CCIT Report #396 September 2002

On the Capacity-Achieving Distribution of the Discrete-Time Non-Coherent and Partially-Coherent AWGN Channels

Michael Katz and Shlomo Shamai (Shitz)*

September 19, 2002

Abstract

We investigate the ultimate communication limits over rapid phase varying channels and consider the capacity of a discrete-time non-coherent additive white Gaussian noise channel under the average power constraint. We obtain necessary and sufficient conditions for the optimal, capacity achieving, input distribution and show that the optimal distribution is discrete and possesses an infinite number of mass points. Using this characterization of the capacity achieving distribution we compute a tight lower bound on the capacity of the channel based on examining sub optimal input distributions. In addition, we provide some easily computable lower and upper bounds on the channel capacity. Finally, we extend some of these results to the partially coherent channel, where it is assumed that a phase locked loop (PLL) is used to track the carrier phase at the receiver, and that an ideal interleaver and de-interleaver are employed - rendering the Tikhonov distributed residual phase errors statistically independent from one symbol interval to another.

Keywords: non-coherent channel, partially-coherent channel, channel capacity, capacity achieving distribution, discrete, Kuhn-Tucker conditions.

1 Introduction

Non-coherent channels emerge in communication systems where it is difficult or even impossible to provide a carrier phase reference at the receiver. In bandpass transmission this situation arises when tracking of the carrier phase is not possible due to phase noise, frequency hopping and various other reasons. In optical communication systems it is quite often not possible to transfer information through the phase of the optical carrier [16]. That is why non-coherent channels play an important role there as well [38].

Results concerning the behavior of the capacity of the non-coherent channel for asymptotically low and high signal to noise ratios (SNR) under the average power constraint date back to Blachman [2] (see also the work of Jelonek [17]), who considered the quadrature additive Gaussian channel

*M. Katz and S. Shamai (Shitz) are with the Department of Electrical Engineering, Thechnion-Israel Institute of Technology, Technion city, Haifa 32000, Israel. Email: michaelk@techunix.technion.ac.il, sshlomo@ee.technion.ac.il. The work was supported by the Fund of Promotion of Research at the Technion.

where any combination of amplitude and phase modulations are employed under either average power or peak power constraints. It was shown in [2] as well that choosing the input amplitude to be positive normal distributed is capacity achieving for asymptotically high SNR. In [38] the non-coherent channel under a peak amplitude constraint was considered, where it was shown that a discrete distribution with finite cardinality is capacity achieving. Furthermore, a discrete capacity achieving distribution was conjectured also for the case of an average power constraint.

In recent years attention has been devoted to the non-coherent channel mainly towards improving non-coherent detection methods [10], [11], [7]. In many cases these methods were shown to achieve a performance approaching that of ideal coherent detection, when one assumed that the unknown (random) carrier phase at the receiver is constant over a large number of consecutive symbols [5], [25], [29], [30], [31] (and references therein).

The capacity of the blockwise non-coherent additive white Gaussian noise (NCAWGN) channel where the input symbols are drawn from an M-ary phase shift keying (MPSK) alphabet has been addressed in [29]. For a sequence of L transmitted symbols, over which the unknown carrier phase is assumed constant, it was shown that capacity is achieved by uniformly and identically distributed MPSK symbols. Moreover, it was shown that for large L the coherent capacity itself is approached. A more general result was obtained in [6] where the capacity of the blockwise NCAWGN channel was considered without imposing any constraints on the input alphabet. It was shown that the capacity achieving distribution in this setting is not Gaussian (as in the coherent case) and is composed of zero mean and uncorrelated components with uniformly distributed phases. As in the MPSK case it was shown that as the length of the block of symbols tends to infinity, the capacity achieving distribution tends to a Gaussian one with independent and identically distributed zero mean components, and the capacity approaches that of a coherent channel. The channel discussed in this paper can in fact be viewed as a specialization of the channel discussed in [6] to a unit block length. A subsequent recent effort [28] extends some of the results of the present work [19] to block non-coherent channels.

It should be noted that the support of a capacity achieving distribution for a continuous input channel need not necessarily include a continuous component. One such example is the additive scalar channel where the noise satisfies certain constraints [35] and the signal is subjected to a peak-power constraint. See also [34] where this result was extended for the Gaussian case to the quadrature (two-dimensional) channel. The same result holds also for some amplitude-limited additive noise channels where the noise has a piecewise constant probability density [27]. See also [12] for some more examples obeying this rule. A common feature of all the above cases is the peak amplitude constraint imposed on the input. However, this behavior of the capacity achieving distribution was demonstrated also in the case of the discrete-time memoryless Rayleigh fading channel [1] under the average power constraint. See also [13] where this result was extended to the discrete-time memoryless Rician fading channel when considered under the second and fourth moment constraints. In [28] the channels treated in [34] and [1] are considered in the multidimensional case, where results of similar flavor are again obtained regarding the topological properties of the capacity achieving distributions. The capacity achieving input vector for the block non-coherent channel was shown to be a product of an isotropic unit vector and a discrete positive norm. Finally, in [9] the capacity achieving distributions of non Gaussian additive noise channels under the average power constraint was considered, where it was shown that when the density of the noise has a tail which decays slower (respectively faster) than the Gaussian, then the capacity achieving distribution has bounded (respectively unbounded) support.

In this paper, using techniques similar to those found in [35] and [1], we prove that the capacity achieving distribution of the discrete-time NCAWGN channel under the average power constraint is discrete with an infinite number of mass points. Some upper and lower bounds on channel capacity are also obtained, where a tight upper bound is obtained for asymptotically high SNR.

