| P_Z) - h(Z) \\
&= \frac{\gamma h(Z) + \frac{\gamma^2}{2} \mathrm{Var}\left[\ln(p_Z(Z))\right] + O(\gamma^3)}{\gamma} + \frac{1-\gamma}{\gamma} \left(\frac{\gamma^2}{2} \mathrm{Var}\left[\ln(p_Z(Z))\right] + O(\gamma^3)\right) - h(Z) \qquad (3.26) \\
&= \gamma \mathrm{Var}\left[\ln(p_Z(Z))\right] + O(\gamma^2) \qquad (3.27)
\end{aligned}$$

where (3.26) follows by (3.22) and (3.25).
Finally, we show 4). Continuity at $\gamma = 0$ follows by (3.25), because the latter implies

$$\lim_{\gamma \downarrow 0} D(P_{\tilde{Z}} \| P_Z) = 0. \qquad (3.28)$$

We next write

$$\begin{aligned}
D(P_{\tilde{Z}} \| P_Z) &= \int_{\mathbb{R}} \alpha p_Z(z)^{1-\gamma} \ln\left(\frac{\alpha p_Z(z)^{1-\gamma}}{p_Z(z)}\right) \mathrm{d}z \\
&= -\gamma\alpha \int_{\mathbb{R}} p_Z(z)^{1-\gamma} \ln(p_Z(z))\mathrm{d}z + \ln(\alpha). \qquad (3.29)
\end{aligned}$$

To prove the desired continuity, it suffices to show that both α and

$$\int_{\mathbb{R}} p_Z(z)^{1-\gamma} \ln(p_Z(z))\mathrm{d}z \qquad (3.30)$$

are continuous in $\gamma \in (0, 1-\zeta)$. The statement for α follows by (3.21). For $k = 0, 1$ the function

$$\gamma \mapsto p_Z(z)^{1-\gamma} \ln(p_Z(z))^k \qquad (3.31)$$

is clearly continuous for every $z \in \mathbb{R}$. For all $\gamma \in (0, 1-\zeta)$, the function

$$z \mapsto p_Z(z)^{1-\gamma} \ln(p_Z(z))^k \tag{3.32}$$

is integrable by Lemma 3.1. Furthermore,

$$\left| p_Z(z)^{1-\gamma} \left(\ln(p_Z(z))\right)^k \right| \leq p_Z(z) \left|\ln(p_Z(z))\right|^k + p_Z(z)^{\zeta} \left|\ln(p_Z(z))\right|^k, \tag{3.33}$$

where the right-hand side is again integrable by Lemma 3.1. By the lemma of continuity under integrals (see Lemma A.1), we conclude that both α and (3.30) are continuous in $\gamma \in (0, 1-\zeta)$, hence so is $D\left(P_{\bar{Z}} \| P_Z\right)$. □

Remark 3.2 *Lemma 3.2 can also be seen as an entropy maximization problem subject to a* $\mathbb{E}[\ln(p_Z(\cdot))]$ *constraint which, in turn, arises from the Kullback-Leibler divergence constraint.*

Proof of Theorem 3.1: Take any code $\mathcal{C}$ of length n. Let $\bar{X}$ denote a random variable such that $P_{\bar{X}}$ is the average input distribution over the secret key, a uniformly drawn message, and the n channel uses, and let $\bar{Y}$ denote the channel output random variable when the input is $\bar{X}$, so $P_{\bar{Y}}$ is the average output distribution in the same sense as $P_{\bar{X}}$:

$$P_{\bar{X}}(\cdot) = \frac{1}{n} \sum_{i=1}^{n} P_{X_{\mathcal{C},i}}(\cdot), \tag{3.34}$$

$$P_{\bar{Y}}(\cdot) = \frac{1}{n} \sum_{i=1}^{n} P_{Y_{\mathcal{C},i}}(\cdot), \tag{3.35}$$

where $X_{\mathcal{C},i}$ and $Y_{\mathcal{C},i}$ are the i^{th} random components of respectively the input and output random vector $X_{\mathcal{C}}^n$ and $Y_{\mathcal{C}}^n$.
Starting with the covertness condition (3.7), similarly to [77] we have:

$$\begin{aligned}
\Delta &\geq D(P_{Y_{\mathcal{C}}^n} \| P_{Z^n}) \\
&= -h(Y_{\mathcal{C}}^n) - \mathbb{E}[\ln\left(p_{Z^n}(Y_{\mathcal{C}}^n)\right)] \\
&= \sum_{i=1}^{n} \left(-h(Y_{\mathcal{C},i}|Y_{\mathcal{C}}^{i-1}) - \mathbb{E}[\ln(p_Z(Y_{\mathcal{C},i}))]\right) \\
&\geq \sum_{i=1}^{n} \left(-h(Y_{\mathcal{C},i}) - \mathbb{E}[\ln(p_Z(Y_{\mathcal{C},i}))]\right) \\
&= \sum_{i=1}^{n} D(P_{Y_{\mathcal{C},i}} \| P_Z) \\
&\geq n\, D(P_{\bar{Y}} \| P_Z),
\end{aligned} \tag{3.36}$$

where the last step follows because the Kullback-Leibler divergence is convex.
We next derive a bound on $M^*(n, \varepsilon, \Delta)$ in terms of $\bar{X}$ and $\bar{Y}$. For each realization $K = k$ of the secret key, we denote by ε_k the average probability of error of the code $\mathcal{C}$ with $K = k$. For each $k \in \mathcal{K}$, we have by Fano's inequality (see Appendix C.2):

$$1 + \varepsilon_k \ln|\mathcal{M}| \geq H(X_{\mathcal{C}}^n | Y_{\mathcal{C}}^n, K = k), \tag{3.37}$$

where the joint distribution on $(X_{\mathcal{C}}^n, Y_{\mathcal{C}}^n)$ is computed according to a uniformly drawn message. Since the message was uniformly selected, $H(X_{\mathcal{C}}^n | K = k) = \ln|\mathcal{M}|$ and (3.37) is rewritten as

$$\ln|\mathcal{M}|\,(1 - \varepsilon_k) - 1 \leq I(X_{\mathcal{C}}^n; Y_{\mathcal{C}}^n | K = k). \tag{3.38}$$

Let ε be the probability of error averaged over the key. By averaging over the key, we obtain

$$\begin{aligned}
\ln|\mathcal{M}|\,(1-\varepsilon)-1 &\le I(X_{\mathcal{C}}^n;Y_{\mathcal{C}}^n|K)\\
&\le I(X_{\mathcal{C}}^n,K;Y_{\mathcal{C}}^n)\\
&= I(X_{\mathcal{C}}^n;Y_{\mathcal{C}}^n) \qquad (3.39)\\
&= \sum_{i=1}^n I(X_{\mathcal{C}}^n;Y_{\mathcal{C},i}|Y^{i-1})\\
&= \sum_{i=1}^n \left(h(Y_{\mathcal{C},i}|Y_{\mathcal{C}}^{i-1}) - h(Y_{\mathcal{C},i}|X_{\mathcal{C}}^n,Y_{\mathcal{C}}^{i-1})\right)\\
&= \sum_{i=1}^n \left(h(Y_{\mathcal{C},i}|Y_{\mathcal{C}}^{i-1}) - h(Y_{\mathcal{C},i}|X_{\mathcal{C},i})\right)\\
&\le \sum_{i=1}^n I(X_{\mathcal{C},i};Y_{\mathcal{C},i})\\
&\le nI(\bar{X};\bar{Y}) \qquad (3.40)\\
&= n(h(\bar{Y})-h(Z)), \qquad (3.41)
\end{aligned}$$

where (3.39) holds by the Markov chain:

$$K \to X_{\mathcal{C}}^n \to Y_{\mathcal{C}}^n; \tag{3.42}$$

and (3.40) because mutual information is concave in the input distribution. By the definition of $M^*(n,\varepsilon,\Delta)$, (3.41) implies

$$\ln\left(M^*(n,\varepsilon,\Delta)\right)(1-\varepsilon)-1 \le n(h(\bar{Y})-h(Z)). \tag{3.43}$$

We shall complete the proof using (3.36), (3.43), and Lemma 3.2. We first treat the special case where Z has the uniform distribution over a subset of $\mathbb{R}$; let us denote this subset by $\mathcal{S}$, so

$$p_Z(z) = \begin{cases} \frac{1}{\lambda(\mathcal{S})}, & z \in \mathcal{S}\\ 0, & \text{otherwise,}\end{cases} \tag{3.44}$$

where λ denotes the Lebesgue measure. In this case, one can show that covert communication is not possible at all, so $L=0$; we provide a proof in Section 3.4.1. It then follows that (3.8) trivially holds in this case. In the rest of this proof, we shall assume that p_Z does not have the form (3.44). Then, given any $\gamma>0$, $P_{\tilde{Z}}$ defined by (3.9) differs from P_Z, therefore $D(P_{\tilde{Z}}\|P_Z)>0$. Recall that the function $\gamma \mapsto D(P_{\tilde{Z}}\|P_Z)$ is continuous on $[0,1-\zeta)$; see Lemma 3.2, part 4). Therefore, there exists a sequence $\{\gamma_n\}$ such that

$$\lim_{n\to\infty}\gamma_n = 0, \tag{3.45}$$

and, for large enough n, the PDFs defined as

$$p_{\tilde{Z}_n}(\tilde{z}) = \alpha_n \cdot p_Z(\tilde{z})^{1-\gamma_n}, \quad \tilde{z}\in\mathbb{R}, \tag{3.46}$$

with

$$\alpha_n = \left(\int_{\mathbb{R}} p_Z(z)^{1-\gamma_n}\mathrm{d}z\right)^{-1}, \tag{3.47}$$

satisfy

$$D(P_{\tilde{Z}_n}\|P_Z) = \frac{\Delta}{n}. \tag{3.48}$$

Using (3.14) we then have

$$\frac{\Delta}{n} = \frac{\gamma_n^2}{2}\mathrm{Var}\left[\ln(p_Z(Z))\right] + O(\gamma_n^3), \tag{3.49}$$

which implies

$$\gamma_n = \sqrt{\frac{2}{\mathrm{Var}\left[\ln(p_Z(Z))\right]}}\sqrt{\frac{\Delta}{n}} + o\left(\frac{1}{\sqrt{n}}\right). \tag{3.50}$$

We now continue from (3.43) as follows:

$$\begin{aligned}
\frac{\ln\left(M^*(n,\varepsilon,\Delta)\right)(1-\varepsilon)-1}{n} &\le h(\bar{Y}) - h(Z) \\
&\le h(\tilde{Z}_n) - h(Z) && (3.51)\\
&= \gamma_n \mathrm{Var}\left[\ln(p_Z(Z))\right] + O(\gamma_n^2) && (3.52)\\
&= \sqrt{2}\sqrt{\mathrm{Var}\left[\ln(p_Z(Z))\right]}\sqrt{\frac{\Delta}{n}} + o\left(\frac{1}{\sqrt{n}}\right), && (3.53)
\end{aligned}$$

where (3.51) follows by (3.13), (3.36), and (3.48); (3.52) by (3.15); and (3.53) by (3.50). Recalling the definition (2.23) and taking $n \to \infty$ and $\varepsilon \to 0$ in (3.53) complete the proof. □

Remark 3.3 *The upper bound in Theorem 3.1 has a similar form to Theorem 2.4, which provides an upper bound on L for DMCs in terms of the output distribution induced by the "off" input symbol, and the capacity-achieving output distribution in the special case where the capacity-achieving output distribution is the uniform distribution (see (2.41) in Remark 2.3).*

Remark 3.4 *Clearly, the right-hand side of (3.8) does not change when the noise Z is scaled by a constant factor c:*

$$\begin{aligned}
\mathrm{Var}\left[\ln(p_{cZ}(cZ))\right] &= \mathrm{Var}\left[\ln\left(\frac{p_Z(Z)}{c}\right)\right] \\
&= \mathrm{Var}\left[\ln(p_Z(Z))\right]. && (3.54)
\end{aligned}$$

This is true in general: for the additive memoryless channel (3.1), scaling the noise by a constant factor c does not affect L, because this effect can be canceled out by multiplying the input X by c at the transmitter. Then, for both the receiver and the eavesdropper, there is no loss in optimality in scaling the output by $1/c$ to recover the same Y as in the original channel.

3.3 Tightness of the upper bound (Achievability)

Under some additional assumptions on the noise PDF p_Z, one can show that the upper bound of Theorem 3.1 is tight.

Assumption 3.1 *Assume that p_Z satisfies the following:*

1) *p_Z is bounded, i.e., there exists some $b > 0$ such that*

$$p_Z(z) \le b, \quad z \in \mathbb{R}; \tag{3.55}$$

2) *$z \mapsto p_Z(z)\ln\left(p_Z(z)\right)$ is uniformly continuous on its support $\mathrm{supp}(p_Z)$, i.e., for all $\varepsilon > 0$, there exists $\eta > 0$ such that, for any $z_1, z_2 \in \mathrm{supp}(p_Z)$, $|z_1 - z_2| \le \eta$, we have*

$$\left|p_Z(z_1)\ln(p_Z(z_1)) - p_Z(z_2)\ln(p_Z(z_2))\right| \le \varepsilon; \tag{3.56}$$

3) *there exists some $\xi \in (0,1)$ such that, for all $\gamma \in [0,\xi)$, there exists a random variable X independent of $Z \sim p_Z$ such that the PDF of $X + Z$ is $p_{\tilde{Z}}$ given by (3.9).*

Theorem 3.2 *For the memoryless additive-noise channel (3.1), if p_Z satisfies (3.2)–(3.4) as well as Assumption 3.1, then*

$$L = \sqrt{2}\sqrt{\mathrm{Var}\left[\ln(p_Z(Z))\right]}. \tag{3.57}$$

Before proving Theorem 3.2, we recall Lévy's convergence theorem concerning weak convergence of real-valued random variables[4] (Theorem A.1), and prove a lemma.

Lemma 3.3 *Consider any p_Z satisfying (3.2)–(3.4) and Assumption 3.1. For every n, let $\gamma_n \in (0, \xi)$, and let X_n be independent of Z and be such that*

$$\tilde{Z}_n = X_n + Z \tag{3.58}$$

has density

$$p_{\tilde{Z}_n}(\tilde{z}) = \alpha_n p_Z(\tilde{z})^{1-\gamma_n}, \quad \tilde{z} \in \mathbb{R}, \tag{3.59}$$

where

$$\alpha_n = \left(\int_{\mathbb{R}} p_Z(z)^{1-\gamma_n} \mathrm{d}z\right)^{-1}. \tag{3.60}$$

If $\gamma_n \to 0$ as $n \to \infty$, then, as $n \to \infty$, $P_{\tilde{Z}_n}$ converges weakly to P_Z, and P_{X_n} converges weakly to δ.

Proof: We first show that $P_{\tilde{Z}_n}$ converges weakly to P_Z. To this end, note that for any bounded continuous function f on $\mathbb{R}$, we have

$$\begin{aligned}
\left|\mathbb{E}\left[f(\tilde{Z}_n)\right] - \mathbb{E}[f(Z)]\right| &= \left|\int_{\mathbb{R}} f(z) p_{\tilde{Z}_n}(z) \mathrm{d}z - \int_{\mathbb{R}} f(z) p_Z(z) \mathrm{d}z\right| \\
&\le \|f\|_\infty \int_{\mathbb{R}} \left|p_{\tilde{Z}_n}(z) - p_Z(z)\right| \mathrm{d}z \\
&= \|f\|_\infty \left\|P_{\tilde{Z}_n} - P_Z\right\|_1 \\
&\le \|f\|_\infty \sqrt{2\, D\left(P_{\tilde{Z}_n} \| P_Z\right)},
\end{aligned} \tag{3.61}$$

where (3.61) follows by Pinsker's inequality (2.15). The right-hand side of (3.61) tends to 0 as $n \to \infty$, because $\gamma_n \to 0$ and by Lemma 3.2. It follows that $\mathbb{E}[f(\tilde{Z}_n)]$ tends to $\mathbb{E}[f(Z)]$ as $n \to \infty$, therefore, by definition, $P_{\tilde{Z}_n}$ converges weakly to P_Z.
The above implies

$$\lim_{n\to\infty} \varphi_{\tilde{Z}_n}(t) = \varphi_Z(t), \quad t \in \mathbb{R}. \tag{3.62}$$

For any n, since X_n is independent of Z, we have [80, Section 16.4]

$$\varphi_{\tilde{Z}_n}(t) = \varphi_Z(t)\varphi_{X_n}(t), \quad t \in \mathbb{R}, \tag{3.63}$$

where φ_{X_n} is the characteristic function of X_n. By properties of characteristic functions [80, Section 16.2], φ_Z is continuous and $\varphi_Z(0) = 1$, hence there exists an interval around 0 on which $\varphi_Z(t) \neq 0$. By (3.62) and (3.63), for any t in this interval,

$$\lim_{n\to\infty} \varphi_{X_n}(t) = 1. \tag{3.64}$$

By Theorem A.2 on the extension of characteristic functions, we know that (3.64) must hold for all $t \in \mathbb{R}$. This and Levy's convergence theorem (Theorem A.1) imply that P_{X_n} converges weakly to δ as $n \to \infty$.
□

Proof of Theorem 3.2: The converse part follows immediately from Theorem 3.1. We shall prove the direct part using techniques similar to [10, Theorem 3], [77, Section V-B], together with Lemma 3.3.

[4]Weak convergence can also be defined for general probability measures, but that is not needed in the present work.

Fix $\chi \in (1, \frac{3}{2})$. For sufficiently large n, let $\tilde{Z}_n$ have the PDF (3.59), with the choice

$$\gamma_n = \sqrt{\frac{2}{\mathrm{Var}\left[\ln(p_Z(Z))\right]}\left(\frac{\Delta}{n} - \frac{1}{n^\chi}\right)}. \tag{3.65}$$

(It will become clear later on that this choice of γ_n is to satisfy the covertness condition.) Further, let X be independent of Z and have the distribution of X_n satisfying (3.58); the existence of such X (for sufficiently large n) is guaranteed by Assumption 3.1.

We generate a random codebook C by picking every codeword i.i.d. according to $P_X^{\otimes n}$ and independently of the other codewords. Let $P_{Y_\mathsf{C}^n}$ denote the corresponding random output distribution:

$$P_{Y_\mathsf{C}^n}(\cdot) = \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} P_{Y^n|X^n}(\cdot|X_{m,k}^n). \tag{3.66}$$

Recall that the whole code of size $|\mathcal{K}| \times |\mathcal{M}|$ should be a resolvability code for the eavesdropper (see Section 2.4) i.e. ensuring covertness, while for a fixed key k, the corresponding subcode of size $|\mathcal{M}|$ should be a good code for the legitimate receiver. We notice that we can decompose the expectation of the Kullback-Leibler divergence with regard to the random codebook as

$$\begin{aligned}
\mathbb{E}_\mathsf{C}[D(P_{Y_\mathsf{C}^n}\|P_{Z^n})] &= \mathbb{E}_\mathsf{C}\left[D\left(P_{Y_\mathsf{C}^n}\middle\|P_{Y^n}\right) + \mathbb{E}\left[\ln\left(\frac{p_{Y^n}(Y_\mathsf{C}^n)}{p_{Z^n}(Y_\mathsf{C}^n)}\right)\right]\right] \\
&= \mathbb{E}_\mathsf{C}\left[D\left(P_{Y_\mathsf{C}^n}\middle\|P_{Y^n}\right)\right] + D\left(P_{Y^n}\|P_{Z^n}\right)
\end{aligned} \tag{3.67}$$

where P_{Y^n} is the output distribution induced by the input distribution P_{X^n} through the channel transition law $P_{Y^n|X^n}$ and the expectation in (3.67) is computed with respect to P_{X^n}. We notice that the first term on the right-hand side of (3.67) goes to 0 with the key length going to infinity (see Theorem 2.1). For the time being, we assume the key to be sufficiently long and therefore allow the first term in (3.67) to be arbitrarily close to zero. We admit that assuming an infinitely long key is unrealistic, but we shall show in Section 3.5 bounds on a sufficient key length. To show that with high probability there exist good realizations of C satisfying the covertness condition (3.7), it now suffices to show that the second term on the right-hand side of (3.67) is less than Δ for sufficiently large n:

$$D\left(P_{Y^n}\|P_{Z^n}\right) \leq \Delta. \tag{3.68}$$

Recalling that $Y^n \sim \tilde{Z}_n^n$, this can be shown as follows:

$$D\left(P_{\tilde{Z}_n}^{\otimes n}\middle\|P_{Z^n}\right) = n\, D\left(P_{\tilde{Z}_n}\middle\|P_Z\right) \tag{3.69}$$

$$= n\frac{\gamma_n^2}{2}\mathrm{Var}\left[\ln(p_Z(Z))\right] + O(n\gamma_n^3) \tag{3.70}$$

$$= \Delta - n^{1-\chi} + O\left(\frac{1}{\sqrt{n}}\right), \tag{3.71}$$

where (3.70) follows by (3.14); and (3.71) by (3.65). From (3.71) it is clear that $D(P_{\tilde{Z}_n}^{\otimes n}\|P_{Z^n}) < \Delta$ for sufficiently large n.

We next look at the mutual information corresponding to this code construction. To simplify notation, we henceforth use Y to denote the single-letter channel output, i.e., $Y = X + Z$; recall that $Y \sim \tilde{Z}_n$ and that the distributions of X and Y both depend on n. Since the inputs are i.i.d., we have

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{\sqrt{n}} I(X^n;Y^n) &= \lim_{n\to\infty} \sqrt{n}\, I(X;Y) \\
&= \lim_{n\to\infty} \sqrt{n}\left(h(Y) - h(Z)\right) \\
&= \lim_{n\to\infty} \sqrt{n}\left(h(\tilde{Z}_n) - h(Z)\right) \\
&= \sqrt{2}\sqrt{\mathrm{Var}[\ln(p_Z(Z))]} \cdot \sqrt{\Delta},
\end{aligned} \tag{3.72}$$

where the last step follows by (3.15) and (3.65).

It now remains to show that this limit of mutual information is operationally achievable. Since the distribution P_X depends on n, we cannot use a standard random coding argument, but we will use an information spectrum approach. Namely, we wish to show

$$\lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{\ln\left(M^*(n,\varepsilon,\Delta)\right)}{\sqrt{n}}\geq\lim_{n\to\infty}\frac{1}{\sqrt{n}}I(X^n;Y^n). \tag{3.73}$$

From Shannon's achievability bound (Theorem 2.2) we know that, for any $\tau>0$ the expectation of the average error probability ε of the above random code is bounded as

$$\mathbb{E}_{\mathsf{C}}[\varepsilon(\mathsf{C})]\leq\mathbb{P}\left[i_{X^n,Y^n}(X^n,Y^n)\leq\ln|\mathcal{M}|+n\tau\right]+e^{-n\tau}, \tag{3.74}$$

where $i_{X^n,Y^n}(x^n,y^n)$ is the information density; see Definition 2.3. Choosing $\tau=n^{-\frac{3}{4}}$, we rewrite (3.74) as

$$\mathbb{E}_{\mathsf{C}}[\varepsilon(\mathsf{C})]\leq\mathbb{P}\left[\frac{i_{X^n,Y^n}(X^n,Y^n)}{\sqrt{n}}\leq\frac{\ln|\mathcal{M}|}{\sqrt{n}}+n^{-\frac{1}{4}}\right]+e^{-n^{\frac{1}{4}}}. \tag{3.75}$$

We note that showing $\mathbb{E}_{\mathsf{C}}[\varepsilon(\mathsf{C})]\to 0$ while $\mathbb{E}_{\mathsf{C}}\left[D\left(P_{Y_{\mathsf{C}}^n}\middle\|P_{Y^n}\right)\right]\to 0$ in (3.67) when $n\to+\infty$ establishes by the Selection lemma (see Lemma A.3) the existence of a code that is both reliable and covert. Letting $n\to\infty$ in (3.75), we know that the expectation of the average probability of error ε will tend to zero provided

$$\lim_{n\to\infty}\frac{\ln|\mathcal{M}|}{\sqrt{n}}<\mathbb{P}\text{-}\liminf_{n\to\infty}\frac{1}{\sqrt{n}}i_{X^n,Y^n}(X^n,Y^n), \tag{3.76}$$

where $\mathbb{P}$-$\liminf$ denotes the *limit inferior in probability*; see Definition A.1. We shall show that the term inside the $\mathbb{P}$-$\liminf$ converges in probability to its expectation

$$\mathbb{E}\left[\frac{1}{\sqrt{n}}i_{X^n,Y^n}(X^n,Y^n)\right]=\frac{1}{\sqrt{n}}I(X^n;Y^n), \tag{3.77}$$

which, combined with (3.72), will yield the desired achievability result. By Chebyshev's inequality (Lemma A.4), for any $a>0$,

$$\mathbb{P}\left[\left|\frac{1}{\sqrt{n}}i_{X^n,Y^n}(X^n,Y^n)-\mathbb{E}\left[\frac{1}{\sqrt{n}}i_{X^n,Y^n}(X^n,Y^n)\right]\right|\geq a\right]\leq\frac{\operatorname{Var}\left(\frac{1}{\sqrt{n}}i_{X^n,Y^n}(X^n,Y^n)\right)}{a^2}. \tag{3.78}$$

Hence, to prove the desired convergence in probability and thereby the desired achievability result, it now only remains to show that the variance on the right-hand side of (3.78) converges to 0 as $n\to\infty$. Recall that Y has the same PDF as $\tilde{Z}_n$, which has the following Taylor expansion

$$p_Y(y)=p_{\tilde{Z}_n}(y)=\alpha_n p_Z(y)\left(1-\gamma_n\ln(p_Z(y))p_Z(y)^{-\theta_n(y)}\right), \tag{3.79}$$

where $\theta_n(y)\in(0,\gamma_n)$ for all $y\in\mathbb{R}$. We now bound the variance in (3.78) as follows:

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{\sqrt{n}}i_{X^n,Y^n}(X^n,Y^n)\right)\\
&\quad=\operatorname{Var}\left(\frac{1}{\sqrt{n}}\ln\left(\frac{p_{Y^n|X^n}(Y^n|X^n)}{p_{Y^n}(Y^n)}\right)\right)\\
&\quad=\operatorname{Var}\left(\ln\left(\frac{p_{Y|X}(Y|X)}{p_Y(Y)}\right)\right) &(3.80)\\
&\quad=\operatorname{Var}\left(\ln\left(\frac{p_Z(Z)}{p_Y(Y)}\right)\right)\\
&\quad=\operatorname{Var}\left(\ln\left(\frac{p_Z(Z)}{\alpha_n p_Z(Y)(1-\gamma_n\ln(p_Z(Y))p_Z(Y)^{-\theta_n(Y)})}\right)\right) &(3.81)
\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{Var}\left(\ln\left(\frac{p_Z(Z)}{p_Z(Y)}\right) - \ln\left(1-\gamma_n \ln(p_Z(Y))p_Z(Y)^{-\theta_n(Y)}\right) - \ln(\alpha_n)\right) \\
&\le \mathbb{E}\left[\left(\ln\left(\frac{p_Z(Z)}{p_Z(Y)}\right) - \ln\left(1-\gamma_n \ln(p_Z(Y))p_Z(Y)^{-\theta_n(Y)}\right)\right)^2\right] \\
&\le 2\,\mathbb{E}\left[\left(\ln\left(\frac{p_Z(Z)}{p_Z(Y)}\right)\right)^2\right] + 2\,\mathbb{E}\left[\left(\ln\left(1-\gamma_n \ln\left(p_Z(Y)\right)p_Z(Y)^{-\theta_n(Y)}\right)\right)^2\right], \qquad (3.82)
\end{aligned}
$$

where (3.80) follows because (X^n, Y^n) are i.i.d.; (3.81) by (3.79); and (3.82) by adding the nonnegative term

$$
\mathbb{E}\left[\left(\ln\left(\frac{p_Z(Z)}{p_Z(Y)}\right) + \ln\left(1-\gamma_n \ln(p_Z(Y))p_Z(Y)^{-\theta_n(Y)}\right)\right)^2\right].
$$

To prove that the variance of the information density approaches zero, it thus suffices to prove that both expectations on the right-hand side of (3.82) approach zero. We start with the second expectation and write it as the sum of two parts:

$$
\begin{aligned}
&\mathbb{E}\left[\left(\ln\left(1-\gamma_n \ln(p_Z(Y))p_Z(Y)^{-\theta_n(Y)}\right)\right)^2\right] \\
&\quad = \mathbb{P}\left[p_Z(Y)\le 1\right]\mathbb{E}\left[\left(\ln\left(1-\gamma_n \ln(p_Z(Y))p_Z(Y)^{-\theta_n(Y)}\right)\right)^2 \middle| p_Z(Y)\le 1\right] \\
&\qquad + \mathbb{P}\left[p_Z(Y)\ge 1\right]\mathbb{E}\left[\left(\ln\left(1-\gamma_n \ln(p_Z(Y))p_Z(Y)^{-\theta_n(Y)}\right)\right)^2 \middle| p_Z(Y)\ge 1\right]. \qquad (3.83)
\end{aligned}
$$

For the second part, we notice that, for any y such that $p_Z(y)\ge 1$ (if such y exists), by (3.55),

$$
\ln(p_Z(y))p_Z(y)^{-\theta_n(y)} \le \ln(p_Z(y)) \le \max\{0, \ln(b)\}, \qquad (3.84)
$$

therefore

$$
\mathbb{E}\left[\left(\ln\left(1-\gamma_n \ln(p_Z(Y))p_Z(Y)^{-\theta_n(Y)}\right)\right)^2 \middle| p_Z(Y)\ge 1\right] \le \left(\ln\left(1-\gamma_n \max\left\{0,\ln(b)\right\}\right)\right)^2, \qquad (3.85)
$$

which clearly tends to zero as $n\to\infty$. We now bound the first term on the right-hand side of (3.83) as follows:

$$
\begin{aligned}
&\mathbb{P}\left[p_Z(Y)\le 1\right]\mathbb{E}\left[\left(\ln\left(1-\gamma_n \ln(p_Z(Y))p_Z(Y)^{-\theta_n(Y)}\right)\right)^2 \middle| p_Z(Y)\le 1\right] \\
&\quad \le \mathbb{P}\left[p_Z(Y)\le 1\right]\mathbb{E}\left[\left(\gamma_n \ln(p_Z(Y))p_Z(Y)^{-\theta_n(Y)}\right)^2 \middle| p_Z(Y)\le 1\right] \\
&\quad = \gamma_n^2 \int_{\mathbb{R}} \mathbb{1}_{\{p_Z(y)\le 1\}} \left(\ln(p_Z(y))\right)^2 p_Z(y)^{-2\theta_n(y)} p_Y(y)\mathrm{d}y \\
&\quad \le \gamma_n^2 \int_{\mathbb{R}} \left(\ln(p_Z(y))\right)^2 p_Z(y)^{-2\theta_n(y)} p_Y(y)\mathrm{d}y \\
&\quad = \gamma_n^2 \int_{\mathbb{R}} \left(\ln(p_Z(y))\right)^2 p_Z(y)^{-2\theta_n(y)} \alpha_n (p_Z(y))^{1-\gamma_n}\mathrm{d}y \\
&\quad = \gamma_n^2 \alpha_n \int_{\mathbb{R}} \left(\ln(p_Z(y))\right)^2 p_Z(y)^{1-\gamma_n-2\theta_n(y)}\mathrm{d}y. \qquad (3.86)
\end{aligned}
$$

By Lemma 3.1, for large enough n, the integral in the above expression is finite. Since $\alpha_n\to 1$ and $\gamma_n\to 0$ as $n\to\infty$, we conclude that the right-hand side of (3.86) tends to zero. We have thus shown that both terms on the right-hand side of (3.83) tend to zero and, therefore, the second expectation on the right-hand side of (3.82) tends to zero as $n\to\infty$.

