The Poisson Channel with Side Information

Shraga Bross
School of Enginerring
Bar-Ilan University, Israel
brosss@macs.biu.ac.il

Amos Lapidoth   Ligong Wang
Signal and Information Processing Laboratory
ETH Zurich, Switzerland
{lapidoth, wang}@isi.ee.ethz.ch

***Abstract*— The continuous-time, peak-limited, infinite-bandwidth Poisson channel with spurious counts is considered. It is shown that if the times at which the spurious counts occur are known noncausally to the transmitter but not to the receiver, then the capacity is equal to that of the Poisson channel with no spurious counts. Knowing the times at which the spurious counts occur only causally at the transmitter does not increase capacity.**

I. Introduction

Communication aided by channel-state information (CSI) at the transmitter provides a rich source for interesting problems in information theory. A discrete memoryless channel (DMC) with independent and identically distributed (i.i.d.) states is described by a transition law

$$\Pr(Y = y|X = x, S = s) = W(y|x, s) \tag{1a}$$

and a state law

$$P_S(s), \tag{1b}$$

where X denotes the channel input, Y denotes the channel output, and S denotes the state of the channel.

Shannon [1] studied the case where the CSI is unknown to the receiver and known *causally* to the transmitter. In this scenario, before sending x_k, the transmitter knows $\{s_1, \ldots, s_k\}$. He showed that the capacity of the channel (1) is given by

$$\sup_{P_U} I(U;Y), \tag{2}$$

where U is a random variable taking value in the set of mappings from channel states to channel inputs, and where $P_{Y|U}$ is given by

$$P_{Y|U}(y|u) = \sum_s P_S(s)W(y|u(s), s).$$

A different scenario is where the CSI is still unknown to the receiver but known *noncausally* to the transmitter. In this case, the transmitter knows the whole state sequence before starting to transmit. The capacity of (1) in this case was found by Gel'fand and Pinsker [2] to be

$$\sup_{U -\circ- (X,S) -\circ- Y} I(U;Y) - I(U;S), \tag{3}$$

where the supremum is over all joint distributions of the form

$$P_U(u)P_{X,S|U}(x, s|u)W(y|x, s)$$

whose marginal distribution on S is $P_S(\cdot)$.

The capacity with noncausal CSI was computed in various cases. The case of writing on binary memory with defects was solved by Kusnetsov and Tsybakov [3] before the discovery of the formula (3); the case of writing on binary memory with defects *and noise* was solved by Heegard and El Gamal [4]; and the case of additive Gaussian noise channel with additive states ("writing on dirty paper") was solved by Costa [5]. In these cases, capacity was found to be equal to the capacity when there are no states at all.

In the present work, we consider the continuous-time, peak-limited Poisson channel with CSI at the transmitter. This channel model without states was studied in [6], [7], and [8]. In [9], Bross and Shamai considered the Poisson arbitrarily varying channel (AVC) with states. In their set-up the states correspond to a nonnegative signal that is added to the channel input.

We consider a model where the channel states correspond to spurious counts at the receiver, and where the receiver does not know which counts are spurious. We study first the case where the CSI is known noncausally to the transmitter, and then the case where the CSI is known causally to the transmitter.

In the noncausal CSI case, we distinguish between two settings. In the first setting we assume that the states are chosen by an adversary subject to a constraint on the average number of spurious counts per second, and we allow the transmitter and the receiver to use random codes. Since the state sequence can be arbitrary, we cannot use Gel'fand and Pinsker's formula (3). Instead, as in [3] and [4], we show that the capacity with no spurious counts can be achieved on this channel with random codes by construction. In the second setting we assume that the spurious counts are random (not necessarily a homogeneous Poisson process). Using the result from the first setting, we show that the capacity with no spurious counts is achievable on this channel with deterministic codes.

For the causal CSI setting, we show that, even if the spurious counts obey a homogeneous Poisson law, causal CSI does not increase the capacity of this channel. Thus, as in [3] and [4], in our channel causal CSI does not increase capacity at all, while noncausal CSI increases it to that of the same channel model but without states.