As opposed to the non-coherent case, there are some communication settings where estimation of the carrier phase at the receiver is a viable option. Usually, some phase tracking device is employed such as a phase locked loop (PLL), in which case the residual phase error at the receiver can be assumed to possess a Tikhonov probability density. The capacity of the scalar partially-coherent AWGN (PCAWGN) channel was considered in [14]-[15] (see also [3]), where it was shown that the capacity achieving distribution is circularly symmetric, and not Gaussian (as in the coherent case). However, a more detailed characterization of the capacity achieving distribution was not found. Assuming that ideal interleaver and de-interleaver are employed in the transmitter and receiver, respectively, causing the phase errors to be statistically independent, one can treat the PCAWGN channel within the framework of memoryless channels. Following this approach, we show that the capacity achieving distribution for the PCAWGN channel has a discrete amplitude and uniform and independent phase (DAUIP).

The paper is organized as follows. In section 2 we formulate the problem and give necessary and sufficient conditions for the capacity achieving distribution. In section 3 we establish our main result regarding the discreteness and infiniteness of the capacity achieving distribution. Section 4 presents some numerical results concerning the capacity and the capacity achieving distribution. In this section we also discuss some lower and upper bounds on the capacity. Section 5 extends some of the results of section 3 to the partially coherent case. A summary concludes the paper where some topics for further research are proposed. Detailed proofs are given in Appendices A - E.

2 Problem Formulation and Preliminary Results

We consider the complex envelope representation of a discrete-time NCAWGN memoryless channel,

$$Y_i = X_i e^{j\theta_i} + N_i \tag{1}$$

where X_i is the complex valued channel input, Y_i is the channel output, N_i are independent circularly complex Gaussian random variables whose real and imaginary parts are independent of each other and are each zero-mean with variance σ^2 (i.e. $\mathbb{E}\left[\mathfrak{Re}(N_i)^2\right] = \mathbb{E}\left[\mathfrak{Im}(N_i)^2\right] = \sigma^2$ and $\mathbb{E}\left[|N_i|^2\right] = 2\sigma^2$) and finally, θ_i are independent random phases each distributed uniformly in the interval $[-\pi,\, \pi)$. The discrete time is designated by the index i, and we assume that the sequences θ_i and N_i are sequences of independent and identically distributed (i.i.d.) random variables. This channel is equivalent [18] (in the sense of capacity) to the envelope detected complex Gaussian noise channel which emerges in the modelling of the optic fiber channel [38], and represented by $Y_i = |X_i + N_i|$, where X_i, Y_i and N_i are as defined above.

The input symbols X_i are average power constrained according to

$$\mathbb{E}\left[|X_i|^2\right] \le P. \tag{2}$$

Since the channel is memoryless, the capacity achieving inputs X_i are i.i.d., and for this reason henceforth we consider the channel (1) with the time index i suppressed. We work with polar coordinates and denote the channel input by $X = x_1 + jx_2 = re^{j\varphi}$ and channel output by $Y = y_1 + jy_2 = Re^{j\psi}$. The output PDF conditioned on the input and on θ is given by,

$$f_{R,\psi\,|\,r,\varphi,\theta}(R,\, \psi\,|\,r,\, \varphi,\, \theta) = \frac{R}{2\pi\sigma^2} \exp\left\{-\frac{R^2 + r^2 - 2Rr\cos(\varphi + \theta - \psi)}{2\sigma^2}\right\}. \tag{3}$$

The output density conditioned on the input alone is obtained by averaging over the uniform distributed phase θ

$$\begin{aligned} f_{R,\psi\,|\,r,\varphi}(R,\, \psi\,|\,r,\, \varphi) &= \int_{-\pi}^{\pi} f_{R,\psi\,|\,r,\varphi,\theta}(R,\, \psi\,|\,r,\, \varphi,\, \theta)\, \frac{d\theta}{2\pi} \\ &= \frac{R}{2\pi\sigma^2} \exp\left\{-\frac{R^2 + r^2}{2\sigma^2}\right\} I_0\left(\frac{rR}{\sigma^2}\right). \end{aligned} \tag{4}$$

Note, that the output conditional PDF is independent of the input phase φ. Strictly speaking, the capacity of the channel (1) subject to the power constraint (2) is given by

$$\sup_{\substack{F_{r,\,\varphi} \\ \mathbb{E}[r^2] \le P}} \; I(R,\, \psi\,;\, r,\, \varphi\, :\, F_{r,\,\varphi}) \tag{5}$$

where $F_{r,\varphi}(r,\, \varphi)$[1] is the joint CDF of the random variables r and φ. However, since the input phase is completely wiped out by the channel (see (4)), it is easily shown [18] that the mutual information in (5) is equal to the mutual information between the input amplitude r and the output amplitude R, $I(R\,;\, r)$. In other words, the distribution of the input phase can be chosen arbitrarily (with no effect on channel capacity) and the capacity is thus given by

$$\sup_{\substack{F_r \\ \mathbb{E}[r^2]\le P}} \; I(R\,;\, r \,:\, F_r) \tag{6}$$

where F_r stands for the CDF of the input amplitude r, and the supremization is performed over all such CDFs obeying the power constraint. In this case the output PDF conditioned on the input is given by

$$\begin{aligned} f_{R\,|\,r}(R\,|\,r) &= \int_{-\pi}^{\pi} f_{R,\,\psi\,|\,r}(R,\, \psi\,|\,r)\, d\psi \\ &= \frac{R}{\sigma^2} \exp\left\{-\frac{R^2+r^2}{2\sigma^2}\right\} I_0\left(\frac{rR}{\sigma^2}\right) \end{aligned} \tag{7}$$

where we have used the fact that $f_{R,\,\psi\,|\,r}(R,\, \psi\,|\,r) = f_{R,\,\psi\,|\,r,\,\varphi}(R,\, \psi\,|\,r,\, \varphi)$.

