On Capacity of Optical Channels with Coherent Detection

Hye Won Chung   Saikat Guha   Lizhong Zheng

***Abstract*—[1] Optical channel with direct detections, also known as Poisson channel, is well studied. With direction detections, only the intensity of the optical signal is measured and used as the carrier of information. In coherent detection, the phase of optical signal is also utilized. Constrained by optical devices, it is hard to directly measure this phase. However, it is possible to mix the input signal with a local reference signal to create output that is phase dependent. Generating such local signals can be based on the instantaneous receiver knowledge, and updated from time to time. We study the channel capacity of such channel, and make connection to the recent development in feedback channels. We are particularly interested in the low SNR regime, and present a capacity result based on a new scaling law.**

I. Introduction: Detection of Optical Signals

We start by describing the optical channel of interest with as little quantum terminology as possible. Over a given period of time $t \in [0,T)$, we first consider a constant input to the channel, which is a *coherent state*, denoted by $|S\rangle$, where $S \in \mathcal{C}$. Here, coherent state can be understood as simply the light generated from a classical laser gun. In a noise-free environment, if one uses a photon counter to receive this optical signal, the output of the photon counter is a Poisson process, with rate $\lambda = |S|^2$, indicating the arrivals of individual photons. Clearly, one can generalize from a constant input to have $|S(t)\rangle$, which results in a non-homogeneous Poisson process at the output. The cost of transmitting such optical signals is naturally the average number of photons, which equals to $\int_0^T |S(t)|^2 dt$. Here, without loss of generality, we set the scaling factors on the rate and photon counts to 1, ignoring issues with linear attenuation and efficiency of optical devices. Such receivers based on photon counters that detect the intensity of the optical signals are called direct detection receivers, and the resulting communication channel is called a Poisson channel. The capacity of the Poisson channel is well studied [9], [5].

Since coherent state optical signal can be described by a complex number S, it is of interest to design coherent receivers, which measure the phase of S, and thus allow information to be modulated on the phase. The following architecture, proposed by Kennedy, is a particular front end of the receiver, the output of which depends on the phase of S.

[1]Hye Won Chung (hwchung@mit.edu) and Lizhong Zheng (lizhong@mit.edu) are with the EECS dept. MIT; Saikat Guha (sguha@bbn.com) is with Raytheon BBN Tech. The authors would like to acknowledge the support by the DARPA Information in a Photon program, contract number HR0011-10-C-0159.

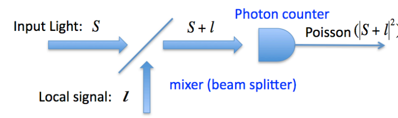

Fig. 1. Coherent Receiver Using Local Feedback Signal

In Figure-1, instead of directly feeding the input optical signal $|S\rangle$ to the photon counter, a local signal $|l\rangle$ is mixed with the input, to generate a coherent state $|S+l\rangle$, and the output of the photon counter is a Poisson process with rate $|S+l|^2$. Note that l can in principle be chosen as an arbitrary complex number, with any desired phase difference from the input signal S. Thus, the output of this processing can be used to extract the phase information in the input. In a sense, the local signal is designed to control the channel through which the optical signal $|S\rangle$ is observed.

Kennedy used this receiver architecture to distinguish between binary hypotheses, i.e., two possible coherent states corresponding to waveforms $S_0(t), S_1(t), t \in [0,T)$, with priori probabilities π_0, π_1, respectively, using a constant control signal l. This work was later generalized by Dolinar [2], where a control waveform $l(t), t \in [0,T)$ was used. The waveform $l(\cdot)$ is chosen adaptively based on the photon arrivals at the output. It was shown that the resulting probability of error for binary hypothesis testing is

$$P_e = \frac{1}{2}\left[1 - \sqrt{1 - 4\pi_0\pi_1 e^{-\int_0^T |S_0(t)-S_1(t)|^2 dt}}\right] \qquad (1)$$

Somewhat surprisingly, this error probability coincides with the lower bound optimized over all possible quantum detectors [3]. The optimality of Dolinar's receiver is an amazing result, as it shows that the minimum probability of error quantum detector for the binary problem can indeed be implemented with the very simple receiver structure in Figure 1. Unfortunately, this result does not generalize to problems with more than 2 hypotheses.

The goal of the current paper is following. We are interested in finding natural generalization of Dolinar's receiver. We would like to consider using such receivers to receive coded transmissions, and thus compute the information rate that can be reliably carried through the optical channel, with the above specific structure of the receiver front end. In stating our observations, we will omit the proofs of some of the results in this version of the paper. In the following, we will start by re-deriving the Dolinar's design of the control waveform $l(t)$ to motivate our approach.