We now consider the first expectation on the right-hand side of (3.82), which can be written as

$$
\mathbb{E}\left[\left(\ln\left(\frac{p_Z(Z)}{p_Z(Y)}\right)\right)^2\right] = \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right] - 2\mathbb{E}\left[\ln(p_Z(Z))\ln(p_Z(Y))\right] + \mathbb{E}\left[\left(\ln(p_Z(Y))\right)^2\right]. \qquad (3.87)
$$

We first consider the third term on the right-hand side:

$$\begin{aligned}\mathbb{E}\left[\left(\ln(p_Z(Y))\right)^2\right] &= \int_{\mathbb{R}} p_Y(y)\left(\ln(p_Z(y))\right)^2 \mathrm{d}y \\ &= \int_{\mathbb{R}} \alpha_n p_Z(y)(1-\gamma_n \ln(p_Z(y)) p_Z(y)^{-\theta_n(y)})(\ln(p_Z(y)))^2 \mathrm{d}y. \end{aligned} \tag{3.88}$$

By Lemma 3.1 and the fact that $\alpha_n \to 1$ as $n \to \infty$, we have

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\ln(p_Z(Y))\right)^2\right] = \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right]. \tag{3.89}$$

We next consider the second term on the right-hand side of (3.87). We can write this expectation as a double Lebesgue integral:

$$\mathbb{E}\left[\ln(p_Z(Z))\ln(p_Z(Y))\right] = \int_{\mathbb{R}}\int_{\mathbb{R}} p_Z(y-x)\ln(p_Z(y-x))\ln(p_Z(y))\mathrm{d}y\,\mathrm{d}P_X(x). \tag{3.90}$$

Since, by Assumption 3.1, $t \mapsto p_Z(t)\ln(p_Z(t))$ is uniformly continuous on $\mathrm{supp}(p_Z)$, we have that the family of functions $\{t \mapsto p_Z(y-t)\ln(p_Z(y-t))\}_y$ is pointwise equicontinuous wherever $y - t \in \mathrm{supp}(p_Z)$. Again by Assumption 3.1, they are also uniformly bounded. By Lemma 3.3 and Theorem A.3, the following limit holds uniformly over $y \in \mathbb{R}$:

$$\lim_{n\to\infty} \int_{\mathbb{R}} p_Z(y-x)\ln(p_Z(y-x))\mathrm{d}P_X(x) = p_Z(y)\ln(p_Z(y)). \tag{3.91}$$

Therefore,

$$\begin{aligned}&\lim_{n\to\infty}\int_{\mathbb{R}}\int_{\mathbb{R}} p_Z(y-x)\ln(p_Z(y-x))\ln(p_Z(y))\mathrm{d}P_X(x)\,\mathrm{d}y \\ &\qquad = \int_{\mathbb{R}} \ln(p_Z(y)) \lim_{n\to\infty}\int_{\mathbb{R}} p_Z(y-x)\ln(p_Z(y-x))\mathrm{d}P_X(x)\,\mathrm{d}y \\ &\qquad = \int_{\mathbb{R}} \ln(p_Z(y))p_Z(y)\ln(p_Z(y))\,\mathrm{d}y \\ &\qquad = \mathbb{E}\left[\left(\ln(p_Z(Z))\right)^2\right]. \end{aligned} \tag{3.92}$$

Combining (3.87), (3.89), and (3.92), we obtain

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\ln\left(\frac{p_Z(Z)}{p_Z(Y)}\right)\right)^2\right] = 0. \tag{3.93}$$

We have now shown that both expectations on the right-hand side of (3.82) tend to zero as $n \to \infty$. Hence the variance in (3.78) tends to zero as $n \to \infty$, establishing (3.73) and completing the proof. □

3.4 Examples

In this section, we apply Theorems 3.1 and 3.2 to characterize the scaling constant L for some well-known noise distributions. As before, we consider an additive memoryless channel of the form (3.1) with the noise sequence being i.i.d.

3.4.1 Uniform noise

Suppose that the noise Z is of the form (3.44) for some subset $\mathcal{S}$ of $\mathbb{R}$. We will show that in this case, $L = 0$. Suppose there exists an input distribution that yields an output distribution P_Y such that

$$D(P_Y \| P_Z) < \infty. \tag{3.94}$$

For this to hold, P_Y needs to be absolutely continuous with respect to P_Z, i.e., $P_Y \ll P_Z$, and we have

$$\begin{aligned} D(P_Y \| P_Z) &= -h(Y) - \int_{\mathcal{S}} p_Y(y) \ln(p_Z(y)) dy \\ &= -h(Y) - \ln\left(\frac{1}{\lambda(\mathcal{S})}\right) \\ &= h(Z) - h(Y). \end{aligned} \tag{3.95}$$

On the other hand, since $Y = X + Z$, we must have

$$\begin{aligned} I(X;Y) &= h(Y) - h(Z) \\ &= -D(P_Y \| P_Z) \end{aligned} \tag{3.96}$$

where the last step follows by (3.95). Since $I(X;Y) \geq 0$, the only way this can happen is to have $P_Z = P_Y$. That is, the only way to satisfy (3.94) is to have $X = 0$ with probability one. It thus follows that covert communication is not possible and therefore, by definition, $L = 0$.

3.4.2 Exponential noise

Let the noise random variable $Z \sim \mathcal{E}\left(\frac{1}{\Lambda}\right)$ with exponential distribution of mean $\frac{1}{\Lambda} > 0$:

$$p_Z(z) = \Lambda e^{-\Lambda z}, \qquad z \in \mathbb{R}^+. \tag{3.97}$$

This distribution satisfies both (3.2)–(3.4) and Assumption 3.1. Verifying (3.2)–(3.4) and Assumption 3.1 part 1) is straightforward. For part 2), we note that $z \mapsto p_Z(z)\ln(p_Z(z))$ has a bounded derivative, therefore, by the mean value theorem, it is uniformly continuous. To check part 3), for any $\gamma > 0$, we notice that $\tilde{Z}$ defined in (3.9) has the exponential distribution $\mathcal{E}\left(\frac{1}{(1-\gamma)\Lambda}\right)$. It was shown in [72] that there exists X independent of Z such that $\tilde{Z} = X + Z$; see Appendix C.3.
We can hence apply Theorem 3.2 to obtain

$$\begin{aligned} L &= \sqrt{2}\sqrt{\mathrm{Var}\left[\ln(p_Z(Z))\right]} \\ &= \sqrt{2}\sqrt{\mathrm{Var}\left[\ln(\Lambda) - \Lambda Z\right]} \\ &= \sqrt{2}\sqrt{\mathrm{Var}\left[\Lambda Z\right]} \\ &= \sqrt{2}. \end{aligned} \tag{3.98}$$

3.4.3 Generalized Gaussian noise

Consider Z having the generalized Gaussian distribution [55, 24, 25] of parameters $p > 0$ and $\sigma > 0$ (see Figure 8) with probability density function

$$p_Z(z) = \frac{c_p}{\sigma} e^{-\frac{|z|^p}{2\sigma^p}}, \qquad z \in \mathbb{R}, \tag{3.99}$$

where

$$c_p = \frac{p}{2^{\frac{p+1}{p}}\Gamma(\frac{1}{p})}. \tag{3.100}$$

Some known properties of additive channels with generalized Gaussian noise (without covertness constraints) are summarized in Appendix C.4.
The problem of characterizing the scaling constant L for this noise distribution was studied in our previous work [C1]; here we shall obtain the same results by directly applying Theorems 3.1 and 3.2.

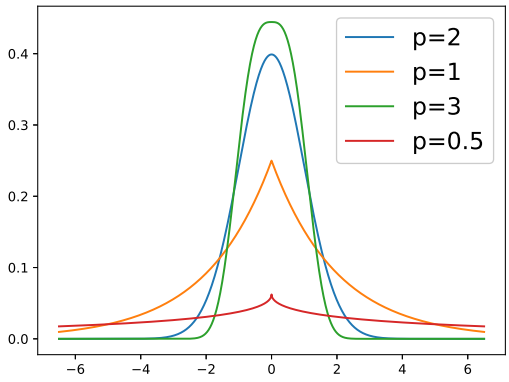

Fig. 8: PDF of generalized Gaussian distributions for $\sigma = 1$.

First we notice that (3.2)–(3.4) are satisfied for all $\sigma > 0$ and $p > 0$. Therefore, we can apply Theorem 3.1 to obtain

$$L \leq \sqrt{2}\sqrt{\mathrm{Var}\left[\frac{|Z|^p}{2\sigma^p}\right]} = \sqrt{\frac{2}{p}}, \tag{3.101}$$

recovering [10, Theorem 2].

When $p = 2$, (3.99) becomes the Gaussian distribution, which satisfies Assumption 3.1. Theorem 3.2 then implies $L = 1$, recovering Theorem 2.5.

When $p \in (0, 1]$, we know that Assumption 3.1 part 3) is satisfied. Indeed, $\tilde{Z}$ defined in (3.9) is also generalized Gaussian with the same p but a different σ; it was shown in [25] that the generalized Gaussian distribution is self-decomposable when $0 < p \leq 1$, meaning that there exists X independent of Z with $\tilde{Z} = X + Z$; see Appendix C.4. The remaining parts of Assumption 3.1 are straightforward to verify. We can therefore apply Theorem 3.2 to obtain

$$L = \sqrt{\frac{2}{p}}, \qquad 0 < p \leq 1, \tag{3.102}$$

recovering [10, Theorem 3].

3.4.4 Generalized gamma noise

Consider Z following the generalized gamma distribution [70, Appendix A.3], [69] of parameters $r, \sigma, \beta > 0$ (see Figure 9):

$$p_Z(z) = \frac{\beta}{\Gamma(r)\sigma^{\beta r}} z^{\beta r - 1} e^{-\left(\frac{z}{\sigma}\right)^\beta} \qquad z \in \mathbb{R}^+. \tag{3.103}$$

One can easily verify that p_Z satisfies (3.2)–(3.4), therefore, by Theorem 3.1, we have

$$L \leq \sqrt{2}\sqrt{\left(r - \frac{1}{\beta}\right)^2 \psi^{(1)}(r) - r + \frac{2}{\beta}}, \tag{3.104}$$

with $\psi^{(1)}$ denoting the first derivative of the digamma function.

Proof: By Theorem 3.1,

$$L \leq \sqrt{2}\sqrt{\mathrm{Var}[\ln(p_Z(Z))]} = \sqrt{2}\sqrt{\mathbb{E}[\ln(p_Z(Z))^2] - \big(h(Z)\big)^2}, \tag{3.105}$$

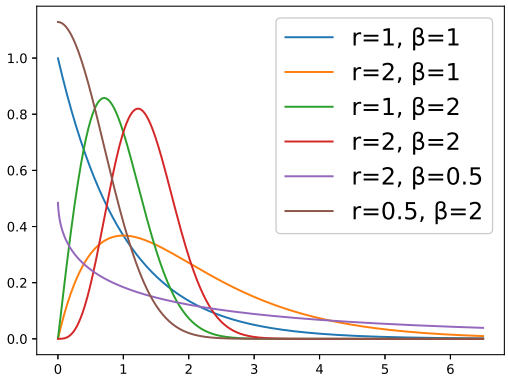

Fig. 9: PDF of generalized gamma distributions for $\sigma = 1$.

where

$$h(Z) = \ln\left(\frac{\Gamma(r)\sigma}{\beta}\right) + r + \left(\frac{1}{\beta} - r\right)\psi(r). \tag{3.106}$$

The other term on the right-hand side of (3.105) can be written as

$$\begin{aligned}
\mathbb{E}[\ln(p_Z(Z))^2] &= \int_0^\infty p_Z(z)\left(\ln\left(\frac{\beta}{\Gamma(r)\sigma^{\beta r}}\right) + (\beta r - 1)\ln(z) - \left(\frac{z}{\sigma}\right)^\beta\right)^2 \mathrm{d}z \\
&= \left(\ln\left(\frac{\beta}{\Gamma(r)\sigma^{\beta r}}\right)\right)^2 + \int_0^\infty p_Z(z)\left(2\ln\left(\frac{\beta}{\Gamma(r)\sigma^{\beta r}}\right)\left((\beta r - 1)\ln(z) - \left(\frac{z}{\sigma}\right)^\beta\right)\right. \\
&\quad \left. + (\beta r - 1)^2\ln(z)^2 - 2(\beta r - 1)\ln(z)\left(\frac{z}{\sigma}\right)^\beta + \left(\frac{z}{\sigma}\right)^{2\beta}\right)\mathrm{d}z \\
&= \left(\ln\left(\frac{\beta}{\Gamma(r)\sigma^{\beta r}}\right)\right)^2 + 2\ln\left(\frac{\beta}{\Gamma(r)\sigma^{\beta r}}\right)\left((\beta r - 1)\mathbb{E}\left[\ln(Z)\right] - \mathbb{E}\left[\left(\frac{Z}{\sigma}\right)^\beta\right]\right) \\
&\quad + (\beta r - 1)^2\mathbb{E}\left[\ln(Z)^2\right] - 2(\beta r - 1)\mathbb{E}\left[\ln(Z)\left(\frac{Z}{\sigma}\right)^\beta\right] + \mathbb{E}\left[\left(\frac{Z}{\sigma}\right)^{2\beta}\right].
\end{aligned} \tag{3.107}$$

Recalling the derivatives of the gamma function

$$\Gamma^{(n)}(k) = \int_0^\infty \ln(t)^n e^{-t} t^{k-1} dt, \qquad \forall n \in \mathbb{N}, \forall k > 0, \tag{3.108}$$

we now proceed to compute all the terms in (3.107):

$$\mathbb{E}\left[\left(\frac{Z}{\sigma}\right)^\beta\right] = r, \tag{3.109}$$

$$\mathbb{E}\left[\left(\frac{Z}{\sigma}\right)^{2\beta}\right] = (r+1)r, \tag{3.110}$$

$$\begin{aligned}
\mathbb{E}[\ln(Z)] &= \int_0^\infty \frac{\beta}{\Gamma(r)\sigma^{\beta r}} z^{\beta r - 1} e^{-\left(\frac{z}{\sigma}\right)^\beta} \ln(z)\mathrm{d}z \\
&= \frac{1}{\sigma}\int_0^\infty \frac{\beta}{\Gamma(r)}\left(\frac{z}{\sigma}\right)^{\beta r - 1} e^{-\left(\frac{z}{\sigma}\right)^\beta} \ln\left(\frac{z}{\sigma}\right)\mathrm{d}z + \ln(\sigma) \\
&= \frac{1}{\beta\Gamma(r)}\int_0^\infty z^{r-1} e^{-z}\ln(z)\mathrm{d}z + \ln(\sigma)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\Gamma^{(1)}(r)}{\beta\Gamma(r)} + \ln(\sigma) \\
&= \frac{1}{\beta}\psi(r) + \ln(\sigma),
\end{aligned} \tag{3.111}
$$

$$
\begin{aligned}
\mathbb{E}[\ln(Z)^2] &= \int_0^\infty \frac{\beta}{\Gamma(r)\sigma^{\beta r}} z^{\beta r-1} e^{-\left(\frac{z}{\sigma}\right)^\beta} \ln(z)^2 \mathrm{d}z \\
&= \frac{1}{\sigma}\int_0^\infty \frac{\beta}{\Gamma(r)} \left(\frac{z}{\sigma}\right)^{\beta r-1} e^{-\left(\frac{z}{\sigma}\right)^\beta} \ln\left(\frac{z}{\sigma}\right)^2 \mathrm{d}z + 2\ln(\sigma)\mathbb{E}[\ln(Z)] - \ln(\sigma)^2 \\
&= \frac{1}{\beta^2\Gamma(r)} \int_0^\infty z^{r-1} e^{-z} \ln(z)^2 \mathrm{d}z + 2\ln(\sigma)\mathbb{E}[\ln(Z)] - \ln(\sigma)^2 \\
&= \frac{\Gamma^{(2)}(r)}{\beta^2\Gamma(r)} + 2\ln(\sigma)\mathbb{E}[\ln(Z)] - \ln(\sigma)^2 \\
&= \frac{1}{\beta^2}\psi^{(1)}(r) + \frac{1}{\beta^2}\psi(r)^2 + 2\frac{\ln(\sigma)}{\beta}\psi(r) + \ln(\sigma)^2,
\end{aligned} \tag{3.112}
$$

$$
\begin{aligned}
\mathbb{E}\left[\ln(Z)\left(\frac{Z}{\sigma}\right)^\beta\right] &= \int_0^\infty \frac{\beta}{\Gamma(r)\sigma^{\beta r}} z^{\beta r-1} e^{-\left(\frac{z}{\sigma}\right)^\beta} \ln(z)\left(\frac{z}{\sigma}\right)^\beta \mathrm{d}z \\
&= \frac{1}{\sigma}\int_0^\infty \frac{\beta}{\Gamma(r)} \left(\frac{z}{\sigma}\right)^{\beta r-1} e^{-\left(\frac{z}{\sigma}\right)^\beta} \ln\left(\frac{z}{\sigma}\right)\left(\frac{z}{\sigma}\right)^\beta \mathrm{d}z + \ln(\sigma)\mathbb{E}\left[\left(\frac{Z}{\sigma}\right)^\beta\right] \\
&= \frac{1}{\beta\Gamma(r)} \int_0^\infty z^r e^{-z} \ln(z)\mathrm{d}z + r\ln(\sigma) \\
&= \frac{\Gamma^{(1)}(r+1)}{\beta\Gamma(r)} + r\ln(\sigma) \\
&= \frac{1}{\beta} r \frac{\Gamma^{(1)}(r)}{\Gamma(r)} + \frac{1}{\beta} + r\ln(\sigma) \\
&= \frac{r}{\beta}\psi(r) + \frac{1}{\beta} + r\ln(\sigma).
\end{aligned} \tag{3.113}
$$

Combining all numbered equations above yields (3.104). □

Remark 3.5 *When $\beta = 1$, (3.103) reduces to the gamma distribution. When $r = \frac{1}{\beta}$, (3.103) reduces to the generalized Gaussian distribution (3.99) with $p = \beta$, but restricted to $\mathbb{R}^+$, and (3.104) becomes the same as (3.101). We cannot apply Theorem 3.2 to obtain an exact characterization of L because the generalized gamma distribution is not known to satisfy Assumption 3.1, in particular, its part 3).*

3.4.5 Cauchy noise

Consider Z following the Cauchy distribution [18, table 17.1] of parameters $\sigma > 0$:

$$
p_Z(z) = \frac{1}{\pi\sigma\left(1 + \left(\frac{z}{\sigma}\right)^2\right)} \qquad z \in \mathbb{R}. \tag{3.114}
$$

One can easily verify that p_Z satisfies (3.2)–(3.4), therefore, by Theorem 3.1, we have

$$
L \leq \sqrt{\frac{2}{3}}\pi. \tag{3.115}
$$

Proof: By Theorem 3.1,

$$
L \leq \sqrt{2}\sqrt{\mathrm{Var}[\ln(p_Z(Z))]} = \sqrt{2}\sqrt{\mathbb{E}[\ln(p_Z(Z))^2] - \big(h(Z)\big)^2}, \tag{3.116}
$$

where the differential entropy is given in [18, table 17.1]:

$$
h(Z) = \ln\left(4\pi\sigma\right). \tag{3.117}
$$

The other term on the right-hand side of (3.116) can be computed as

$$\begin{aligned}
\mathbb{E}\left[\ln\left(p_Z(Z)\right)^2\right] &= \int_{-\infty}^{+\infty} \frac{1}{\pi\sigma\left(1+\left(\frac{z}{\sigma}\right)^2\right)}\left(\ln\left(\pi\sigma\left(1+\left(\frac{z}{\sigma}\right)^2\right)\right)\right)^2 \mathrm{d}z \\
&= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\left(\ln\left(\pi\sigma\left(1+\tan(\theta)^2\right)\right)\right)^2}{\pi}\mathrm{d}\theta \qquad (3.118) \\
&= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\left(\ln\left(\pi\sigma\right)-2\ln\left(\cos(\theta)\right)\right)^2}{\pi}\mathrm{d}\theta \\
&= \left(\ln\left(\pi\sigma\right)\right)^2 - 4\ln\left(\pi\sigma\right)\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\ln\left(\cos(\theta)\right)}{\pi}\mathrm{d}\theta + 4\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\left(\ln\left(\cos(\theta)\right)\right)^2}{\pi}\mathrm{d}\theta \\
&= \ln(\pi\sigma)^2 + 4\ln(2)\ln(\pi\sigma) + \frac{\pi^2}{3} + 4\ln(2)^2 \qquad (3.119)
\end{aligned}$$

where (3.118) follows by the variable change $\frac{z}{\sigma} = \tan(\theta)$ and (3.119) by [34, Section 4.224 eq. 6 and 8]. □

3.5 Bounds on the key length

In Section 3.3, we assumed that an arbitrarily long key was shared between the transmitter and the receiver. In this section, we provide an upper bound in terms of n on the required key length to achieve the optimal L given in Theorem 3.2 and refine this upper bound when the noise has a Gaussian or exponential distribution. Our proofs use channel resolvability techniques (see Section 2.4) to ensure that, on average, the output distribution of the code approaches the "covert process" $\{\tilde{Z}_n\}$[5].

Proposition 3.1 *For the memoryless additive-noise channel (3.1), if p_Z satisfies (3.2)–(3.4) as well as Assumption 3.1, then there exists a sequence of codes that asymptotically achieves the optimal scaling factor L of Theorem 3.2 with key lengths satisfying*

$$\ln|\mathcal{K}| = O(n). \qquad (3.120)$$

Proposition 3.2 *For P_Z being a Gaussian or exponential distribution, (3.120) can be strengthened to*

$$\ln|\mathcal{K}| = o(\sqrt{n}). \qquad (3.121)$$

Remark 3.6 *For Gaussian noise, the result of Proposition 3.2 is essentially known, albeit in settings with several technical differences from ours; for example, apply [7, Theorem 6] to the special case where (in the notation therein) $P_0 = Q_0$ and $P_1 = Q_1$. There are mainly two differences: [7, Theorem 6] assumes that the channel input can only take two different values, whereas we allow any input value in $\mathbb{R}$, furthermore the covertness constraint is different in [7], requiring that the variational distance goes to 0 when n goes to infinity. If we only require the total variation distance to be upper bounded by a constant rather than vanishing, as in [7, Section VII.A] then (also in the notation therein) $\omega_n = O(1)$ and the key length is $o(\sqrt{n})$ since ξ can be chosen to be arbitrarily small. Moreover, if the eavesdropper and legitimate receiver observe two different AWGN channels and the eavesdropper's channel is noisier than the receiver's channel, then one can show that no key is needed (see Appendix D.1). This would be similar to [7, Theorem 6] with $D(Q_1\|Q_0) < D(P_1\|P_0)$. Some other related results on the key length in Gaussian covert communication can be found in [86, 87, 78].*

In order to prove Propositions 3.1 and 3.2, we consider the same random code construction as in the proof of Theorem 3.2 in Section 3.3, where the codewords are generated i.i.d. according to $P_X^{\otimes n}$ such

[5] See discussion in [7, Section III.A].

that $X \sim P_X$ satisfies (3.58). (As before, to avoid cumbersome notation, we do not explicitly write the distributions of X and Y as functions of n.) We denote by C the random codebook and by $P_{Y_{\mathsf{C}}^n}$ the output distribution conditional on C. Recalling (3.68) and choosing $Y \sim \tilde{Z}_n$ where $\tilde{Z}_n$ has PDF (3.59), we shall find sufficient conditions on $\ln|\mathcal{K}|$ such that

$$\lim_{n\to\infty} \mathbb{E}_{\mathsf{C}}\left[D\left(P_{Y_{\mathsf{C}}^n} \| P_{Y^n}\right)\right] = 0, \tag{3.122}$$

which will ensure that C has the desired covertness property with high probability. (The existence with high probability comes from Markov's inequality; Lemma A.2.) By Lemma A.3, this further guarantees the existence of a good deterministic code $\mathcal{C}$ that is covert while achieving L when Theorem 3.2 is applicable.