Given the channel description (7) and the power constraint (2), which now assumes the form $\mathbb{E}r^2 \le P$, we follow the approach used in [35], [34] and [1], and denote by $\mathcal{F}$ the set of all probability distribution functions. Let $\Omega \subset \mathcal{F}$ be the set of all input CDFs $F(r)$ defined for $r \ge 0$ and satisfying the power constraint, i.e.

$$\begin{aligned} F(0^-) &= 0 \\ \int_0^\infty r^2\, dF(r) &\le P. \end{aligned} \tag{8}$$

Note that this definition of the set Ω includes discrete input distributions. Now, for every $F \in \Omega$ define the mutual information functional

$$I(R\,;\, r \,:\, F) = \int_0^\infty \int_0^\infty f(R\,|\,r) \log \frac{f(R\,|\,r)}{f(R \,:\, F)}\, dR\, dF(r) \tag{9}$$

where $f(R \,:\, F)$ is the channel output PDF induced by the input distribution F

$$f(R \,:\, F) = \int_0^\infty f(R\,|\,r)\, dF(r). \tag{10}$$

The capacity is then given by

$$C = \sup_{F\in\Omega} I(R\,;\, r \,:\, F). \tag{11}$$

[1]We use the notations F_x, $F_x(\alpha)$ and $F(x)$ interchangeably throughout the paper to signify the cumulative distribution function of the random variable x.

Existence and uniqueness of a solution to (11) follow from an optimization theorem based on [24, Sec. 5.10] and proved in [1].

Theorem 1 [1] *If J is a real-valued, weak* continuous functional on a weak* compact set $\Omega \subseteq X^*$, then J* achieves *its maximum on Ω. If furthermore Ω is convex, and J is strictly concave, then the maximum*

$$C = \max_{F \in \Omega} J(F)$$

is achieved by a unique F_0 in Ω.

The fact that the set Ω is convex and weak* compact was shown in [1]. In Appendix A it is shown that the mutual information functional in (9) is continuous and strictly concave. Therefore, by theorem 1 we have shown existence and uniqueness. With that in hand it is possible to use the following optimization theorem[2] (see [1] and [24, Sec. 8.3]).

Theorem 2 [24] *Let X be a linear vector space, Z a normed space, $\mathcal{F}$ a convex subspace of X, and P the positive cone in Z. Assume that P contains an interior point. Let f be a real valued concave functional on $\mathcal{F}$ and g a convex mapping from $\mathcal{F}$ to Z. Assume the existence of a point $F_1 \in \mathcal{F}$ for which $g(F_1) < 0$. Let*

$$C = \sup_{\substack{F \in \mathcal{F} \\ g(F) \leq 0}} f(F) \tag{12}$$

and assume C is finite. Then there is an element $z_0^* \geq 0$ in Z^* such that

$$C = \sup_{F \in \mathcal{F}} \{ f(F) - z_0^* \left(g(F) \right) > \} . \tag{13}$$

Furthermore, if the supremum is achieved in (12) *at F_0, it is achieved by F_0 in* (13) *and*

$$z_0^* \left(g(F_0) \right) = 0. \tag{14}$$

As in [1] we apply the theorem to our problem where the functional f is the mutual information functional (9) and $\mathcal{F}$ is the set of distribution functions of nonnegative random variables. It is easily shown that $\mathcal{F}$ is convex. Moreover, in Appendix A it was shown that $I(R; r : F)$ is concave over $\mathcal{F}$. Define the mapping $g(F) : \mathcal{F} \to \mathbb{R}$

$$g(F) = \int_0^\infty r^2 dF(r) - P$$

[2]The expression $x^*(x)$ denotes the value of the functional x^* belonging to the dual space X^* evaluated for some element x belonging to the vector space X.

where $P > 0$. The mapping $g(F)$ is linear and hence convex. The positive cone in $\mathbb{R}$ has an interior point, and if F_1 is the unit step, $g(F_1) < 0$. By theorem 2, there exists $\gamma \geq 0$ such that

$$C = \sup_{F \in \mathcal{F}} \{I(R; r : F) - \gamma g(F)\}. \tag{15}$$

Moreover, since capacity is achieved for some F_0, the supremum is achieved in (15) by F_0 and

$$\gamma g(F_0) = 0. \tag{16}$$

As in [1] and in [34], we follow the footsteps of Smith [35] and use the following optimization theorem, which relies on the notion of *weak differentiability* [35], in order to obtain necessary and sufficient conditions for F_0 to achieve the supremum in (15).

Theorem 3 [35] *Assume a weakly differentiable functional f on a convex set $\mathcal{F}$ achieves its maximum.*

1) *If f achieves its maximum at F_0 then $f'_{F_0}(F) \leq 0$ for all $F \in \mathcal{F}$.*

2) *If f is concave, then $f'_{F_0}(F) \leq 0$ for all $F \in \mathcal{F}$ implies that f achieves its maximum at F_0.*

In Appendix C it is shown that the functional $I(R; r : F) - \gamma g(F)$ is weakly differentiable in $\mathcal{F}$. Furthermore, since $g(F)$ is linear and $I(R; r : F)$ is concave, $I(R; r : F) - \gamma g(F)$ is also concave. Thus, we apply theorem 3 and obtain necessary and sufficient conditions for F_0 to achieve the supremum in (15), namely

$$I'_{F_0}(R; r : F) - \gamma g'_{F_0}(F) \quad \leq \quad 0, \qquad \forall F \in \mathcal{F}$$

which by the results of Appendix C can be written equivalently as

$$\begin{aligned} \int_0^\infty \left[i(r\,; F_0) - \gamma r^2\right] dF(r) \quad &\leq \quad I(R; r : F_0) - \gamma \int_0^\infty r^2\, dF_0(r) \\ &= \quad C - \gamma P, \end{aligned} \tag{17}$$

where

$$i(r\,; F_0) = \int_0^\infty f(R\,|\,r) \log \frac{f(R\,|\,r)}{f(R : F_0)}\, dR$$

and where the last assertion in (17) follows from the fact that the Lagrange multiplier assumes the value of zero if the power constraint is not met with equality.