II. Binary Hypothesis Testing

We consider the binary hypothesis testing problem with two possible input signals, $|S_0(t)\rangle, |S_1(t)\rangle$, under hypotheses $H = 0, 1$ respectively, and denote $\pi_0(t)$ and $\pi_1(t)$ as the posterior distribution over the two hypotheses, conditioned on the output of the photon counter up to time t. For simplicity, we assume that $S_0, S_1 \in \mathcal{R}$, and generalize to the complex valued case later. Based on the receiver knowledge, we choose the control signal $l(t)$, to be applied in an arbitrarily short interval $[t, t+\Delta)$. After observing the output during this interval, the receiver can update the posterior probabilities to obtain $\pi_0(t+\Delta)$ and $\pi_1(t+\Delta)$, and then follow the same procedure to choose the control signal in the next interval, and so on. As we pick Δ to be arbitrarily small, we can restrict the control signal $l(t)$ in such a short interval to be a constant l. In the following, we focus on solving the single step optimization of l in the above recursion, and drop the dependence on t to simplify the notation.

We first observe that the optimal value of l must be real, as having a non-zero imaginary part in l simply adds a constant rate to the two Poisson processes corresponding to the two hypotheses, and does not improve the quality of observation. We write $\lambda_i = (S_i + l)^2, i = 0, 1$ to denote the rate of the resulting Poisson processes. Over a very short period of time, the realized Poisson processes can have, with a high probability, either 0 or 1 arrival, with probabilities $1 - \lambda_i\Delta, \lambda_i\Delta$, resp. [2] Now over this short period of time, the receiver front end can be thought as a binary channel as shown in Figure 2. Note that the channel parameters λ_i's depend on the value of the control signal l. Our goal is to pick an l for each short interval such that they contribute to the overall decision in the best way.

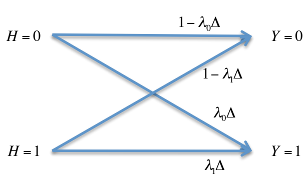

Fig. 2. Effective binary channel between the input hypotheses and the observation over a Δ period of time

The difficulty here is that it is not obvious how we should quantify the "contribution" of the observation over a short period of time to the overall decision making. An intuitive approach one can use is to choose l that maximizes the mutual information over the binary channel. For convenience, we write the input to the channel as H and the output of the channel as $Y \in \{0, 1\}$, indicating either 0 or 1 photon arrival. The following result gives the solution to this optimization problem.

Lemma 1: The optimal choice that maximizes the mutual information $I(H; Y)$ for the effective binary channel is

$$l^* = \frac{S_0\pi_0 - S_1\pi_1}{\pi_1 - \pi_0}. \tag{2}$$

Note that l^* is independent of Δ.
With this choice of the control signal, the following relation holds

$$\pi_0\sqrt{\lambda_0} = \pi_1\sqrt{\lambda_1}. \tag{3}$$

The relation in (3) gives some useful insights. If $\pi_0 > \pi_1$, we have $\lambda_1 > \lambda_0$, and vice versa. That is, by switching the sign of the control signal l, we always make the Poisson rate corresponding to the hypothesis with the higher probability smaller. In the short interval where this control is applied, with a high probability we would observe no photon arrival, in which case we would confirm the more likely hypothesis. For a very small value of Δ, this occurs with a dominating probability, such that the posterior distribution moves only by a very small amount. On the other hand, when there is indeed an arrival, i.e. $Y = 1$, we would be quite surprised, and the posterior distribution of the hypotheses moves away from the prior. Consider this latter case, the updated distribution over the hypotheses can be written as

$$\frac{\Pr(H = 1|Y = 1)}{\Pr(H = 0|Y = 1)} = \frac{\pi_1 \cdot \lambda_1\Delta}{\pi_0 \cdot \lambda_0\Delta} = \frac{\pi_0}{\pi_1}$$

The posterior distribution under the case of 0 or 1 arrival turns out to be inverse to each other. In other words, the larger one of the two probabilities of the hypotheses remains the same no matter if there is an arrival in the interval or not. As we apply such optimal control signals recursively, this larger value progresses towards 1 at a predictable rate, regardless of when and how many arrivals are observed. *The random photon arrivals only affect the decision on which is the more likely hypothesis, but does not affect the quality of this decision.* The next Lemma describes this recursive control signal and the resulting performance. Without loss of generality, we assume that at $t = 0$, the prior distribution satisfies $\pi_0 \geq \pi_1$. Also we write $N(t)$ be the number of arrivals observed in $[0, t)$

Lemma 2: Let $g(t)$ satisfy, $g(0) = \pi_0/\pi_1$, and

$$g(t) = g(0) \cdot \exp\left[\int_0^t \frac{(S_0(t) - S_1(t))^2(g(\tau) + 1)}{g(\tau) - 1} d\tau\right].$$

The recursive mutual-information maximization procedure described above yields a control signal

$$l^*(t) = \begin{cases} l_0(t) & \text{if } N(t) \text{ is even} \\ l_1(t) & \text{if } N(t) \text{ is odd} \end{cases}$$

where

$$l_0(t) = \frac{S_1(t) - S_0(t)g(t)}{g(t) - 1}, \quad l_1(t) = \frac{S_0(t) - S_1(t)g(t)}{g(t) - 1}.$$

[2] One has to be careful in using the above approximation of the binary channel. As we are optimizing over the control signal, it is not obvious that the resulting λ_i's are bounded, In other word, the mean of the Poisson distributions, $\lambda_i\Delta$, might not be small. Thus, the assumption of either 0 or 1 arrival, and the approximation in the corresponding probabilities, need to be justified. More detail on this step can be found in the full version of this paper.