The rest of this paper is arranged as follows: in Section II we recall some important results regarding the peak-limited Poisson channel; in Section III we state and prove the capacity results for the noncausal CSI case; and in Section IV we discuss the causal CSI case.

II. Preliminary: the Poisson Channel

In this section we recall some important results regarding the peak-limited Poisson channel. This channel's time-t input $x(t)$ is a nonnegative real, and its time-t output $y(t)$ is a nonnegative integer. For a given input signal $x(t)$, $t \in \mathbb{R}$, the output $Y(t)$, $t \in \mathbb{R}$ is a Poisson process whose time-t intensity is $x(t)+\lambda$, where λ is a nonnegative constant called the *dark current*.

We impose a peak-power constraint on the input so

$$x(t) \leq A, \quad t \in \mathbb{R}, \tag{4}$$

where $A > 0$ is the maximal allowed input power. We denote the capacity of this channel (in bits per second) by $C(A,\lambda)$.

The value of $C(A,\lambda)$ was first found by Kabanov [6] and Davis [7] using martingale techniques. Later, Wyner [8] showed that $C(A,\lambda)$ can be achieved by dividing the channel into small time-slots and then looking at the resulting DMC.

To derive an achievability result, we follow Wyner and discretize this channel in time with every time-slot being Δ seconds long. We further restrict the input distribution so that within each time-slot the input is constant: either A or 0. The receiver front-end produces 0 if there were no counts in the time-slot, and 1 if there were one or more counts. This discrete-time channel is memoryless and for small Δ its law can be approximated by

$$W(1|x) = \begin{cases} \lambda\Delta, & x = 0 \\ (A+\lambda)\Delta, & x = 1. \end{cases} \tag{5}$$

We denote the capacity of the channel (5) (in bits per channel use) by $C_\Delta(A,\lambda)$. Wyner [8] showed that

$$\begin{aligned} C(A,\lambda) &= \lim_{\Delta \downarrow 0} \frac{C_\Delta(A,\lambda)}{\Delta} \\ &= \max_{p\in(0,1)} \Big\{ p(A+\lambda)\log(A+\lambda) + (1-p)\lambda\log\lambda \\ &\qquad\qquad - (pA+\lambda)\log(pA+\lambda) \Big\}. \end{aligned}$$

We note that $C(A,\lambda)$ can also be written as

$$C(A,\lambda) = \max_{p\in(0,1)} (pA+\lambda) D\left(\mathrm{Ber}\left(\frac{p(A+\lambda)}{pA+\lambda}\right) \middle\| \mathrm{Ber}\,(p) \right), \tag{6}$$

where $\mathrm{Ber}\,(\pi)$ denotes the Bernoulli distribution of parameter π.

III. Noncausal CSI

We now consider the continuous-time Poisson channel as described in Section II, but with spurious counts occurring at the receiver. We assume that the times of these occurrences are known noncausally to the transmitter but not to the receiver. We consider two settings: for Section III-A we assume that the spurious counts are generated by a malicious adversary and that the transmitter and the receiver are allowed to use random codes; for Section III-B we assume that the spurious counts occur randomly according to a law known to both transmitter and receiver and that only deterministic codes are allowed.

A. Random Codes against Arbitrary States

Consider the Poisson channel as described in Section II but with an adversary who generates spurious counts at the receiver and reveals the times at which they occur to the transmitter before communication begins. These spurious counts can be modeled by a counting signal $s(\cdot)$, which is a nonnegative, integer-valued, monotonically increasing function. Thus, conditional on the input being $x(\cdot)$, the output is given by

$$Y(t) = Z(t) + s(t), \quad t \in \mathbb{R}, \tag{7}$$

where $Z(t)$ is a Poisson process whose time-t intensity is $x(t)+\lambda$. We assume that the adversary is subject to the restriction that, within each transmission block, the average number of spurious counts per second cannot exceed a certain constant ν which is known to both transmitter and receiver.