To establish (3.122), we apply the channel resolvability bound for the Kullback-Leibler divergence by Hayashi and Matsumoto (Theorem 2.1), which asserts that, for $\rho \in (0, 1]$,

$$\mathbb{E}_{\mathsf{C}}\left[D\left(P_{Y_{\mathsf{C}}^n} \| P_{Y^n}\right)\right] \leq \frac{1}{\rho} \ln\left(1 + e^{-\rho \ln(|\mathcal{K}|\times|\mathcal{M}|) + n\Psi(\rho|P_{Y|X}, P_X)}\right) \tag{3.123}$$

where

$$\Psi(\rho|P_{Y|X}, P_X) = \ln\left(\mathbb{E}\left[\left(\frac{p_{Y|X}(Y|X)}{p_Y(Y)}\right)^{\rho}\right]\right) \tag{3.124}$$

and $P_{Y_{\mathsf{C}}^n}$ is defined as in (3.66). (A similar resolvability technique based on Theorem 2.1 was proposed in [87, Lemma 13].)

Proof of Proposition 3.1: For sufficiently small ρ, the expectation in (3.124) can be upper-bounded as

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{p_{Y|X}(Y|X)}{p_Y(Y)}\right)^{\rho}\right] &= \mathbb{E}\left[\left(\frac{p_Z(Y-X)}{p_Y(Y)}\right)^{\rho}\right] \\
&\leq \mathbb{E}\left[\left(\frac{b}{p_Y(Y)}\right)^{\rho}\right] && (3.125) \\
&= b^{\rho}\int_{\mathbb{R}} p_Y(y)^{1-\rho}\mathrm{d}y \\
&= b^{\rho}\int_{\mathbb{R}} \alpha_n^{1-\rho} p_Z(y)^{1-\gamma_n(1-\rho)-\rho}\mathrm{d}y \\
&= \alpha_n^{1-\rho} b^{\rho}\int_{\mathbb{R}} p_Z(y)^{1-\rho}\left(1-\gamma_n(1-\rho)\ln(p_Z(y))p_Z(y)^{-\theta_n(y)}\right)\mathrm{d}y && (3.126) \\
&= b^{\rho}\int_{\mathbb{R}} p_Z(y)^{1-\rho}\mathrm{d}y + O\left(\frac{1}{\sqrt{n}}\right) && (3.127)
\end{aligned}$$

where (3.125) follows by (3.55); (3.126) by the Taylor expansion of $p_Y(y)$ with the Lagrange form of the remainder (3.79); and (3.127) by (3.21), (3.65), and Lemma 3.1. For sufficiently small ρ, Lemma 3.1 ensures that the integral in (3.127) is finite, which in turn implies that (3.124) is bounded. Therefore choosing $\ln|\mathcal{K}| = O(n)$ ensures that the right-hand side of (3.123) goes to 0 as $n \to \infty$, establishing (3.122). □

Proof of Proposition 3.2:

Gaussian noise Let $Z \sim \mathcal{N}(0, \sigma^2)$ with $\sigma > 0$. It is easy to check that the input and output distributions are $X \sim \mathcal{N}\left(0, \frac{\sigma^2\gamma_n}{1-\gamma_n}\right)$ and $Y \sim \mathcal{N}\left(0, \frac{\sigma^2}{1-\gamma_n}\right)$, respectively, and

$$\begin{aligned}
\Psi(\rho|P_{Y|X}, P_X) &= \ln\left(\int_{\mathbb{R}}\int_{\mathbb{R}} p_{X,Y}(x,y)\left(\frac{p_{Y|X}(y|x)}{p_Y(y)}\right)^{\rho}\mathrm{d}x\,\mathrm{d}y\right) \\
&= \ln\left(\int_{\mathbb{R}}\int_{\mathbb{R}} p_Z(y-x)p_X(x)\left(\frac{p_Z(y-x)}{p_Y(y)}\right)^{\rho}\mathrm{d}x\,\mathrm{d}y\right) \\
&= \ln\left(\int_{\mathbb{R}}\int_{\mathbb{R}}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^2\sqrt{\frac{1-\gamma_n}{\gamma_n}}e^{-\frac{(y-x)^2}{2\sigma^2}}e^{-\frac{x^2}{2\sigma^2}\frac{\gamma_n}{1-\gamma_n}}\right.
\end{aligned}$$

$$\times \left(\frac{1}{\sqrt{1-\gamma_n}} \frac{e^{-\frac{(y-x)^2}{2\sigma^2}}}{e^{-\frac{y^2}{2\sigma^2}\frac{1}{1-\gamma_n}}} \right)^{\rho} \mathrm{d}x\,\mathrm{d}y \Bigg)$$

$$= \ln \left(\int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^2 \frac{(1-\gamma_n)^{\frac{1}{2}-\frac{\rho}{2}}}{\sqrt{\gamma_n}} e^{-\frac{\left(y-\frac{1+\rho}{1+\rho\gamma_n}x\right)^2}{2\sigma^2\frac{1}{1+\rho\gamma_n}}} \right.$$

$$\left. \times e^{-\frac{x^2}{2\sigma^2}\frac{(1-\gamma_n)(1-\rho^2\gamma_n)}{\gamma_n(1+\rho\gamma_n)}} \mathrm{d}y\,\mathrm{d}x \right) \tag{3.128}$$

$$= \ln \left(\frac{(1-\gamma_n)^{\frac{1}{2}-\frac{\rho}{2}}}{\sqrt{\gamma_n}} \frac{1}{\sqrt{1+\rho\gamma_n}} \sqrt{\frac{\gamma_n(1+\rho\gamma_n)}{(1-\gamma_n)(1-\rho^2\gamma_n)}} \right)$$

$$= \ln \left(\frac{(1-\gamma_n)^{-\frac{\rho}{2}}}{\sqrt{1-\rho^2\gamma_n}} \right), \tag{3.129}$$

where (3.128) follows by Fubini's theorem.

If the message rate scales according to the optimal scaling constant L as in Theorem 3.2, i.e., $\lim_{n\to\infty} \frac{\ln|\mathcal{M}|}{\sqrt{\Delta}\sqrt{n}} = 1$, then there exists a positive sequence $\{\xi_n\}$ such that $\xi_n = o(\sqrt{n})$ and $\ln|\mathcal{M}| \geq \sqrt{\Delta}\sqrt{n} - \xi_n$. Let $\{\rho_n\}, \rho_n \in (0,1)$ for any n, be such that $\rho_n \to 0$ and $\rho_n\xi_n \to \infty$ when $n \to \infty$. Continuing from (3.129), we obtain

$$\begin{aligned}
\Psi(\rho_n|P_{Y|X}, P_X) &= -\frac{\rho_n}{2}\ln(1-\gamma_n) - \frac{1}{2}\ln\left(1-\rho_n^2\gamma_n\right) \\
&= -\frac{\rho_n}{2}\left(-\gamma_n + O(\gamma_n^2)\right) - \frac{1}{2}\left(-\rho_n^2\gamma_n + O(\rho_n^4\gamma_n^2)\right) \\
&= \left(\frac{\rho_n}{2} + \frac{\rho_n^2}{2}\right)\gamma_n + O\left(\rho_n\gamma_n^2\right) \\
&= \rho_n(1+\rho_n)\sqrt{\frac{\Delta}{n}} + O\left(\frac{\rho_n}{n}\right),
\end{aligned} \tag{3.130}$$

where (3.130) follows by recalling the expression of γ_n in (3.65). We have now established the upper bound

$$\begin{aligned}
\mathbb{E}_{\mathsf{C}}\left[D\left(P_{Y_{\mathsf{C}}^n}\|P_{Y^n}\right)\right] &\leq \frac{1}{\rho_n}\ln\left(1 + e^{-\rho_n(\ln|\mathcal{K}|+\ln|\mathcal{M}|)+\rho_n(1+\rho_n)\sqrt{\Delta}\sqrt{n}+O(\rho_n)}\right) \\
&\leq \frac{1}{\rho_n}e^{-\rho_n\left(\ln|\mathcal{K}|-\rho_n\sqrt{\Delta}\sqrt{n}-\xi_n+O(1)\right)}.
\end{aligned} \tag{3.131}$$

A key length $\ln|\mathcal{K}| = \rho_n\sqrt{\Delta}\sqrt{n} + 2\xi_n$ would be sufficient to ensure (3.122). Since $\rho_n \to 0$ and $\xi_n = o(\sqrt{n})$, it follows that $\ln|\mathcal{K}| = o(\sqrt{n})$.

Exponential noise For Z having the exponential distribution (3.97) of mean $\frac{1}{\Lambda} > 0$, the target output distribution P_Y is an exponential distribution of mean $\frac{1}{(1-\gamma_n)\Lambda}$. The input distribution that induces this output distribution is a mixture of a point mass at 0 and an exponential distribution [72]; see Appendix C.11:

$$\mathbb{P}[X=0] = 1-\gamma_n, \tag{3.132}$$

$$\mathbb{P}[X > x | X > 0] = e^{-(1-\gamma_n)\Lambda x}. \tag{3.133}$$

We compute $\Psi(\rho|P_{Y|X}, P_X)$ as follows:

$$\begin{aligned}
\Psi(\rho|P_{Y|X}, P_X) &= \ln\left(\int_{\mathbb{R}}\int_{\mathbb{R}} p_Z(y-x)\left(\frac{p_Z(y-x)}{p_Y(y)}\right)^{\rho}\mathrm{d}P_X(x)\,\mathrm{d}y\right) \\
&= \ln\left((1-\gamma_n)\int_{\mathbb{R}^+}\Lambda e^{-\Lambda y}\left(\frac{1}{1-\gamma_n}\frac{e^{-\Lambda y}}{e^{-\Lambda(1-\gamma_n)y}}\right)^{\rho}\mathrm{d}y\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad + \gamma_n \int_{\mathbb{R}^+} \int_0^y \Lambda^2 (1-\gamma_n) e^{-\Lambda(y-x)} e^{-\Lambda(1-\gamma_n)x} \left(\frac{1}{1-\gamma_n} \frac{e^{-\Lambda(y-x)}}{e^{-\Lambda(1-\gamma_n)y}} \right)^{\rho} \mathrm{d}x \, \mathrm{d}y \Bigg) \\
&= \ln \Bigg((1-\gamma_n)^{1-\rho} \int_{\mathbb{R}^+} \Lambda e^{-\Lambda(1+\rho\gamma_n)y} \mathrm{d}y \\
&\quad + \gamma_n \int_{\mathbb{R}^+} \int_0^y \Lambda^2 (1-\gamma_n)^{1-\rho} e^{-\Lambda(1+\gamma_n\rho)y} e^{\Lambda(\gamma_n+\rho)x} \mathrm{d}x \, \mathrm{d}y \Bigg) \\
&= \ln \left(\frac{(1-\gamma_n)^{1-\rho}}{1+\rho\gamma_n} + \gamma_n \int_{\mathbb{R}^+} \Lambda \frac{(1-\gamma_n)^{1-\rho}}{\gamma_n+\rho} e^{-\Lambda(1+\gamma_n\rho)y} \left(e^{\Lambda(\gamma_n+\rho)y} - 1 \right) \mathrm{d}y \right) \\
&= \ln \left(1 - \gamma_n + \gamma_n \frac{(1-\gamma_n)^{1-\rho}}{\gamma_n+\rho} \left(\frac{1}{1+\gamma_n\rho-\gamma_n-\rho} - \frac{1}{1+\gamma_n\rho} \right) + o(\gamma_n) \right) \\
&= \frac{\rho}{1-\rho} \gamma_n + o(\gamma_n) \\
&= \frac{\sqrt{2}\rho}{1-\rho} \sqrt{\frac{\Delta}{n}} + o\left(\frac{1}{\sqrt{n}} \right), \qquad (3.134)
\end{aligned}
$$

where (3.134) follows by recalling the expression of γ_n in (3.65). By a similar reasoning to the Gaussian case, it follows that $\ln |\mathcal{K}| = o(\sqrt{n})$ suffices to ensure (3.122). □

Remark 3.7 *The proof of Proposition 3.2 requires the characterization of the optimal input P_X, which can be challenging for more general noise distributions.*

3.6 Concluding Remarks

There are not many examples of additive-noise channels whose capacity under a certain input cost constraint [33, 74] admits a closed-form expression, notably the AWGN channel with a second-moment constraint, the exponential-noise channel with a first-moment constraint [72], the channel with Cauchy noise and logarithmic constraint [26], and some generalized Gaussian-noise channels [24]. Some works provide bounds on the capacity or show properties of optimal input distributions, e.g., [22, 27].

In contrast, we were able to derive a simple expression for the scaling constant L for covert communication over rather general additive-noise channels. This is because, in a sense, the covertness condition (2.6) translates to a constraint on the output that is naturally "fitted" to the noise distribution[6], which in turn allows for a precise characterization of (or an elegant upper bound on) L.

It is not clear to us how to generalize Theorems 3.1 and 3.2 to scenarios where the receiver and the eavesdropper face different noise distributions, because the cost constraint would now be fitted to the eavesdropper's noise, not the receiver's. A special case where such generalization is straightforward is the Gaussian channel where the receiver and the eavesdropper are corrupted by Gaussian noise with different variances σ^2 and σ_e^2, respectively. In this case, it is well known that $L = \sigma_e^2/\sigma^2$. This example is presented in detail in Appendix D.1. See also [78, Theorem 2]. We can not derive such a simple computation of L for other types of noise but are able to derive an upper bound for exponential channels, involving only the ratio between the two means of the noise; see Appendix D.2.

Validity of the formula (3.57) is limited by Assumption 3.1, notably its part 3). For example, 3) is not known to be true for some generalized Gaussian distributions [25]. However, 3) being false (or not known to be true or false) for a certain noise PDF does not necessarily imply that the formula (3.57) cannot hold for such noise. This is because the input distribution to achieve L does not need to be unique, i.e., we do not need the PDF of $X+Z$ to be exactly (3.9). For example, for the AWGN channel, (3.9) is a Gaussian distribution, and X should also be Gaussian in order to induce this output distribution. However, choosing X to take only two values, $\pm a$ for some a approaching zero as $n \to \infty$, can also attain L [76]. If one can show that similar input distributions can attain L on other additive-noise channels, in

[6] For example, for Gaussian noise, it translates into a second-moment constraint, and for exponential noise into a first-moment constraint.

particular, those that do not satisfy Assumption 3.1, then one may be able to further extend the validity of (3.57).

4 Second-order asymptotics of covert communication

In this section, we present the original contribution presented in part in [C2].
Up to now, we have considered covert communication in the asymptotic regime where the number of channel uses tends to infinity. In this chapter, we will consider the finite blocklength regime and provide some bounds for the first and second-order asymptotics of covert communication over additive memoryless Gaussian channels.

4.1 Introduction

Channel coding in the finite blocklength regime (without covertness constraint) In the context of channel coding, the seminal paper of Polyanskiy, Poor, and Verdù [59] characterized the maximal channel coding rate achievable for a given average or maximal probability of error ε and blocklength n. A code $\mathcal{C}$ of blocklength n, message size $M = |\mathcal{M}|$ and average (respectively maximal) error probability ε is called an (n, M, ε)-code. We denote

$$M^*(n, \varepsilon) = \max \left\{ M \mid \exists (n, M, \varepsilon)\text{-code} \right\}. \tag{4.1}$$

In particular [59] showed that for discrete memoryless channels, both for average and maximal probability of error, the maximum code size asymptotically scales like [59, eq. (223)]

$$\ln\left(M^*(n, \varepsilon)\right) = nC - \sqrt{nV}\; Q^{-1}(\varepsilon) + O\left(\ln(n)\right) \tag{4.2}$$

where C is the capacity and V is a characteristic of the channel referred to as *channel dispersion*. The capacity C is known as the first-order asymptotics and $\sqrt{\frac{V}{n}}Q^{-1}(\varepsilon)$ as the second-order asymptotic. An expression of the form (4.2) also holds for Gaussian channels for maximal probability of error under maximal, equal, and average power constraint. However, for the average probability of error, it holds only under an equal and maximal power constraint. In fact, for average probability of error and average power constraint, it turns out that the strong converse does not hold, i.e. the ε-capacity [61, Definition 19.3]

$$C_\varepsilon = \liminf_{n \to +\infty} \frac{\ln\left(M^*(n, \varepsilon)\right)}{n} \tag{4.3}$$

is not equal to the capacity for $\varepsilon > 0$. In this setting, the ε-capacity has been characterized as follows:

Theorem 4.1 *[58, Theorem 77] For the AWGN channel with average power constraint, for SNR P and average probability of error $0 < \varepsilon < 1$ we have*

$$\ln\left(M^*(n, \varepsilon)\right) = \frac{n}{2} \ln\left(1 + \frac{P}{1 - \varepsilon}\right) + O(n^{\frac{2}{3}}). \tag{4.4}$$

Covert communication in the finite blocklength regime We define the maximal probability of decoding error as

$$\varepsilon_{\max} = \max_{m \in \mathcal{M}, k \in \mathcal{K}} \mathbb{P}\left[g(y^n, k) \neq m \mid x^n = f(m, k)\right] \tag{4.5}$$

and the average probability of decoding error

$$\varepsilon = \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} \mathbb{P}\left[g(y^n, k) \neq m \mid x^n = f(m,k)\right]. \tag{4.6}$$

A code $\mathcal{C}$ of blocklength n, message size $M = |\mathcal{M}|$, average (respectively maximal) error probability ε which satisfies the covertness constraint (2.6) is called an $(n, M, \varepsilon, \Delta)_{\text{avg}}$-code (respectively an $(n, M, \varepsilon, \Delta)_{\max}$-code). We denote

$$M^*(n, \varepsilon, \Delta) = \max\left\{M \mid \exists (n, M, \varepsilon, \Delta)_{\text{avg}}\text{-code}\right\}, \tag{4.7}$$

$$\bar{M}^*(n, \varepsilon, \Delta) = \max\left\{M \mid \exists (n, M, \varepsilon, \Delta)_{\max}\text{-code}\right\}. \tag{4.8}$$

Covert communication in finite blocklength was first investigated by Tahmasbi and Bloch [71] who characterized the first and second-order asymptotics for discrete memoryless channels with a given maximal probability of error ε and found that the first-order asymptotics scale like the square root of the blocklength while the second-order asymptotics scale like the fourth root of the blocklength. Furthermore, under a maximal error probability constraint, the first-order asymptotics do not depend on ε [71].
However, under an average error probability constraint, the strong converse does not hold and the first-order asymptotics of covert communication depend on the error probability ε. In fact, as explained in [71, Appendix A], assuming that $\ln(M^*(n, \varepsilon, \Delta)) = f(\Delta)\sqrt{n} + o(\sqrt{n})$ for some function f which is strictly increasing in Δ and does not depend on ε leads to a contradiction. Essentially, this follows from the fact that one can add a certain amount of all zero-codewords which increases the size of the codebook without affecting the covertness constraint.
In this chapter, we show that for an AWGN channel, the covertness constraint (2.43) implies an average power constraint. This allows us to establish an upper bound for the second-order asymptotics of covert communication over an AWGN channel when considering a maximal error probability constraint. A similar upper bound and matching lower bound were stated in [84, 83]; however, no rigorous proof of covertness is given for the lower bound and the upper bound's proof relies on a maximal power constraint in [84] and on a second-moment constraint in [83].
When considering an average error probability constraint, we show that similarly to the case of DMCs, the strong converse does not hold and the first-order asymptotics depend on ε. Similarly to the notion of ε-capacity, we can define an ε-scaling constant L_ε for the first-order asymptotics, for which we establish a straightforward upper bound derived directly from Fano's inequality and new lower bounds.

Notation and preliminaries Channel inputs can be subjected to one of three types of constraints:

- *equal power constraint*: $M_e^*(n, \varepsilon, P)$ denotes the maximal size of the message set $\mathcal{M}$, for n uses of the channel, and average probability of error ε such that each input $f(m) \in \mathcal{X}^n$, for all $m \in \mathcal{M}$ satisfies

$$\|f(m)\|_2^2 = nP, \tag{4.9}$$

- *maximal power constraint*: $M_m^*(n, \varepsilon, P)$ denotes the maximal size of the message set $\mathcal{M}$, for n uses of the channel, and average probability of error ε such that each input $f(m) \in \mathcal{X}^n$, for all $m \in \mathcal{M}$ satisfies

$$\|f(m)\|_2^2 \leq nP, \tag{4.10}$$

- *average power constraint*: $M_a^*(n, \varepsilon, P)$ denotes the maximal size of the message set $\mathcal{M}$, for n uses of

the channel, and average probability of error ε such the codebook satisfies

$$\frac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|}\|f(m)\|_2^2 \leq nP. \tag{4.11}$$

Similarly, for a maximal probability of error ε, we denote by $\bar{M}_e^*(n,\varepsilon,P)$, $\bar{M}_m^*(n,\varepsilon,P)$, and $\bar{M}_a^*(n,\varepsilon,P)$ the maximum cardinalities of a code of blocklength n under the constraints (4.9), (4.10) and (4.11) respectively.

Lemma 4.1 *[58, Lemma 65][67, Section XIII] For any $P' > P > 0$, given the maximal probability of decoding error $0 \leq \varepsilon \leq 1$, the inequalities*

$$\bar{M}_e^*(n,\varepsilon,P) \leq \bar{M}_m^*(n,\varepsilon,P) \leq \bar{M}_e^*(n+1,\varepsilon,P) \tag{4.12}$$

and

$$\bar{M}_m^*(n,\varepsilon,P) \leq \bar{M}_a^*(n,\varepsilon,P) \leq \frac{1}{1-\frac{P}{P'}}\bar{M}_m^*\left(n,\varepsilon,P'\right) \tag{4.13}$$

hold. Moreover, in the average probability of error formalism (4.12) holds without change, while (4.13) becomes

$$M_m^*(n,\varepsilon,P) \leq M_a^*(n,\varepsilon,P) \leq \frac{1}{1-\frac{P}{P'}}M_m^*\left(n,\frac{\varepsilon}{1-\frac{P}{P'}},P'\right) \tag{4.14}$$

which holds provided that $\frac{\varepsilon}{1-\frac{P}{P'}} \leq 1$.

Proof:

1) The left-hand sides of (4.12), (4.13) and (4.14) are trivial since a code satisfying the equal power constraint (4.9) also satisfies the maximal power constraint (4.10), and a code satisfying the maximal power constraint (4.10) also satisfies the average power constraint (4.11).
2) The right-hand side of (4.12) is shown with the following argument: given an (n,M,ε)-code under the maximal power constraint P, we can construct a new code with blocklength $n+1$, same error probability ε and equality constraint P by adding to each codeword a further coordinate. Furthermore, the probability of error for the new code is at most as great as that of the first code, since the added coordinate can only improve the decoding process; e.g. the decoder could ignore the last coordinate and then recover the same probability of error.
3) The right-hand side of (4.13) is shown with the following: consider an (n,M,ε)-code under average power constraint P and with maximal probability of decoding error ε. We can construct from the (n,M,ε)-code a new code under a maximal power constraint P' with the same maximal probability of error ε by removing any codeword c such that $\|c\|_2^2 > nP'$. The probability to find a codeword such that $\|c\|_2^2 > nP'$ is bounded through Markov's inequality (Lemma A.2) as

$$\begin{aligned}\mathbb{P}\left[\|c\|_2^2 > nP'\right] &\leq \frac{\mathbb{E}\left[\|c\|_2^2\right]}{nP'} \\ &= \frac{P}{P'}.\end{aligned} \tag{4.15}$$

Furthermore removing codewords will not increase the maximal probability of error, allowing us to conclude.
4) Finally the right-hand side of (4.14) is shown with the following. Let $P' > P > 0$, consider an (n,M,ε)-code $\mathcal{C}$ under average power constraint P and average probability of error ε. As previously we can construct a new code under maximal power constraint P' by removing any codeword $c \in \mathcal{C}$ such that $\|c\|_2^2 > nP'$. We denote $\mathcal{C}_1 = \left\{c \in \mathcal{C} : \|c\|_2^2 > nP'\right\}$ and $\mathcal{C}_2 = \left\{c \in \mathcal{C} : \|c\|_2^2 \leq nP'\right\}$. The cardinality of the code $\mathcal{C}_2$ after expurgation is greater or equal to $M(1-P/P')$. Let ε_1 be the

average error for the codewords in $\mathcal{C}_1$, and ε_2 the average error for codewords in $\mathcal{C}_2$. We now bound the average probability of error ε_2 of the new code after expurgation. By definition,

$$\varepsilon = \varepsilon_1 \frac{|\mathcal{C}_1|}{|\mathcal{C}|} + \varepsilon_2 \frac{|\mathcal{C}_2|}{|\mathcal{C}|}. \tag{4.16}$$

Thus, the new error probability ε_2 is upper bounded as

$$\varepsilon_2 \leq \varepsilon \frac{|\mathcal{C}|}{|\mathcal{C}_2|} = \frac{\varepsilon}{1 - \frac{P}{P'}}. \tag{4.17}$$

This concludes the proof.

□

Lemma 4.2 *[61, Theorem 19.4] For any $0 < \tau < 1$, $P > 0$, the following inequalities hold:*

$$\tau M_m^*(n, \varepsilon(1-\tau), P) \leq \bar{M}_m^*(n, \varepsilon, P) \leq M_m^*(n, \varepsilon, P) \tag{4.18}$$

$$\tau M_e^*(n, \varepsilon(1-\tau), P) \leq \bar{M}_e^*(n, \varepsilon, P) \leq M_e^*(n, \varepsilon, P). \tag{4.19}$$

Proof: The right-hand side inequalities are obvious. To prove the inequalities on the left-hand side, consider an (n, M, ε)-code, and define the error probability for the i^{th} codeword as

$$\lambda_i = \mathbb{P}\left[\hat{m} \neq m_i | m = m_i\right]. \tag{4.20}$$

Then by Markov's inequality (Lemma A.2), we have

$$\begin{aligned} \frac{\left|\left\{i \text{ s.t. } \lambda_i > \frac{\varepsilon}{1-\tau}\right\}\right|}{M} &= \mathbb{P}\left[\lambda_i > \frac{\varepsilon}{1-\tau}\right] \\ &\leq 1 - \tau. \end{aligned} \tag{4.21}$$

Now by removing those codewords whose λ_i exceeds $\frac{\varepsilon}{1-\tau}$, we can extract from the (n, M, ε)-code a new $(n, M', \frac{\varepsilon}{1-\tau})_{\max}$-code. Note that removing codewords does not affect equal power constraint or maximal power constraint. Furthermore, the constructed code has size satisfying

$$\tau M \leq M', \tag{4.22}$$

completing the proof. □

4.2 Upper bound on the first and second-order asymptotics for maximal probability of error

We prove the following upper bound for the first and second-order asymptotics of covert communication over an AWGN channel for a maximal probability of error ε. Note that due to the square root law, the convergence to L is slow (see Figure 10).

Theorem 4.2 *Consider the AWGN channel (2.42) subject to the covertness constraint (2.43). For the maximal probability of error $0 < \varepsilon < 1$, the maximum code size admits the following upper bound:*

$$\ln(\bar{M}^*(n, \varepsilon, \Delta)) \leq \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}} n^{\frac{1}{4}} Q^{-1}(\varepsilon) + O\left(\ln\left(n\right)\right). \tag{4.23}$$

First, we show that for the Gaussian channel, the covertness condition (2.43) induces an average power constraint.