Furthermore, at time T, the decision of the hypothesis testing problem is $\hat{H} = 0$ if $N(T)$ is even, and $\hat{H} = 1$ otherwise. The resulting probability of error coincides with (1).

Figure 3 shows an example of the optimal control signal. The plot is for a case where $S_i(t)$'s are constant on-off-keying waveforms. As shown in the plot, the control signal $l(t)$ jumps between two prescribed curves, l_0, l_1, corresponding to the cases $\pi_0 > \pi_1$ and $\pi_0 < \pi_1$, resp. With the proper choice of the control signal, each time when there is a photon arrival, the receiver is so surprised that it flips its choice of $\hat{H}$. However, $g(t) = \max\{\pi_0(t), \pi_1(t)\}/\min\{\pi_0(t), \pi_1(t)\}$ indicating how much the receiver is committed to the more likely hypothesis, increases at a prescribed rate regardless of the arrivals.

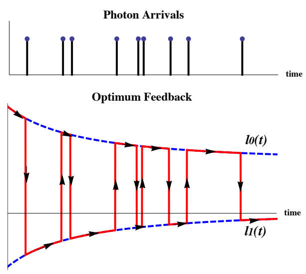

Fig. 3. An example of the control signal that achieves the minimum probability of error.

The success in the binary hypothesis testing problem reveals some useful insights for general dynamic communication problems. Regardless of the physical channel that one communicates over, one can always have a "slow motion" understanding of the process by studying how the posterior distribution over the messages evolves over time. Over the process of communications, this posterior distribution, conditioned on more and more observations of the receiver, should move from the prior towards a deterministic distribution, allowing the receiver to "lock in" on a particular message. This viewpoint is more general than the conventional setup in information theory, and particularly useful in understanding dynamic problems, as it is not based on any notion of sufficient statistics, block codes, or any predefined notion of reliability. As we measure how far the posterior distribution moves at each time, we can quantify how the communication process at each time point contributes to the overall decision making.

III. Coded Transmissions and Scaling Laws

We now turn our attention to the problem of coded transmissions over the optical channel with coherent receivers. We are interested in finding the classical capacity of such channels, i.e., the number of information bits that can be modulated into the optical signals, and reliably decoded with a receiver architecture shown in Figure 1. We are particularly interested in the case where the average number of photon transmitted is small, and hence a high photon efficiency, in bits/photon, is achieved.

The capacity of the same channel without the constraint in the receiver architecture is studied in [1], [4]. It is shown [11] that the capacity of the channel is given by

$$C_{\mathsf{Holevo}}(\mathcal{E}) = (1+\mathcal{E})\log(1+\mathcal{E}) - \mathcal{E}\log\mathcal{E} \text{ bits/use} \qquad (4)$$

where $\mathcal{E}$ is the average number of photon transmitted per channel use. To achieve this data rate, an optimal joint quantum measurement over a long sequences of symbols must be used. In practice, however, such measurement is very hard to implement. We are therefore interested in finding the achievable data rate when a simple receiver structure is adopted. Nevertheless, (4) serves as a performance benchmark. In the regime of interests where $\mathcal{E} \to 0$, it is useful to approximate (4) as

$$C_{\mathsf{Holevo}}(\mathcal{E}) = \mathcal{E}\log\frac{1}{\mathcal{E}} + \mathcal{E} + o(n). \qquad (5)$$

As another performance reference, we also consider the capacity when a direct detection receiver is used. The capacity of this channel is studied in [5], [9], and the regime of low average photon numbers is studied in [10]. For our purpose of performance comparison, we actually need a more precise scaling law of performance. The following Lemma describes such a result.

Lemma 3 (Capacity of Direct Detection): As $\mathcal{E} \to 0$, the optimal input distribution to the optical channel with direct detection is on-off-keying, with

$$|S\rangle = \begin{cases} |0\rangle, & \text{with prob. } 1-p^* \\ |\sqrt{\mathcal{E}/p^*}\rangle, & \text{with prob. } p^* \end{cases}$$

where $\lim_{\mathcal{E}\to 0} \frac{p^*}{\frac{\mathcal{E}}{2}\log\frac{1}{\mathcal{E}}} = 1$, and the resulting capacity is

$$C_{\mathsf{DD}}(\mathcal{E}) = \mathcal{E}\log\frac{1}{\mathcal{E}} - \mathcal{E}\log\log\frac{1}{\mathcal{E}} + O(n) \qquad (6)$$

Comparing (5) and (6), we observe that the two capacities have the same leading term. This means as $\mathcal{E} \to 0$, the optimal photon efficiency of $\log(1/\mathcal{E})$ bits/photon can be achieved even with a very simple direct detection receiver.