In a (T,R) (deterministic) code, the encoder maps the message $M \in \{1,\ldots,2^{RT}\}$ and the channel state $s(t)$, $t \in [0,T]$, to an input signal $x(t)$, $t \in [0,T]$, and the decoder guesses the message M from the channel output $y(t)$, $t \in [0,T]$. A (T,R) *random code* is a probability distribution on all deterministic (T,R) codes.[1]

A rate R (in bits per second) is said to be *achievable with random codes* on the channel (7) if, for every $T > 0$, there exists a random (T,R) code such that, as T tends to infinity, the average probability of a guessing error tends to zero for *all possible* $s(\cdot)$. The *random coding capacity* of this channel is defined as the supremum over all rates achievable with random codes.

Since the adversary can choose not to introduce any spurious counts, the random coding capacity of the channel (7) is upper-bounded by $C(A,\lambda)$, which is given in (6). Our first result is that this bound is tight:

Theorem 1: For any positive A, λ and ν, the random coding capacity of the channel (7), where $s(\cdot)$ is known noncausally to the transmitter but unknown to the receiver, is equal to $C(A,\lambda)$.

Proof: We only need to prove the lower bound. Namely, we need to show that any rate below the right-hand side (RHS) of (6) is achievable with random codes on the channel (7). To this end, for fixed positive constants T, R and α, we shall construct a block coding scheme to transmit RT bits of information using $(1+\alpha)T$ seconds. (Later we shall choose α arbitrarily small.) We divide the block into two phases, first the *training phase* and then the *information phase*, where the training phase is αT seconds long and the information phase is T seconds long. Within each phase we make the same discretization in time as in Section II, where every time-slot is Δ seconds long. We choose Δ to be small enough so that $\lambda\Delta$, $A\Delta$ and $\nu\Delta$ are all small compared to one. In this case, our channel is reduced to a DMC whose transition law can be

[1] For more explicit formulations of random and deterministic codes, see [10] and references therein.

approximated by:

$$W(1|x,s)=\begin{cases}\lambda\Delta, & x=0, s=0\\ (A+\lambda)\Delta, & x=1, s=0\\ 1, & s=1.\end{cases} \qquad (8)$$

Here $x=1$ means that the continuous-time channel input in the time-slot is A, and $x=0$ means that the continuous-time channel input is zero; $y=1$ means that at least one count is detected at the receiver in the time-slot, and $y=0$ means that no counts are detected; $s=1$ means that there is at least one spurious count in the time-slot, so the output is stuck at 1, and $s=0$ means that there are no spurious counts in the time-slot, so the channel is the same as (5). From now on we shall refer to time-slots where $s=1$ as "stuck slots."

Denote the total number of stuck slots in the information phase by μT. Then, by the state constraint,

$$\mu \leq (1+\alpha)\nu. \qquad (9)$$

In the training phase the transmitter tells the receiver the value of μT. To do this, the transmitter and the receiver use the channel as an AVC. Namely, in this phase the transmitter ignores his knowledge about the times of the spurious counts. Since, by the state constraint, the total number of stuck slots in the whole transmission block cannot exceed $(1+\alpha)\nu T$, we know that the total number of stuck slots in the training phase also cannot exceed $(1+\alpha)\nu T$. It can be easily verified using the formula for random coding capacity of the AVC with state constraints [11] that the random coding capacity of the AVC (8) under this constraint is proportional to Δ for small Δ. Thus, the amount of information that can be reliably transmitted in the training phase is proportional to αT for large T and small Δ. On the other hand, according to (9), we only need $\log\big((1+\alpha)\nu T\big)$ bits to describe μT. Thus we conclude that, for any $\alpha>0$, for large enough T and small enough Δ, the training phase can be accomplished successfully with high probability.

We next describe the information phase (which is T seconds long). If the training phase is successful, then in the information phase, both the transmitter and the receiver know the total number of stuck slots μT, but only the transmitter knows their positions. With such knowledge, they can use the following random coding scheme to transmit RT bits of information in this phase:

