Limits of low-probability-of-detection communication over a discrete memoryless channel

Ligong Wang, Gregory Wornell, Lizhong Zheng

Limits of Low-Probability-of-Detection Communication over a Discrete Memoryless Channel

Ligong Wang
ETIS, CNRS UMR 8051
ENSEA, Université de Cergy-Pontoise
Cergy-Pontoise, France
ligong.wang@ensea.fr

Gregory W. Wornell
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA, USA
gww@mit.edu

Lizhong Zheng
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA, USA
lizhong@mit.edu

***Abstract*—This paper considers the problem of communication over a discrete memoryless channel subject to the constraint that the probability that an adversary who observes the channel outputs can detect the communication is low. Specifically, the relative entropy between the output distributions when a codeword is transmitted and when no input is provided to the channel must be sufficiently small. For a channel whose output distribution induced by the zero input symbol is not a mixture of the output distributions induced by other input symbols, it is shown that the maximum number of bits that can be transmitted under this criterion scales like the square root of the blocklength. Exact expressions for the scaling constant are also derived.**

***Index Terms*—Low probability of detection, covert communication, information-theoretic security, Fisher information.**

I. Introduction

In many secret-communication applications, it is required not only that the adversary should not know the content of the message being communicated, as in [1], but also that it should not know whether the legitimate parties are communicating at all or not. Such problems are often referred to as communication with *low probability of detection (LPD)* or covert communication. Depending on the application, they can be formulated in various ways.

In [2] the authors consider a wiretap channel model [3], and refer to this LPD requirement as *stealth*. They show that stealth can be achieved without sacrificing communication rate or using an additional secret key. In their scheme, when not sending a message, the transmitter sends some random noise symbols to simulate the distribution of a codeword. There are many scenarios, however, where this cannot be done, because the transmitter must be switched off when not transmitting a message. Indeed, the criterion is often that the adversary should not be able to tell whether the transmitter is on or off, rather than whether it is sending anything meaningful or not. It is the former criterion that is considered in the current paper.

Part of this work was conducted while L. Wang was with the Research Laboratory of Electronics at MIT.

This work was supported in part by AFOSR under Grant No. FA9550-11-1-0183, and by NSF under Grant No. CCF-1319828.

Our work is closely related to the recent works [4]–[6]. In [4] the authors consider the problem of communication over an additive white Gaussian noise (AWGN) channel with the requirement that a wiretapper should not be able to tell with high confidence whether the transmitter is sending a codeword or the all-zero sequence. It is observed that the maximum number of bits that can be transmitted under this requirement scales like the *square root* of the blocklength. In [5] the authors consider a similar problem for the binary symmetric channel and show that the "square-root law" also holds. One difference between [4] and [5] is that in the former the transmitter and the receiver use a secret key to generate their codebook, whereas in the latter no key is used. More recently, [6] studies the LPD problem from a resolvability perspective and improves upon [4] in terms of secret-key length.

In the current paper, we show that the square-root law holds for a broad class of discrete memoryless channels (DMCs).[1] Furthermore, we provide exact characterizations for the scaling constant of the number of bits with respect to the square root of the blocklength, which is not done in [4]–[6].

The square-root law has been observed in various scenarios in *steganography* [7]–[9]. Due to space limitation, we cannot discuss the connections and differences between steganography and LPD communications here.

Our setting can be briefly described as follows:

- We consider a DMC whose input alphabet contains an "off" symbol. When the transmitter is switched off, it always sends this symbol.
- The transmitter and the receiver share a secret key that is sufficiently long.
- The adversary observes the same channel outputs as the intended receiver, i.e., there is no wiretap structure.
- The LPD criterion is that the relative entropy between the output distributions when a codeword is transmitted and when the all-zero sequence is transmitted must be sufficiently small [10].

[1]The achievability part of the square-root law, but not the converse, is independently derived in [6].

We assume that the receiver does know when the transmitter is sending a message. This is a realistic assumption because the transmitter and the receiver can use part of their secret key to perform synchronization prior to transmission.