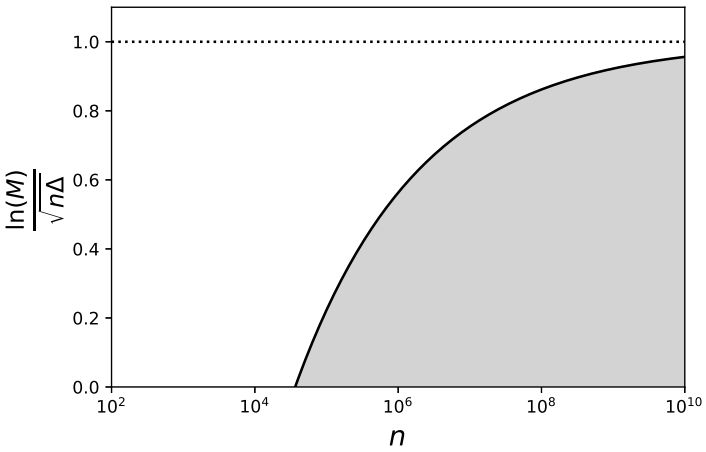

Fig. 10: Upper bound on the first and second-order asymptotics for $\Delta = 10^{-2}$ and maximal error probability $\varepsilon = 10^{-3}$ as a function of the blocklength n.

Lemma 4.3 *For the AWGN channel (2.42), the covertness constraint (2.43) implies the average power constraint:*

$$\rho_n \leq 2\sigma^2 \sqrt{\frac{\Delta}{n}} + O\left(\frac{1}{n}\right), \tag{4.24}$$

where

$$\rho_n = \frac{1}{n} \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} \|f(m,k)\|_2^2 \tag{4.25}$$

is the average power of the code.

Proof: Take any code $\mathcal{C}$ of length n satisfying the covertness constraint (2.43). As in equation (3.34), let $\bar{X}$ denote a random variable such that $P_{\bar{X}}$ is the average input distribution of the code over the secret key, a uniformly drawn message, and the n channel uses. Let $\bar{Y}$ denote the corresponding channel output as in equation (3.35) so that $P_{\bar{Y}}$ is the average output distribution of the code. We notice that

$$\begin{aligned}
\mathbb{E}\left[\bar{Y}^2\right] &= \int_{\mathbb{R}} y^2 p_{\bar{Y}}(y) \mathrm{d}y \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}} y^2 p_{Y_i}(y) \mathrm{d}y \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} \int_{\mathbb{R}} y^2 p_{Y|X}(y|(f(m,k)_i) \mathrm{d}y \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} (\sigma^2 + (f(m,k)_i^2) \\
&= \rho_n + \sigma^2,
\end{aligned} \tag{4.26}$$

where $f(m,k)_i$ is the i^{th} component of the input vector $f(m,k)$. Starting with the covertness condition (2.43), similarly to [77, eq (13)] we have:

$$\begin{aligned}
\Delta &\geq D(P_{Y^n} \| P_{Z^n}) \\
&\geq n\, D(P_{\bar{Y}} \| P_Z) \qquad (4.27) \\
&= n \left(-h(\bar{Y}) - \mathbb{E}\left[\ln(p_Z(\bar{Y}))\right]\right) \\
&\geq n \left(-\frac{1}{2} \ln(2\pi(\rho_n + \sigma^2)e) - \mathbb{E}\left[\ln(p_Z(\bar{Y}))\right]\right) \qquad (4.28)
\end{aligned}$$

$$
\begin{aligned}
&= n\left(-\frac{1}{2}\ln(2\pi(\rho_n+\sigma^2)e)+\frac{1}{2}\ln(2\pi\sigma^2)+\frac{1}{2\sigma^2}(\rho_n+\sigma^2)\right)\\
&= n\left(-\frac{1}{2}\ln\left(\frac{\rho_n+\sigma^2}{\sigma^2}\right)+\frac{1}{2}\frac{\rho_n}{\sigma^2}\right)\\
&\geq n\left(-\frac{1}{2}\left(\frac{\rho_n}{\sigma^2}-\frac{1}{2}\frac{\rho_n^2}{\sigma^4}+\frac{1}{3}\frac{\rho_n^3}{\sigma^6}\right)+\frac{1}{2}\frac{\rho_n}{\sigma^2}\right)\\
&= n\left(\frac{1}{4}\frac{\rho_n^2}{\sigma^4}-\frac{1}{6}\frac{\rho_n^3}{\sigma^6}\right) \qquad (4.29)
\end{aligned}
$$

where (4.27) follows from the same steps as (3.36); (4.28) holds since the Gaussian maximizes entropy among all distributions with the same second moments. Finally, we conclude that covertness imposes the average power constraint on the code (4.24) by noticing that in order to satisfy (4.28) ρ_n needs to vanish with the blocklength n and computing a Taylor series. □

Second, we characterize an upper bound for the first and second-order asymptotics under a maximal probability of error and with an *equal* power constraint which is vanishing in n. This intermediate result will be used to prove Theorem 4.2.

Lemma 4.4 *Consider a code for message set $\mathcal{M}$ over the AWGN channel (2.42) subject to the equality power constraint*

$$
\|f(m)\|_2^2 = nP, \qquad \forall m \in \mathcal{M}, \tag{4.30}
$$

where

$$
P = 2\sigma^2\sqrt{\frac{\Delta}{n}}\left(1+\eta_n\right), \qquad \eta_n = \frac{\ln(n)}{\sqrt{n}}. \tag{4.31}
$$

For the maximal probability of error $0<\varepsilon<1$, the maximum code size admits the following upper bound:

$$
\ln(\bar{M}_e^*(n,\varepsilon,P)) \leq \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}(\varepsilon) + O\left(\ln\left(n\right)\right). \tag{4.32}
$$

Proof: We consider an (n,M,ε)-code $\mathcal{C}$ for message set $\mathcal{M}$ under the maximal probability of error ε and the equal power constraint (4.30). We fix Q_{Y_n} to be a centered Gaussian distribution with variance σ^2+P. We now consider hypothesis testing between the output distributions $P_{Y^n|X^n=x^n}$ and Q_{Y^n}. Recall that by the Neyman-Pearson Lemma B.1, given $\varepsilon\in[0,1]$, there exists $\gamma\in\mathbb{R}$ such that the hypothesis test

$$
\begin{aligned}
T_\gamma : \mathcal{Y}^n &\to \{0,1\}\\
y^n &\mapsto \mathbb{1}_{\left\{\frac{\mathrm{d}P_{Y^n|X^n=x^n}}{\mathrm{d}Q_{Y^n}}(y^n)\geq\gamma\right\}}
\end{aligned} \tag{4.33}
$$

(where 1 means the test chooses $P_{Y^n|X^n=x^n}$) achieves the minimal false positive error under Q_{Y^n}

$$
\beta_{1-\varepsilon}(P_{Y^n|X^n=x^n},Q_{Y^n}) = \min_{\int_{\mathcal{Y}^n}T_\gamma(y^n)\ \mathrm{d}P_{Y^n|X^n=x^n}\geq 1-\varepsilon}\int_{\mathcal{Y}^n}T_\gamma(y)\ \mathrm{d}Q_{Y^n} \tag{4.34}
$$

such that the probability of false negative error under $P_{Y^n|X^n=x^n}$ is not larger than ε.

Developing the expression for $\beta_{1-\varepsilon}$, we find that for any $x^n\in\mathbb{R}^n$ satisfying the equal power constraint (4.65),

$$
\begin{aligned}
&\beta_{1-\varepsilon}(P_{Y^n|X^n=x^n},Q_{Y^n})\\
&\quad= Q_{Y^n}\left[\ln\left(\frac{\frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\prod_{i=1}^n e^{-\frac{(Y_i-x_i)^2}{2\sigma^2}}}{\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+P)^{\frac{n}{2}}}\prod_{i=1}^n e^{-\frac{Y_i^2}{2(\sigma^2+P)}}}\right)\geq\ln(\gamma)\right]\\
&\quad= Q_{Y^n}\left[\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right)-\frac{P}{2\sigma^2(\sigma^2+P)}\sum_{i=1}^n Y_i^2+\sum_{i=1}^n\frac{1}{\sigma^2}Y_ix_i-\sum_{i=1}^n\frac{1}{2\sigma^2}x_i^2\geq\ln(\gamma)\right]
\end{aligned}
$$

$$= Q_{Y^n}\left[\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right)+\frac{n}{2}-\frac{P}{2\sigma^2(\sigma^2+P)}\sum_{i=1}^{n}\left(Y_i-\frac{\sigma^2+P}{P}x_i\right)^2\geq\ln(\gamma)\right], \tag{4.35}$$

where

$$\begin{aligned}1-\varepsilon &\leq P_{Y^n|X^n=x^n}\left[\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right)+\frac{n}{2}-\frac{P}{2\sigma^2(\sigma^2+P)}\sum_{i=1}^{n}\left(Y_i-\frac{\sigma^2+P}{P}x_i\right)^2\geq\ln(\gamma)\right]\\ &= \mathbb{P}\left[\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right)+\frac{n}{2}-\frac{P}{2\sigma^2(\sigma^2+P)}\sum_{i=1}^{n}\left(Z_i-\frac{\sigma^2}{P}x_i\right)^2\geq\ln(\gamma)\right]. \end{aligned}\tag{4.36}$$

Recalling that the cumulative distributive function of a non-central chi-squared random variable depends only on the norm $\|\cdot\|_2$ of the means (see Appendix A.6); we notice that

$$\beta_{1-\varepsilon}(P_{Y^n|X^n=x^n},Q_{Y^n})=\beta_{1-\varepsilon} \tag{4.37}$$

does not depend on x^n. Therefore by Theorem B.2 we know that

$$|\mathcal{M}|\leq\frac{1}{\beta_{1-\varepsilon}} \tag{4.38}$$

and without loss of generality, we can choose $x^n=(\sqrt{P},\sqrt{P},\dots,\sqrt{P})$. éass Furthermore, as a consequence of (B.16) we have: for any $\gamma>0$,

$$\beta_{1-\varepsilon}\geq\sup_{\gamma>0}\frac{1}{\gamma}\left(1-\varepsilon-P_{Y^n|X^n=x^n}\left[\frac{\mathrm{d}P_{Y^n|X^n=x^n}}{\mathrm{d}Q_{Y^n}}(Y^n|x^n)\geq\gamma\right]\right). \tag{4.39}$$

Consider any $\xi_n\in\mathbb{R}$ and

$$\gamma=\exp\left(\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right)-\xi_n\right), \tag{4.40}$$

then (4.39) can be rewritten as

$$\begin{aligned}\beta_{1-\varepsilon}\geq{}&\exp\left(-\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right)+\xi_n\right)\\ &\times\left(1-\varepsilon-P_{Y^n|X^n=x^n}\left[\ln\left(\frac{\mathrm{d}P_{Y^n|X^n=x^n}}{\mathrm{d}Q_{Y^n}}(Y^n|x^n)\right)\geq\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right)-\xi_n\right]\right).\end{aligned}\tag{4.41}$$

We can lower bound the right-hand side of (4.41) noticing that

$$\begin{aligned}&P_{Y^n|X^n=x^n}\left[\ln\left(\frac{\mathrm{d}P_{Y^n|X^n=x^n}}{\mathrm{d}Q_{Y^n}}(Y^n|x^n)\right)\geq\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right)-\xi_n\right]\\ &\quad= P_{Y^n|X^n=x^n}\left[\ln\left(\frac{\frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\prod_{i=1}^{n}e^{-\frac{(Y_i-\sqrt{P})^2}{2\sigma^2}}}{\frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2+P)^{\frac{n}{2}}}\prod_{i=1}^{n}e^{-\frac{Y_i^2}{2(\sigma^2+P)}}}\right)\geq\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right)-\xi_n\right]\\ &\quad= P_{Y^n|X^n=x^n}\left[\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{-P}{\sigma^2+P}Y_i^2+2Y_i\sqrt{P}-P\right)\geq-\xi_n\right]\\ &\quad= \mathbb{P}\left[\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{-P}{\sigma^2+P}(\sqrt{P}+Z_i)^2+2\sqrt{P}Z_i+P\right)\geq-\xi_n\right] \qquad (4.42)\\ &\quad= \mathbb{P}\left[\frac{1}{2\sigma^2(\sigma^2+P)}\sum_{i=1}^{n}\left(-PZ_i^2+2\sigma^2\sqrt{P}Z_i+\sigma^2P\right)\geq-\xi_n\right], \qquad (4.43)\end{aligned}$$

where (4.42) follows by (4.65). Then we upper bound (4.43) and lower bound (4.41) with the following steps. For simplicity, we denote the random expression inside (4.43) by

$$W_i=\frac{1}{2\sigma^2(\sigma^2+P)}\left(-PZ_i^2+2\sigma^2\sqrt{P}Z_i+\sigma^2P\right)\qquad\forall i=1,\dots,n. \tag{4.44}$$

We notice that for all $i = 1, \dots, n$,

$$\mathbb{E}[W_i] = 0 \tag{4.45}$$

and

$$\begin{aligned}
\mathrm{Var}[W_i] &= \frac{1}{4\sigma^4(\sigma^2+P)^2}\mathbb{E}\left[\left(-PZ_i^2 + 2\sigma^2\sqrt{P}Z_i + \sigma^2 P\right)^2\right] \\
&= \frac{1}{4\sigma^4(\sigma^2+P)^2}\mathbb{E}\left[P^2Z_i^4 - 4\sigma^2\sqrt{P}PZ_i^3 + (4\sigma^4 P - 2\sigma^2P^2)Z_i^2 + 4\sigma^4\sqrt{P}PZ_i + \sigma^4P^2\right] \\
&= \frac{(2\sigma^2+P)P}{2(\sigma^2+P)^2}.
\end{aligned} \tag{4.46}$$

To simplify notation, we set

$$V(P) = \frac{(2\sigma^2+P)P}{2(\sigma^2+P)^2}. \tag{4.47}$$

We apply the Berry-Esseen Theorem A.5 to the random variables W_i then take out the absolute values, and find that for any $\lambda \in \mathbb{R}$,

$$\mathbb{P}\left[\sum_{i=1}^{n} W_i \geq \lambda\sqrt{nV(P)}\right] \leq \frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}} + Q(\lambda) \tag{4.48}$$

where

$$B(P) = 6\mathbb{E}\left[|W_i|^3\right]. \tag{4.49}$$

Let $\alpha_n > 0$ such that $\xi_n = -\sqrt{nV(P)}Q^{-1}(\alpha_n)$. Starting from (4.41) and combining (4.43) with (4.48), we obtain:

$$\beta_{1-\varepsilon} \geq \exp\left(-\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right) - \sqrt{nV(P)}Q^{-1}(\alpha_n)\right)\left(1-\varepsilon-\frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}} - \alpha_n\right), \tag{4.50}$$

where we chose λ in (4.48) such that $\lambda = Q^{-1}(\alpha_n)$. Moreover, we can choose $\alpha_n = 1-\varepsilon-2\frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}}$ and (4.50) becomes

$$\beta_{1-\varepsilon} \geq \exp\left(-\frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right) - \sqrt{nV(P)}Q^{-1}\left(1-\varepsilon-2\frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}}\right)\right)\frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}}. \tag{4.51}$$

Combining (4.38) and (4.51) we deduce that

$$\ln|\mathcal{M}| \leq \frac{n}{2}\ln\left(1+\frac{P}{\sigma^2}\right) + \sqrt{nV(P)}Q^{-1}\left(1-\varepsilon-2\frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}}\right) - \ln\left(\frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}}\right). \tag{4.52}$$

Recalling (4.49) with (4.47) and (4.44), we compute

$$\frac{B(P)}{V(P)^{\frac{3}{2}}} = 12\sqrt{\frac{2}{\pi}} + o(1), \tag{4.53}$$

by noticing that we have the upper bound

$$\begin{aligned}
\mathbb{E}\left[|W_i|^3\right] &= \frac{1}{8\sigma^6(\sigma^2+P)^3}\mathbb{E}\left[\left|-PZ_i^2 + 2\sigma^2\sqrt{P}Z_i + \sigma^2P\right|^3\right] \\
&\leq \frac{1}{8\sigma^6(\sigma^2+P)^3}\mathbb{E}\left[\left(PZ_i^2 + 2\sigma^2\sqrt{P}|Z_i| + \sigma^2P\right)^3\right] \\
&= \frac{1}{8\sigma^6(\sigma^2+P)^3}\mathbb{E}\Big[P^3Z_i^6 + 6\sigma^2\sqrt{P}P^2|Z_i|^5 + (3\sigma^2P^3 + 12\sigma^4P^2)Z_i^4 \\
&\quad + (12\sigma^4\sqrt{P}P^2 + 8\sigma^6\sqrt{P}P)|Z_i|^3 + (3\sigma^4P^3 + 12\sigma^6P^2)Z_i^2 + 6\sigma^6\sqrt{P}P^2|Z_i| \\
&\quad + \sigma^6P^3\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{P^{\frac{3}{2}}\mathbb{E}\left[|Z_i|^3\right] + o(P^{\frac{3}{2}})}{(\sigma^2+P)^3} \\
&= \frac{2\sqrt{\frac{2}{\pi}}\sigma^3 P^{\frac{3}{2}} + o(P^{\frac{3}{2}})}{(\sigma^2+P)^3} \\
&= 2\sqrt{\frac{2}{\pi}}\frac{P^{\frac{3}{2}}}{\sigma^3} + o(P^{\frac{3}{2}}).
\end{aligned} \tag{4.54}
$$

and similarly, we have the lower bound

$$
\begin{aligned}
\mathbb{E}\left[|W_i|^3\right] &\geq \frac{1}{8\sigma^6(\sigma^2+P)^3}\mathbb{E}\left[\left(-PZ_i^2 + \left|2\sigma^2\sqrt{P}Z_i + \sigma^2 P\right|\right)^3\right] \\
&\geq \frac{1}{8\sigma^6(\sigma^2+P)^3}\mathbb{E}\left[\left(-PZ_i^2 + 2\sigma^2\sqrt{P}\,|Z_i| - \sigma^2 P\right)^3\right] \\
&= \frac{1}{8\sigma^6(\sigma^2+P)^3}\mathbb{E}\Big[- P^3Z_i^6 + 6\sigma^2\sqrt{P}P^2\,|Z_i|^5 - (3\sigma^2P^3 + 12\sigma^4P^2)Z_i^4 \\
&\quad + (12\sigma^4\sqrt{P}P^2 + 8\sigma^6\sqrt{P}P)\,|Z_i|^3 - (3\sigma^4P^3 + 12\sigma^6P^2)Z_i^2 + 6\sigma^6\sqrt{P}P^2\,|Z_i| \\
&\quad - \sigma^6P^3\Big] \\
&= \frac{P^{\frac{3}{2}}\mathbb{E}\left[|Z_i|^3\right] + o(P^{\frac{3}{2}})}{(\sigma^2+P)^3} \\
&= 2\sqrt{\frac{2}{\pi}}\frac{P^{\frac{3}{2}}}{\sigma^3} + o(P^{\frac{3}{2}}).
\end{aligned} \tag{4.55}
$$

In addition, the Taylor expansion of Q^{-1} at $1-\varepsilon$ with the Lagrange form of the remainder ensures there exists θ in $\left(1-\varepsilon-2\frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}}, 1-\varepsilon\right)$ such that

$$
Q^{-1}(\alpha_n) = Q^{-1}(1-\varepsilon) - 2\frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}}\frac{dQ^{-1}}{dx}\bigg|_{x=\theta}. \tag{4.56}
$$

Injecting (4.53) and (4.56) in (4.52), we obtain

$$
\begin{aligned}
\ln|\mathcal{M}| &\leq \sqrt{\Delta}\sqrt{n}(1+\eta_n) + \sqrt{nV(P)}\left(Q^{-1}(1-\varepsilon) - 2\frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}}\frac{dQ^{-1}}{dx}\bigg|_{x=\theta}\right) + O(\ln(n)) \\
&\leq \sqrt{\Delta}\sqrt{n}(1+\eta_n) + \sqrt{nV(P)}Q^{-1}(1-\varepsilon) - 2\frac{B(P)}{V(P)}\min_{\theta\in\left(\frac{1-\varepsilon}{2},1-\varepsilon\right)}\frac{dQ^{-1}}{dx}\bigg|_{x=\theta} + O(\ln(n)) \qquad (4.57) \\
&= \sqrt{\Delta}\sqrt{n} + \sqrt{nV(P)}Q^{-1}(1-\varepsilon) - 2\frac{B(P)}{V(P)}\min_{\theta\in\left(\frac{1-\varepsilon}{2},1-\varepsilon\right)}\frac{dQ^{-1}}{dx}\bigg|_{x=\theta} + O(\ln(n)) \qquad (4.58) \\
&= \sqrt{\Delta}\sqrt{n} + \sqrt{n}\sqrt{2\sqrt{\frac{\Delta}{n}} + o\left(\frac{1}{\sqrt{n}}\right)}Q^{-1}(1-\varepsilon) + O(\ln(n)) \qquad (4.59) \\
&= \sqrt{\Delta}\sqrt{n} + \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}(1-\varepsilon) + O(\ln(n)) \\
&= \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}(\varepsilon) + O(\ln(n)) \qquad (4.60)
\end{aligned}
$$

where (4.57) follows for large enough n because $\frac{dQ^{-1}}{dx}$ is continuous on $(0,1)$; (4.58) follows by recalling the expression of η_n in (4.65); (4.59) follows by (4.47) and (4.61). □

Proof of Theorem 4.2: We consider an $(n, M, \varepsilon, \Delta)_{\max}$ code $\mathcal{C}$ for message set $\mathcal{M}$ and key set $\mathcal{K}$. Due to Lemma 4.3 this codebook must satisfy the average power constraint (4.24). As in equation (4.31), we denote

$$
P = 2\sigma^2\sqrt{\frac{\Delta}{n}}\,(1+\eta_n)\,, \qquad \eta_n = \frac{\ln(n)}{\sqrt{n}} \tag{4.61}
$$

and we omit the dependence of P from n for readability. Notice that for large enough n, $P > \rho_n$ with

ρ_n defined in (4.24). We denote by $\mathcal{D}$ the subset of all codewords from $\mathcal{C}$ such that

$$\|c\|_2^2 \leq nP. \tag{4.62}$$

As in (4.15), we know that the cardinality of $\mathcal{D}$ is bounded as

$$(M \times K)\left(1 - \frac{\rho_n}{P}\right) \leq |\mathcal{D}|\,, \tag{4.63}$$

where $K = |\mathcal{K}|$. By the pigeonhole principle, there must be at least one sub-codebook $\mathcal{C}_k \subset \mathcal{C}$ of size M containing the codewords from $\mathcal{C}$ indexed by the key $k \in \mathcal{K}$, such that

$$M\left(1 - \frac{\rho_n}{P}\right) \leq |\mathcal{D} \cap \mathcal{C}_k|. \tag{4.64}$$

As in the proof of Lemma 4.1 part 2), by adding an extra coordinate to each codeword of $\mathcal{D} \cap \mathcal{C}_k$, we can obtain a new code of dimension $n+1$ with size $|\mathcal{D} \cap \mathcal{C}_k|$, maximal probability of decoding error ε and satisfying the *equal* power constraint

$$\|c\|_2^2 = (n+1)P, \qquad \forall c \in \mathcal{D} \cap \mathcal{C}_k. \tag{4.65}$$

By Lemma 4.4 and Lemma 4.1 we deduce that

$$\begin{aligned}
\ln|\mathcal{D} \cap \mathcal{C}_k| &\leq \ln\left(\bar{M}_e^*(n+1, \varepsilon, P)\right) \\
&\leq \sqrt{\Delta}\sqrt{n+1} - \sqrt{2}\Delta^{\frac{1}{4}}(n+1)^{\frac{1}{4}}Q^{-1}(\varepsilon) + O(\ln(n)) \\
&= \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}(\varepsilon) + O(\ln(n)).
\end{aligned} \tag{4.66}$$

Recalling (4.64) we deduce that the maximum size of the message set $\mathcal{M}$ under the average power constraint (4.24) is

$$\begin{aligned}
\ln(\bar{M}_a^*(n, \varepsilon, \rho_n)) &\leq \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}(\varepsilon) + O\left(\ln(n)\right) - \ln\left(1 - \frac{\rho_n}{P}\right) \\
&= \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}(\varepsilon) + O\left(\ln(n)\right)
\end{aligned} \tag{4.67}$$

where (4.67) follows by the definition of ρ_n in (4.24) and P in (4.61). Recalling that the covertness constraint (2.43) implies the average power constraint (4.24) we deduce that

$$\ln(\bar{M}^*(n, \varepsilon, \Delta)) \leq \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}(\varepsilon) + O\left(\ln(n)\right). \tag{4.68}$$

□

4.3 Lower bound on the first-order asymptotics for average probability of error

As shown in [71], even for DMCs, if we consider covert communication under an average error probability constraint ε, the strong converse does not hold and the first-order asymptotics depend on ε. We will show that this is also true for Gaussian channels.

We define the corresponding scaling constant by

$$L_\varepsilon = \liminf_{n \to +\infty} \frac{\ln\left(M^*(n, \varepsilon, \Delta)\right)}{\sqrt{n\Delta}}. \tag{4.69}$$

First we notice that Fano's inequality (Appendix C.2) averaged over the keys as in (3.41), implies that

$$\begin{aligned}
L_\varepsilon &\leq \frac{L}{1-\varepsilon} \\
&\leq \frac{\sqrt{2}\sqrt{\mathrm{Var}\left[\ln(p_Z(Z))\right]}}{1-\varepsilon}
\end{aligned} \tag{4.70}$$

for all additive noise channels such that Theorem 3.1 holds. In particular, for the AWGN channel, we have

$$L_\varepsilon \leq \frac{1}{1-\varepsilon}. \tag{4.71}$$

In the following theorem, we establish a lower bound for L_ε in the case of an AWGN channel. Then in Figure 11, we show the different possible values for L_ε.

Theorem 4.3 *Consider the AWGN channel (2.42) subject to the covertness constraint (2.43). For an average probability of error $0 < \varepsilon < 1$, there exists an $(n, M, \varepsilon, \Delta)_{\text{avg}}$-code with*

$$\ln(M) \geq \sqrt{\frac{\Delta}{1-\varepsilon}}\sqrt{n} + O\left(n^{\frac{1}{3}}\right). \tag{4.72}$$

In particular

$$L_\varepsilon \geq \frac{1}{\sqrt{1-\varepsilon}}. \tag{4.73}$$

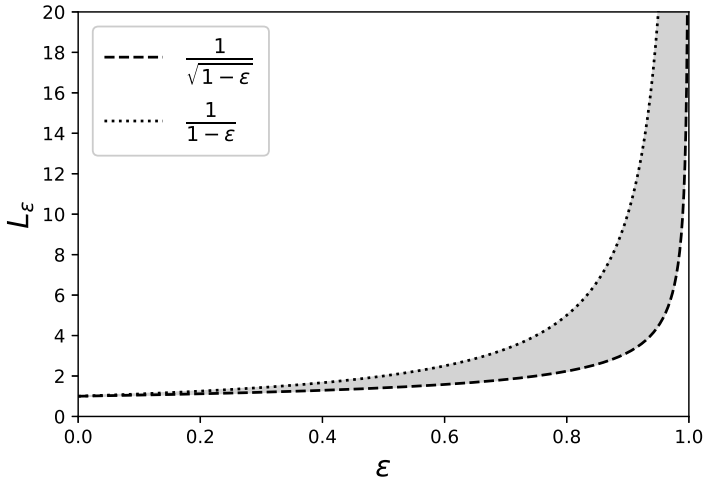

Fig. 11: Upper and lower bounds on L_ε.