- **Codebook:** Generate $2^{(R+R')T}$ codewords independently such that the symbols within every codeword are chosen i.i.d. Ber (p). Label the codewords as $\mathbf{x}(m,k)$ where $m\in\{1,\dots,2^{RT}\}$ and $k\in\{1,\dots,2^{R'T}\}$.
- **Encoder:** For a given message $m\in\{1,\dots,2^{RT}\}$ and a state sequence $\mathbf{s}$, find a k such that

$$\sum_{j=1}^{T/\Delta} \mathrm{I}\Big\{\big(x_j(m,k),s_j\big)=(1,1)\Big\} \geq (1-\epsilon)\frac{p(A+\lambda)}{pA+\lambda}\mu T,$$

where $\mathrm{I}\{\cdot\}$ denotes the indicator function so the left-hand side (LHS) is the number of slots where $\mathbf{x}(m,k)$ and $\mathbf{s}$ are both one. Send $\mathbf{x}(m,k)$. If no such k can be found, send an arbitrary sequence.
- **Decoder:** Find a codeword $\mathbf{x}(m',k')$ in the codebook such that, for the observed $\mathbf{y}$,

$$\sum_{j=1}^{T/\Delta} \mathrm{I}\Big\{\big(x_j(m',k'),y_j\big)=(1,1)\Big\} \geq (1-\epsilon)\frac{p(A+\lambda)}{pA+\lambda}(pA+\lambda+\mu)T.$$

Output m'. If no such codeword can be found, declare an error.

We next analyze the error probability of this random codebook. There are three types of errors which we analyze separately:

- The encoder cannot find a k such that $\mathbf{x}(m,k)$ meets the requirement. We know that the total number of stuck slots in this phase is μT. Since the codebook is generated independently of the stuck slots, we know that the symbols of a particular codeword at these slots are drawn i.i.d. according to Ber (p). By Sanov's theorem [12], [13], for large T, the probability that a particular codeword satisfies the requirement, i.e., has at least $(1-\epsilon)\frac{p(A+\lambda)}{pA+\lambda}\mu T$ ones in these μT slots, is approximately $2^{\mu T D\left(\mathrm{Ber}\left((1-\epsilon)\frac{p(A+\lambda)}{pA+\lambda}\right)\middle\|\mathrm{Ber}(p)\right)}$. Therefore, when T tends to infinity, the probability of this error tends to zero if we choose

$$R' > \mu D\left(\mathrm{Ber}\left((1-\epsilon)\frac{p(A+\lambda)}{pA+\lambda}\right)\middle\|\,\mathrm{Ber}\,(p)\right). \qquad (10)$$

- There are less than $(1-\epsilon)\frac{p(A+\lambda)}{pA+\lambda}(pA+\lambda+\mu)T$ slots where the transmitted codeword and the output are both equal to one. The probability of this error tends to zero as T tends to infinity by the law of large numbers.
- There is some $\mathbf{x}'$ which is not the transmitted codeword such that there are at least $(1-\epsilon)\frac{p(A+\lambda)}{pA+\lambda}(pA+\lambda+\mu)T$ slots where $x'_j=1$ and $y_j=1$. To analyze the probability of this error, we first note that, by the law of large numbers, when T is large, the number of slots where $y_j=1$ is close to $(pA+\lambda+\mu)T$. We also note that a particular codeword that is not transmitted is drawn i.i.d. Ber (p) independently of $\mathbf{y}$. Therefore, again by Sanov's theorem, the probability that this codeword has at least $(1-\epsilon)\frac{p(A+\lambda)}{pA+\lambda}(pA+\lambda+\mu)T$ ones at the approximately $(pA+\lambda+\mu)T$ slots where $y_j=1$ is approximately $2^{(pA+\lambda+\mu)TD\left(\mathrm{Ber}\left((1-\epsilon)\frac{p(A+\lambda)}{pA+\lambda}\right)\middle\|\mathrm{Ber}(p)\right)}$. Thus, when T tends to infinity, this probability tends to zero if we choose

$$R+R' < (pA+\lambda+\mu)\cdot D\left(\mathrm{Ber}\left((1-\epsilon)\frac{p(A+\lambda)}{pA+\lambda}\right)\middle\|\,\mathrm{Ber}\,(p)\right). \qquad (11)$$

By combining (10) and (11) and noting that ϵ can be chosen to be arbitrarily small, we conclude that, for every $p\in(0,1)$, when T is large and when Δ is small, successful transmission in the information phase can be achieved with high probability as long as

$$R < (pA + \lambda) D\left(\text{Ber}\left(\frac{p(A+\lambda)}{pA+\lambda}\right) \middle\| \text{Ber}(p) \right). \tag{12}$$