The rest of this paper is arranged as follows. In Section II we formulate the problem and briefly analyze the case where the "off" input symbol induces an output distribution that can be written as a mixture of the other output distributions; the remaining sections focus on the case where it cannot. In Section III we derive formulas that can be used to characterize the maximum number of bits that can be transmitted. In Section IV we derive a simpler expression for this number under some conditions that are satisfied by many channels in practice. We conclude the paper with some remarks in Section V. The key ideas for proving our theorems are included. A full-length paper with complete proofs is in preparation [11].

II. Problem Formulation

Consider a DMC of finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, and of transition law $W(\cdot|\cdot)$. Throughout this paper, we use the letter P to denote input distributions on $\mathcal{X}$ and the letter Q to denote output distributions on $\mathcal{Y}$. Let $0 \in \mathcal{X}$ be the "off" input symbol; i.e., when the transmitter is not sending a message, it always transmits 0. Denote

$$Q_0(\cdot) \triangleq W(\cdot|0). \tag{1}$$

Without loss of generality, we assume that no two input symbols induce the same output distribution; in particular, $Q_0(\cdot) = W(\cdot|x)$ implies $x = 0$.

A (deterministic) code of blocklength n for message set $\mathcal{M}$ consists of an encoder $\mathcal{M} \to \mathcal{X}^n$, $m \mapsto x^n$ and a decoder $\mathcal{Y}^n \to \mathcal{M}$, $y^n \mapsto \hat{m}$. The transmitter and the receiver choose a *random* code of blocklength n for message set $\mathcal{M}$ using a secret key shared between them. The adversary is assumed to know the distribution according to which the transmitter and the receiver choose the random code, but not their actual choice.

The random code, together with a message M uniformly drawn from $\mathcal{M}$, induces a distribution $Q^n(\cdot)$ on $\mathcal{Y}^n$. We require that, for some constant $\delta > 0$,

$$D\left(Q^n \| Q_0^{\times n}\right) \le \delta. \tag{2}$$

Here $Q_0^{\times n}$ denotes the n-fold product distribution of Q_0, i.e., the output distribution over n channel uses when the transmitter is off. Note that condition (2), together with the assumption that the adversary does not know the secret key that is used to choose the random code, provides a limit on the adversary's probability of successfully detecting whether the transmitter is on or off. For example, the total variation distance between Q^n and $Q_0^{\times n}$ can be bounded via Pinsker's inequality.

At this point, we observe that an input symbol x with $\mathsf{supp}(W(\cdot|x)) \not\subseteq \mathsf{supp}(Q_0)$, where $\mathsf{supp}(\cdot)$ denotes the support of a distribution, should never be used by the transmitter. Indeed, using such an input symbol with nonzero probability would result in $D\left(Q^n \| Q_0^{\times n}\right)$ being infinity. Hence we can drop all such input symbols, as well as all output symbols that do not lie in $\mathsf{supp}(Q_0)$, reducing the channel to one where

$$\mathsf{supp}(Q_0) = \mathcal{Y}. \tag{3}$$

Throughout this paper we assume that (3) is satisfied.[2]

Our goal is to find the maximum possible value for $\log|\mathcal{M}|$ for which a random codebook of length n exists that satisfies condition (2), and whose average probability of error is at most ϵ. (Later we shall require that ϵ be arbitrarily small.) We denote this maximum value by $K_n(\delta, \epsilon)$.

We call an input symbol x *redundant* if $W(\cdot|x)$ can be written as a mixture of the other output distributions, i.e., if

$$W(\cdot|x) \in \mathsf{conv}\left\{W(\cdot|x') \colon x' \in \mathcal{X}, x' \ne x\right\}, \tag{4}$$

where conv denotes the convex hull. As we shall show, $K_n(\delta, \epsilon)$ can increase either linearly with the blocklength n or like $\sqrt{n}$, depending on whether 0 is redundant or not.