In practice, however, the two performances have significant difference. For example, if one wishes to achieve a photon efficiency of 10 bits/photon, one can solve for $\mathcal{E}$ that satisfies $C(\mathcal{E})/\mathcal{E} = 10$ bits/photon in both cases, and get $\mathcal{E}_{\mathsf{Holevo}} \approx 0.0027$ and $\mathcal{E}_{\mathsf{DD}} \approx 0.00010$. The resulting capacities also differ by more than 1 order of magnitude. This example says that although (5) and (6) have the same limit as $\mathcal{E} \to 0$, the rates at which this limit is approached are quite different, which is of practical importance. Similar phenomenon has also been observed for wideband wireless channels [6], [7].

As a result, the 2nd term in the capacity results cannot be ignored. In fact, any reasonable scheme with coherent processing should at least achieve a rate higher than that with direct detection, and thus should have the leading term as $\mathcal{E}\log\frac{1}{\mathcal{E}}$. It is the second term in the achievable rate that indicates whether a new scheme is making a significant step towards achieving the Holevo capacity limit. In the following, we will study the achievable rates over the optical channel with receiver front end as shown in Figure 1, and evaluate the performance according to this scaling law.

Now, we will consider the problem of coded transmission and finding the maximum information rate that can be conveyed through an optical channel with a coherent processing receiver. There are exponentially many, possible messages for the coded transmission. The first question one might ask is how to optimize the control signal to adjust the channel and to help the reliable communication over the channel. When communicating with a long block of N symbols, there is no issue of a pressing deadline of decision of the right message for the most of the time. Therefore, it makes sense to always use the mutual information maximization to decide which control signal to apply. A straightforward generalization of the Dolinar's receiver can be described as follows:

First, at each time instance $i \in \{1, \ldots, N\}$, the encoding map can be written $f_i : \{1, 2, \ldots, M = 2^{NR}\} \to X_i \in \mathcal{X}$, where X_i is the symbol transmitted in the i^{th} use of the channel. This map ensures that X_i has a desired input distribution P_X, computed under the assumption that all messages are equally likely. That is, $\frac{1}{2^{NR}}|\{m : f_i(m) = x\}| = P_X(x), \quad \forall x \in \mathcal{X}$.

The receiver keeps track of the posterior distribution over the messages. Given the distribution over the messages conditioned on the previous observations, $P_{M|Y^{i-1}}(\cdot|y^{i-1})$, one can compute the effective input distribution $P'_X(x) = \sum_{m:f_i(m)=x} P_{M|Y^{i-1}}(m|y^{i-1})$. Using this as the prior distribution of the transmitted symbol, one can apply the control signal that maximizes the mutual information.

Upon observing the output Poisson process in the i^{th} symbol period, denoted as $Y_i = y_i$, the receiver computes the posterior distribution of the transmitted symbol $P''_X(x) = P_{X_i|Y_i}(x|y_i)$[3], and uses that to update its knowledge of the messages:

$$P_{M|Y^i}(m|y^i) = P_{M|Y^{i-1}}(m|y^{i-1}) \cdot \frac{P''_X(x)}{P'_X(x)},$$

for all m such that $f_i(m) = x$.

Repeating this process, we have a coherent-processing receiver based on updating the receiver knowledge. There are two further simplifications that make the analysis of this scheme even simpler.

First, we observe that with exponentially many messages, for a dominating fraction of the time when the block code is transmitted, the receivers knowledge, $P_{M|Y^i}$, satisfies that the probability of any message, including the correct one, is exponentially small. Thus, with a random coding map f_i, P'_X is very close to P_X. Thus, the step of updating the receiver's knowledge is in fact not important. This assumption starts to fail only when the receiver starts to lock in a specific message, i.e., when $P_{M|Y^i}(m)$ is not exponentially small for some m. It is shown in [8] that the fraction of time when this happens is indeed very small, and can thus be ignored when a long term average performance metric such as the data rate is of concern.

[3]We omit the conditioning on the history Y^{i-1} here to emphasize that the update is based on the observations in a single symbol period

Secondly, suppose we choose the optimal input distribution, which maximizes the photon efficiency, over a short period Δ of time. After using this input for Δ time, the receiver would update the posterior distribution, which makes the effective input distribution on X deviate from the optimum. This is undesirable. One can avoid this problem by using very short symbol periods. That is, after transmitting for a very short time period, the transmitter should re-shuffle the messages so that the distribution of the transmitted symbols, conditioned on the receiver's knowledge, is re-adjusted back to the optimal choice. This is precisely the same argument we used in classical communication over wideband channels. As a result, we do not have to worry about updating the receiver's knowledge and the control signals even within a symbol period. Instead, we are interested only in the photon efficiency over a short time period. In other words, we can focus only on a thin slice on the left end of Figure 3.

Based on these observations, we are now ready to state our results in the photon efficiency of the optical channel of interests.