The following theorem (which is proved in [18]) formulates the necessary and sufficient conditions for the optimal distribution in a more convenient form.

Theorem 4 *Let E_0 denote the set of increase points of a distribution F_0. Then*

$$\int_0^\infty \left[i(r\,;\,F_0) - \gamma r^2\right] dF(r) \leq C - \gamma P \tag{18}$$

for all $F \in \mathcal{F}$ if and only if

$$\begin{aligned} i(r\,;\,F_0) &\leq C + \gamma(r^2 - P), \qquad \forall r \geq 0 & (19)\\ i(r\,;\,F_0) &= C + \gamma(r^2 - P), \qquad \forall r \in E_0. & (20)\end{aligned}$$

We refer to (19) and (20) as the Kuhn-Tucker conditions (KTC).

3 Main Result

In this section we present our main result concerning the capacity achieving distribution of the NCAWGN channel, namely that it is discrete and possesses an infinite number of mass points. We do not prove that discreteness and infiniteness hold directly. Rather, we show that all other options are impossible. The main result is thus stated in the following theorem.

Theorem 5 *The optimal input distribution $F_0(r)$ achieving the supremum in* (11) *is discrete with an* infinite *set of mass points*[3]*, but with only a* finite *number of mass points over any* bounded *interval.*

Proof: It is obvious that the optimal input must have one of the following properties.

1) Its support contains an infinite number of mass points on some bounded interval.

2) It is discrete, infinite, but with only a finite number of mass points over any bounded interval.

3) It is discrete and finite.

Following Smith [35] and Shamai and Bar-David [34] we use the KTC to prove that the first and third cases are not possible. Assume that case 1 holds. Since any closed interval is a compact set, it follows by the Bolzano Weirstrass theorem that this interval contains an accumulation point of the set of mass points. Now consider the function

$$g(z) = i(z\,;\,F_0) - C - \gamma(z^2 - P) \tag{21}$$

for any $z \in \mathbb{C}$ satisfying $\mathfrak{Re}(z) > 0$. The following proposition is proved in Appendix B.

[3] A "mass point" or "increase point" x_0 of a probability distribution function $F_X(\alpha)$ satisfies $F_X(x_0 + \epsilon) - F_X(x_0 - \epsilon) > 0, \;\; \forall \epsilon > 0$.

Proposition *The function $g(z)$ is analytic for $z \in \mathbb{C}$ satisfying $\mathfrak{Re}(z) > 0$.*

By the KTC (20) it follows that $g(z)$ equals zero for $z \in E_0$. Thus, we have an analytic function over a domain $D = \{z \in \mathbb{C} : \mathfrak{Re}(z) > 0\}$ which equals zero over an infinite set of points having an accumulation point in D. By the identity theorem [4, Sec. 2-8] the function $g(z)$ equals zero throughout the domain D. This implies, in particular, that the KTC hold with equality for all $r > 0$, or

$$\int_0^\infty f(R\,|\,r) \log \frac{f(R\,|\,r)}{f(R\,:\,F_0)}\, dR = C + \gamma(r^2 - P), \quad \forall r > 0. \tag{22}$$

After substituting (7) in (22) and performing the change of variables $R = \sigma^2 x$ one obtains after some manipulations

$$-\int_0^\infty x \left[e^{-\frac{1}{2}\sigma^2 x^2} \log \frac{\tilde{f}(x\,:\,F_0)}{\sigma^2 x} \right] J_0\left(xjr\right) dx =$$
$$= -\int_0^\infty x e^{-\frac{1}{2}\sigma^2 x^2} J_0\left(xjr\right) \log J_0\left(xjr\right) dx + \left(\alpha + \beta r^2\right) e^{\frac{r^2}{2\sigma^2}} \tag{23}$$

where we have denoted the function $\tilde{f}(x\,:\,F_0) = f(\sigma^2 x\,:\,F_0)$ and defined the two constants $\alpha = \frac{1}{\sigma^2}(1 + C + \log \sigma^2 -$ and $\beta = \frac{1}{\sigma^2}\left(\gamma + \frac{1}{\sigma^2}\right)$ and where we have used the fact that $I_0(z) = J_0(jz)$.

We interpret the term on the left hand side (LHS) of (23) as an application of the Hankel transform (see [32, p. 9-3], [36, Sec. 5.3] and [40]). Referring to [36], the Hankel transform pair, namely

$$F_\nu(s) \triangleq \mathcal{H}_\nu\left\{f(t)\ ;\ t \to s\right\} = \int_0^\infty t f(t) J_\nu(st)\, dt$$
$$f(t) = \mathcal{H}_\nu^{-1}\{F_\nu(s)\ ;\ s \to t\} = \int_0^\infty s F_\nu(s) J_\nu(st)\, ds \tag{24}$$

where $t \to s$ designates transformation from the t domain to the s domain and vice verse, is well defined for $\nu > -\frac{1}{2}$ and for a function $f(t)$ for which $\sqrt{t} f(t)$ is piecewise continuous and absolutely integrable on the positive real line. In Appendix D it is verified that the term in square brackets on the LHS of (23) satisfies the above conditions for the Hankel transform, implying that the Hankel transform in (23) exists, and it's inverse is given according to (24). Notice, however, that the Hankel transform in (23) is in fact evaluated along the imaginary axis.