Recall that at the first order, when we require vanishing average probability of error, BPSK inputs with amplitude $\frac{\sqrt{2}\sigma\Delta^{\frac{1}{4}}}{n^{\frac{1}{4}}}$ are optimal over the AWGN channel; see Section 2.8. We now consider random coding with BPSK inputs under an average probability of error ε. First, we prove an achievability result using BPSK inputs under average probability of error. Then by adding all zero-codewords to this code, we construct a new code achieving the first-order asymptotics of Theorem 4.3.

Lemma 4.5 *Over the AWGN channel (2.42) under the covertness constraint (2.43), given $0 < \varepsilon < 1$ there exists an $(n, M, \varepsilon, \Delta)_{\text{avg}}$-code with average probability of error ε such that*

$$\ln(M) \geq \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}\left(\frac{\varepsilon}{2}\right) + O(\ln(n)). \tag{4.74}$$

Proof: We denote

$$P = 2\sigma^2\sqrt{\frac{\Delta}{n}}, \tag{4.75}$$

and consider the uniform distribution P_X on $\{-\sqrt{P}, \sqrt{P}\}$. We denote P_Y the corresponding output distribution through $P_{Y|X}$:

$$p_Y(y) = \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-\sqrt{P})^2}{2\sigma^2}} + \frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y+\sqrt{P})^2}{2\sigma^2}}. \tag{4.76}$$

We use the notation $P_{X^n} = P_X^{\otimes n}$ and $P_{Y^n} = P_Y^{\otimes n}$ for the corresponding i.i.d. product distributions. We generate a random code C by picking every codeword i.i.d. from $P_X^{\otimes n}$. We denote by $X_{m,k}^n$ for $m = 1, \dots, |\mathcal{M}|, k = 1, \dots, |\mathcal{K}|$ the random codewords, by $X_{m,k,i}$ the i^{th} component of $X_{m,k}^n$, and by $P_{Y_\mathsf{C}^n}$ the output statistics of the random code:

$$P_{Y_\mathsf{C}^n}(\cdot) = \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} P_{Y^n|X^n}(\cdot|X_{m,k}^n). \tag{4.77}$$

We want to show that this code is covert with high probability. First, we notice that

$$\mathbb{P}\left[D\left(P_{Y_\mathsf{C}^n} \| P_{Z^n}\right) > \Delta\right] = \mathbb{P}\left[D\left(P_{Y_\mathsf{C}^n} \| P_{Y^n}\right) + \mathbb{E}_{P_{Y_\mathsf{C}^n}}\left[\ln\left(\frac{p_{Y^n}(Y_\mathsf{C}^n)}{p_{Z^n}(Y_\mathsf{C}^n)}\right)\right] > \Delta\right]. \tag{4.78}$$

We compute

$$\begin{aligned}
&\mathbb{E}_{P_{Y_\mathsf{C}^n}}\left[\ln\left(\frac{p_{Y^n}(Y_\mathsf{C}^n)}{p_{Z^n}(Y_\mathsf{C}^n)}\right)\right] \\
&= \int_{\mathcal{Y}^n} p_{Y_\mathsf{C}^n}(y^n) \ln\left(\prod_{i=1}^n \frac{\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{2}\left(e^{-\frac{(y_i-\sqrt{P})^2}{2\sigma^2}} + e^{-\frac{(y_i+\sqrt{P})^2}{2\sigma^2}}\right)}{\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y_i^2}{2}}}\right) \mathrm{d}y^n \\
&= \int_{\mathcal{Y}^n} p_{Y_\mathsf{C}^n}(y^n) \left(\sum_{i=1}^n \ln\left(\frac{1}{2}\left(e^{+\frac{y_i\sqrt{P}}{\sigma^2}} + e^{-\frac{y_i\sqrt{P}}{\sigma^2}}\right)\right) - \frac{nP}{2\sigma^2}\right) \mathrm{d}y^n \\
&\leq \int_{\mathcal{Y}^n} p_{Y_\mathsf{C}^n}(y^n) \sum_{i=1}^n \left(\frac{y_i^2 P}{2\sigma^4} - \frac{y_i^4 P^2}{12\sigma^8} + \frac{y_i^6 P^3}{45\sigma^{12}}\right) - \frac{nP}{2\sigma^2} \mathrm{d}y^n \qquad (4.79) \\
&= \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} \int_{\mathcal{Y}^n} \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - X_{m,k,i})^2}{2}} \sum_{i=1}^n \left[\frac{y_i^2 P}{2\sigma^4} - \frac{y_i^4 P^2}{12\sigma^8} + \frac{y_i^6 P^3}{45\sigma^{12}}\right] \mathrm{d}y^n - \frac{nP}{2\sigma^2} \\
&= \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{m=1}^{|\mathcal{M}|} \sum_{i=1}^n \left(\left(\sigma^2 + X_{m,k,i}^2\right) \frac{P}{2\sigma^4} - \left(3\sigma^4 + 6\sigma^2 X_{m,k,i}^2 + X_{m,k,i}^4\right) \frac{P^2}{12\sigma^8} \right. \\
&\quad \left. + \left(15\sigma^6 + 45\sigma^4 X_{m,k,i}^2 + 15\sigma^2 X_{m,k,i}^4 + X_{m,k,i}^6\right) \frac{P^3}{45\sigma^{12}} \right) - \frac{nP}{2\sigma^2} \qquad (4.80)
\end{aligned}$$

where (4.79) follows by the inequality $\ln(\cosh(x)) \leq \frac{x^2}{2} - \frac{x^4}{12} + \frac{x^6}{45}$. Thus

$$\begin{aligned}
\mathbb{P}\left[D\left(P_{Y_\mathsf{C}^n} \| P_{Z^n}\right) > \Delta\right] &= \mathbb{P}\left[D\left(P_{Y_\mathsf{C}^n} \| P_{Y^n}\right) + n\left(\left(\sigma^2 + P\right) \frac{P}{2\sigma^4} - \left(3\sigma^4 + 6\sigma^2 P + P^2\right) \frac{P^2}{12\sigma^8} \right.\right. \\
&\quad \left.\left. + \left(15\sigma^6 + 45\sigma^4 P + 15\sigma^2 P^2 + P^3\right) \frac{P^3}{45\sigma^{12}} \right) - \frac{nP}{2\sigma^2} > \Delta \right] \qquad (4.81) \\
&= \mathbb{P}\left[D\left(P_{Y_\mathsf{C}^n} \| P_{Y^n}\right) > \frac{nP^3}{6\sigma^6} - \frac{11}{12}\frac{nP^4}{\sigma^8} - \frac{nP^5}{3\sigma^{10}} - \frac{nP^6}{45\sigma^{12}}\right] \qquad (4.82) \\
&\leq \frac{\mathbb{E}\left[D\left(P_{Y_\mathsf{C}^n} \| P_{Y^n}\right)\right]}{\frac{4\Delta^{\frac{3}{2}}}{3\sqrt{n}} + O\left(\frac{1}{n}\right)} \qquad (4.83) \\
&= O\left(\sqrt{n}e^{-n}\right), \qquad (4.84)
\end{aligned}$$

where (4.81) stands because $\mathbb{P}\left[X_{m,k,i}^2 = P\right] = 1$ for any i, m, k; (4.82) by (4.75); (4.83) follows by the Markov's inequality (Lemma A.2) and (4.75); (4.84) by Hayashi's bound (Theorem 2.1) which ensures that choosing $\ln|\mathcal{K}| = O(n)$ implies that (4.84) tends to 0 when $n \to +\infty$.

We now prove the existence of a code achieving (4.74). From Shannon's achievability bound (Theorem

2.2), we know that for all $\gamma > 0$,

$$\mathbb{E}_{\mathsf{C}}[P_e(\mathsf{C})] \leq \mathbb{P}\left[i_{X^n,Y^n}(X^n,Y^n) \leq \ln|\mathcal{M}| + n\gamma\right] + \exp(-n\gamma), \tag{4.85}$$

where $P_e(\mathsf{C})$ is the average error probability of the random code C. We now assume that $\ln|\mathcal{M}|$ is chosen so that

$$\mathbb{E}_{\mathsf{C}}[P_e(\mathsf{C})] \leq \frac{\varepsilon}{2}. \tag{4.86}$$

A sufficient condition for this will be derived later. Then, by Markov's inequality (Lemma A.2),

$$\mathbb{P}\left[P_e(\mathsf{C}) \leq \varepsilon\right] \geq \frac{1}{2} \tag{4.87}$$

and it follows that there exists a code $\mathcal{C}$ with average error probability $P_e \leq \varepsilon$ which satisfies the covertness condition (2.43), since

$$\begin{aligned}
\mathbb{P}\left[\{P_e(\mathsf{C}) \leq \varepsilon\} \cap \left\{D\left(P_{Y_{\mathsf{C}}^n}\|P_{Z^n}\right) \leq \Delta\right\}\right] &\geq 1 - \mathbb{P}\left[P_e(\mathsf{C}) > \varepsilon\right] - \mathbb{P}\left[D\left(P_{Y_{\mathsf{C}}^n}\|P_{Z^n}\right) > \Delta\right] \\
&\geq \frac{1}{2} - \mathbb{P}\left[D\left(P_{Y_{\mathsf{C}}^n}\|P_{Z^n}\right) > \Delta\right] \\
&> 0,
\end{aligned} \tag{4.88}$$

where (4.88) follows by (4.84) for large enough n.
We now find a sufficient condition for (4.86) to hold. We notice that for any $x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n$,

$$\begin{aligned}
i_{X^n,Y^n}(x^n,y^n) &= \ln\left(\frac{p_{Y^n|X^n}(y^n|x^n)}{p_{Y^n}(y^n)}\right) \\
&= \ln\left(\frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-x_i)^2}{2\sigma^2}}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{2}\left(e^{-\frac{(y_i-\sqrt{P})^2}{2\sigma^2}} + e^{-\frac{(y_i+\sqrt{P})^2}{2\sigma^2}}\right)}\right) \\
&= \sum_{i=1}^n \left(\ln\left(\frac{e^{\frac{x_i y_i}{\sigma^2}}}{\frac{1}{2}\left(e^{\frac{\sqrt{P}y_i}{\sigma^2}} + e^{-\frac{\sqrt{P}y_i}{\sigma^2}}\right)}\right) + \frac{P - x_i^2}{2\sigma^2}\right) \\
&= \sum_{i=1}^n \left(\frac{x_i y_i}{\sigma^2} - \ln\left(\cosh\left(\frac{\sqrt{P}y_i}{\sigma^2}\right)\right) + \frac{P - x_i^2}{2\sigma^2}\right) \\
&\geq \sum_{i=1}^n \left(\frac{x_i y_i}{\sigma^2} - \frac{P y_i^2}{2\sigma^4} + \frac{P - x_i^2}{2\sigma^2}\right),
\end{aligned} \tag{4.89}$$

where (4.89) follows by the inequality $\ln(\cosh(x)) \leq \frac{x^2}{2}$. Therefore

$$\begin{aligned}
&\mathbb{P}\left[i_{X^n,Y^n}(X^n,Y^n) \leq \ln|\mathcal{M}| + n\gamma\right] \\
&\quad\leq \mathbb{P}\left[\sum_{i=1}^n \left(\frac{X_i Y_i}{\sigma^2} - \frac{P Y_i^2}{2\sigma^4}\right) \leq \ln|\mathcal{M}| + n\gamma\right] \qquad (4.90) \\
&\quad= \mathbb{P}\left[\sum_{i=1}^n \left(\frac{X_i^2 + X_i Z_i}{\sigma^2} - \frac{P(X_i^2 + Z_i^2 + 2X_i Z_i)}{2\sigma^4}\right) \leq \ln|\mathcal{M}| + n\gamma\right] \\
&\quad= \mathbb{P}\left[\sum_{i=1}^n \left(\frac{X_i Z_i}{\sigma^2} - \frac{P(Z_i^2 + 2X_i Z_i)}{2\sigma^4}\right) + \frac{nP}{\sigma^2} - \frac{nP^2}{2\sigma^4} \leq \ln|\mathcal{M}| + n\gamma\right] \\
&\quad= \mathbb{P}\left[\sum_{i=1}^n \left(\left(\frac{1}{\sigma^2} - \frac{P}{\sigma^4}\right) X_i Z_i - \frac{P}{2\sigma^4} Z_i^2\right) + \frac{nP}{\sigma^2} - \frac{nP^2}{2\sigma^4} \leq \ln|\mathcal{M}| + n\gamma\right], \qquad (4.91)
\end{aligned}$$

where (4.90) follows because $\mathbb{P}\left[X_i^2 = P\right] = 1$. For simplicity, we denote the random expression inside

(4.91) by

$$W_i = \left(\frac{1}{\sigma^2} - \frac{P}{\sigma^4}\right) X_i Z_i - \frac{P}{2\sigma^4} Z_i^2 \qquad \forall i = 1, \dots, n. \tag{4.92}$$

We notice that for any $i = 1, \dots, n$,

$$\mathbb{E}\left[W_i | X_i\right] = -\frac{P}{2\sigma^2} \tag{4.93}$$

and

$$\begin{aligned}
\mathrm{Var}\left[W_i | X_i\right] &= \mathbb{E}\left[\left(\frac{1}{\sigma^2} - \frac{P}{\sigma^4}\right)^2 X_i^2 Z_i^2 + 2\left(\frac{1}{\sigma^2} - \frac{P}{\sigma^4}\right)\frac{P}{2\sigma^4} X_i Z_i^3 + \left(\frac{P}{2\sigma^4}\right)^2 Z_i^4 \middle| X_i\right] - \left(\frac{P}{2\sigma^2}\right)^2 \\
&= \sigma^2 X_i^2 \left(\frac{1}{\sigma^2} - \frac{P}{\sigma^4}\right)^2 + 3\sigma^4 \left(\frac{P}{2\sigma^4}\right)^2 - \left(\frac{P}{2\sigma^2}\right)^2 \\
&= \frac{X_i^2}{\sigma^2} - \frac{2PX_i^2}{\sigma^4} + \frac{P^2 X_i^2}{\sigma^6} + \frac{3P^2}{4\sigma^4} - \frac{P^2}{4\sigma^4},
\end{aligned} \tag{4.94}$$

with

$$\mathbb{P}\left[\mathrm{Var}\left[W_i | X_i\right] = \frac{P}{\sigma^2} - \frac{3P^2}{2\sigma^4} + \frac{P^3}{\sigma^6}\right] = 1. \tag{4.95}$$

We denote

$$V(P) = \frac{P}{\sigma^2} - \frac{3P^2}{2\sigma^4} + \frac{P^3}{\sigma^6} \tag{4.96}$$

and similarly

$$B(P) = 6\mathbb{E}\left[\left|W_i - \mathbb{E}\left[W_i | X_i\right]\right|^3 \middle| X_i\right] = O\left(P^{\frac{3}{2}}\right). \tag{4.97}$$

Injecting (4.91) in (4.85), we obtain

$$\begin{aligned}
\mathbb{E}_{\mathsf{C}}[P_e(\mathsf{C})] &\leq \mathbb{P}\left[\sum_{i=1}^{n}\left(W_i - \mathbb{E}\left[W_i | X_i\right] + \frac{P}{2\sigma^2} - \frac{P^2}{2\sigma^4}\right) \leq \ln|\mathcal{M}| + n\gamma\right] + e^{-n\gamma} \\
&= \mathbb{E}\left[\mathbb{P}\left[\sum_{i=1}^{n}\left(W_i - \mathbb{E}\left[W_i | X_i\right]\right) \leq \ln|\mathcal{M}| + n\gamma - \frac{nP}{2\sigma^2} + \frac{nP^2}{2\sigma^4} \middle| X^n\right]\right] + e^{-n\gamma} \qquad (4.98) \\
&= 1 - \mathbb{E}\left[\mathbb{P}\left[\sum_{i=1}^{n}\left(W_i - \mathbb{E}\left[W_i | X_i\right]\right) \geq \ln|\mathcal{M}| + n\gamma - \frac{nP}{2\sigma^2} + \frac{nP^2}{2\sigma^4} \middle| X^n\right]\right] \\
&\quad + e^{-n\gamma}. \qquad (4.99)
\end{aligned}$$

Then choosing

$$\ln|\mathcal{M}| = -n\gamma + \frac{nP}{2\sigma^2} - \frac{nP^2}{2\sigma^4} + Q^{-1}\left(1 - \frac{\varepsilon}{2} + \frac{1}{n^{\frac{3}{8}}}\right)\sqrt{n}\sqrt{V(P)} \tag{4.100}$$

and applying the Berry-Esseen Theorem A.5 to the random variables W_i ensures that the right-hand side of (4.99) hence $\mathbb{E}_{\mathsf{C}}[P_e(\mathsf{C})]$ is upper-bounded as follows

$$\begin{aligned}
\mathbb{E}_{\mathsf{C}}[P_e(\mathsf{C})] &\leq 1 - \left(1 - \frac{\varepsilon}{2} + \frac{1}{n^{\frac{3}{8}}} - \frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}}\right) + e^{-n\gamma} \\
&= \frac{\varepsilon}{2} - \frac{1}{n^{\frac{3}{8}}} + \frac{B(P)}{\sqrt{n}V(P)^{\frac{3}{2}}} + e^{-n\gamma}
\end{aligned} \tag{4.101}$$

Choosing,

$$\gamma = \frac{\ln(n)}{n} \tag{4.102}$$

in (4.101), we deduce that (4.100) ensures

$$\mathbb{E}_{\mathsf{C}}[P_e(\mathsf{C})] \leq \frac{\varepsilon}{2} - \frac{1}{n^{\frac{3}{8}}} + O\left(\frac{1}{\sqrt{n}}\right), \tag{4.103}$$

where (4.103) follows by recalling (4.75), (4.96) and (4.97). Therefore (4.100) guarantees (4.86) for n large enough. Then injecting (4.75) and (4.96) in (4.100), we deduce that there exists a code of average probability of error ε and such that the covertness condition holds, satisfying

$$\begin{aligned}\ln|\mathcal{M}| &= \sqrt{\Delta}\sqrt{n} + \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}\left(1 - \frac{\varepsilon}{2} + \frac{1}{n^{\frac{3}{8}}}\right) + O\left(\ln(n)\right) \\ &= \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}\left(\frac{\varepsilon}{2} - \frac{1}{n^{\frac{3}{8}}}\right) + O\left(\ln(n)\right)\end{aligned} \tag{4.104}$$

In addition, the Taylor expansion of Q^{-1} at $\frac{\varepsilon}{2}$ with the Lagrange form of the remainder ensures there exists θ in $\left(\frac{\varepsilon}{2} - \frac{1}{n^{\frac{1}{4}}}, \frac{\varepsilon}{2}\right)$ such that

$$\begin{aligned}\ln|\mathcal{M}| &= \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}\left(\frac{\varepsilon}{2}\right) + \sqrt{2}\Delta^{\frac{1}{4}}\frac{1}{n^{\frac{1}{8}}}\left.\frac{dQ^{-1}}{dx}\right|_{x=\theta} + O\left(\ln(n)\right) \\ &= \sqrt{\Delta}\sqrt{n} - \sqrt{2}\Delta^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}\left(\frac{\varepsilon}{2}\right) + O\left(\ln(n)\right)\end{aligned} \tag{4.105}$$

where (4.105) follows because $\left|\frac{dQ^{-1}}{dx}\right|$ is bounded on $\left(\frac{\varepsilon}{2} - \frac{1}{n^{\frac{1}{4}}}, \frac{\varepsilon}{2}\right)$ by e.g. $\left|\frac{dQ^{-1}}{dx}\right|\Big|_{x=\frac{\varepsilon}{4}}$ for n large enough. □

Proof of Theorem 4.3: The proof follows the ideas of [71, Appendix A] and [58, Theorem 77]. Let $\varepsilon > \varepsilon' > 0$ and

$$\Delta' = \frac{1-\varepsilon'}{1-\varepsilon}\Delta. \tag{4.106}$$

Lemma 4.5 shows the existence of an $(n, M, \varepsilon', \Delta')_{\text{avg}}$-code $\mathcal{C}'$ such that

$$\ln(M) \geq \sqrt{\Delta'}\sqrt{n} - \sqrt{2}\Delta'^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}\left(\frac{\varepsilon'}{2}\right) + O(\ln(n)). \tag{4.107}$$

We now consider a new code $\mathcal{C}$ of size $M\frac{1-\varepsilon'}{1-\varepsilon}$ obtained by adding αM all-zero codewords (independently of the value of the key), where $\alpha = \frac{\varepsilon-\varepsilon'}{1-\varepsilon}$. Then given than $\varepsilon > \varepsilon'$ the average probability of error of this new code P_e admits ε as an upper bound:

$$\begin{aligned}P_e &\leq \frac{1}{1+\alpha}\varepsilon' + \frac{\alpha}{1+\alpha} \\ &\leq \varepsilon.\end{aligned} \tag{4.108}$$

Furthermore by convexity of the Kullback-Leibler divergence,

$$\begin{aligned}D(P_{Y_{\mathcal{C}}^n}\|P_{Z^n}) &\leq \frac{1}{1+\alpha}D(P_{Y_{\mathcal{C}'}^n}\|P_{Z^n}) + \frac{\alpha}{1+\alpha}D(P_{Z^n}\|P_{Z^n}) \\ &= \frac{1}{1+\alpha}\Delta' \\ &\leq \frac{1-\varepsilon}{1-\varepsilon'}\Delta' \\ &= \Delta,\end{aligned} \tag{4.109}$$

and $\mathcal{C}$ is an $(n, M\frac{1-\varepsilon'}{1-\varepsilon}, \varepsilon, \Delta)_{\text{avg}}$-code such that

$$\begin{aligned}\ln\left(M\frac{1-\varepsilon'}{1-\varepsilon}\right) &\geq \sqrt{\Delta'}\sqrt{n} - \sqrt{2}\Delta'^{\frac{1}{4}}n^{\frac{1}{4}}Q^{-1}\left(\frac{\varepsilon'}{2}\right) + O(\ln(n)) + \ln\left(\frac{1-\varepsilon'}{1-\varepsilon}\right) \\ &= \sqrt{\Delta'}\sqrt{n} + O\left(n^{\frac{1}{3}}\right) + \ln\left(\frac{1-\varepsilon'}{1-\varepsilon}\right).\end{aligned} \tag{4.110}$$

Finally, note that (4.110) still holds if we take

$$\varepsilon' = n^{-1/6}\,: \tag{4.111}$$

the proof follows the same step as the proof of Lemma 4.5 until (4.98). Then we check that (4.98) can be upper bounded by $\frac{n^{-\frac{1}{6}}}{2}$ when choosing

$$\ln|\mathcal{M}| = -n\gamma + \frac{nP'}{2\sigma^2} - \frac{nP'^2}{2\sigma^4} - n^{\frac{1}{3}}, \tag{4.112}$$

with

$$P' = 2\sigma^2\sqrt{\frac{\Delta'}{n}}, \tag{4.113}$$

and

$$\gamma = \frac{\ln(4n^{\frac{1}{6}})}{n}. \tag{4.114}$$

First, we prove that

$$\mathbb{E}\left[\mathbb{P}\left[\sum_{i=1}^{n}\left(W_i - \mathbb{E}\left[W_i|X_i\right]\right) \leq -n^{\frac{1}{3}}\middle| X^n\right]\right] \leq \frac{1}{4}n^{-\frac{1}{6}}. \tag{4.115}$$

This follows from a result in large deviation theory (Theorem A.4). We check that the hypotheses of Theorem A.4 are verified. We recall from (4.92) that for any $i = 1, \ldots, n$, W_i is equal to

$$W_i' = \left(\frac{1}{\sigma^2} - \frac{P'}{\sigma^4}\right)\sqrt{P'}Z_i - \frac{P'}{2\sigma^4}Z_i^2 \qquad \forall i = 1, \ldots, n, \tag{4.116}$$

with probability $1/2$ and to

$$W_i'' = -\left(\frac{1}{\sigma^2} - \frac{P'}{\sigma^4}\right)\sqrt{P'}Z_i - \frac{P'}{2\sigma^4}Z_i^2 \qquad \forall i = 1, \ldots, n \tag{4.117}$$

with probability $1/2$. Furthermore $Z_i \sim \mathcal{N}(0, \sigma^2)$ follows the same distribution as $-Z_i$, so $W_i' \sim W_i''$. Therefore we can restrict ourselves to the case

$$W_i = -\left(\frac{1}{\sigma^2} - \frac{P'}{\sigma^4}\right)\sqrt{P'}Z_i - \frac{P'}{2\sigma^4}Z_i^2 \qquad \forall i = 1, \ldots, n. \tag{4.118}$$

We compute

$$\begin{aligned}
f(\lambda) &= \ln\left(\mathbb{E}\left[e^{\lambda(W_i - \mathbb{E}[W_i])}\right]\right) \\
&= \ln\left(\mathbb{E}\left[e^{\lambda\left(-\left(\frac{1}{\sigma^2}-\frac{P'}{\sigma^4}\right)\sqrt{P'}Z_i - \frac{P'}{2\sigma^4}Z_i^2 + \frac{P'}{2\sigma^2}\right)}\right]\right) \\
&= \ln\left(\mathbb{E}\left[e^{\lambda\left(-\frac{P'}{2\sigma^4}\left(Z_i + \left(\frac{\sigma^2}{P'}-1\right)\sqrt{P'}\right)^2 + \frac{1}{2} - \frac{P'}{2\sigma^2} + \frac{P'^2}{2\sigma^4}\right)}\right]\right) \\
&= \lambda\left(\frac{1}{2} - \frac{P'}{2\sigma^2} + \frac{P'^2}{2\sigma^4}\right) - \frac{1}{2}\ln\left(1 + 2\frac{P'}{2\sigma^4}\lambda\right) - \frac{\frac{P'}{2\sigma^4}\lambda\left(\frac{\sigma^2}{P'}-1\right)^2 P'}{1 + 2\frac{P'}{2\sigma^4}\lambda} \qquad (4.119) \\
&= \lambda\left(\frac{1}{2} - \frac{P'}{2\sigma^2} + \frac{P'^2}{2\sigma^4}\right) - \frac{1}{2}\ln\left(1 + \frac{P'}{\sigma^4}\lambda\right) - \frac{\lambda\left(\frac{1}{2} - \frac{P'}{\sigma^2} + \frac{P'^2}{2\sigma^4}\right)}{1 + \frac{P'}{\sigma^4}\lambda}, \qquad (4.120)
\end{aligned}$$

where (4.119) follows by the moment generating function of a non-central chi-squared random variable [42, eq. (29.6)']. We deduce that (4.120) is $< \infty$ in a ball around the origin. Let $\Gamma = (-\infty, -1]$, and

$a_n = n^{\frac{1}{3}}$, then by Theorem A.4,

$$\begin{aligned}
\mathbb{P}\left[\sum_{i=1}^{n}(W_i - \mathbb{E}[W_i]) < -n^{\frac{1}{3}}\right] &= \mathbb{P}\left[n^{-\frac{1}{3}}\sum_{i=1}^{n}(W_i - \mathbb{E}[W_i]) < -1\right] \\
&\leq O\left(e^{-\frac{1}{2V(P')n^{\frac{1}{3}}}}\right) \\
&= O\left(e^{-\frac{\sqrt{1-\varepsilon}}{4\sqrt{1-\varepsilon'}\sqrt{\Delta}}n^{\frac{1}{6}}}\right) \qquad (4.121) \\
&\leq O\left(e^{-\frac{\sqrt{1-\varepsilon}}{4\sqrt{\Delta}}n^{\frac{1}{6}}}\right) \\
&= o(n^{-\frac{1}{6}}), \qquad (4.122)
\end{aligned}$$

where (4.121) follows by (4.96), (4.113) and (4.106). (4.122) implies (4.115), then combining (4.115) and (4.98) ensures

$$\begin{aligned}
\mathbb{E}_{\mathsf{C}}\left[P_e(\mathsf{C})\right] &\leq \mathbb{E}\left[\mathbb{P}\left[\sum_{i=1}^{n}(W_i - \mathbb{E}\left[W_i\right]) \leq -n^{\frac{1}{3}}\middle| X^n\right]\right] + e^{-n\gamma} \\
&\leq \frac{1}{2}n^{-\frac{1}{6}} \qquad (4.123)
\end{aligned}$$

and (4.110) with (4.111). Finally, (4.110) can be rewritten as

$$\begin{aligned}
\ln\left(M\frac{1-\varepsilon'}{1-\varepsilon}\right) &\geq \sqrt{\frac{1-\varepsilon'}{1-\varepsilon}}\sqrt{\Delta}\sqrt{n} + O\left(n^{\frac{1}{3}}\right) + \ln\left(\frac{1-\varepsilon'}{1-\varepsilon}\right) \qquad (4.124) \\
&= \sqrt{\frac{\Delta}{1-\varepsilon}}\sqrt{n} + O\left(n^{\frac{1}{3}}\right), \qquad (4.125)
\end{aligned}$$

where (4.124) follows by (4.106); (4.125) follows by (4.111).