IV. Capacity Results of Coherent Receiver

Theorem 4: For the optical channel with a receiver front end as shown in Figure 1, and sequentially updated control signals, suppose that the transmitted symbols are drawn from a finite alphabet, i.e., at each time the transmitted optical signal $|X_i\rangle$ is chosen from $X_i \in \mathcal{X} \subset \mathcal{C}$, with $|\mathcal{X}|$ finite. Then the achieved photon efficiency is upper bounded by

$$\frac{C(\mathcal{E})}{\mathcal{E}} \leq \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} + O(1) \quad (7)$$

This says that essentially the achievable photon efficiency with coherent receivers is not significantly different from that of direct detection receivers. In this paper, we will show a proof technique of this theorem for binary input $X = 0, 1$ channel whose input signals are $|S_0\rangle, |S_1\rangle \in \mathcal{C}$, respectively. This result can be generalized for any finite input channels, which will be discussed in a longer version of this paper.

In Section A, we first assume that both the input singals and the receiver's control signal are fixed during a short symbol time, Δ, and then calculate the maximum achievable information rate with optimized input distribution and the control signal. In Section B, any infinite bandwidth coherent receiver is allowed, meaning the receiver can update its control signal infinitely many times within the symbol time Δ based on output observations at each instant, and the corresponding capacity is calculated.

A. Capacity of Binary Input Channel

We first consider a binary input channel with the coherent receiver. Either $|S_0\rangle$ or $|S_1\rangle \in \mathcal{C}$ is transmitted for the input symbol $X = 0, 1$ respectively, with probabilities $P_X = [1 - p, p]$. Each input symbol has duration of short time Δ, so that it is fixed within the Δ time. This assumption makes sure that the channel bandwidth is wide enough, but not infinite. We write $\lambda_i \Delta = (S_i + l)^2 \Delta, i = 0, 1$ to denote the rate of the resulting output Poisson processes for the Δ time when the control signal mixed with the received signal equals $l \in \mathcal{C}$. In principle, since infinite bandwidth can be allowed at the receiver side, meaning l can be updated infinitely many times even within the Δ, it is not required to fix l for each symbol detection. However, we will first calculate the capacity of coherent channel under the assumption of the fixed control signal for each symbol detection, and then generalize it to include the infinite bandwidth receiver.

To calculate the mutual information between the input X and the output Y of this channel, we incorporate another assumption that the output of this channel is either 0 or 1 with probabilities $P_{Y|X=i} = [e^{-\lambda_i \Delta}, 1 - e^{-\lambda_i \Delta}]$ for $i = 0, 1$. [4] Then, the mutual information of this channel is

$$\begin{aligned} I(X;Y) =& H_B((1-p)e^{-\lambda_0 \Delta} + pe^{-\lambda_1 \Delta}) \\ & - (1-p) \cdot H_B(e^{-\lambda_0 \Delta}) - p \cdot H_B(e^{-\lambda_1 \Delta}) \end{aligned} \tag{8}$$

where $H_B(x) = -x \log x - (1-x)\log(1-x)$.

When $\mathcal{E}$ denotes the average number of photon per channel use, it can also be written in terms of the input symbol time, i.e., $\mathcal{E} = n\Delta$ when n is a fixed power constraint. Then the input signals should satisfy the following average power constraint,

$$(1-p)|S_0|^2 + p|S_1|^2 = n. \tag{9}$$

Also, since the control signal l can move the mean of two signals to any value without power constraint, it is reasonable to choose the mean of S_0 and S_1 as 0 to utilize the limited signal power efficiently.

$$(1-p)S_0 + pS_1 = 0 \tag{10}$$

From (9) and (10), the optimum input signals can be represented as functions of n and p, and for the optimum signals, based on Eq. (2), the control signal which maximizes mutual information of the channel for Δ-time can be calculated.

$$S_0^*(p) = -\sqrt{\frac{np}{1-p}} \tag{11}$$

$$S_1^*(p) = \sqrt{\frac{n(1-p)}{p}} \tag{12}$$

$$l^*(p) = \frac{S_0(1-p) - S_1 p}{p - (1-p)} = \frac{2\sqrt{np(1-p)}}{1-2p} \tag{13}$$

By plugging the above results into Eq. (8), $I(X;Y)$ can be written as a function of n and p.

$$I(X;Y) = \frac{n\Delta}{1-2p} \log \frac{1-p}{p} - \frac{(n\Delta)^2}{2p} \log \frac{1-p}{p} + o(\Delta) \tag{14}$$

Here, we keep the second term of Eq. (14) explicitly and claim that for small enough p, this term cannot be ignored as $o(\Delta)$. It will be justified by substituting p^* which maximizes Eq. (14) into the term.

Lemma 5: The optimal choice of p^* *that maximizes Eq. (14) satisfies* $\lim_{\mathcal{E}\to 0} \frac{p^*}{\frac{n\Delta}{2}\log\frac{1}{n\Delta}} = 1$, *and the resulting capacity is*

$$\begin{aligned} \frac{C_{\mathsf{Coh\_B}}(\mathcal{E})}{\mathcal{E}} &= \max_p \frac{I(X;Y)}{\mathcal{E}} \\ &\le \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + O(\mathcal{E}) \end{aligned} \tag{15}$$

This result shows that for the binary input channel with coherent receiver of finite bandwidth, the second term of the capacity dose not improve compared to that of direct detection.