Next, we identify the second term on the right hand side (RHS) of (23) as a Hankel transform of some function, also evaluated along the imaginary axis. It is easily shown [18] that this term can in fact be expressed as

$$\left(\alpha+\beta r^2\right) e^{\frac{r^2}{2\sigma^2}} = \left(\alpha-\beta t^2\right) e^{-\frac{t^2}{2\sigma^2}}\Big|_{t=jr} =$$

$$= \quad \mathcal{H}_0\left\{\alpha\sigma^2 e^{-\frac{1}{2}\sigma^2x^2} - 2\beta\sigma^4\left(1-\frac{1}{2}\sigma^2x^2\right)e^{-\frac{1}{2}\sigma^2x^2} \quad ; \quad x\to t\right\}\Bigg|_{t=jr}. \tag{25}$$

Equation (23) can now be written as

$$-\mathcal{H}_0\left\{e^{-\frac{1}{2}\sigma^2x^2}\log\frac{\tilde{f}(x\,:\,F_0)}{\sigma^2 x} \quad ; \quad x\to t\right\}\Bigg|_{t=jr} =$$

$$-\int_0^\infty xe^{-\frac{1}{2}\sigma^2x^2} J_0\left(xt\right)\log J_0\left(xt\right)dx\Big|_{t=jr} +$$

$$+\mathcal{H}_0\left\{\alpha\sigma^2 e^{-\frac{1}{2}\sigma^2x^2} - 2\beta\sigma^4\left(1-\frac{1}{2}\sigma^2x^2\right)e^{-\frac{1}{2}\sigma^2x^2} \quad ; \quad x\to t\right\}\Bigg|_{t=jr}. \tag{26}$$

Since all the terms in (26) are analytic functions of t, we can conclude (by the identity theorem) that equation (26) holds for all $t\geq 0$. Taking the inverse Hankel transform on both sides of (26) and rearranging we arrive at

$$\tilde{f}(x\,:\,F_0) =$$

$$= Axe^{-\beta\sigma^6x^2}\exp\left\{e^{\frac{1}{2}\sigma^2x^2}\int_0^\infty\!\!\int_0^\infty tue^{-\frac{1}{2}\sigma^2u^2}J_0\left(ut\right)\log\left\{J_0\left(ut\right)\right\}J_0\left(xt\right)\,du\,dt\right\}, \tag{27}$$

where $A=\sigma^2 e^{2\beta\sigma^4-\alpha\sigma^2}$, which is a scaled version of the channel output PDF.

To invalidate the existence of a valid output PDF it is enough to show that this scaled version is not a real function of x. We do this by showing that the imaginary part of $\tilde{f}(x\,:\,F_0)$ is not identically zero. Assume that it is identically zero. This implies that the phase of (27) is identically zero for all x (since the PDF must be nonnegative). The phase of $\tilde{f}(x\,:\,F_0)$ is given by

$$\arg\tilde{f}(x\,:\,F_0) = \mathfrak{Im}\left\{e^{\frac{1}{2}\sigma^2x^2}\int_0^\infty\!\!\int_0^\infty tue^{-\frac{1}{2}\sigma^2u^2}J_0\left(ut\right)\log\left\{J_0\left(ut\right)\right\}J_0\left(xt\right)\,du\,dt\right\}.$$

The phase is identically zero if and only if

$$\int_0^\infty\!\!\int_0^\infty tue^{-\frac{1}{2}\sigma^2u^2}J_0\left(ut\right)\arg\left\{J_0\left(ut\right)\right\}J_0\left(xt\right)\,du\,dt$$

$$= \int_0^\infty zJ_0\left(z\right)\arg\left\{J_0\left(z\right)\right\}\left[\int_0^\infty\frac{1}{t}e^{-\frac{1}{2}\sigma^2\frac{z^2}{t^2}}J_0\left(xt\right)\,dt\right]dz \equiv 0.$$

Since the last term is identically zero by our assumption, its derivative with respect to x is also identically zero. Integrating that derivative along the positive real axis must also then yield zero. however,

$$\begin{aligned}
&\int_0^\infty \frac{d}{dx}\left\{\int_0^\infty zJ_0(z)\arg\{J_0(z)\}\left[\int_0^\infty \frac{1}{t}e^{-\frac{1}{2}\sigma^2\frac{z^2}{t^2}}J_0(xt)\,dt\right]dz\right\}dx \\
= &-\int_0^\infty zJ_0(z)\arg\{J_0(z)\}\left[\int_0^\infty e^{-\frac{1}{2}\sigma^2\frac{z^2}{t^2}}\int_0^\infty J_1(xt)\,dx\,dt\right]dz \\
= &\int_0^\infty\int_0^\infty -zJ_0(z)\arg\{J_0(z)\}\,e^{-\frac{1}{2}\sigma^2\frac{z^2}{t^2}}\frac{1}{t}\,dt\,dz > 0
\end{aligned}$$

where the last inequality holds because of the non-negativity of the integrand. This contradicts our assumption that the function $\tilde{f}(x : F_0)$ is real, and invalidates the output PDF. Thus the assumption of an infinite number of mass points over some bounded interval is false, proving that the optimal distribution is discrete, either finite or infinite, with a finite number of mass points over any bounded interval.

We now proceed to the last part of the proof where we wish to prove that the discrete optimal distribution is infinite[4]. Assume, therefore, that case 3 holds, i.e. that the optimal input distribution is discrete and finite.