Furthermore, since we have shown that the training phase can be accomplished with any positive α, the overall transmission rate, which is equal to $\frac{R}{1+\alpha}$, can also be made arbitrarily close to the RHS of (12). Optimization over p implies that we can achieve all rates up to the RHS of (6) with random coding. ■

B. Deterministic Codes against Random States

We next consider the case where, rather than being an arbitrary sequence chosen by an adversary, the spurious counts are random. Such random counts can be modeled by a random counting process $S(\cdot)$ which is independent of the message, so the channel output is given by

$$Y(t) = Z(t) + S(t), \quad t \in \mathbb{R}, \tag{13}$$

where $Z(\cdot)$ is a Poisson process whose time-t intensity is $x(t)+\lambda$, and conditional on $x(\cdot)$, $Z(\cdot)$ is independent of $S(\cdot)$. We assume that $S(0) = 0$ with probability one and

$$\varlimsup_{t\to\infty} \frac{\mathsf{E}[S(t)]}{t} < \infty. \tag{14}$$

Note that these conditions are satisfied, for example, when $S(\cdot)$ is a homogeneous Poisson process. We also assume that the law of $S(\cdot)$ is known to both transmitter and receiver (and, in particular, the code may depend on the law of $S(\cdot)$), while the realization of $S(\cdot)$ is known noncausally to the transmitter but unknown to the receiver. A rate is said to be achievable on this channel if, for every $T > 0$, there exists a deterministic (T, R) code such that the average probability of error averaged over $S(\cdot)$ tends to zero as T tends to infinity. The capacity of this channel is defined as the supremum over all achievable rates.

Theorem 2: The capacity of the channel (13), where the realization of $S(\cdot)$ is known noncausally to the transmitter but unknown to the receiver, is equal to $C(A, \lambda)$, irrespective of the law of $S(\cdot)$.

Proof: We first observe that the capacity of (13) cannot exceed $C(A, \lambda)$. This is because we can mimic the channel (13) over a channel without spurious counts as follows: The transmitter and the receiver use common randomness (which does not help on the single user channel without states) to generate $S(\cdot)$ and then the receiver ignores its realization.

We shall next show that any rate below $C(A, \lambda)$ is achievable on (13). Fix any $\epsilon > 0$ and $R < C(A, \lambda)$. Let

$$\zeta \triangleq \varlimsup_{t\to\infty} \frac{\mathsf{E}[S(t)]}{t}.$$

Since $\zeta < \infty$, there exists a t_0 such that

$$\mathsf{E}[S(t)] \le 2\zeta, \quad t > t_0.$$

Using this and Markov's inequality we have

$$\Pr\left[S(t) \ge \frac{2\zeta}{\epsilon}\right] \le \epsilon, \quad t > t_0.$$

Thus the error probability of a (T, R) random code where $T > t_0$ can be bounded as

$$\begin{aligned} \Pr[\text{error}] &\le \Pr\left[\text{error}, S(T) \ge \frac{2\zeta}{\epsilon}\right] + \Pr\left[\text{error}, S(T) < \frac{2\zeta}{\epsilon}\right] \\ &\le \Pr\left[S(T) \ge \frac{2\zeta}{\epsilon}\right] \\ &\quad + \Pr\left[\text{error} \middle| S(T) < \frac{2\zeta}{\epsilon}\right] \cdot \Pr\left[S(T) < \frac{2\zeta}{\epsilon}\right] \\ &\le \epsilon + \Pr\left[\text{error} \middle| S(T) < \frac{2\zeta}{\epsilon}\right], \quad T > t_0. \end{aligned} \tag{15}$$

To bound the second term on the RHS of (15) we use Theorem 1 which says that there exists t_1 such that, for any $T > t_1$, there exists a random (T, R) code whose average error probability conditional on any realization of $S(\cdot)$ satisfying $s(T) < \frac{2\zeta}{\epsilon}$ is not larger than ϵ. Therefore, for such codes,

$$\Pr\left[\text{error} \middle| S(T) < \frac{2\zeta}{\epsilon}\right] \le \epsilon, \quad T > t_1. \tag{16}$$

Combining (15) and (16) yields that for any $T > \max\{t_0, t_1\}$ there exists a random (T, R) code for which

$$\Pr[\text{error}] \le 2\epsilon.$$

Since this is true for all $\epsilon > 0$ and $R < C(A, \lambda)$, we conclude that all rates below $C(A, \lambda)$ are achievable on (13) with random codes.