Remark 1. This lemma essentially says that even with a feedback control signal which can adjust the channel according to the output observations, the resulting capacity is not significantly different from the case without feedback. As discussed in previous sections, the feedback control signal tunes channel to maximize mutual information between input and output according to the updated posterior distribution of input symbols, which is the effective input distribution of channel. For decision problems such as binary hypothesis testing, this feedback works well in terms of minizing probability of decision error. However, for communication problems which incorporates coding, it turns out that updating the feedback control signal is not as powerful as before, since coding itself can refresh the channel input distribution to be optimum for every new transmission of symbols. In other words, coding which can hold the input distribution at the optimum is a much more powerful technique than the feedback control signal which can adjust channel distribution according to a deviated input distribution from the optimum.

Remark 2. Another role of the control signal l is that it moves the mean of two input signals S_0 and S_1 to any complex value with no power constraint. For modulations with peak power constraint, such as Binary Phase Shift Keying (BPSK), this role of control signal is significantly important in increasing the transmitted photon efficiency. However, for modulations without peak power constraint, the role of control signal in shifting the mean of two input signals does not help much in increasing the photon efficiency and the resulting optimized control signal is $\lim_{\mathcal{E}\to 0} l(p^*) = 0$, which can be shown by plugging $\lim_{\mathcal{E}\to 0} \frac{p^*}{\frac{n\Delta}{2}\log\frac{1}{n\Delta}} = 1$ into (13).

However, note that, in this section, the feedback control signal is assumed to be constant for each symbol detection. Now, we will eliminate this assumption and state the capacity results for infinite bandwidth coherent receiver.

B. Capacity of Infinite Bandwidth Coherent Receiver

In Section A, when we calculate the capacity of a binary channel, it is assumed that the feedback control signal l is fixed during each symbol time Δ. However, even though input symbol cannot be changed during the Δ-time due to a

[4] To prove that this assumption does not change the second term of capacity which is of our interest, we need to show that the optimum input signal and the control signal does still guarantee the condition of finite $\lambda_i \Delta$, $i = 0, 1$ so that the probability of more than one photon arrivals at the output is $o(\Delta)$. More details on this step can be found in the full version of this paper.

finite bandwidth restriction of channel, in principle, coherent receiver itself can update the feedback signal l infinitely many times during the Δ-time based on output observations at each instant.

From Eq. (15), we can observe that higher photon efficiency can be achieved as $\mathcal{E}$, the average number of photons per channel, decreases. For example, if we are allowed to use channel twice and at each time transmit average $\frac{\mathcal{E}}{2}$ photons, compared to the case of transmitting the same $\mathcal{E}$ photons by one channel use, we can gain about $\log 2$ nats/photon from the first dominant term of Eq. (15). Of course, since the channel is used twice for the first case, the spectral efficiency, bits/channel use, gets worse compared to the later case. We can see the trade-off between photon efficiency and spectral efficiency.

When the infinite bandwidth coherent receiver observes very tiny piece of a transmitted symbol at each time and instantaneously updates the control signal l, the receiver essentially divides $\mathcal{E}$ into many pieces and observes them sequentially with feedback signal. Therefore, gain in photon efficiency can be expected. However, it also has to be recognized that even though one symbol with energy $\mathcal{E}$ is divided into infinitely many pieces, among the pieces, there exists certain amount of correlation so that each observation cannot be independent, meaning by observing previous pieces, information about the next observation is already partially known at the receiver. The gain from division of one symbol into many pieces and the loss from the correlation among the pieces will be analyzed to evaluate the maximum achievable rate with infinite bandwidth coherent receiver.

To understand how the partial observation of a symbol does affect the information for the rest part of a symbol, we will start with a simple example that the coherent receiver updates the control signal once after observing half of the symbol, meaning l is updated after observing a symbol for the first $\frac{\Delta}{2}$. The mutual information between channel input X and output Y is

$$I(X;Y) = I(X;Y^{(1)}_{\frac{\Delta}{2}}) + I(X;Y^{(2)}_{\frac{\Delta}{2}}|Y^{(1)}_{\frac{\Delta}{2}}) \tag{16}$$

where $Y^{(i)}_{\frac{\Delta}{2}}$, $i = 1, 2$ is the receiver's observation at the i^{th} $\frac{\Delta}{2}$-time. The two mutual information terms in (16) will be calculated and compared to see how the first partial observation $Y^{(1)}_{\frac{\Delta}{2}}$ affects the next observation of the same input signal.

For binary input X with distribution $P_X = [1-p, p]$, when input signals and the control signal which is applied for the first half of Δ are optimized as in (11)-(13) as independent of Δ, $I(X;Y^{(1)}_{\frac{\Delta}{2}})$ can be calculated by substituting $\frac{\Delta}{2}$ in the place of Δ in (14).

$$I(X;Y^{(1)}_{\frac{\Delta}{2}}) = \frac{n\Delta}{2(1-2p)}\log\frac{1-p}{p} - \frac{(n\Delta)^2}{8p}\log\frac{1-p}{p} + o(\Delta) \tag{17}$$

This rate is greater than a half of (14), which reflects the fact that the photon efficiency increases as the number of photons per channel use is reduced. At the receiver's point of view, it first detects average $\frac{\mathcal{E}}{2}$-number of photons and partially decodes X with some gain in photon efficiency compared to the case of using $\mathcal{E}$ to decode the whole X.