Referring to (19) we begin with the necessary and sufficient conditions for the optimal input distribution, namely

$$\int_0^\infty f(R\,|\,r)\log\frac{f(R\,|\,r)}{f(R\,:\,F_0)}\,dR \le C + \gamma(r^2 - P), \quad \forall r > 0 \tag{28}$$

where equality in (28) holds for $r \in E_0$, E_0 now denoting the set of mass points of the discrete optimal input distribution $F_0(r)$. Assume that the set of mass points $\{r_i\}_{i=1}^N$ is such that $r_1 < r_2 < \cdots < r_N$. Also, let the set of corresponding probabilities be $\{p_i\}_{i=1}^N$ where $p_i = \text{Prob}\{r = r_i\}$. Now, $f(R : F_0)$ can be expressed by

$$f(R\,:\,F_0) = \sum_{i=1}^N p_i f(R\,|\,r_i). \tag{29}$$

Substituting (29) in (28) yields after some simple manipulation

$$\int_0^\infty f(R\,|\,r)\log\left[\sum_{i=1}^N p_i e^{-\frac{r_i^2}{2\sigma^2}} I_0\left(\frac{Rr_i}{\sigma^2}\right)\right]dR \;\ge\; \left(\frac{1}{2\sigma^2}-\gamma\right)r^2 - C_w(r) - C + \gamma P \tag{30}$$

[4]By "infinite distribution" we refer to a distribution with an infinite number of mass points.

where $C_w(r)$ is defined as

$$C_w(r) \quad \triangleq \quad -\int_0^\infty f(R\,|\,r)\log\frac{f(R\,|\,r)}{R}\,dR - \log\sigma^2 e. \tag{31}$$

Equation (31) is the Wyner polyphase capacity whose behavior for small and large signal to noise ratios (SNRs) was studied in [41].

We first claim that $\gamma \leq \frac{1}{2\sigma^2}$. To see why this is so consider the following chain of arguments. The parameter γ is a Lagrange multiplier. As such it represents the slope of the optimum of the object function as a function of the constraint value. In this case, the Kuhn-Tucker equation (30) is stated for the case of a fixed noise variance of $2\sigma^2$ and a variable power constraint P. Therefore, γ is actually the slope of the curve $C_\sigma(P)$ at P. Now consider the capacity of the complex Gaussian channel $C_\sigma^G(P) = \log(1+\frac{P}{2\sigma^2})$. This capacity is evidently an upper bound to the capacity of the non-coherent channel. Therefore, when P tends to infinity, the slope of $C_\sigma(P)$ must also tend to zero. If that were not the case, then we would have the non-coherent capacity curve $C_\sigma(P)$ exceed the capacity curve of the complex Gaussian channel at some P. From the convexity and monotonicity of channel capacity we can safely say that the largest slopes of both capacity curves are located at $P=0$. The slope of $C_\sigma^G(P)$ at $P=0$ is $\frac{1}{2\sigma^2}$. If $\gamma > \frac{1}{2\sigma^2}$, we could find a small enough P such that the capacity of the non-coherent channel would exceed the capacity of the complex Gaussian channel. As this is not possible, we conclude that $\gamma \leq \frac{1}{2\sigma^2}$ for all $P \geq 0$. In fact, for $P=0$ there actually holds an equality between the slopes of the capacity curves of the complex Gaussian channel and the non-coherent channel. This is because the slope of the capacity curve at $P=0$ is inversely proportional to the minimum $\frac{E_b}{N_0}$ which is the same for both channels [39].

Now, evaluate the LHS of (30)

$$\begin{aligned} \text{LHS} \quad &< \quad \int_0^\infty f(R\,|\,r)\log\left[\sum_{i=1}^N p_i e^{-\frac{r_i^2}{2\sigma^2}} I_0\left(\frac{Rr_N}{\sigma^2}\right)\right]dR \\ &= \quad \int_0^\infty f(R\,|\,r)\log I_0\left(\frac{Rr_N}{\sigma^2}\right)dR + \log\left[\sum_{i=1}^N p_i e^{-\frac{r_i^2}{2\sigma^2}}\right] \end{aligned} \tag{32}$$

where the second term in (32) is some constant independent of r. Focus on the first term which is further bounded by

$$\begin{aligned} \int_0^\infty f(R\,|\,r)\log I_0\left(\frac{Rr_N}{\sigma^2}\right)dR \quad &\leq \quad \int_0^\infty f(R\,|\,r)\log\exp\left\{\frac{Rr_N}{\sigma^2}\right\}dR \\ &= \quad \frac{r_N}{\sigma^2}\int_0^\infty f(R\,|\,r)R\,dR. \end{aligned}$$

We show that the last integral grows at most linearly with r.

$$\begin{aligned}
\int_0^{\infty} f(R\,|\,r) R\, dR &= \int_0^{r} f(R\,|\,r) R\, dR + \int_r^{\infty} f(R\,|\,r) R\, dR \\
&\leq r \int_0^{r} f(R\,|\,r)\, dR + \int_r^{\infty} f(R\,|\,r) R\, dR \\
&\leq r + \mathbb{E}_{R\,|\,r}\left\{ R I_{R>r} \right\} \\
&\leq r + \sqrt{\mathbb{E}_{R\,|\,r}\left\{R^2\right\} \mathbb{E}_{R\,|\,r}\left\{I_{R>r}^2\right\}} \qquad (33) \\
&= r + \sqrt{(r^2 + 2\sigma^2)\mathrm{Prob}\{R > r\,|\,r\}} \\
&\leq r + \sqrt{(r^2 + 2\sigma^2)\frac{r^2 + 2\sigma^2}{r^2}} \qquad (34) \\
&= 2r + \frac{2\sigma^2}{r} \qquad (35)
\end{aligned}$$

where we have used the Cauchy Schwartz inequality in (33) and the Chebysev inequality in (34). To conclude, we have shown that the LHS of (30) is upper bounded by

$$\mathrm{LHS} \;<\; \frac{2r_N}{\sigma^2} r + \frac{2r_N}{r} + \log\left[\sum_i p_i e^{-\frac{r_i^2}{2\sigma^2}}\right]. \qquad (36)$$

However, (30) must hold for all $r \geq 0$. This can not be, since the RHS of (30) grows faster than the LHS. We therefore conclude that case 3 does not hold, and we are left with the only possibility that the optimal input distribution achieving the channel capacity is discrete with infinitely many mass points, but with only a finite number of mass points over any bounded interval.