A. Case 1: input symbol 0 is redundant

This is the case where there exists some distribution P on $\mathcal{X}$ such that

$$P(0) = 0 \tag{5a}$$

$$\sum_{x \in \mathcal{X}} P(x) W(\cdot|x) = Q_0(\cdot). \tag{5b}$$

It can be seen that, for any $\delta \ge 0$,

$$\lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{K_n(\delta, \epsilon)}{n} = \max I(P, W), \tag{6}$$

where the maximum is taken over input distribution P that satisfies (5). Indeed, a random codebook generated independently and identically distributed (IID) according to P that satisfies (5) yields $D(Q^n \| Q_0^{\times n}) = 0$. By the standard typicality argument [12], the probability of a decoding error can be made arbitrarily small as n goes to infinity. Conversely, for a codebook whose empirical input distribution does not satisfy (5b), $D(Q^n \| Q_0^{\times n})$ grows linearly in n and is hence unbounded as n goes to infinity. Finally, a codebook that uses the symbol 0 in this case can be shown to be suboptimal, i.e., the empirical distribution of an optimal codebook should satisfy (5a).

As an example, consider a binary symmetric channel with an additional "off" symbol as shown in Fig. 1. Its optimal input distribution is uniform on $\{-1, 1\}$, and its capacity under the LPD constraint (2) is the same as its capacity without this constraint, and equals $1 - H_\mathrm{b}(p)$, where $H_\mathrm{b}(\cdot)$ is the binary entropy function.

B. Case 2: input symbol 0 is not redundant

This is the case where no P satisfying (5) can be found. It is the focus of the rest of this paper. We shall show that, in this case, K_n grows like $\sqrt{n}$. Let[3]

$$L \triangleq \lim_{\epsilon \downarrow 0} \varlimsup_{n \to \infty} \frac{K_n(\delta, \epsilon)}{\sqrt{n\delta}}. \tag{7}$$

[2]For channels that cannot be reduced to one that satisfies (3), such as the binary erasure channel, nontrivial LPD communication is not possible.

[3]By definition L can be infinity, as it is in Case 1.

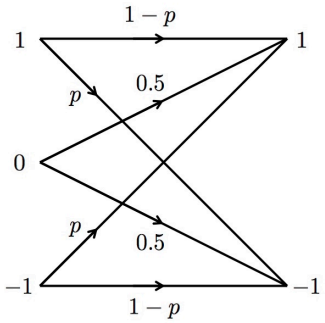

Fig. 1. A binary symmetric channel on the alphabet $\{-1, 1\}$ with cross-over probability p, with an additional "off" input symbol 0 which induces a uniform output distribution.

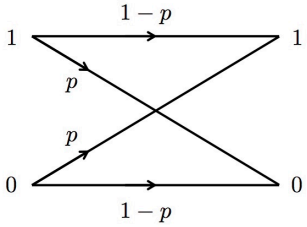

Fig. 2. The binary symmetric channel with cross-over probability p.

We shall characterize L in the following sections.

Before proceeding with mathematical analyses, we provide some intuition why positive communication rates cannot be achieved in this case. We note that, to achieve a positive rate, a necessary condition is that a non-vanishing proportion of input symbols used in the codebook should be different from the "off" symbol 0. This would mean that the average marginal distribution $\bar{P}$ on $\mathcal{X}$ has a positive probability at values other than 0 and, since Q_0 cannot be written as a mixture of output distributions produced by nonzero input symbols, $\bar{Q}$ must be different from Q_0 so $D(\bar{Q}\|Q_0) > 0$. This implies that $D(Q^n\|Q_0^{\times n})$ must grow without bound as n tends to infinity.

A simple example for this case is the binary symmetric channel in Fig. 2. Later we compute L for this channel.

III. General Expressions for L

We have the following natural but nontrivial result.[4]

Theorem 1: For any DMC,

$$L = \lim_{n\to\infty} \sqrt{\frac{n}{\delta}} \max_{P_n} I(P_n, W) \tag{8}$$

where the maxima are taken over joint distributions on $\mathcal{X} \times \mathcal{Y}$ induced by input distributions P_n and channel W, whose marginals Q_n on $\mathcal{Y}$ satisfy

$$D(Q_n\|Q_0) \le \frac{\delta}{n}. \tag{9}$$

[4]The results in [4] are derived based on the assumption that a similar formula holds for the AWGN channel, but it does not prove such a formula. We provide a complete proof in [11].