□

4.4 Concluding remarks

In this section, we showed that for the memoryless Gaussian channel, the covertness constraint implies an average power constraint. Using converse bounds for channel coding under a maximal error probability constraint in [59], we derived upper bounds on the first and second asymptotics for covert communication over an AWGN channel under the maximal error probability criterion. In the covert setup, we observe different scaling behaviors for the first and second-order asymptotics compared to the classical non-covert setup. We recall that the difference in the first asymptotics follows from the square root law of covert communication.

In order to establish an achievability result for maximal error, one would need to show the existence of a code that is simultaneously reliable and covert. Note that Feinstein's Lemma (Remark 3.74) ensures the existence of a reliable code, but such a code may not be covert. In the case of DMCs, Tahmasbi and Bloch [71] prove the existence of a covert and reliable code with a random coding argument. The probability that this code is reliable with respect to maximal error is actually vanishingly small, but the authors show that the probability that the code is covert tends to 1 very fast, thanks to a double-exponential concentration result for the Kullback-Leibler divergence [71, Lemma 2]. Unfortunately, [71, Lemma 2] relies on the hypothesis of finite alphabets and generalizing it to continuous alphabets is a non-trivial problem[7].

[7]For the variational distance, such a concentration result can be established as a corollary of [52, Theorem 31].

For covert communication with an average probability of error ε, the first-order asymptotics depend on ε. Since an average error probability constraint allows more freedom in the code construction compared to a maximum error probability constraint, we introduced the quantity L_ε which serves as the counterpart to the ε-capacity in the context of covert communication. Fano's inequality gives an immediate upper bound for L_ε and we derived a lower bound in the case of an AWGN channel. Finally, we conclude that allowing a positive average probability of error increases the amount of covert information that can be asymptotically shared over an AWGN channel.

5 Conclusions and perspectives

In this thesis, we have considered the problem of covert communication over continuous channels. We highlight our main contributions as well as some open problems and perspectives for future work.

General formula for the scaling constant over additive memoryless channels We have derived a general formula for the scaling constant L of the square root law for covert communication over general additive memoryless channels. The three technical assumptions (3.2)–(3.4) are mild requirements that ensure an upper bound on L for a wide range of noise distributions. In particular, one can show that all α-stable distributions verify these integrability assumptions[8]. However, ensuring the achievability of the general formula for L is a hard problem since we only know a few self-decomposable distributions that satisfy Assumption 3.1. Nevertheless, we notice that this self-decomposable property is not always necessary as there could exist other input distributions that achieve L, as we have seen for BPSK inputs in the case of the Gaussian channel.

General noise with memory We have shown that for an additive Gaussian channel the presence of memory does not improve the quantity of information that can be sent reliably and covertly; see Section 2.9. We have shown that the covertness constraint induces an average power constraint on the input for an AWGN channel (see eq (3.50) in Section 3.2). For the sake of comparison, if we consider a Gaussian channel with memory with no covert constraint under an average power constraint on the input, then memory can be exploited by *water-filling* [18, Section 9.5] hence the capacity is larger than without memory. On the contrary, we showed that covert communications do not improve with noise memory. Intuitively, in the context of covert communication, any advantage the legitimate receiver can exploit is also accessible to the eavesdropper. The generalization of this finding to non-Gaussian noise is still an open problem.

Second-order asymptotics We also considered covert communication over AWGN channels in the finite blocklength regime under both a maximal error probability constraint and an average error probability constraint. For maximal error, we derived an upper bound on the second-order asymptotics of covert communication. Furthermore, allowing a positive average error probability increases the first-order asymptotic of covert communication. For average error ϵ, we showed that the strong converse does not hold, and the scaling constant L_ϵ is larger than L. The exact characterization of L_ϵ is still an open question.

Possible extensions of this work One could extend this work to the case where the eavesdropper and the legitimate receiver do not see the same channel outputs. Although the self-decomposability condition in Assumption 3.1 makes it hard to study, the special case where the eavesdropper and the legitimate receiver's channels are both AWGN is easily obtained because we know the optimal input minimizing the Kullback-Leibler divergence of the covertness constraint on the eavesdropper's channel while maximizing the entropy on the legitimate receiver's channel; see Appendix D.1. Moreover, we can show that no key is needed if the eavesdropper's channel is noisier than the legitimate receiver's; see Appendix D.1.2. Nevertheless, in the general case such an optimization would be difficult, as illustrated for the exponential degraded channel where we are only able to derive an upper bound on L; see Appendix

[8] Malcolm Egan, personal communication.

D.2. Another avenue for extension would be a multi-user scenario with multiple continuous channels along the lines of [11].

An alternative research direction could be the challenging problem of code design for covert communication. Over DMCs, several code designs have been proposed, such as polar codes [31] which have low complexity but are suboptimal due to the low speed of polarization, pulse position modulation [8] or the concatenated coding scheme [85] which has polynomial complexity. Over the AWGN, a new scheme combining pulse position modulation, multilevel coding, and amplitude scaling [43] was shown to achieve optimal scaling.

A Measure theory tools

In this section, we recall basic definitions and theorems of measure theory that will be used throughout this thesis.

A.1 Limit inferior in probability

Definition A.1 (Limit inferior in probability [73]**)** *If $\{X_n\}$ is a sequence of random variables, its limit inferior in probability is the supremum of all the reals α for which $\mathbb{P}[X_n \leq \alpha] \to 0$ as $n \to +\infty$. Similarly, its limit superior in probability is the infimum of all the reals β for which $\mathbb{P}[X_n \geq \beta] \to 0$ as $n \to +\infty$.*

A.2 Continuity under integral sign

Lemma A.1 (Continuity under integral sign [64, Section 11.4]**)** *Let $t_0 \in \mathbb{R}$, a measured space $(\Omega, \mathcal{T}, \mu)$, a function $f : \Omega \times \mathbb{R} \to \mathbb{R}$ such that $f(\cdot, t)$ is $\mathcal{T}$-measurable and $\mathcal{L}^1(\mu)$ for all $t \in \mathbb{R}$. If*

1) *for all $\omega \in \Omega$, $f(\omega, \cdot)$ is continuous at t_0,*
2) *there exists $\varepsilon > 0$ and $g : \Omega \to \mathbb{R}$ a $\mathcal{T}$-measurable function such that $|f(\omega, t)| \leq g(\omega)$ for any $t \in (t_0 - \varepsilon, t_0 + \varepsilon)$ and g is $\mathcal{L}^1(\mu)$,*

then the function

$$t \to \int_\Omega f(\omega, t)\ \mathrm{d}\mu \tag{A.1}$$

is continuous at t_0.

A.3 Weak convergence and Lévy's theorem

Definition A.2 (Characteristic function [80, Section 16.1]**)** *Let X be a real-valued random variable. The characteristic function of X, $\varphi_X \colon \mathbb{R} \to \mathbb{C}$, $t \mapsto \varphi_X(t)$, is given by*

$$\varphi_X(t) = \mathbb{E}\left[e^{itX}\right], \quad t \in \mathbb{R}. \tag{A.2}$$

Definition A.3 (Weak convergence [80, Section 17.1]**)** *Let $\{X_n\}$ be a sequence of random variables and X be another random variable. The probability distributions $\{P_{X_n}\}$ converge weakly to P_X as $n \to \infty$ if, for every bounded continuous function f on $\mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{E}\left[f(X_n)\right] = \mathbb{E}\left[f(X)\right]. \tag{A.3}$$

Since both the real and the imaginary parts of the function $x \mapsto e^{itx}$ are bounded and continuous for every $t \in \mathbb{R}$, if $\{P_{X_n}\}$ converge weakly to P_X as $n \to \infty$, then the characteristic functions of $\{X_n\}$ must converge pointwise to the characteristic function of X. The reverse is also true:

Theorem A.1 (Lévy's convergence theorem [80, Section 18.1]**)** *Consider a sequence of random variables $\{X_n\}$ with respective characteristic functions $\{\varphi_{X_n}\}$. If, as $n \to \infty$, φ_{X_n} converges pointwise to some function φ that is continuous at 0, then φ is the characteristic function of some random variable X, and P_{X_n} converges weakly to P_X.*

Theorem A.2 (Extension of characteristic function [45, Theorem 9.6.4]**)** *Let $\varphi_n, n = 1, 2, \ldots,$ be the characteristic functions of the distribution functions $\{F_n\}$. Suppose that φ_n converges when $n \to +\infty$ in an interval $(-\eta, \eta)$ to a function φ.*

1) *If $z \mapsto \varphi(z)$ is analytic and bounded in $|t| \leq \eta$ where $z = t + i\tau$ with $0 < \tau \leq r$ for a chosen $r > 0$ and*

2) *if φ is continuous at $t = 0$, or more generally, the Fourier series of φ is summable to 1 at $t = 0$,*

then F_n converges to a distribution function F and φ is uniquely extended to the characteristic function of F on $(-\infty, +\infty)$.

A.4 Uniform convergence of measures

Definition A.4 (Equicontinuity) *Let $\mathcal{X}$ and $\mathcal{Y}$ be two metric spaces. We shall denote by $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ the respective metrics of these spaces. A class of functions $\mathfrak{A}$ from $\mathcal{X}$ to $\mathcal{Y}$ is said pointwise equicontinuous or equicontinuous if for any $x \in \mathcal{X}$, for any $\varepsilon > 0$, there exists $\delta > 0$ for which $d_{\mathcal{Y}}(f(x), f(y)) < \varepsilon$ for all $y \in \mathcal{Y}$ such that $d_{\mathcal{X}}(x, y) < \delta$ and all $f \in \mathfrak{A}$.*

Theorem A.3 (Uniform convergence of measures [63, Theorem 3.1][9, Theorem 8.2.18]**)** *Let $\mathfrak{A}$ be a class of continuous functions on the separable metric space $\mathcal{X}$ possessing the following properties:*

1) *$\mathfrak{A}$ is uniformly bounded i.e. there exists a constant M such that $|f(x)| \leq M$ for all $f \in \mathfrak{A}$ and $x \in \mathcal{X}$,*

2) *$\mathfrak{A}$ is equicontinuous,*

then for any sequences of measures $\{\mu_n\}$, $\{\mu_n\}$ converges weakly to the measure μ if and only if for each family $\mathfrak{A}$ satisfying the two previous conditions, we have

$$\lim_{n \to +\infty} \sup_{f \in \mathfrak{A}} \left| \int_{\mathcal{X}} f \mathrm{d}\mu_n - \int_{\mathcal{X}} f \mathrm{d}\mu \right| = 0. \tag{A.4}$$

A.5 Concentration inequalities

Lemma A.2 (Markov's inequality [80, Section 6.4]**)** *Consider a real and positive random variable X. For any $\varepsilon > 0$,*

$$\mathbb{P}\left[X > \varepsilon\right] \leq \frac{\mathbb{E}[X]}{\varepsilon}. \tag{A.5}$$

Lemma A.3 (Selection lemma [6, Lemma 2.2]**)** *Let X_n be a random variable taking values in $\mathcal{X}_n$, let $\mathcal{F}$ be a finite set of functions $f : \mathcal{X}_n \mapsto \mathbb{R}^+$ such that the cardinality of $\mathcal{F}$ does not depend on n and*

$$\mathbb{E}_{X_n}[f(X_n)] \to 0, \quad n \to +\infty, \qquad \forall f \in \mathcal{F}. \tag{A.6}$$

Then for any $\varepsilon > 0$, for large enough n, there exists a specific realization x_n of X_n such that

$$f(x_n) \leq \varepsilon, \qquad \forall f \in \mathcal{F}. \tag{A.7}$$

Proof: Let $\varepsilon > 0$. Consider n large enough such that $\mathbb{E}\left[f(X_n)\right] < \varepsilon |\mathcal{F}|$, for any $f \in \mathcal{F}$. Using the union bound and Markov's inequality, we obtain

$$\begin{aligned} \mathbb{P}\left[\cup_{f \in \mathcal{F}} \left\{f(X_n) \geq \varepsilon\right\}\right] &\leq \sum_{f \in \mathcal{F}} \mathbb{P}\left[f(X_n) \geq \varepsilon\right] \\ &\leq \sum_{f \in \mathcal{F}} \frac{\mathbb{E}\left[f(X_n)\right]}{\varepsilon} \\ &< 1. \end{aligned} \tag{A.8}$$

□

Lemma A.4 (Chebyshev's inequality [80, Section 7.3]**)** *Consider a random variable X of non-zero finite variance and any $a > 0$, then*

$$\mathbb{P}\left[|X - \mathbb{E}[X]| > a\right] \leq \frac{\mathrm{Var}[X]}{a^2}. \tag{A.9}$$

Theorem A.4 (Moderate Deviations [23, Theorem 3.7.1]**)** *Let $X_1, \dots, X_n$ be a sequence of $\mathbb{R}$-valued i.i.d. random variables with variance σ^2 such that for any $i = 1, \dots, n$, $f(\lambda) = \ln(\mathbb{E}[e^{\lambda X_i}]) < \infty$ in some ball around the origin and $\mathbb{E}[X_i] = 0$. Fix $a_n \to 0$ such that $n \times a_n \to +\infty$ as $n \to +\infty$, and let $Z_n = \sqrt{\frac{a_n}{n}} \sum_{i=1}^{n} X_i$. Then, for every measurable set Γ,*

$$\limsup_{n \to +\infty} a_n \ln\left(\mathbb{P}\left[Z_n \in \Gamma\right]\right) \leq -\frac{1}{2} \inf_{x \in \bar{\Gamma}} \frac{x^2}{\sigma^2}, \tag{A.10}$$

where $\bar{\Gamma}$ is the closure of the set Γ.

We recall here the Berry-Esseen theorem, which can be seen as a quantitative version of the Central Limit Theorem.

Theorem A.5 (Berry-Esseen Theorem [29, Theorem 2, Ch. XVI.5]**)** *Consider $\{X_k\}$ a sequence of independent random variables, with non-zero variance and finite third moment. For all $\lambda \in \mathbb{R}$,*

$$\left| \mathbb{P}\left[\sum_{k=1}^{n} \left(X_k - \mathbb{E}\left[X_k\right]\right) \geq \lambda \sqrt{\sum_{k=1}^{n} \mathrm{Var}(X_k)} \right] - Q(\lambda) \right| \leq \frac{6 \sum_{i=1}^{n} \mathbb{E}\left[\left|X_k - \mathbb{E}\left[X_k\right]\right|^3\right]}{\left(\sum_{k=1}^{n} \mathrm{Var}(X_k)\right)^{\frac{3}{2}}}. \tag{A.11}$$

A.6 Tail bounds for non-central chi-squared random variables

A *non-central chi-squared* random variable X with n degrees of freedom is the sum of the squares of n independent standard Gaussian random variables with means respectively $\mu_1, \dots, \mu_n$:

$$X = \sum_{i=1}^{n} N_i^2 \tag{A.12}$$

where $N_i \sim \mathcal{N}(\mu_i, 1)$ for all i.
We denote $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$, then the tail of X [40, Section VII, eq 2.17] is characterized as follows:

$$\mathbb{P}\left[X \geq t\right] = Q_{\frac{n}{2}}\left(\sqrt{\|\boldsymbol{\mu}\|_2}, \sqrt{t}\right) \tag{A.13}$$

where $Q_{\frac{n}{2}}$ is the Marcum Q-function of order $\frac{n}{2}$ [40, Section VII, eq 2.18].

B Hypothesis testing

In this appendix, we state the Neyman-Pearson lemma and recall some known bounds for the probability of missed detection and false alarm in hypothesis testing. We consider the general problem of hypothesis testing between two possible distributions P and Q associated with a random variable Y on the measurable space $(\mathcal{Y}, \mathcal{T})$ such that $P \ll Q$ and define their associated hypotheses:

$$\begin{aligned} H_0 &: Y \sim Q \\ H_1 &: Y \sim P. \end{aligned} \tag{B.1}$$

We consider the following test with output 0 or 1 when choosing respectively hypothesis H_0 or H_1:

$$Z : \mathcal{Y} \to \{0, 1\}. \tag{B.2}$$

The probability of false positive error (false alarm), meaning the probability of deciding for P i.e. H_1 when in fact $Y \sim Q$, is

$$\beta = \int_{\mathcal{Y}} Z(y) \, \mathrm{d}Q \tag{B.3}$$

and the probability of false negative error (missed detection), meaning the probability of deciding for H_0 when actually $Y \sim P$ is

$$\kappa = 1 - \int_{\mathcal{Y}} Z(y) \, \mathrm{d}P. \tag{B.4}$$

B.1 Neyman-Pearson lemma

We know that the optimal test to distinguish between the two hypotheses is a maximum likelihood test involving the threshold of the Radon-Nikodym derivative [36, p. 128] of P with respect to Q:

Lemma B.1 (Neyman-Pearson lemma [59, 62]) *For any $\alpha \in [0, 1]$ there exists $\gamma^* > 0$ such that the test*

$$\begin{aligned} T_{\gamma^*} &: \mathcal{Y} \to \{0, 1\} \\ &\quad y \mapsto \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) \geq \gamma^*\right\}} \end{aligned} \tag{B.5}$$

achieves the optimal performance

$$\beta_\alpha(P, Q) = \int_{\mathcal{Y}} T_{\gamma^*}(y) \, \mathrm{d}Q = \min_{\int_{\mathcal{Y}} Z(y) \, \mathrm{d}P \, \geq \, \alpha} \int_{\mathcal{Y}} Z(y) \, \mathrm{d}Q \tag{B.6}$$

where $Z : \mathcal{Y} \to \{0, 1\}$ stands for any test deciding on H_1 with output 1 and H_0 with output 0. The constant γ^* is uniquely determined by solving the equation (B.6). Moreover, any other test Z satisfying $\int_{\mathcal{Y}} Z(y) \, \mathrm{d}P \geq \alpha$ either differs from T_{γ^*} only on the set $\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) = \gamma^*\right\}$ or is strictly larger with respect to Q i.e. $\int_{\mathcal{Y}} Z(y) \, \mathrm{d}Q > \int_{\mathcal{Y}} T_{\gamma^*}(y) \, \mathrm{d}Q$.

Thus the Neyman-Pearson lemma B.1 allows to restrict the set of tests to the maximum likelihood tests. In particular the Neyman-Pearson optimal test (B.5) achieves the following minimum error:

Lemma B.2 ([51, Theorem 13.1.1]) *We consider the following family of maximum likelihood tests indexed by $\gamma > 0$:*

$$\begin{aligned} T_\gamma &: \mathcal{Y} \to \{0, 1\} \\ &\quad y \mapsto \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) \geq \gamma\right\}} \end{aligned} \tag{B.7}$$

where the output 0 *indicates that the test chooses* H_0 *and the output* 1, H_1. *We denote the probability of detecting* H_1 *when in fact* $Y \sim Q$ *by*

$$\beta = \int_{\mathcal{Y}} T_\gamma(y)\ \mathrm{d}Q \tag{B.8}$$

and the probability of missing H_1 *when actually* $Y \sim P$ *by*

$$\kappa = 1 - \int_{\mathcal{Y}} T_\gamma(y)\ \mathrm{d}P. \tag{B.9}$$

Then

$$\min_\gamma(\beta + \kappa) = 1 - \frac{1}{2} d_{TV}\left(P, Q\right). \tag{B.10}$$

Proof: We notice that combining (B.8) and (B.9), we obtain

$$\beta + \kappa = 1 + \int_{\mathcal{Y}} T_\gamma(y)\left(\mathrm{d}Q - \mathrm{d}P\right), \tag{B.11}$$

therefore the minimum of (B.11) is obtained by setting $T_\gamma(y) = 1$ when $\frac{\mathrm{d}P}{\mathrm{d}Q}(y) > 1$ and $T_\gamma(y) = 0$ when $\frac{\mathrm{d}P}{\mathrm{d}Q} < 1$. On the set $\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) = 1\right\}$, it does not matter how T is defined as the corresponding integral is equal to zero. Then we can write

$$\min_\gamma(\beta + \kappa) = 1 + \int_{\mathcal{Y}} \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) > 1\right\}}\left(\mathrm{d}Q - \mathrm{d}P\right) \tag{B.12}$$

By symmetry, we have

$$\min_\gamma(\beta + \kappa) = 1 + \int_{\mathcal{Y}} \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) < 1\right\}}\left(\mathrm{d}P - \mathrm{d}Q\right). \tag{B.13}$$

The equality can be obtained by noticing that

$$\begin{aligned} d_{TV}\left(P, Q\right) &= 2 \sup_{y \in \mathcal{Y}} |P(y) - Q(y)| \\ &= \int_{\mathcal{Y}} \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) > 1\right\}}\left(\mathrm{d}P - \mathrm{d}Q\right) + \int_{\mathcal{Y}} \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) < 1\right\}}\left(\mathrm{d}Q - \mathrm{d}P\right). \end{aligned} \tag{B.14}$$

□

B.2 General properties of β_α

The optimal false positive error β_α in (B.6) given by the Neyman-Pearson Lemma B.1 satisfies the two following properties [59, p. 2316][61, Theorem 14.10].

1) For any $\gamma > 0$, $\tau \in \mathbb{R}$,

$$\begin{aligned} P\left[\frac{\mathrm{d}P}{\mathrm{d}Q}(Y) \geq \tau\right] - \gamma Q\left[\frac{\mathrm{d}P}{\mathrm{d}Q}(Y) \geq \tau\right] &= \int_{\mathcal{Y}} \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) \geq \tau\right\}}\left(\mathrm{d}P - \gamma\ \mathrm{d}Q\right) \\ &\leq \int_{\mathcal{Y}} \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) \geq \tau\right\}} \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) \geq \gamma\right\}}\left(\mathrm{d}P - \gamma\ \mathrm{d}Q\right) \\ &\leq P\left[\frac{\mathrm{d}P}{\mathrm{d}Q}(Y) \geq \tau, \frac{\mathrm{d}P}{\mathrm{d}Q}(Y) \geq \gamma\right] \\ &\leq P\left[\frac{\mathrm{d}P}{\mathrm{d}Q}(Y) \geq \gamma\right]. \end{aligned} \tag{B.15}$$

In particular, for any $\alpha \geq 0$ and $\gamma > 0$, we obtain

$$\alpha \leq \mathbb{P}\left[\frac{\mathrm{d}P}{\mathrm{d}Q}(Y) \geq \gamma\right] + \gamma \beta_\alpha(P, Q); \tag{B.16}$$

where we chose $\tau = \gamma^*$ given by the Neyman Pearson Lemma B.1 in (B.15) such that it forms the Neyman-Pearson test achieving

$$P\left[\frac{\mathrm{d}P}{\mathrm{d}Q}(Y) \geq \gamma^*\right] \geq \alpha \tag{B.17}$$

and

$$Q\left[\frac{\mathrm{d}P}{\mathrm{d}Q}(Y) \geq \gamma^*\right] = \beta_\alpha(P,Q). \tag{B.18}$$

2) For any $0 \leq \alpha \leq 1$ and $\gamma > 0$ satisfying

$$\int_{\mathcal{Y}} \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) \geq \gamma\right\}} \mathrm{d}P \geq \alpha, \tag{B.19}$$

we have

$$\begin{aligned}
\beta_\alpha(P,Q) &= \min_{T:\mathcal{Y}\mapsto[0,1] \text{ s.t. } \int_{\mathcal{Y}} T_\gamma(y)\ \mathrm{d}P \geq \alpha} \int_{\mathcal{Y}} T_\gamma(y)\ \mathrm{d}Q \\
&\leq \int_{\mathcal{Y}} \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) \geq \gamma\right\}} \mathrm{d}Q \\
&\leq \frac{1}{\gamma} \int_{\mathcal{Y}} \mathbb{1}_{\left\{\frac{\mathrm{d}P}{\mathrm{d}Q}(y) \geq \gamma\right\}} \mathrm{d}P \\
&\leq \frac{1}{\gamma}.
\end{aligned} \tag{B.20}$$

Furthermore, β_α satisfies the following data processing inequality.

Theorem B.1 (Data processing inequality for $\boldsymbol{\beta_\alpha}$ [60, Section V]) *We consider two input distributions P_X and Q_X on $\mathcal{X}$ of the same channel with law $P_{Y|X}$ and the respective output distributions P_Y and Q_Y on $\mathcal{Y}$. The data-processing inequality for β_α states that for any $\alpha \geq 0$*

$$\beta_\alpha(P_X, Q_X) \leq \beta_\alpha(P_Y, Q_Y). \tag{B.21}$$

Proof:

$$\begin{aligned}
\beta_\alpha(P_Y, Q_Y) &= \min_{T:\mathcal{Y}\mapsto[0,1] \text{ s.t. } \int_{\mathcal{Y}} T(y)\ \mathrm{d}P_Y \geq \alpha} \int_{\mathcal{Y}} T(y)\ \mathrm{d}Q_Y \\
&= \min_{T:\mathcal{Y}\mapsto[0,1] \text{ s.t. } \int_{\mathcal{X}} \int_{\mathcal{Y}} T(y)\ \mathrm{d}P_{Y|X}\ \mathrm{d}P_X \geq \alpha} \int_{\mathcal{X}} \int_{\mathcal{Y}} T(y)\ \mathrm{d}P_{Y|X}\ \mathrm{d}Q_X \\
&\geq \min_{T':\mathcal{X}\mapsto[0,1] \text{ s.t. } \int_{\mathcal{X}} T'(x)\ \mathrm{d}P_X \geq \alpha} \int_{\mathcal{X}} T'(x)\ \mathrm{d}Q_X
\end{aligned} \tag{B.22}$$

where (B.22) follows by the theorem of Fubini-Tonelli for measure product [5, Section 7.3.6] and (B.22) follows because all T' are not necessarily of the form $\int_{\mathcal{Y}} T(y)\mathrm{d}P_{Y|X}$. □

B.3 Converse bounds in the finite blocklength regime

A general upper bound on the maximum size of the message set for the maximal probability of decoding error ε was first derived in [59]:

Theorem B.2 *[59, Theorem 31] Consider a channel $P_{Y|X}$ and any code for message set $\mathcal{M}$ with the maximal probability of error $\varepsilon \in (0,1)$, subject to the constraint $f(m) \in \mathcal{F}$, for all $m \in \mathcal{M}$. Fix a probability measure Q_Y on $\mathcal{Y}$. Suppose that for any $0 < \alpha < 1$, $\beta_\alpha\left(P_{Y|X=x}, Q_Y\right) = \beta_\alpha$ has the same value for any $x \in \mathcal{F}$. Then*

$$|\mathcal{M}| \leq \frac{1}{\beta_{1-\varepsilon}}. \tag{B.23}$$

A similar converse bound holds for the average probability of error:

Theorem B.3 *[59, Theorem 28] Consider a channel $P_{Y|X}$ and any code for message set $\mathcal{M}$ with the average probability of error $\varepsilon \in (0,1)$, subject to the constraint $f(m) \in \mathcal{F}$, for all $m \in \mathcal{M}$. Fix a probability measure Q_Y on $\mathcal{Y}$. Suppose that for any $0 < \alpha < 1$, $\beta_\alpha\left(P_{Y|X=x}, Q_Y\right) = \beta_\alpha$ has the same value for any $x \in \mathcal{F}$. Then*

$$|\mathcal{M}| \leq \frac{1}{\beta_{1-\varepsilon}}. \tag{B.24}$$

C Information theory tools

In this section, we recall some well-known results in information theory that are used in this thesis.

