On Capacity of Optical Channels with Coherent Detection

Hye Won Chung   Saikat Guha   Lizhong Zheng

***Abstract*—[1] We study the general coherent-state hypothesis testing problem and the capacity of the pure-loss optical channel with a *coherent processing* receiver (a receiver that uses coherent feedback control and direct detection). We describe the binary hypothesis minimum probability of error receiver as optimizing the communication efficiency at each instant, based on recursively updated knowledge of the receiver. Using this viewpoint, we give a natural generalization of the designs to general M-ary hypothesis testing problems. We analyze the information capacity with coherent receivers, and compare the result with that with direct detection receivers and with arbitrary quantum receivers (the Holevo limit), using the appropriate scalings in the low photon number regime.**

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].

תוחלת של מ״מ פואסוני
שווה ל- λ. מאחר
והתהליך הפואסוני משתנה
בזמן, כלומר לא הומוגני,
על מנת לקבל את התוחלת
נעשה אינטגרל על ה t.
כלומר:

$E = \int_t \lambda(t)dt = \int_t |S(t)|^2 dt$

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.

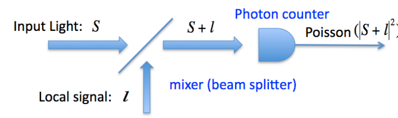

Fig. 1. Coherent Receiver Using Local Feedback Signal

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

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

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 twofold. We are interested in finding natural generalization of Dolinar's receiver to general hypothesis testing problems with more than two possible signals. In addition, 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

π_1, π_2 - Prior Probabillity
$\pi_1(t), \pi_2(t)$ - Posterior Probabilities

end. In stating our observations, we will omit the proofs of all 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.

איך פתאום
$\lambda_i\Delta$ הוא
הסתברות?

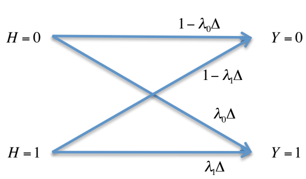

Fig. 2. Effective binary channel between the input hypotheses and the observation over a Δ period of time

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

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

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

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

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 Poission 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, $g(t) = \max\{\pi_0, \pi_1\}/\min\{\pi_0, \pi_1\}$ indicating how much the receiver is committed to the more likely hypothesis, increases at a prescribed rate regardless of the arrivals.

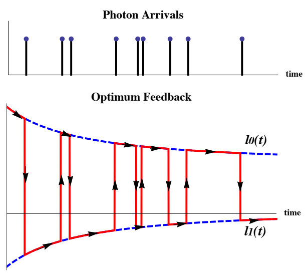

Fig. 3. An example of the control signal that achieves the minimum probability of error.

III. Generalization to M-ary Hypothesis Testing

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.

The optimality result in Lemma 2 is however difficult to duplicate for general M-ary problems. Of course we can always mimic the procedure, to choose the control signal that maximizes the mutual information over an M-input-binary-output channel at each time. The result does not always give a minimum probability of error receiver in general. The reason for that is intuitive. There is a fundamental difference between maximizing mutual information and minimizing the probability of error. In other words, on a general M-ary alphabet, a posterior distribution with a lower entropy does not necessarily have a lower probability of detection errors. These two coincide only for the binary case, since the posterior distribution over two messages live in a single dimensional space. In general, the goal of decision making favors the posterior distributions, over the messages, with a dominating largest element; maximizing mutual information, however, does not distinguish between what kind of information is conveyed.

Consequently, it is hard to define a metric on the efficiency of communication over a time interval in the middle of a communication session that precisely measures how well this interval serves the overall purpose. Even if one can define a precise metric, it is often hard to imagine that analytical solution of the optimal control signal or the resulting performance can be computed from optimizing such metrics. Moreover, such metrics should be time varying, depending on how much time is left before the decision is made. Intuitively, at a early time point, since the observation will be combined with a lot more information yet to come, we are more willing to take any kind of information, and hence it makes sense to maximize mutual information. On the other hand, as the deadline of decision making approaches, the system becomes more "picky", and demands only on the information that helps the receiver to lock in to one particular message. Thus, the control signal should be optimized accordingly.

To put this intuition to test, we restrict our attention to the family of Renyi entropy. Renyi entropy of order α of a given distribution P over an alphabet $\mathcal{X}$ is defined as

$$H_\alpha(P) = \frac{1}{1-\alpha} \log \left(\sum_{x \in \mathcal{X}} P^\alpha(x) \right).$$

It is easy to verify that H_1 is the Shannon entropy, and $H_\infty(P) = -\log\left(\max_{x\in\mathcal{X}} P(x)\right)$, which is a measure of the probability of error in guessing the value of X, with distribution P, as $\hat{X} = \arg\max_x P(x)$.