Proof Sketch: The converse part follows from standard techniques. We use Fano's inequality to show that the random codebook must satisfy

$$K_n \le I(X^n; Y^n) + \sqrt{n}\,\epsilon_n \tag{10}$$

where ϵ_n tends to zero as n tends to infinity. Recall that the distribution Q^n on Y^n must satisfy (2). Let $\bar{P}$ and $\bar{Q}$ denote the average marginal distributions on $\mathcal{X}$ and $\mathcal{Y}$, respectively, averaged over the codebook and over the n channel uses. Clearly, $\bar{Q}$ is the output distribution induced by $\bar{P}$ via W. By the chain rule and the concavity of $I(P, W)$ in P, we have

$$I(X^n; Y^n) \le nI(\bar{P}, W). \tag{11}$$

On the other hand, by the convexity of $D(\cdot\|\cdot)$ and the fact that $Q_0^{\times n}$ is a product distribution, we have

$$D(Q^n\|Q_0^{\times n}) \ge nD(\bar{Q}\|Q_0). \tag{12}$$

The converse part of Theorem 1 is then established.

The achievability part requires new proof techniques. Let $\{P_n\}$ be a sequence of input distributions such that the induced output distributions $\{Q_n\}$ satisfy (9). For every n, we randomly generate a codebook by choosing the codewords IID according to P_n. It is clear that the output distribution on $\mathcal{Y}^{\times n}$ for this code is $Q_n = Q_n^{\times n}$ and that (2) is satisfied. It remains to show that, provided that the size of the codebook is smaller than $2^{nI(P_n,W)-\sqrt{n}\epsilon_n}$ for some ϵ_n tending to zero as n tends to infinity, the probability of a decoding error can be made arbitrarily small. This cannot be proven using the asymptotic equipartition property [12], because P_n depends on n. Neither can we directly apply the information-spectrum method [13], [14], which is used when the communication rate is positive and when no single-letter expression is required. In our proof, we start with a one-shot achievability bound as in [15], [16], and then carefully analyze the probability that the information density deviates from $I(P_n, W)$. Details of the proof are in [11]. ∎

Using Theorem 1 we can derive the following expression for L.

Theorem 2: For any DMC satisfying (3), whose "off" input symbol 0 is not redundant, and which has at least one input symbol other than 0,[5] L is positive and finite, and is given by

$$L = \max_{\tilde{P}:\, \tilde{P}(0)=0} \frac{\sum_{x\in\mathcal{X}} \tilde{P}(x) D\left((W(\cdot|x)\|\, Q_0\right)}{\sqrt{\dfrac{1}{2}\displaystyle\sum_{y\in\mathcal{Y}} \frac{(\tilde{Q}(y) - Q_0(y))^2}{Q_0(y)}}}, \tag{13}$$

where $\tilde{Q}$ is the output distribution induced by $\tilde{P}$ via W.

Proof Sketch: For large n, the right-hand side of (9) is close to zero, which, together with the assumption that 0 is not redundant, requires that $P_n(0)$ be close to one. We let $\tilde{P}$ be P_n conditional on $\{X \neq 0\}$. The rest of the proof consists of

[5]By our assumption, this input symbol induces an output distribution different from Q_0.

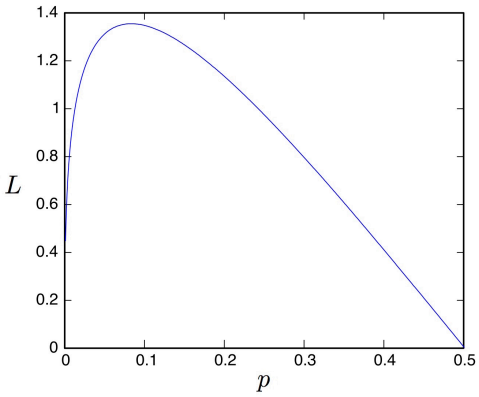

Fig. 3. The value of L for the binary symmetric channel in Fig. 2 as a function of p.

finding approximate expressions for $I(P_n, W)$ and $D(Q^n\|Q_0)$ in terms of $\tilde{P}$ and $\tilde{Q}$, respectively. ■

For some channels (13) is very easy to compute.