We next observe that we do not need to use random codes. Indeed, picking for every T and R the best deterministic (T, R) code yields at worst the same average error probability as that of any random (T, R) code. Thus we conclude that any rate below $C(A, \lambda)$ is achievable on (13) with deterministic codes, and hence the capacity of (13) is equal to $C(A, \lambda)$. ■

IV. Causal CSI

In this section we shall argue that causal CSI does not increase the capacity of a peak-limited Poisson channel. We look at the simplest case where the spurious counts occur as a homogeneous Poisson process of intensity μ. We shall be imprecise regarding the definition of causality in continuous time by directly looking at the DMC (8).[2] Since the continuous-time state $S(\cdot)$ is a Poisson process, the discrete-time state sequence $\mathbf{S}$ is i.i.d. with each component taking the value 1 with probability $\mu\Delta$ and taking the value 0 otherwise.

To argue that having causal CSI does not increase capacity, we shall show that every mapping u from channel states to input symbols (as in (2)) is equivalent to a *deterministic input symbol* in the sense that it induces the same output distribution as the latter. Indeed, since when $s = 1$ the input symbol x has no influence on the output Y, we know that the value of $u(1)$ does not influence the output distribution. Therefore $u\colon s \mapsto u(s)$ is equivalent to the mapping that maps both 0 and 1 to $u(0)$, and is thus also equivalent to the deterministic input symbol $x = u(0)$.

[2]For formulations of causality in continuous time see [14] or [15].

In a more explicit argument, we use the capacity expression (2). For any distribution P_U on $\mathcal{U}$, we let

$$P_X(x) = \sum_{u:\ u(0)=x} P_U(u).$$

Then, for the above P_U and P_X,

$$I(U;Y) = I(X;Y).$$

Therefore we have

$$\sup_{P_U} I(U;Y) \leq \sup_{P_X} I(X;Y),$$

where the LHS is the capacity of the channel with causal CSI, and where the RHS is the capacity of the channel with no CSI. Thus we conclude that the capacity of our channel model with causal CSI is not larger than that of the channel with no CSI.

References

[1] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Research and Development*, vol. 2, pp. 289–293, 1958.

[2] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Prob. Contr. and Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[3] A. V. Kusnetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Probl. Peredachi Inf.*, vol. 10, no. 2, Apr.–June 1974.

[4] C. D. Heegard and A. A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. Inform. Theory*, vol. 29, no. 5, pp. 731–739, Sept. 1983.

[5] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[6] Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probability and Its Appl.*, vol. 23, pp. 143–147, 1978.

[7] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inform. Theory*, vol. 26, pp. 710–715, Nov. 1980.

[8] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel — part I," *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1449–1461, Nov. 1988.

[9] S. Bross and S. Shamai (Shitz), "Capacity and decoding rules for the Poisson arbitrarily varying channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 3076–3093, Nov. 2003.

[10] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[11] I. Csiszár and P. Narayan, "Arbitrarily varying channels with constrained inputs and states," *IEEE Trans. Inform. Theory*, vol. 34, pp. 27–34, Jan. 1988.

[12] I. N. Sanov, "On the probability of large deviations of random variables," *Mat. Sbornik*, vol. 42, pp. 11–44, 1957.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.

[14] A. Lapidoth, "On the reliability function of the ideal Poisson channel with noiseless feedback," *IEEE Trans. Inform. Theory*, vol. 39, pp. 491–503, Mar. 1993.

[15] T. T. Kadota, M. Zakai, and J. Ziv, "Mutual information of the white Gaussian channel with and without feedback," *IEEE Trans. Inform. Theory*, vol. 17, no. 4, pp. 368–371, July 1971.