4 Numerical Results

4.1 Lower Bounds

One of the serious implications of theorem 5 is that it prevents the numerical computation of the capacity achieving distribution as one can not solve (numerically) an optimization problem with an infinite number of design variables (mass points' locations and probabilities). This situation is in contrast to the cases studied in [35], [34], [1] and [13] where the set of mass points of the optimal input distribution was shown to be finite.

Of course, for any finite set of mass points, an optimization of the mutual information over that set yields a sub-optimal input distribution and a corresponding lower bound on capacity, which can only improve with the addition of more mass points to the optimization process. We take this approach, and compute (numerically) a "tight" lower bound on capacity by optimizing over a sufficient number of mass points. By "sufficient" we mean that adding more mass points does not alter the mutual information significantly. In what follows all numerical results are obtained with $\sigma^2 = 1$ and SNR is defined as the ratio $\frac{P}{2\sigma^2}$.

Figure 1 depicts the locations of the non zero mass points belonging to the sub-optimal input distributions. It should be noted, however, that a mass point at zero is always present. Each line corresponds to some mass point, and shows the dependence of its location as function of the SNR. Notice how at low SNR the locations of the mass points further away from zero "escape" to infinity. As SNR increases the mass points closer to zero tend to fixed values (more or less equally spaced) and new mass points arrive from infinity.

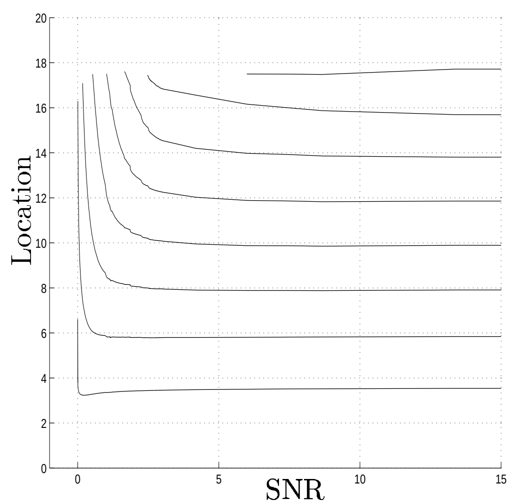

Figure 1: Non-zero mass points of sub-optimal input distributions

Figure 2 shows the probabilities of all mass points, including the zero mass point, versus SNR. Higher probabilities correspond to mass point closer to zero. Notice how the probability of all mass points (other than the zero mass point) tends to zero at low SNR. Also note that at any SNR the probabilities of mass points further away from zero tend rapidly to zero. This intuitively suggest that the mass points above a certain number do not convey a significant contribution to the mutual information.

It is observed that for very low SNR values there are effectively only two mass points. One at zero, and the other away from zero, with an appropriately small probability, so as to satisfy the average power constraint. As SNR increases, the probability of the positive mass point increases, and its location decreases. When SNR increases further, additional mass points join the input distribution allowing for more signal levels. The additional masses emerge from infinity, with appropriately scaled probabilities, so as to meet the overall power constraint, resembling the Rayleigh

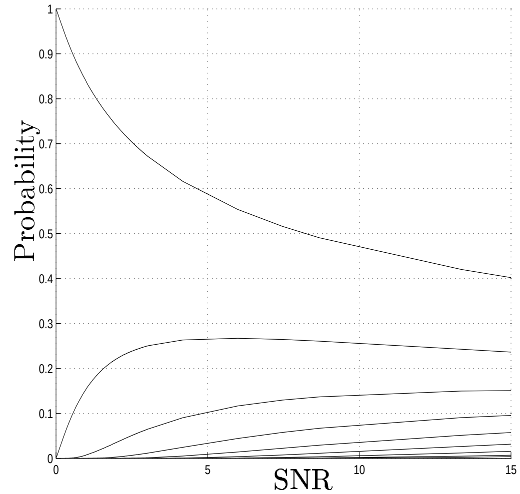

Figure 2: Probabilities of mass points of sub-optimal input distributions

fading channel treated in [1]. The "tight" lower bound on capacity obtained from evaluating the average mutual information for the sub-optimal distributions is depicted in figure 3. Note that for higher SNR the numerical optimization becomes increasingly complex because more and more significant mass points emerge in the optimal input distribution. Therefore, one must resort to other (less tighter) bounds which can be computed more easily.

Three such lower bounds are obtained by evaluating the average mutual information for three continuous distributions. The first is the geometric distribution where the input amplitude is distributed according to

$$f_r(\alpha) = \sqrt{\frac{2}{P}} \exp\left\{ -\sqrt{\frac{2}{P}}\alpha \right\}, \quad \alpha \geq 0. \tag{37}$$

For the second lower bound we choose a Rayleigh distributed input amplitude. This input distribution corresponds to a complex Gaussian input to the channel (1). The bounds are depicted in figure 3. As can be seen in figure 3, the geometric lower bound shows tighter behavior for SNR lower than 15 dB, however for higher SNR the Rayleigh lower bound displays better performance, and has a tighter trend as SNR increases further. The Rayleigh input lower bound coincides with similar results by Colavolpe and Raheli [6] and Hou, Belzer and Fischer [15]. For the third lower bound we compute the average mutual information for an input amplitude distributed according to the positive normal distribution, namely,

$$f_r(\alpha) = \sqrt{\frac{2}{\pi P}} \exp\left\{ -\frac{\alpha^2}{2P} \right\}, \quad \alpha \geq 0. \tag{38}$$

The positive normal distribution (38) was discovered to be capacity achieving for high SNR by

Blachman [2]. Recently, it was also shown by Lapidoth [21] that this lower bound is indeed asymptotically tight for large SNR. The bound is plotted in figure 3.