Example 1: Binary symmetric channel.

Consider the binary symmetric channel in Fig. 2. Clearly, the only possible choice for $\tilde{P}$ in (13) is $\tilde{P}(1) = 1$. We thus obtain the value of L as a function of p, which we plot in Fig. 3. Not surprisingly, when p approaches 0.5, L approaches zero, as does the capacity of the channel. It is however interesting to notice that, when p approaches zero, L also approaches zero, even though the capacity of the channel approaches 1 bit per use. This is because, when p is very small, it is very easy to distinguish the two input symbols 0 and 1 at the receiver end. Hence the LPD criterion requires that the transmitter must use 1 very sparsely, limiting the number of information bits it can send. The maximum of L is approximately 1.35, achieved at $p = 0.083$.

IV. A Simpler but Less General Expression for L

In this section we consider channels that satisfy the following condition.

Condition 1: There exists a capacity-achieving input distribution that uses all the input symbols.

Note that Condition 1 implies that no input symbol is redundant; in particular, 0 is not redundant.

We next give a simple upper bound on L under Condition 1. Later we provide an additional condition under which this bound is tight.

Theorem 3: Consider a DMC that satisfies Condition 1. Denote its capacity-achieving output distribution by Q^*, then

$$L \le \sqrt{2\,\mathsf{var}_{Q_0}\left(\log\frac{Q_0(Y)}{Q^*(Y)}\right)}, \tag{14}$$

where $\mathsf{var}_{Q_0}(\cdot)$ denotes the variance of a function of Y where Y has distribution Q_0.

The proof of Theorem 3 utilizes the following lemma.

Lemma 1: Let Q^* denote the capacity-achieving output distribution for a DMC $W(\cdot|\cdot)$ of capacity C. Let P' be any input distribution, and let Q' denote the output distribution induced by P' via W. Then

$$I(P', W) \le C - D(Q'\|Q^*), \tag{15}$$

where equality holds if $\mathsf{supp}(P') \subseteq \mathsf{supp}(P^*)$ for some capacity-achieving input distribution P^*.

Proof: We have the following identity (see [17]):

$$I(P', W) = \sum_{x\in\mathcal{X}} P'(x) D(W(\cdot|x)\|Q^*) - D(Q'\|Q^*). \tag{16}$$

By the Kuhn-Tucker conditions for channel capacity [18],

$$D(W(\cdot|x)\|Q^*)) \le C \tag{17}$$

where equality holds if $x \in \mathsf{supp}(P^*)$. We hence have

$$\begin{aligned} C &= \sum_{x\in\mathcal{X}} P^*(x) D(W(\cdot|x)\|Q^*) \\ &\ge \sum_{x\in\mathcal{X}} P'(x) D(W(\cdot|x)\|Q^*), \end{aligned} \tag{18}$$

where equality holds if $\mathsf{supp}(P') \subseteq \mathsf{supp}(P^*)$. Combining (16) and (18) proves the lemma. ■

Proof Sketch for Theorem 3: Note that under Condition 1, equality always holds in (15). Using this together with Theorem 1 we obtain

$$L = \lim_{n\to\infty} \sqrt{\frac{n}{\delta}} \left(C - \min D(Q_n\|Q^*)\right), \tag{19}$$

where the minimum is over $Q_n \in \mathsf{conv}\{W(\cdot|x)\colon x \in \mathcal{X}\}$ satisfying (9). To find an upper bound on L, we drop the condition that $Q_n \in \mathsf{conv}\{W(\cdot|x)\colon x \in \mathcal{X}\}$. Then the minimum of $D(Q_n\|Q^*)$ for a fixed $D(Q_n\|Q_0)$ is well known to be achieved by a distribution of the form [19]