Now, based on the partial observation of X, we can evaluate the posterior distribution $P_{X|Y^{(1)}_{\frac{\Delta}{2}}}$, which is a new effective input distribution at the receiver's point of view, and update the feedback control signal to match with the evolving distribution. $P_{X|Y^{(1)}_{\frac{\Delta}{2}}}$ moves away from P_X depending on the realization of the first output. When we write $P_{X|Y^{(1)}_{\frac{\Delta}{2}}=y} = [1-p', p']$, for some y, the optimum control signal l applied for the next half of Δ is updated as

$$l^*(p') = \frac{S_0(p)\cdot(1-p') - S_1(p)\cdot p'}{p' - (1-p')} = \frac{\sqrt{n}(p + p' - 2pp')}{(1-2p')\sqrt{p(1-p)}}. \tag{18}$$

Note that the input signals $S_0(p), S_1(p)$ are still the values optimized for the initial input distribution $P_X = [1-p, p]$. The mutual information between input X and output $Y^{(2)}_{\frac{\Delta}{2}}$ given $Y^{(1)}_{\frac{\Delta}{2}}$ is therefore calculated with the original input signals under new distribution $P_{X|Y^{(1)}_{\frac{\Delta}{2}}}$ and the updated control signal.

$$I(X;Y^{(2)}_{\frac{\Delta}{2}}|Y^{(1)}_{\frac{\Delta}{2}}) = \frac{n\Delta}{2(1-2p)}\log\frac{1-p}{p} - \frac{3(n\Delta)^2}{8p}\log\frac{1-p}{p} + o(\Delta) \tag{19}$$

Compared to (17), the mutual information in (19) has lower rate for the same average photons $\frac{\mathcal{E}}{2}$. It is because the input distribution shown at the receiver side, $P_{X|Y^{(1)}_{\frac{\Delta}{2}}}$, is deviated from P_X with which the input signals are optimized. Even though feedback adjusts the channel distribution according to the evolving posterior distribution, it does not help enough to recover the loss. By summing (17) and (19), the total rate is

$$I(X;Y) = \frac{n\Delta}{(1-2p)}\log\frac{1-p}{p} - \frac{(n\Delta)^2}{2p}\log\frac{1-p}{p} + o(\Delta), \tag{20}$$

which is the same until $o(\Delta)$ with (14) where the feedback control signal is fixed during Δ. From this example, we observe that updating l once at the middle of Δ based on the previous observation does not significantly change the capacity of coherent receiver.

Now, we will allow infinite bandwidth coherent receiver. The question here is again whether there will be enough gain or not in photon efficiency to improve the second dominant term. To analyze this problem, we start with a simple chain rule. When receiver observes tiny pieces of the received signal and updates the control signal based on the observations at every $\frac{i\Delta}{M}$, $i = 0, 1, \ldots, M-1$ where M can go to infinity, the mutual information between input X and output $Y_1^M = \{Y_1, \ldots, Y_M\}$ where Y_i denotes output observation during $[\frac{i\Delta}{M}, \frac{(i+1)\Delta}{M})$ is

$$I(X;Y_1^M) = I(X;Y_1) + I(X;Y_2|Y_1) + \cdots + I(X;Y_M|Y_1^{M-1}) \tag{21}$$

where

$$I(X;Y_i|Y_1^{i-1}) = \sum_{y_1^{i-1}\in\mathcal{Y}_1^{i-1}} P_{Y_1^{i-1}}(y_1^{i-1}) I(X;Y_i|Y_1^{i-1}=y_1^{i-1}). \quad (22)$$

To analyze Eq. (22), we use the following observations. First, when we define a map,

$$\mathsf{F}: P_X \to I(P_X;Y) \text{ with the optimum } l^*,$$

it can be shown that F is strictly concave, and there exists a maximum point P_X^*. When the input distribution is set to be $P_X = P_X^*$, as we observe Y_1^{i-1}, $P_{X|Y_1^{i-1}=y_1^{i-1}}$ for different y_1^{i-1} deviates from P_X. More importantly, the deviation is associated with $I(P_X;Y_1^{i-1})$. As more information we extract from the observation history, the distribution for the next observation moves further apart, and since F is concave, $\mathsf{E}_{Y_1^{i-1}}[I(P_{X|Y_1^{i-1}};Y_i)]$ is bounded away from the optimum.