C.1 Data-processing inequality for the Kullback-Leibler divergence

Theorem C.1 *[61, Theorem 2.16] Consider two input distributions P_X and Q_X for the same channel with law $P_{Y|X}$ and the respective output distributions P_Y and Q_Y. The data-processing inequality for the Kullback-Leibler divergence states that*

$$D\left(P_Y||Q_Y\right) \leq D\left(P_X||Q_X\right). \tag{C.1}$$

C.2 Fano's inequality

Theorem C.2 *[18, Theorem 2.10.1] For any estimator $\hat{X}$ such that $X \to Y \to \hat{X}$, with probability of error $P_e = \mathbb{P}\left\{\hat{X} \neq X\right\}$, we have*

$$H(P_e) + P_e \ln|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y). \tag{C.2}$$

This inequality can be weakened to

$$1 + P_e \ln|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y). \tag{C.3}$$

C.3 Channels with exponential noise

This section outlines some known results about the exponential noise channel. First, we notice that under a mean constraint, an exponential distribution maximizes the entropy [18, Section 12], which implies the following.

Theorem C.3 *[72, Theorem 1] Consider $P, \Lambda > 0$. Let Z be exponentially distributed with mean Λ. Consider the following mixture of a point mass and an exponential distribution:*

$$\mathbb{P}\left[X = 0\right] = \frac{\Lambda}{\Lambda + P}, \tag{C.4}$$

$$\mathbb{P}\left[X > x \mid X > 0\right] = e^{-\frac{x}{\Lambda+P}} \tag{C.5}$$

Assuming X and Z are independent, then $Y = X + Z$ follows an exponential distribution $\mathcal{E}\left(\Lambda + P\right)$ with mean $\Lambda + P$ and in particular

$$I(X;Y) = \ln\left(1 + \frac{P}{\Lambda}\right). \tag{C.6}$$

Furthermore for any nonnegative random variable $\tilde{X}$, independent of Z, with mean P,

$$I(\tilde{X}; \tilde{X} + Z) \leq I(X;Y), \tag{C.7}$$

with equality only if $\tilde{X} = X$.

Proof: We check that $Y = X + Z$ follows an exponential output with mean $\Lambda + P$ using Laplace transforms. The Laplace transform of Z is

$$t \mapsto \int_0^{+\infty} e^{-tz} \frac{1}{\Lambda} e^{-\frac{z}{\Lambda}} \mathrm{d}z = \frac{1}{1 + \Lambda t} \tag{C.8}$$

and the Laplace transform of X is

$$t \mapsto \frac{\Lambda}{\Lambda + P} + \frac{P}{\Lambda + P} \frac{1}{1 + (\Lambda + P)t}. \tag{C.9}$$

We recover the Laplace transform of Y as

$$\mathbb{E}[e^{-tY}] = \mathbb{E}[e^{-tX}] \times \mathbb{E}[e^{-tZ}] = \frac{1}{1 + (\Lambda + P)t}. \tag{C.10}$$

□

Consider the following channel

$$Y_i = X_i + Z_i, \qquad i = 1, \ldots, n, \tag{C.11}$$

where all Z_i are i.i.d. exponential random variables with mean Λ; and any codeword $(X_1, \ldots, X_n)$ is constrained to satisfy

$$X_i \geq 0, \qquad i = 1, \ldots, n, \text{ and } \frac{1}{n} \sum_{i=1}^{n} X_i \leq P. \tag{C.12}$$

Recalling Theorem C.3, the following theorem presents the optimal rate for the channel (C.11) achieved by the well-defined input distribution (C.4).

Theorem C.4 *[72, Theorem 3] The capacity of the additive memoryless exponential noise channel* (C.11) *under the constraint* (C.12) *is equal to*

$$\ln\left(1 + \frac{P}{\Lambda}\right). \tag{C.13}$$

C.4 Channels with generalized Gaussian noise

This section outlines some known results about the additive channel with generalized Gaussian noise. A generalized Gaussian random variable X [55, 24, 25] of parameters $p, \mu, \sigma > 0$ is denoted $X \sim \mathcal{N}_p(0, \sigma^p)$ and has the PDF:

$$p_X(x) = \frac{c_p}{\sigma} e^{-\frac{|x-\mu|^p}{2\sigma^p}}, \qquad x \in \mathbb{R}, \tag{C.14}$$

where

$$c_p = \frac{p}{2^{\frac{p+1}{p}} \Gamma(\frac{1}{p})}. \tag{C.15}$$

Note that

$$\mathbb{E}\left[|X|^p\right] = \frac{2\sigma^p}{p} \text{ and } h(X) = \ln\left(\frac{\sigma}{c_p}\right) + \frac{1}{p}. \tag{C.16}$$

First, we notice that under an absolute p-moment constraint, a generalized Gaussian distribution of parameter p maximizes the entropy [18, Section 12]: let $a > 0$, any variable $\tilde{X}$ such that $\mathbb{E}\left[\left|\tilde{X}\right|^p\right] \leq a$ satisfies

$$h(\tilde{X}) \leq \frac{1}{p} \ln\left(\frac{a \times p}{2}\right) - \ln\left(c_p\right) + \frac{1}{p}, \tag{C.17}$$

where the right-hand side of (C.17) is the entropy of $X \sim \mathcal{N}_p\left(0, a^{\frac{p}{2}}\right)$.

Definition C.1 (Self-decomposability [25, Definition 8]**)** *Consider a random variable $Y \sim \mathcal{N}_p(0, 1)$. Y is self-decomposable if for every $\alpha \geq 1$ there exists a random variable X_α independent of $Z \sim \mathcal{N}_p(0, 1)$ such that*

$$\alpha Y = X_\alpha + Z. \tag{C.18}$$

Theorem C.5 *[25, Theorem 6] Consider $Y \sim \mathcal{N}_p(0,1)$. Y is self-decomposable for $p \in (0,1] \cup \{2\}$.*

Consider the following channel

$$Y_i = X_i + Z_i, \qquad i = 1, \ldots, n, \tag{C.19}$$

where all Z_i are i.i.d. $\sim \mathcal{N}_p(0,1)$, $p > 0$. The capacity is not known except for special cases ($p = 2$ corresponding to Gaussian noise and $p = 1$ corresponding to Laplace noise) but an upper bound was given in [24, Proposition 7].

Remark C.1 *In [C1], we showed that for the additive generalized Gaussian channel the covertness constraint implies an output constraint on the p-th moment [10, eq (24)].*
If we consider a p-th moment constraint on the output, then for $p \in (0,1] \cup \{2\}$ the self-decomposability property shows that it is possible to obtain an output with generalized Gaussian distribution, which maximizes the entropy and thus the mutual information; and the channel capacity is equal to

$$C = \frac{1}{p} \ln \left(\frac{\mathbb{E}[|Y|^p] \times p}{2} \right). \tag{C.20}$$

D Special cases of degraded channels

In this appendix, we consider covert communication scenarios where the eavesdropper's channel is not the same as the channel of the legitimate receiver. We compute the exact value of the scaling constant L in the case of degraded AWGN channels and recover the result previously shown in [78, Theorem 2] of the scaling constant L using different techniques. Then we provide bounds for the key length and show how no key is needed when the noise variance for the eavesdropper's channel is strictly greater than that of the main channel. This last finding is similar to [7, Theorem 6] with the exception that [7] only considers binary discrete input distributions. Finally, we present an upper bound for L in the case of degraded exponential channels.

D.1 Gaussian degraded channel

We consider here scenario in equation (2.45) where the eavesdropper and the legitimate receiver observe the outputs of two different AWGN channels with noise power σ_e^2 for the eavesdropper and σ^2 for the legitimate receiver; see Figure 5. In this setup, the covertness constraint (2.6) can be written in the form (2.46).

D.1.1 Proof of Theorem 2.6

First, we introduce the following lemma.

Lemma D.1 *For any centered Gaussian random variable $Z \sim \mathcal{N}(0, \sigma^2)$, for any random variable Y, we have the two following inequalities*

$$h(Y) \leq \frac{1}{2} \ln(2\pi \mathbb{E}[Y^2]) + \frac{1}{2}, \tag{D.1}$$

$$D(P_Y || P_Z) \geq \frac{1}{2} \ln\left(\frac{\sigma^2}{\mathbb{E}[Y^2]}\right) + \frac{1}{2}\left(\frac{\mathbb{E}[Y^2]}{\sigma^2} - 1\right). \tag{D.2}$$

Furthermore, the inequalities (D.1) and (D.2) hold as equalities when Y follows a centered Gaussian distribution.

Proof: First, we notice that (D.1) comes directly from the maximization of the entropy for a fixed second moment: the maximization is achieved by a Gaussian distribution.
We then show (D.2) via the following:

$$\begin{aligned} D(P_Y||P_Z) &= -h(Y) - \int_{\mathbb{R}} p_Y(y) \ln(p_Z(y)) \mathrm{d}y \\ &= -h(Y) + \int_{\mathbb{R}} p_Y(y) \left(\ln\left(\sqrt{2\pi}\sigma\right) + \frac{y^2}{2\sigma}\right) \mathrm{d}y \\ &= -h(Y) + \ln\left(\sqrt{2\pi}\sigma\right) + \frac{\mathbb{E}[Y^2]}{2\sigma^2} \\ &\geq -\frac{1}{2} \ln\left(2\pi \mathbb{E}[Y^2]\right) - \frac{1}{2} + \ln\left(\sqrt{2\pi}\sigma\right) + \frac{\mathbb{E}[Y^2]}{2\sigma^2} \\ &= \frac{1}{2} \ln\left(\frac{\sigma^2}{\mathbb{E}[Y^2]}\right) + \frac{1}{2}\left(\frac{\mathbb{E}[Y^2]}{\sigma^2} - 1\right) \end{aligned} \tag{D.3}$$

which is the desired inequality. Note that (D.3) follows from (D.1), and that it holds with equality when Y follows a centered Gaussian distribution. □

Proof of Theorem 2.6, converse part: Take any code $\mathcal{C}$ of length n. Let $\bar{X}$ denote a random variable such that $P_{\bar{X}}$ is the average input distribution over the secret key, a uniformly drawn message, and the n

channel uses. Let $\bar{Y}$ and $\bar{Y}_e$ denote the corresponding outputs through the legitimate receiver's channel and the eavesdropper's channel respectively:

$$P_{\bar{X}}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} P_{X_{\mathcal{C},i}}(\cdot), \tag{D.4}$$

$$P_{\bar{Y}}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} P_{Y_{\mathcal{C},i}}(\cdot), \tag{D.5}$$

$$P_{\bar{Y}_e}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} P_{Y_{e,\mathcal{C},i}}(\cdot), \tag{D.6}$$

where $X_{\mathcal{C},i}$, $Y_{e,\mathcal{C},i}$, and $Y_{\mathcal{C},i}$ are the i^{th} components of respectively $X_{\mathcal{C}}^n$, $Y_{e,\mathcal{C}}^n$, and $Y_{\mathcal{C}}^n$. Note that $P_{\bar{Y}}$ and $P_{\bar{Y}_e}$ are the average output distributions respectively for the legitimate receiver and the eavesdropper. Starting with the condition (2.46), we have:

$$\begin{aligned}
\Delta &\geq D(P_{Y_{e,\mathcal{C}}^n}||P_{Z_e^n}) \\
&\geq nD(P_{\bar{Y}_e}||P_{Z_e}) \qquad \text{(D.7)} \\
&\geq n\left(\frac{1}{2}\ln\left(\frac{\sigma_e^2}{\mathbb{E}[\bar{Y}_e^2]}\right) + \frac{1}{2}\left(\frac{\mathbb{E}[\bar{Y}_e^2]}{\sigma_e^2} - 1\right)\right) \\
&\geq n\left(-\frac{1}{2}\ln\left(1 + \frac{\mathrm{Var}[\bar{X}]}{\sigma_e^2}\right) + \frac{1}{2}\frac{\mathrm{Var}[\bar{X}]}{\sigma_e^2}\right), \qquad \text{(D.8)}
\end{aligned}$$

where (D.7) follows from the same steps as (3.36), and (D.8) follows from inequality (D.2) in Lemma D.1.

From (D.8), we notice that as $n \to +\infty$, $\mathrm{Var}[\bar{X}]$ must approach zero, and:

$$\mathrm{Var}[\bar{X}] = O\left(\frac{1}{\sqrt{n}}\right); \tag{D.9}$$

then from (D.8), we have

$$\begin{aligned}
\frac{\Delta}{n} &\geq -\frac{1}{2}\left(\frac{\mathrm{Var}[\bar{X}]}{\sigma_e^2} - \frac{1}{2}\left(\frac{\mathrm{Var}[\bar{X}]}{\sigma_e^2}\right)^2 + o\left(\frac{1}{\sqrt{n}}\right)\right) + \frac{1}{2}\frac{\mathrm{Var}[\bar{X}]}{\sigma_e^2} \\
&\geq \frac{1}{4}\left(\frac{\mathrm{Var}[\bar{X}]}{\sigma_e^2}\right)^2 + o\left(\frac{1}{\sqrt{n}}\right) \qquad \text{(D.10)}
\end{aligned}$$

i.e.

$$\mathrm{Var}[\bar{X}] \leq 2\sqrt{\frac{\Delta}{n}}\sigma_e^2 + o\left(\frac{1}{\sqrt{n}}\right). \tag{D.11}$$

We next derive a bound on $M^*(n, \varepsilon, \Delta)$ in terms of $\bar{X}$ and $\bar{Y}$. For each realization k of the key K, we denote by ε_k its probability of error. Let ε be the average probability of error over the random codebook. For each k, we have by Fano's inequality:

$$\ln|\mathcal{M}|(1-\varepsilon_k) - 1 \leq I(X_{\mathcal{C}}^n; Y_{\mathcal{C}}^n|K = k). \tag{D.12}$$

By averaging over the random code, we obtain

$$\begin{aligned}
\ln|\mathcal{M}|(1-\varepsilon) - 1 &\leq I(X_{\mathcal{C}}^n; Y_{\mathcal{C}}^n|K) \\
&\leq n(h(\bar{Y}) - h(Z)), \qquad \text{(D.13)}
\end{aligned}$$

where (D.13) follows from the same step as (3.41). By the definition of $M^*(n, \varepsilon, \Delta)$, (D.13) implies

$$\ln\left(M^*(n, \varepsilon, \Delta)\right)(1-\varepsilon) - 1 \leq n(h(\bar{Y}) - h(Z)). \tag{D.14}$$

By the inequality (D.1) in Lemma D.1, we know that:

$$\begin{aligned}\ln\left(M^*(n,\varepsilon,\Delta)\right)(1-\varepsilon)-1 &\leq n\left(\frac{1}{2}\ln\left(2\pi\mathbb{E}[\bar{Y}^2]\right)+\frac{1}{2}-\left(\frac{1}{2}\ln(2\pi\sigma^2)+\frac{1}{2}\right)\right)\\ &= \frac{n}{2}\ln\left(\frac{\mathbb{E}[\bar{Y}^2]}{\sigma^2}\right)\\ &= \frac{n}{2}\ln\left(1+\frac{\mathrm{Var}[\bar{X}]}{\sigma^2}\right)\\ &\leq \frac{n}{2}\frac{\mathrm{Var}[\bar{X}]}{\sigma^2}. \end{aligned} \tag{D.15}$$

Then by injecting (D.11) in (D.15) we obtain

$$\begin{aligned}\ln\left(M^*(n,\varepsilon,\Delta)\right)(1-\varepsilon)-1 &\leq \frac{\sigma_e^2}{\sigma^2}\frac{n}{2}\sqrt{4\frac{\Delta}{n}+o\left(\frac{1}{\sqrt{n}}\right)}\\ &= \sqrt{n}\frac{\sigma_e^2}{\sigma^2}\sqrt{\Delta}+o\left(\sqrt{n}\right). \end{aligned} \tag{D.16}$$

Finally recalling the definition (2.23) of L, taking $n\to\infty$ and $\varepsilon\to 0$ in (D.16), we obtain the desired upper bound:

$$L\leq\frac{\sigma_e^2}{\sigma^2}. \tag{D.17}$$

□

Proof of Theorem 2.6, achievability part: We consider a random code C for n channel uses in which every codeword is i.i.d. according to $P_X^{\otimes n}$, with P_X a normal centered distribution:

$$X\sim\mathcal{N}\left(0,2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right) \tag{D.18}$$

and X is independent of $Z\sim\mathcal{N}(0,\sigma^2)$ and $Z_e\sim\mathcal{N}_p(0,\sigma_e^2)$. We denote by $X^n=(X_1,\dots,X_n)$ the associated i.i.d. input sequence; by $Y^n=(Y_1,\dots,Y_n)$ the associated i.i.d. output sequence for the receiver with $Y_i\sim\mathcal{N}\left(0,\sigma^2+2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)$ for all $1\leq i\leq n$; and by $Y_e^n=(Y_{e,1},\dots,Y_{e,n})$ the associated i.i.d. output sequence for the eavesdropper with $Y_{e,i}\sim\mathcal{N}\left(0,\sigma_e^2+2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)$ for all $1\leq i\leq n$. Note that the distributions of every input and output symbol depend on n.

We check that the random code C satisfies the covertness condition (2.46). Similarly to the proof for Theorem 3.2 (eq (3.67)), we notice that

$$\mathbb{E}_{\mathsf{C}}\left[D(P_{Y_e^n|\mathsf{C}}||P_{Z_e^n})\right]=D(P_{Y_e^n}||P_{Z_e^n})+\mathbb{E}_{\mathsf{C}}\left[D(P_{Y_e^n|\mathsf{C}}||P_{Y_e^n})\right] \tag{D.19}$$

We assume that the key length is large enough so that $\mathbb{E}_{\mathsf{C}}\left[D(P_{Y_e^n|\mathsf{C}}||P_{Y_e^n})\right]$ is arbitrarily small. (The actual sufficient key length will be characterized in Section D.1.2.) Therefore, ensuring the covertness condition (2.46) on the eavesdropper's channel amounts to checking the following:

$$\begin{aligned} D(P_{Y_e^n}||P_{Z_e^n}) &= n\,D(P_{Y_e}||P_{Z_e}) && \text{(D.20)}\\ &= n\left(\frac{1}{2}\ln\left(\frac{\sigma_e^2}{\mathbb{E}[Y_e^2]}\right)+\frac{1}{2}\left(\frac{\mathbb{E}[Y_e^2]}{\sigma_e^2}-1\right)\right) && \text{(D.21)}\\ &= n\left(-\frac{1}{2}\ln\left(1+\frac{\mathbb{E}[X^2]}{\sigma_e^2}\right)+\frac{1}{2}\frac{\mathbb{E}[X^2]}{\sigma_e^2}\right)\\ &\leq n\frac{1}{4}\left(\frac{\mathbb{E}[X^2]}{\sigma_e^2}\right)^2 && \text{(D.22)}\\ &= \Delta, && \text{(D.23)} \end{aligned}$$

where $Y \sim \mathcal{N}\left(0, \sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)$ and $Y_e \sim \mathcal{N}\left(0, \sigma_e^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)$; (D.20) follows because both Y_e^n and Z_e^n are i.i.d.; (D.21) because we have equality in (D.2); and (D.22) because $\ln(1+a) \geq a - \frac{a^2}{2}$, $a > -1$. It now remains to show that this limit of mutual information is operationally achievable. Namely, we have

$$\lim_{\varepsilon \to 0} \liminf_{n \to \infty} \frac{\ln\left(M^*(n, \varepsilon, \Delta)\right)}{\sqrt{n}} \geq \lim_{n \to \infty} \frac{1}{\sqrt{n}} I(X^n; Y^n), \tag{D.24}$$

following the same steps (3.74)–(3.78) of the proof of Theorem 3.2 by showing

$$\operatorname{Var}\left(\frac{1}{\sqrt{n}} i_{X^n, Y^n}(X^n, Y^n)\right) \to 0, \qquad n \to +\infty. \tag{D.25}$$

We show (D.25) by computing

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{\sqrt{n}} i_{X^n, Y^n}(X^n, Y^n)\right) \\
&= \operatorname{Var}\left(\frac{1}{\sqrt{n}} \ln\left(\frac{p_{Y^n|X^n}(Y^n|X^n)}{p_{Y^n}(Y^n)}\right)\right) \\
&= \operatorname{Var}\left(\frac{1}{\sqrt{n}} \ln\left(\frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{Z_i^2}{2\sigma^2}}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2}} e^{-\frac{Y_i^2}{2\left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)}}}\right)\right) \\
&= \frac{1}{n} \operatorname{Var}\left(-\sum_{i=1}^n \frac{Z_i^2}{2\sigma^2} + \sum_{i=1}^n \frac{Y_i^2}{2\left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)}\right) \\
&= \frac{1}{n} \sum_{i=1}^n \operatorname{Var}\left(\frac{Y_i^2}{2\left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)} - \frac{Z_i^2}{2\sigma^2}\right) \\
&= \operatorname{Var}\left(\frac{Y^2}{2\left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)} - \frac{Z^2}{2\sigma^2}\right) \\
&\leq \frac{1}{4\left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)^2 \sigma^4} \mathbb{E}\left[\left(\sigma^2 Y^2 - \left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right) Z^2\right)^2\right] \\
&= \frac{1}{4\left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)^2 \sigma^4} \left(\sigma^4 \mathbb{E}\left[Y^4\right] - 2\sigma^2 \left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right) \mathbb{E}\left[Y^2 Z^2\right]\right. \\
&\quad \left. + \left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)^2 \mathbb{E}\left[Z^4\right]\right) \\
&= \frac{1}{4\left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)^2 \sigma^4} \left(\sigma^4 \left(3\sigma^4 + 6\sigma^2 \mathbb{E}\left[X^2\right] + \mathbb{E}\left[X^4\right]\right)\right. \\
&\quad \left. - 2\sigma^2 \left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right) \left(\sigma^2 \mathbb{E}\left[X^2\right] + 3\sigma^4\right) + 3\sigma^4 \left(\sigma^2 + 2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)^2\right) \\
&= o\left(\frac{1}{\sqrt{n}}\right).
\end{aligned} \tag{D.26}$$

This establishes (D.25) and (D.24). We continue from (D.24) to complete the proof:

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{\ln\left(M^*(n,\varepsilon,\Delta)\right)}{\sqrt{n}\sqrt{\Delta}}\\
&\geq \liminf_{n\to\infty}\frac{I(X^n,Y^n)}{\sqrt{n}\sqrt{\Delta}}\\
&= \liminf_{n\to\infty}\frac{\sqrt{n}}{\sqrt{\Delta}}I(X,Y)\\
&= \liminf_{n\to\infty}\frac{\sqrt{n}}{\sqrt{\Delta}}\left(h(Y)-h(Z)\right)\\
&= \lim_{n\to\infty}\frac{\sqrt{n}}{\sqrt{\Delta}}\left(\frac{1}{2}\ln(2\pi\mathbb{E}[Y^2])+\frac{1}{2}-\left(\frac{1}{2}\ln(2\pi\sigma^2)+\frac{1}{2}\right)\right)\\
&= \lim_{n\to\infty}\frac{\sqrt{n}}{\sqrt{\Delta}}\frac{1}{2}\ln\left(\frac{\mathbb{E}[Y^2]}{\sigma^2}\right)\\
&= \lim_{n\to\infty}\frac{\sqrt{n}}{\sqrt{\Delta}}\frac{1}{2}\ln\left(1+\frac{\mathbb{E}[X^2]}{\sigma^2}\right)\\
&= \frac{\sigma_e^2}{\sigma^2}, \qquad\qquad (D.27)
\end{aligned}$$

where (D.27) follows by recalling (D.18); which is the desired lower bound. □

D.1.2 Bounds on the key length

Proposition D.1 *In the degraded AWGN setting of equation (2.45), if $\sigma_e^2 > \sigma^2$, no key is needed. Otherwise a sufficient key length is $\ln|\mathcal{K}| = O(\sqrt{n})$.*