$$Q_n(y) = \frac{Q_0(y)^{1-\lambda_n} Q^*(y)^{\lambda_n}}{\sum_{y'\in\mathcal{Y}} Q_0(y')^{1-\lambda_n} Q^*(y')^{\lambda_n}} \tag{20}$$

for some positive λ_n which tends to zero as n tends to infinity. It remains to compute $D(Q_n\|Q_0)$ and $D(Q_n\|Q^*)$ for Q_n of the form (20) for small λ_n. In fact, when λ_n is close to zero, $D(Q_n\|Q_0)$ is approximated by the Fisher Information [20] which, in this case, equals the variance in (14):

$$D(Q_n\|Q_0) = \frac{\lambda_n^2}{2} \mathsf{var}_{Q_0}\left(\log\frac{Q_0(Y)}{Q^*(Y)}\right) + o(\lambda_n^2). \tag{21}$$

This together with (9) implies that

$$\lambda_n \le \sqrt{\frac{2\delta}{n\,\mathsf{var}_{Q_0}\left(\log\frac{Q_0(Y)}{Q^*(Y)}\right)}} + o(n^{-1/2}). \tag{22}$$

On the other hand, one can show that $D(Q_n\|Q^*)$ satisfies

$$C - D(Q_n\|Q^*) = \lambda_n \mathsf{var}_{Q_0}\left(\log\frac{Q_0(Y)}{Q^*(Y)}\right) + o(\lambda_n). \tag{23}$$

Combining (19), (22), and (23) proves the theorem. ■

The bound (14) is tight for many channels, e.g., the binary symmetric channel in Fig. 2. We next provide a sufficient condition for (14) to be tight.

Let $\mathbf{s}$ be the $|\mathcal{Y}|$-dimensional vector given by

$$s(y) = Q_0(y)\left(\log\frac{Q^*(y)}{Q_0(y)} + C\right), \quad y \in \mathcal{Y}. \tag{24}$$

Consider the linear system with unknowns α_x, $x \in \mathcal{X} \setminus \{0\}$:

$$\sum_{x\in\mathcal{X}\setminus\{0\}} \alpha_x \left(W(\cdot|x) - Q_0\right) = \mathbf{s}. \tag{25}$$

Solving (25) is a simple problem in linear algebra.

Theorem 4: Suppose Condition 1 is met. If (25) has nonnegative solutions, namely, if there exist constants $\alpha_x \geq 0$, $x \in \mathcal{X} \setminus \{0\}$ that satisfy (25), then (14) holds with equality.

Proof Sketch: The vector $\mathbf{s}$ represents the tangent of the curve $Q_n(y)$ given in (20) as a function of λ_n at $\lambda_n = 0$. That (25) has nonnegative solutions means that $\mathbf{s}$ lies in the convex cone generated by $\{W(\cdot|x) - Q_0, x \in \mathcal{X} \setminus \{0\}\}$. This further implies that, for small enough λ_n, Q_n of the form given in (25) is a valid output distribution, which, as can be seen in the proof of Theorem 3, guarantees (14) to hold with equality. ■

Example 2: A k-array uniform-error channel.

Consider a channel with $\mathcal{X} = \mathcal{Y} = \{0, 1, \ldots, k-1\}$ and

$$W(y|x) = \begin{cases} 1-p, & y = x \\ \dfrac{p}{k-1}, & y \neq x \end{cases} \tag{26}$$

where $p \in (0, 1)$. Such a channel appears in direct-detection optical communication with temporal or spatial pulse-position modulation [21], [22]. Clearly, its capacity-achieving output distribution Q^* is uniform. It is easy to check that (25) has nonnegative solutions. We can hence use Theorem 4 to obtain

$$L = \sqrt{2v(k,p)} \tag{27}$$

where

$$v(k,p) = (1-p)\left(\log\frac{1}{1-p}\right)^2 + p\left(\log\frac{k-1}{p}\right)^2 - \left((1-p)\log\frac{1}{1-p} + p\log\frac{k-1}{p}\right)^2. \tag{28}$$

V. Concluding Remarks

A DMC in practice often represents discretization of a continuous-alphabet channel. For example, Figs. 1 and 2 can result from two different discretizations of the same AWGN channel. In this sense, our results suggest that the optimal discretization may depend heavily on whether there is an LPD requirement or not.