Now, we will explicitly write the relationship between deviation of $P_{X|Y_1^{i-1}=y_1^{i-1}}$ for different y_1^{i-1}'s and $I(P_X;Y_1^{i-1})$. When $P_X = [1-p,p]$, we can write $P_{X|Y_1^{i-1}=y_1^{i-1}} = [1-p\cdot\alpha_i(y_1^{i-1}),\ p\cdot\alpha_i(y_1^{i-1})]$ for some α depending on y_1^{i-1}. Then,

$$\begin{aligned} &I(P_X;Y_1^{i-1}) \\ &= \sum_{y_1^{i-1}} P_{Y_1^{i-1}}(y_1^{i-1}) D(P_{X|Y_1^{i-1}}||P_X) \\ &= \sum_{y_1^{i-1}} P_{Y_1^{i-1}}(y_1^{i-1}) \frac{(p\cdot\alpha_i(y_i^{i-1})-p)^2}{2p(1-p)} \\ &\quad + o(p^2\cdot\sum_{y_1^{i-1}} P_{Y_1^{i-1}}(y_1^{i-1})(\alpha_i(y_1^{i-1})-1)^2) \end{aligned} \quad (23)$$

Remark 3. The above equation shows that how much the posterior distributions are deviated from each other for different observation history is proportional to the extracted information from the observations, $I(P_X;Y_1^{i-1})$. Since F is concave, as the variance of the posterior distribution grows, $\mathsf{E}_{Y_1^{i-1}}[I(P_{X|Y_1^{i-1}};Y_i)]$ decreases. Therefore, there exists a trade-off between $I(P_X;Y_1^{i-1})$ and $\mathsf{E}_{Y_1^{i-1}}[I(P_{X|Y_1^{i-1}};Y_i)]$. Now the question is how to balance these mutual informations to maximize the sum rates for the infinite observations. This question can be answered by calculating the concavity of the map F in P_X and analytically solve the optimum P_X^* for the sum rate.

The following lemma summarizes the capacity result for infinite bandwidth coherent receiver.

Lemma 6: Suppose that the coherent receiver splits a received binary symbol into infinitely many pieces and observes them sequentially with the updating optimum feedback l. The optimal choice of p^* that maximizes the capacity in Eq. (21) when $M \to \infty$ satisfies $\lim_{\mathcal{E}\to 0} \frac{p^*}{\frac{n\Delta}{2}\log\frac{1}{n\Delta}} = 1$, and the resulting capacity is

$$\begin{aligned} \frac{C_{\mathsf{Coh\_BW}}(\mathcal{E})}{\mathcal{E}} &= \max_p \frac{I(X;Y)}{\mathcal{E}} \\ &\le \log\frac{1}{\mathcal{E}} - \log\log\frac{1}{\mathcal{E}} + O(\mathcal{E}) \end{aligned} \quad (24)$$

This lemma claims that infinite bandwidth coherent receiver does not gain enough to improve the second dominant term of the resulting capacity.

Based on the spirit of Lemma 5 and 6, the main theorem of this paper stated at the beginning of this chapter can be proved. The important observation from the theorem is that even when the coherent receiver can move the mean of any finite number of input signals without any power constraint and can update the control signal instantaneously based on the observations at the output, in the low power regime, the gain from this coherent processing is not enough to improve the second dominant term of capacity compared to that of direct detection.

V. Conclusion

While the main theorem is negative by nature, it is of practical importance. It implies that in order to achieve the photon efficiency predicted by the Holevo limit, it is necessary to resort to quantum processing that introduces non-classical states, such as entangled or squeezed states. The approach of mixing coherent states and applying a local control signal would not yield significant improvement in terms of photon efficiency. However, the proposed sequential receiver designs and the updating scheme for the control signal of coherent receiver can be applied for the other dynamic communication problems.

References

[1] A. S. Holevo, "The capacity of the quantum channel with general signal states," IEEE Trans. Information Theory, **44**, pp. 269-273, 1998.

[2] S. J. Dolinar, Jr., "An optimum receiver for the binary coherent state quantum channel," MIT Research Laboratory of Electronics Quarterly Progress Report 111, Massachusetts Institute of Technology, Cambridge, Massachusetts, pp. 115-120, Oct. 1973.

[3] C. W. Helstrom, "Quantum detection and estimation theory," New York: Academic Press, 1976.

[4] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," Phys. Rev. A **56**, 10.1103, 1997.

[5] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel-Part I," IEEE Trans. Information Theory, **34**, pp. 1449-1461, 1988.

[6] S. Verdu, "Spectral Efficiency in the Wideband Regime", IEEE Trans. Information Theory, **4**8, June, 2002, p. 1319.

[7] L. Zheng, D. Tse and M. Medard, "Channel Coherence in the Low SNR Regime", IEEE Trans. Information Theory, **53**, March 2007, pp. 976-97

[8] I. Csiszar and J. Korner, "Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd Edition", Cambridge, University Press, 2011

[9] S. Shamai (Shitz), "Capacity of a pulse amplitude modulated direct detection photon channel", Proc. Inst. Elec. Eng. Vol. 137, no. 6, pp. 424-30, December, 1990

[10] A. Lapidoth, J. H. Shapiro, V. Venkatesan, and L. Wang, "The Poisson Channel at Low Input Powers" IEEE Trans. Information Theory, **55**, January 2009

[11] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J.H. Shapiro, H.P. Yuen, "Classical capacity of the lossy bosonic channel:the exact solution", Phys. Rev. Letter, **92**, 027902, 2004