Proof: We proceed in a similar way to the proof of Proposition 3.2. We consider the same random code construction as in the proof of Theorem 2.6 where the codewords are generated i.i.d. according to $P_X^{\otimes n}$ such that (D.18) holds. (As before, we omit the dependence in n of the distributions of X and Y_e.) We denote by C the random codebook and by $P_{Y_e^n|\mathsf{C}}$ the corresponding output distribution for the eavesdropper. Referring back to the expression (D.19), we find sufficient conditions for $\ln|\mathcal{K}|$ such that $\mathbb{E}_{\mathsf{C}}\left[D\left(P_{Y_e^n|\mathsf{C}}||P_{Y^n}\right)\right]$ vanishes, which ensures the existence of at least one good deterministic code. To establish this result, we apply the channel resolvability bounds of Theorem 2.1. Let $\rho\in(0,1]$, Theorem 2.1 states that

$$\mathbb{E}_{\mathsf{C}}\left[D\left(P_{Y_e^n|\mathsf{C}}||P_{Y_e^n}\right)\right] \leq \frac{1}{\rho}\ln\left(1+e^{-\rho\ln(|\mathcal{K}|\times|\mathcal{M}|)+n\Psi(\rho|P_{Y_e|X},P_X)}\right) \qquad (D.28)$$

where $\Psi(\rho|P_{Y_e|X},P_X)$ is defined as in (2.21). Therefore studying $\Psi(\rho|P_{Y_e|X},P_X)$ will allow us to derive a sufficient condition on the key length to ensure that (D.28) vanishes.
The direct computation of Ψ gives a bound for the key length. We recall that $Z_e\sim\mathcal{N}(0,\sigma_e^2)$, $X\sim\mathcal{N}\left(0,2\sqrt{\frac{\Delta}{n}}\sigma_e^2\right)$, $Y_e\sim\mathcal{N}\left(0,\left(1+2\sqrt{\frac{\Delta}{n}}\right)\sigma_e^2\right)$. If the message rate scales according to the optimal scaling constant L as in Theorem 2.6, i.e., $\lim_{n\to\infty}\frac{\ln|\mathcal{M}|}{\sqrt{\Delta}\sqrt{n}}=\frac{\sigma_e^2}{\sigma^2}$, then there exists a positive sequence $\{\xi_n\}$ such that $\xi_n=o(\sqrt{n})$ and $\ln|\mathcal{M}|\geq\frac{\sigma_e^2}{\sigma^2}\sqrt{\Delta}\sqrt{n}-\xi_n$. Let $\{\rho_n\},\rho_n\in(0,1)$ for any n, be such that $\rho_n\to 0$ and $\rho_n\xi_n\to\infty$ when $n\to\infty$. We compute

$$\begin{aligned}
\Psi(\rho_n|P_{Y_e|X},P_X) &= \ln\left(\int_{\mathbb{R}}\int_{\mathbb{R}}p_{X,Y_e}(x,y)\left(\frac{p_{Y_e|X}(y|x)}{p_{Y_e}(y)}\right)^{\rho_n}\mathrm{d}x\,\mathrm{d}y\right)\\
&= \ln\left(\int_{\mathbb{R}}\int_{\mathbb{R}}p_{Z_e}(y-x)p_X(x)\left(\frac{p_{Z_e}(y-x)}{p_{Y_e}(y)}\right)^{\rho_n}\mathrm{d}x\,\mathrm{d}y\right)\\
&= \ln\left(\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}\sigma_e}e^{-\frac{(y-x)^2}{2\sigma_e^2}}\frac{1}{\sqrt{2\pi}\sqrt{2\sqrt{\frac{\Delta}{n}}}\sigma_e}e^{-\frac{x^2}{4\sqrt{\frac{\Delta}{n}}\sigma_e^2}}\right.
\end{aligned}$$

$$\times \left(\sqrt{1+2\sqrt{\frac{\Delta}{n}}} \frac{e^{-\frac{(y-x)^2}{2\sigma_e^2}}}{e^{-\frac{y^2}{2\left(1+2\sqrt{\frac{\Delta}{n}}\right)\sigma_e^2}}} \right)^{\rho_n} \mathrm{d}x\,\mathrm{d}y \Bigg)$$

$$= \ln \left(\int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{1}{\sqrt{2\pi}\sigma_e} \right)^2 \frac{\left(1+2\sqrt{\frac{\Delta}{n}}\right)^{\frac{\rho_n}{2}}}{\sqrt{2}\left(\frac{\Delta}{n}\right)^{\frac{1}{4}}} \times e^{-\frac{\left(y-\frac{(1+\rho_n)\left(1+2\sqrt{\frac{\Delta}{n}}\right)}{1+2(1+\rho_n)\sqrt{\frac{\Delta}{n}}}x\right)^2}{2\sigma_e^2 \frac{1+2\sqrt{\frac{\Delta}{n}}}{1+2(1+\rho_n)\sqrt{\frac{\Delta}{n}}}}} \right.$$

$$\left. \times e^{-\frac{x^2}{2\sigma_e^2} \frac{-(1+\rho_n)^2\left(2\sqrt{\frac{\Delta}{n}}+4\frac{\Delta}{n}\right)+\left(1+2(1+\rho_n)\sqrt{\frac{\Delta}{n}}\right)^2}{2\sqrt{\frac{\Delta}{n}}\left(1+2(1+\rho_n)\sqrt{\frac{\Delta}{n}}\right)}} \mathrm{d}y\,\mathrm{d}x \right) \tag{D.29}$$

$$= \ln \left(\frac{\left(1+2\sqrt{\frac{\Delta}{n}}\right)^{\frac{\rho_n}{2}}}{\sqrt{2}\left(\frac{\Delta}{n}\right)^{\frac{1}{4}}} \sqrt{\frac{1+2\sqrt{\frac{\Delta}{n}}}{1+2(1+\rho_n)\sqrt{\frac{\Delta}{n}}}} \sqrt{\frac{2\sqrt{\frac{\Delta}{n}}\left(1+2(1+\rho_n)\sqrt{\frac{\Delta}{n}}\right)}{1+2(1-\rho_n^2)\sqrt{\frac{\Delta}{n}}}} \right)$$

$$= \ln \left(\frac{\left(1+2\sqrt{\frac{\Delta}{n}}\right)^{\frac{1}{2}+\frac{\rho_n}{2}}}{\left(1+2(1-\rho_n^2)\sqrt{\frac{\Delta}{n}}\right)^{\frac{1}{2}}} \right)$$

$$= \rho_n\left(1+\rho_n\right)\sqrt{\frac{\Delta}{n}} + O\left(\frac{\rho_n}{n}\right). \tag{D.30}$$

We have now established the upper bound

$$\begin{aligned} \mathbb{E}_{\mathsf{C}}\left[D\left(P_{Y_e^n|\mathsf{C}}||P_{Y_e^n}\right)\right] &\leq \frac{1}{\rho_n} \ln\left(1+e^{-\rho_n(\ln|\mathcal{K}|+\ln|\mathcal{M}|)+\rho_n(1+\rho_n)\sqrt{\Delta}\sqrt{n}+O(\rho_n)}\right) \\ &\leq \frac{1}{\rho_n} e^{-\rho_n\left(\ln|\mathcal{K}|+\left(\frac{\sigma_e^2}{\sigma^2}-1\right)\sqrt{\Delta}\sqrt{n}-\rho_n\sqrt{\Delta}\sqrt{n}-\xi_n+O(1)\right)}. \end{aligned} \tag{D.31}$$

We notice that setting $\ln|\mathcal{K}| = \left(2-\frac{\sigma_e^2}{\sigma^2}\right)\sqrt{\Delta}\sqrt{n}$ ensures that (D.31) hence (D.28) goes to 0. Furthermore, we notice that if $\sigma_e > \sigma$, (D.31) hence (D.28) goes to 0 without the need of a key. This is in agreement with [7] which showed that if the noise variance of the eavesdropper is worse than the one of the legitimate receiver, no key is needed for covert communication. In the special case where $\sigma_e = \sigma$, we obtain $o(\sqrt{n})$ as stated in Proposition D.1. □

D.2 Exponential degraded channel

We consider the exponential channel (illustrated in Figure 12) with i.i.d. exponential noise with mean Λ_e denoted Z_e for the eavesdropper and i.i.d. exponential noise with mean Λ denoted Z for the legitimate receiver:

$$\begin{aligned} Y_i &= X_i + Z_i, \qquad Z_i \sim \mathcal{E}(\Lambda), \quad \Lambda > 0, \qquad i = 1,2,\ldots,n, \\ Y_{e,i} &= X_{e,i} + Z_{e,i}, \qquad Z_{e,i} \sim \mathcal{E}(\Lambda_e), \quad \Lambda_e > 0, \qquad i = 1,2,\ldots,n. \end{aligned} \tag{D.32}$$

In this setup, the covertness constraint (2.6) can be written in the form (2.46), where $P_{Z_e^n}$ denotes the distribution of the noise vector Z_e^n, and $P_{Y_{e,\mathcal{C}}^n}$ that of the output sequence averaged over the messages and the key:

$$p_{Y_{e,\mathcal{C}}^n}(y^n) = \frac{1}{|\mathcal{K}| \times |\mathcal{M}|} \sum_{m=1}^{|\mathcal{M}|} \sum_{k=1}^{|\mathcal{K}|} \frac{1}{\Lambda_e^n} e^{-\frac{\|y^n - f(m,k)\|_1}{\Lambda_e}}. \tag{D.33}$$

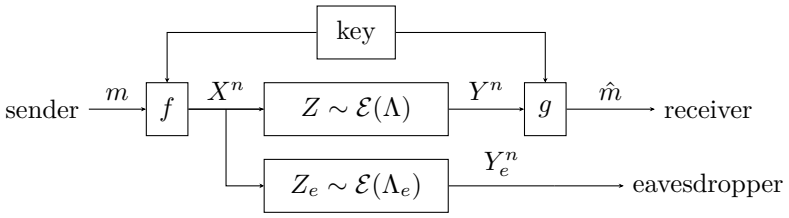

Fig. 12: Covert communication over the degraded exponential channel.

D.2.1 A loose upper bound on L

Theorem D.1 *The maximum amount of information that can be sent covertly and reliably to the legitimate receiver on the channel* (D.32) *satisfying the covertness condition* (2.46) *is upper bounded as*

$$L \leq \sqrt{2}\,\frac{\Lambda_e}{\Lambda}. \tag{D.34}$$

In order to show Theorem D.1 we first need to introduce the following lemma.

Lemma D.2 *For any exponential random variable Z with mean Λ, for any non negative random variable Y, we have the two following inequalities*

$$h(Y) \leq 1 + \ln(\mathbb{E}[Y]), \tag{D.35}$$

$$D(P_Y||P_Z) \geq \ln\left(\frac{\Lambda}{\mathbb{E}[Y]}\right) + \frac{\mathbb{E}[Y]}{\Lambda} - 1. \tag{D.36}$$

Furthermore, the inequalities (D.35) *and* (D.36) *hold as equalities when Y follows an exponential distribution.*

Proof: First, we notice that (D.35) comes directly from the maximization of the entropy for a non-negative random variable and for a fixed first moment: the maximization is achieved by an exponential distribution [18, Section 12].
We then show (D.36) via the following:

$$\begin{aligned} D(P_Y||P_Z) &= -h(Y) - \int_{\mathbb{R}^+} p_Y(y)\ln(p_Z(y))\mathrm{d}y \\ &= -h(Y) + \int_{\mathbb{R}^+} p_Y(y)\left(\ln(\Lambda) + \frac{y}{\Lambda}\right)\mathrm{d}y \\ &= -h(Y) + \ln(\Lambda) + \frac{\mathbb{E}[Y]}{\Lambda} \\ &\geq -1 - \ln(\mathbb{E}[Y]) + \ln(\Lambda) + \frac{\mathbb{E}[Y]}{\Lambda} \\ &= \ln\left(\frac{\Lambda}{\mathbb{E}[Y]}\right) + \frac{\mathbb{E}[Y]}{\Lambda} - 1 \end{aligned} \tag{D.37}$$

which is the desired inequality. Note that (D.37) follows from (D.35), and that it holds with equality when Y follows an exponential distribution. □

Proof of Theorem D.1: Take any code $\mathcal{C}$ of length n. As in the previous section, let $\bar{X}$ denote a random variable such that $P_{\bar{X}}$ is the average input distribution over the secret key, a uniformly drawn message, and the n channel uses. Let $\bar{Y}$ denote the channel output random variable through the legitimate receiver's channel, and let $\bar{Y}_e$ denote the channel output random variable through the eavesdropper's channel. We denote $P_{\bar{X}}$, $P_{\bar{Y}}$ and $P_{\bar{Y}_e}$ as in (D.4),(D.5) and (D.6).

Starting from the condition (2.46), we have:

$$\begin{aligned}
\Delta &\geq D(P_{Y^n_{\mathcal{C},e}}||P_{Z^n_e}) \\
&\geq nD(P_{\bar{Y}_e}||P_{Z_e}) && \text{(D.38)} \\
&\geq n\left(\ln\left(\frac{\Lambda_e}{\mathbb{E}[\bar{Y}_e]}\right)+\frac{\mathbb{E}[\bar{Y}_e]}{\Lambda_e}-1\right) && \text{(D.39)} \\
&= n\left(-\ln\left(1+\frac{\mathbb{E}[\bar{X}]}{\Lambda_e}\right)+\frac{\mathbb{E}[\bar{X}]}{\Lambda_e}\right), && \text{(D.40)}
\end{aligned}$$

where (D.38) follows from the same steps as (3.36), and (D.39) from inequality (D.36) in Lemma D.2. From (D.40), we notice that as $n \to +\infty$, $\mathbb{E}[\bar{X}]$ must approach zero, and:

$$\mathbb{E}[\bar{X}] = O\left(\frac{1}{\sqrt{n}}\right). \tag{D.41}$$

Then from (D.40), we have

$$\frac{\Delta}{n} \geq \frac{\mathbb{E}[\bar{X}]^2}{2\Lambda_e^2} + O\left(\frac{1}{n^{\frac{3}{2}}}\right) \tag{D.42}$$

i.e.

$$\mathbb{E}[\bar{X}] \leq \sqrt{2}\sqrt{\frac{\Delta}{n}}\Lambda_e + O\left(\frac{1}{n}\right). \tag{D.43}$$

We next derive a bound on $M^*(n,\varepsilon,\Delta)$ in terms of $\bar{X}$ and $\bar{Y}$. As in (D.14), we have by Fano's inequality

$$\ln\left(M^*(n,\varepsilon,\Delta)\right)(1-\varepsilon) - 1 \leq n(h(\bar{Y}) - h(Z)). \tag{D.44}$$

By (D.35) in Lemma D.2, we know that:

$$\begin{aligned}
\ln\left(M^*(n,\varepsilon,\Delta)\right)(1-\varepsilon) - 1 &\leq n\left(1+\ln(\mathbb{E}[\bar{Y}]) - (1+\ln(\Lambda))\right) \\
&= n\ln\left(\frac{\mathbb{E}[\bar{Y}]}{\Lambda}\right) \\
&= n\ln\left(1+\frac{\mathbb{E}[\bar{X}]}{\Lambda}\right) \\
&\leq n\frac{\mathbb{E}[\bar{X}]}{\Lambda}; && \text{(D.45)}
\end{aligned}$$

then by injecting (D.43) in (D.45) we obtain

$$\begin{aligned}
\ln\left(M^*(n,\varepsilon,\Delta)\right)(1-\varepsilon) - 1 &\leq n\frac{\Lambda_e}{\Lambda}\sqrt{2}\sqrt{\frac{\Delta}{n}} + O(1) \\
&= \sqrt{2}\frac{\Lambda_e}{\Lambda}\sqrt{\Delta}\sqrt{n} + O(1)
\end{aligned} \tag{D.46}$$

Finally recalling the definition (2.23) of L, taking $n \to \infty$ and $\varepsilon \to 0$ in (D.46), we obtain the desired result.

□

References

[1] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, February 2012.

[2] N. C. Beaulieu and D. J. Young, "Designing time-hopping ultrawide bandwidth receivers for multiuser interference environments," *Proc. of the IEEE*, vol. 97, no. 2, pp. 255–284, February 2009.

[3] E. V. Belmega and A. Chorti, "Protecting secret key generation systems against jamming: Energy harvesting and channel hopping approaches," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2611–2626, November 2017.

[4] A. Bendary, A. Abdelaziz, and C. E. Koksal, "Achieving positive covert capacity over MIMO AWGN channels," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 1, pp. 149–162, March 2021.

[5] S. K. Berberian, *Fundamentals of real analysis*. Springer Science & Business Media, 2012.

[6] M. Bloch and J. Barros, *Physical-layer security: from information theory to security engineering*. Cambridge University Press, 2011.

[7] M. R. Bloch, "Covert communication over noisy channels: a resolvability perspective," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, February 2016.

[8] M. R. Bloch and S. Guha, "Optimal covert communications using pulse-position modulation," in *Proc. of IEEE International Symposium on Information Theory*, Aachen, Germany, June 25–30 2017.

[9] V. I. Bogachev and M. A. S. Ruas, *Measure theory*. Springer, 2007.

[10] C. Bouette, L. Luzzi, and L. Wang, "Covert communication over two types of additive noise channels," in *Proc. of IEEE Information Theory Workshop*, Saint-Malo, France, April 23–28 2023.

[11] A. Bounhar, M. Sarkiss, and M. Wigger, "Covert multi-access communication with a non-covert user," in *Proc. of IEEE International Conference on Communications*, Denver, CO, USA, June 9–13 2024.

[12] ——, "Whispering secrets in a crowd: Leveraging non-covert users for covert communications," 2024. [Online]. Available: https://arxiv.org/abs/2408.12962

[13] C. Cachin, "An information-theoretic model for steganography," in *Proc. of International Workshop on Information Hiding*, Portland, OR, USA, April 14–17 1998.

[14] P. Cardieri, "Modeling interference in wireless ad hoc networks," *IEEE Communications Surveys & Tutorials*, vol. 12, no. 4, pp. 551–572, May 2010.

[15] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *Proc. of IEEE Information Theory Workshop*, Hobart, Australia, November 2–5 2014.

[16] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: hiding messages in noise," in *Proc. of IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 7–13 2013.

[17] L. Clavier, T. Pedersen, I. Larrad, M. Lauridsen, and M. Egan, "Experimental evidence for heavy tailed interference in the IoT," *IEEE Communications Letters*, vol. 25, no. 3, pp. 692–695, March 2021.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons Inc., 2006.

[19] I. Csiszar and J. Korner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[20] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, November 2013.

[21] P. J. Davis, "Leonhard Euler's integral: A historical profile of the gamma function," *The American Mathematical Monthly*, vol. 66, no. 10, pp. 849–869, 1959.

[22] M. L. De Freitas, M. Egan, L. Clavier, A. Goupil, G. W. Peters, and N. Azzaoui, "Capacity bounds for additive symmetric α-stable noise channels," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 5115–5123, August 2017.

[23] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer Verlag, 1998.

[24] A. Dytso, R. Bustin, H. V. Poor, and S. Shamai Shitz, "On additive channels with generalized Gaussian noise," in *Proc. of IEEE International Symposium on Information Theory*, Aachen, Germany, June 25–30 2017.

[25] ——, "Analytical properties of generalized Gaussian distributions," *Journal of Statistical Distributions and Applications*, vol. 5, no. 1, pp. 1–40, December 2018.

[26] J. Fahs and I. Abou-Faycal, "A Cauchy input achieves the capacity of a Cauchy channel under a logarithmic constraint," in *Proc. of IEEE International Symposium on Information Theory*, Honolulu, HI, USA, June 30–July 5 2014.

[27] ——, "On properties of the support of capacity-achieving distributions for additive noise channel models with

input cost constraints," *IEEE Transactions on Information Theory*, vol. 64, no. 2, pp. 1178–1198, February 2018.

[28] A. Feinstein, "A new basic theorem of information theory," *Transactions of the IRE Professional Group on Information Theory*, vol. 4, no. 4, pp. 2–22, September 1954.

[29] W. Feller, *An introduction to probability theory and its applications*, 2nd ed. John Wiley & Sons Inc., 1971, vol. 2.

[30] T. Filler, A. D. Ker, and J. Fridrich, "The square root law of steganographic capacity for Markov covers," in *Proc. of the International Society for Optical Engineering*, San Jose, CA, USA, August 2–5 2009.

[31] G. Frèche, M. R. Bloch, and M. Barret, "Polar codes for covert communications over asynchronous discrete memoryless channels," in *Proc. of the 51st Annual Conference on Information Sciences and Systems*, March 22–24 2017.

[32] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, 2009.

[33] R. G. Gallager, "Energy limited channels: coding, multiaccess, and spread spectrum," *Laboratory for Information and Decision Systems, MIT*, vol. LIDS-P, no. 1714, November 1987.

[34] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Academic Press, 2007.

[35] K. Gulati, B. L. Evans, J. G. Andrews, and K. R. Tinsley, "Statistics of co-channel interference in a field of Poisson and Poisson-Poisson clustered interferers," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6207–6222, December 2010.

[36] P. R. Halmos, *Measure theory*. Springer, 2013, vol. 18.

[37] T. Han and S. Verdù, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[38] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1562–1575, April 2006.

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

[40] C. W. Helstrom, *Statistical theory of signal detection*, 2nd ed. Pergamon Press, 1968.

[41] J. Hou and G. Kramer, "Effective secrecy: Reliability, confusion and stealth," in *Proc. of IEEE International Symposium on Information Theory*, Honolulu, HI, USA, 29 June – 04 July 2014.

[42] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions*, 2nd ed. John Wiley & Sons Inc., 1995, vol. 2.

[43] I. A. Kadampot, M. Tahmasbi, and M. R. Bloch, "Codes for covert communication over additive white gaussian noise channels," in *Proc. of IEEE International Symposium on Information Theory*, Paris, France, July 7–12 2019, pp. 977–981.

[44] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*. Springer-Verlag, 1988.

[45] T. Kawata, *Fourier Analysis in Probability Theory*. Academic Press, 1972.

[46] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*, 1st ed. Prentice Hall, 1998.

[47] J. H. B. Kemperman, "On the optimum rate of transmitting information," *The Annals of Mathematical Statistics*, vol. 40, no. 6, pp. 2156–2177, December 1969.

[48] A. D. Ker, "A capacity result for batch steganography," *IEEE Signal Processing Letters*, vol. 14, no. 8, pp. 525–528, August 2007.

[49] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79 – 86, March 1951.

[50] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 7, pp. 1195–1205, October 2015.

[51] E. L. Lehmann, J. P. Romano, and G. Casella, *Testing statistical hypotheses*, 3rd ed. Springer, 2005, vol. 3.

[52] J. Liu, P. Cuff, and S. Verdú, "E_γ-resolvability," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2629–2658, May 2017.

[53] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[54] M. Mitev, A. Chorti, E. V. Belmega, and H. V. Poor, "Protecting physical layer secret key generation from active attacks," *Entropy*, vol. 23, no. 8, p. 960, July 2021.

[55] S. Nadarajah, "A generalized normal distribution," *Journal of Applied statistics*, vol. 32, no. 7, pp. 685–694, September 2005.

[56] R. Pappu, B. Recht, J. Taylor, and N. Gershenfeld, "Physical one-way functions," *Science*, vol. 297, no. 5589, pp. 2026–2030, September 2002.
[57] M. S. Pinsker, *Information and information stability of random variables and processes*. Holden-Day, 1964.
[58] Y. Polyanskiy, *Channel coding: Non-asymptotic fundamental limits*. Princeton University, 2010.
[59] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
[60] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *Proc. of Allerton Conference on Communication, Control, and Computing*, Allerton, IL, USA, September 29–October 1 2010.
[61] Y. Polyanskiy and Y. Wu, *Information Theory: From Coding to Learning*. To be published by Cambridge University Press, 2024. [Online]. Available: https://people.lids.mit.edu/yp/homepage/data/itbook-export.pdf
[62] H. V. Poor, *An introduction to signal detection and estimation*, 2nd ed. Springer-Verlag, 2013.
[63] R. Ranga Rao, "Relations between weak and uniform convergence of measures with applications," *The Annals of Mathematical Statistics*, vol. 33, no. 2, pp. 659–680, June 1962.
[64] R. L. Schilling, *Measures, integrals and martingales*, 2nd ed. Cambridge University Press, 2017.
[65] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, October 1949.
[66] ——, "Certain results in coding theory for noisy channels," *Information and control*, vol. 1, no. 1, pp. 6–25, September 1957.
[67] ——, "Probability of error for optimal codes in a Gaussian channel," *The Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, May 1959.
[68] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread spectrum communications handbook*, electronic ed. McGraw-Hill Incorporation, 2002.
[69] E. W. Stacy, "A generalization of the gamma distribution," *The Annals of mathematical statistics*, vol. 33, no. 3, pp. 1187–1192, September 1962.
[70] F. Steutel and K. van Harn, *Infinite divisibility of probability distributions on the real line*. Marcel Dekker Incorporation, 2004.
[71] M. Tahmasbi and M. R. Bloch, "First and second order asymptotics in covert communication," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2190–2212, April 2019.
[72] S. Verdú, "The exponential distribution in information theory," *Problems of Information Transmission*, vol. 32, no. 1, pp. 86–95, March 1996.
[73] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.
[74] S. Verdú, "On channel capacity per unit cost," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 1019–1030, September 1990.
[75] ——, "Total variation distance and the distribution of relative information," in *Proc. of Information Theory and Applications Workshop*, San Diego, CA, USA, February 9–14 2014.
[76] L. Wang, "On Gaussian covert communication in continuous time," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–10, December 2019.
[77] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3493–3503, April 2016.
[78] S. Y. Wang and M. R. Bloch, "Covert MIMO communications under variational distance constraint," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 4605–4620, September 2021.
[79] X. Wang, P. Hao, and L. Hanzo, "Physical-layer authentication for wireless security enhancement: Current challenges and future developments," *IEEE Communications Magazine*, vol. 54, no. 6, pp. 152–158, June 2016.
[80] D. Williams, *Probability with martingales*. Cambridge University Press, 1991.
[81] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, October 1975.
[82] C. Yeang, *Transforming noise: a history of its science and technology from disturbing sounds to informational errors, 1900-1955*. Oxford University Press, 2023.
[83] X. Yu, S. Wei, S. L. Huang, and X. P. Zhang, "On the second order asymptotics of covert communications over AWGN channels," in *Proc. of IEEE International Conference on Communications*, Denver, CO, USA, June 9–13 2024.
[84] X. Yu, S. Wei, and Y. Luo, "Finite blocklength analysis of Gaussian random coding in AWGN channels under covert constraint," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1261–1274,

October 2020.

[85] Q. Zhang, M. Bakshi, and S. Jaggi, "Covert communication with polynomial computational complexity," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1354–1384, March 2020.

[86] Q. E. Zhang, M. R. Bloch, M. Bakshi, and S. Jaggi, "Undetectable radios: covert communication under spectral mask constraints," in *Proc. of IEEE International Symposium on Information Theory*, Paris, France, July 8–13 2019.

[87] ——, "Undetectable radios: covert communication under spectral mask constraints," *arXiv preprint arXiv:2001.01201*, 2020.

Index

Information-theoretic limits of covert communication over additive-noise channels

Abstract: Physical-layer security aims to exploit the randomness inherent in communication channels in order to guarantee confidentiality even against computationally unlimited attackers. In this thesis, we focus on covert communication, also known as "communication with low probability of detection". This is a scenario where a transmitter and receiver try to prevent an eavesdropper from making a good guess on whether a communication is ongoing or not. We assume that the receiver and the eavesdropper share the same channel and therefore see the same outputs. In this scenario, the transmitter shares a key with the legitimate receiver to provide the necessary advantage to detect the communication and reliably decode the message. Asymptotically, the amount of information that can be sent reliably and covertly scales like the square root of the number of channel uses; this is known as the *square root law* of covert communication.
We study the corresponding scaling constant of the square root law for general memoryless channels including non-Gaussian additive noise, as well as Gaussian channels with memory. In the latter case, we show that the scaling constant is the same as over the memoryless Gaussian channel. For continuous memoryless channels, under mild integrability conditions, we show that the scaling constant is upper bounded by a simple expression which only involves the variance of the logarithm of the probability density function of the noise. Moreover, we show that under some additional assumptions, this upper bound is tight. Furthermore, we provide upper bounds on the length of the secret key required to achieve covertness.
The second objective of this work is to investigate the limits of covert communication over continuous memoryless channels in the finite blocklength regime. We provide bounds for the first and second-order asymptotics for covert communication over an AWGN channel under a maximal error probability criterion and for the first-order asymptotics under an average error probability criterion.

Limites fondamentales des communications dissimulées sur canaux à bruit additif

Résumé : La sécurité de la couche physique vise à exploiter le caractère aléatoire inhérent aux canaux de communication afin de garantir la confidentialité même contre des attaquants avec une capacité de calcul illimitée. Dans cette thèse, nous nous concentrons sur les communications dissimulées, également appelées "communications à faible probabilité de détection". Il s'agit d'un scénario où l'émetteur et le récepteur tentent d'empêcher un espion de deviner si une communication est en cours ou non. Nous supposons que le récepteur et l'espion observent les mêmes sorties de canal. Dans ce scénario, l'émetteur partage une clé avec le récepteur légitime afin de lui donner l'avantage nécessaire pour détecter la communication et décoder le message. Asymptotiquement, la quantité d'informations qui peut être envoyée de manière fiable et dissimulée est proportionnelle à la racine carrée du nombre d'utilisations du canal de communication; ce phénomène est connu comme la *square root law* des communications dissimulées.
Nous étudions la constante de proportionnalité correspondant à la square root law pour une classe générale de canaux de communication avec bruit additif non gaussien, ainsi que des canaux gaussiens avec mémoire. Dans ce dernier cas, nous montrons que la constante de proportionnalité est la même que pour le canal gaussien sans mémoire. Pour les canaux continus sans mémoire, sous certaines conditions d'intégrabilité, nous montrons que la constante de proportionnalité admet comme borne supérieure une expression simple qui dépend seulement de la variance du logarithme de la densité de probabilité du bruit. De plus, nous montrons que sous certaines hypothèses supplémentaires, cette borne supérieure devient une égalité. Nous prouvons également des bornes supérieures sur la longueur de la clé secrète nécessaire pour communiquer de manière dissimulée.
Le deuxième objectif de ce travail est d'étudier les limites fondamentales des communications dissimulées pour des canaux continus sans mémoire en longueur finie. Nous fournissons des bornes pour les asymptotiques de premier et de second ordre des communications dissimulées sur un canal AWGN selon un critère de probabilité d'erreur maximale, ainsi que pour l'asymptotique de premier ordre selon un critère de probabilité d'erreur moyenne.