The current paper focuses on DMCs, but the methods introduced here can be applied to the AWGN channel as well. In [11], we show that $L = 1$ for the AWGN channel irrespectively of the noise power.

In practice, LPD communication systems of positive data rates often can be implemented even when the channel model does not seem to allow positive rates. Indeed, in such applications, the concern is often not that the transmitted signal should be sufficiently weak, but rather that it should have a wide spectrum and resemble white noise [23]. We believe that one of the reasons why such systems may work is that realistic channels often have memory. For example, on a channel whose noise level varies with a coherence time that is longer than the length of a codeword, the transmitter and the receiver can use the adversary's ignorance of the actual noise level to communicate without being detected. One way to formulate this scenario is to assume that the channel has an unknown parameter that is fixed. This is discussed for the binary symmetric channel in [24]. Further addressing this scenario is part of ongoing research.

References

[1] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Techn. J.*, vol. 28, pp. 656–719, 1949.

[2] J. Hou and G. Kramer, "Effective secrecy: reliability, confusion and stealth," in *Proc. IEEE Int. Symp. Inform. Theory*, (Honolulu, HI, USA), June 29–July 4 2014.

[3] A. D. Wyner, "The wiretap channel," *Bell System Techn. J.*, vol. 54, pp. 1355–1387, 1975.

[4] B. A. Bash, D. Goekel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Select. Areas Commun.*, vol. 31, pp. 1921–1930, Sept. 2013.

[5] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. IEEE Int. Symp. Inform. Theory*, (Istanbul, Turkey), July 10–15 2013.

[6] M. Bloch, "Covert communication over noisy channels: A resolvability perspective." Subm. to *IEEE Trans. Inform. Theory*, 2015, arXiv:1503.08778.

[7] A. D. Ker, "A capacity result for batch steganography," *IEEE Signal Processing Lett.*, vol. 14, pp. 525–528, Aug. 2007.

[8] J. Fridrich, *Steganography in Digital Media: principles, Algorithms, and Applications*. Cambridge University Press, 2009.

[9] T. Filler and J. Fridrich, "Fisher information determines capacity of ϵ-secure steganography," in *Information Hiding*, Lecture Notes in Computer Science, 2009.

[10] C. Cachin, "An information-theoretic model for steganography," in *Information Hiding*, Lecture Notes in Computer Science, 1998.

[11] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of covert communication." in preparation, 2015.

[12] C. E. Shannon, "A mathematical theory of communication," *Bell System Techn. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[13] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.

[14] T. S. Han, *Information Spectrum Methods in Information Theory*. Springer Verlag, 2003.

[15] L. Wang, R. Colbeck, and R. Renner, "Simple channel coding bounds," in *Proc. IEEE Int. Symp. Inform. Theory*, (Seoul, Korea), pp. 1804–1808, June 28–July 3, 2009.

[16] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, pp. 2307–2359, May 2010.

[17] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Sci. Math. Hungar.*, vol. 2, pp. 291–292, 1967.

[18] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[19] I. Csiszár and F. Matúš, "Information projections revisited," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1474–1490, June 2003.

[20] S. Kullback, *Information Theory and Statistics*. John Wiley & Sons, 1959.

[21] L. Wang and G. W. Wornell, "A refined analysis of the Poisson channel in the high-photon-efficiency regime," *IEEE Trans. Inform. Theory*, vol. 60, pp. 4299–4311, July 2014.

[22] Y. Kochman and G. W. Wornell, "On high-efficiency optical communication and key distribution," in *Proc. Inf. Theory and Appl. Workshop*, (San Diego, CA, USA), Feb. 5–10 2012.

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

[24] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *Proc. Inform. Theory Workshop (ITW)*, (Hobart, Australia), Nov. 2–5, 2014.