Now for a general M-ary hypothesis testing problem, we consider a recursive design of the control signal l similar to that introduced in section II, except that at each time, rather than maximizing the mutual information over the effective channel, which is equivalent to minimizing the conditional Shannon entropy of the messages, we instead minimize the average Renyi α entropy , i.e., we solve the optimization problem

$$\min_l \sum_y P_Y(y) \cdot H_\alpha(P_{H|Y=y}(\cdot)) \tag{4}$$

Intuitively, for $\alpha \in [1, \infty)$, the larger α is, optimization in (4) tends more in favor of posterior distributions that are concentrated on a single entry. Smaller values of α, on the other hand, correspond to receivers that are more agnostic to what type of information is obtained. A good design should use smaller values of α at the beginning of the communication session, and increase α as time passes by. In the numerical example in Figure 4, to illustrate the point, we compare the cases where α is chosen to be fixed throughout the time $t \in [0, T]$. We observe that using $\alpha = 1$ gives better performance if T is longer, and choosing a larger α yields better error probability when T is small. This experiment confirms our intuition.

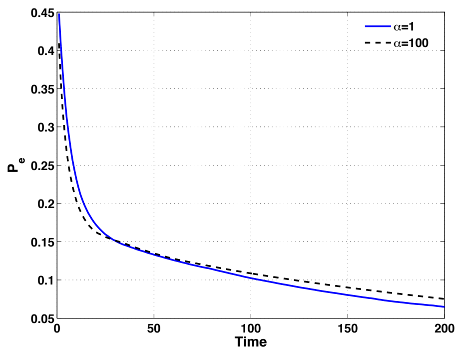

Fig. 4. Probability of detection error using control signals that minimizes the average Renyi α-entropy for different values of α

IV. Coded Transmissions and Capacity Results

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

$$C_{\mathsf{Holevo}}(n) = (1+n)\log(1+n) - n\log n \text{ bits/use} \qquad (5)$$

where n 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, (5) serves as a performance benchmark. In the regime of interests where $n \to 0$, it is useful to approximate (5) as

$$C_{\mathsf{Holevo}}(n) = n\log\frac{1}{n} + n + o(n). \qquad (6)$$

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 $n \to 0$, the optimal input distribution to the optical channel with direct detection is on-off-keying, with

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

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

$$C_{\mathsf{DD}}(n) = n\log\frac{1}{n} - n\log\log\frac{1}{n} + O(n) \qquad (7)$$

Comparing (6) and (7), we observe that the two capacities have the same leading term. This means as $n \to 0$, the optimal photon efficiency of $\log(1/n)$ 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 n that satisfies $C(n)/n = 10$ bits/photon in both cases, and get $n_{\mathsf{Holevo}} \approx 0.0027$ and $n_{\mathsf{DD}} \approx 0.00010$. The resulting capacities also differ by more than 1 order of magnitude. This example says that although (6) and (7) have the same limit as $n \to 0$, 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 $n\log\frac{1}{n}$. 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.

The problem of coded transmission and finding the maximum information rate that can be conveyed through an optical channel with a coherent processing receiver is in fact easier than that of hypothesis testing, even though there are exponentially many, possible messages. One first observation is when communicating with a long block of N symbols, there is no issue of a pressing deadline of decision for 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.

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 4.

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

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

$$\frac{C(n)}{n} \leq \log\frac{1}{n} - \log\log\frac{1}{n} + O(1) \tag{8}$$

This says that essentially the achievable photon efficiency with coherent receivers is not significantly different from that of direct detection receivers.

This theorem is a useful step in understanding more general coherent receivers, with joint processing of multiple symbols. Here, we describe a general optical receiver with *classical processing* as follows. Let the codeword transmitted by a sequence of coherent states $|X_1\rangle|X_2\rangle\ldots|X_N\rangle$, where each X_i is drawn from a finite alphabet. The receiver forms $|Y_1\rangle, |Y_2\rangle, \ldots, |Y_M\rangle$ and uses photon counter to observe them separately. Each Y_j is formed by a linear passive mixing of the X_i's and an arbitrary control signal l_j: $Y_j = \sum_{i=1}^N \alpha_{ij} X_i + l_j$, where α_{ij} satisfy $\sum_j |\alpha_{ij}|^2 \leq 1, \forall i$ and $\sum_i |\alpha_{ij}|^2 \leq 1, \forall j$, which ensure the physical constraint of energy conservation and the fact that duplication or noiseless amplification of coherent states are not possible. The mixing parameters and the control signals can be decided sequentially based on the earlier observations. Following the spirit of Theorem 4, we state the following conjecture.

Conjecture 5: The achievable photon efficiency by an optical receiver with classical processing satisfies (8).

While this conjecture 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.

[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

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