4.2 Upper Bounds

In order to gain some more insight as to the behavior of the capacity curve we provide some upper bounds on the capacity as well, adopting several approaches. Our first approach, which is motivated by the work of Taricco and Elia [37] on the Rayleigh fading channel, is to maximize the mutual information with respect to *output densities*. Since the set of output densities is a superset of the set of output densities induced by legitimate input distributions, by maximizing over the set of output densities one obtains an upper bound on capacity.

First note that the channel model (1) can be equivalently expressed as $Y = re^{j\theta} + N$ where r is the nonnegative channel input, and N, θ and Y are as defined in section 2. This is because we can consider only nonnegative real inputs due to the complete phase loss. In this case, the mutual information between the input r and output Y can be expressed as

$$I(r\,;Y) = I(r,\,\theta\,;Y) - I(\theta\,;Y\,|\,r). \tag{39}$$

Now, we would like to maximize the first term on the RHS of (39), over all *output densities*, while constraining the second term on the RHS of (39) to a fixed value, Λ, and then to maximize the result over Λ to obtain an upper bound on the channel capacity.

Denoting $Y = Re^{j\psi}$ and $f_R(R)$ the density of the output amplitude, we have that the first term on the RHS of (39) is given by

$$\begin{aligned} I(r,\,\theta\,;Y) &= H(Y) - H(Y\,|\,r,\,\theta) \\ &= -\int_0^\infty f_R(R)\log\frac{f_R(R)}{R}\,dR - \log\sigma^2 e \end{aligned} \tag{40}$$

where (40) is justified by the uniform output phase which is also independent of the output amplitude. Further, denoting by $F_r(r)$ the CDF of the input r, the second term on the RHS of (39) can be expressed by

$$\begin{aligned} I(\theta\,;Y\,|\,r) &= \int_0^\infty I(\theta\,;Y\,|\,r=a)\,dF_r(a) \\ &= \int_0^\infty C_w(a)\,dF_r(a) \end{aligned} \tag{41}$$

where $C_w(a)$ is as defined in (31).

Now, substitute (40) and (41) into (39) and consider the following optimization problem the solution of which when optimized over Λ is none other than the channel capacity.

$$\max_{\substack{F_r \\ \int_0^\infty dF_r(r)=1 \\ \int_0^\infty r^2\, dF_r(r)\le P \\ \int_0^\infty C_w(r)\, dF_r(r)\le \Lambda}} \left\{-\int_0^\infty f_R(R)\log\frac{f_R(R)}{R}\, dR - \log\sigma^2 e - \int_0^\infty C_w(r)\, dF_r(r)\right\}. \tag{42}$$

Note that by replacing the function $C_w(r)$ in (42) with a function $\alpha(r)$ which satisfies $\alpha(r)\le C_w(r),\ \ r\ge 0$ we obtain an optimization problem the solution of which upper bounds channel capacity. This is because the new object function would be greater than or equal to the old one, and because by doing so the constraints are relaxed. We select

$$\alpha(r) = \int_0^\infty f(R\,|\,r)\left\{\log R - \frac{1}{2}\left[\log 2\sigma^2 - C_E\right]\right\} dR \tag{43}$$

where C_E is Euler's constant. It is possible to show that the condition $\alpha(r)\le C_w(r),\ \ r\ge 0$ holds (for details see [18]). Notice that the third term of the object function in (42) (which appears also in the constraints) when $C_w(r)$ is replaced with $\alpha(r)$ assumes the form

$$\int_0^\infty \alpha(r)\, dF_r(r) \;=\; \int_0^\infty \left\{\log R - \frac{1}{2}\left[\log 2\sigma^2 - C_E\right]\right\} f_R(R)\, dR.$$

After substituting the last term into (42) we arrive at the following optimization problem the solution of which upper bounds channel capacity

$$\max_{\substack{f_R \\ \int_0^\infty f_R(R)\, dR=1 \\ \int_0^\infty R^2 f_R(R)\, dR\le P+2\sigma^2 \\ \int_0^\infty \log R f_R(R)\, dR\le \Lambda+\frac{1}{2}[\log 2\sigma^2 - C_E]}} \left\{-\int_0^\infty f_R(R)\log f_R(R)\, dR + \frac{1}{2}\left[\log\frac{2}{\sigma^2 e^2} - C_E\right]\right\}. \tag{44}$$

This problem can be solved (for details see [18]) for any fixed Λ using the calculus of variations as was done in [37]. The solution to this problem is given by the following pair of equations

$$\begin{aligned} C_{ub}(\mu) &= \log\Gamma(\mu) - \mu\left(\frac{\Gamma'(\mu)}{\Gamma(\mu)} - 1\right) + \Lambda - C_E - 1 \\ \Lambda &= \frac{1}{2}\log\left(1+\frac{P}{2\sigma^2}\right) + \frac{1}{2}\left(\frac{\Gamma'(\mu)}{\Gamma(\mu)} - \log\mu\right) + \frac{1}{2}C_E. \end{aligned} \tag{45}$$

It can be verified that optimizing the last result over Λ yields the upper bound

$$\begin{aligned} C_{ub} &= \frac{1}{2}\log\left(1+\frac{P}{2\sigma^2}\right) + \frac{1}{2}\left[\log\frac{2\pi}{e} - C_E\right] \\ &\simeq \frac{1}{2}\log\left(1+\frac{P}{2\sigma^2}\right) + 0.13. \end{aligned} \tag{46}$$

This upper bound is plotted in figure 3 as the mutual information upper bound. Although it is not at all tight, it is however very easily computed and has a closed form analytic expression for any SNR.

Recently, it was shown that the capacity of the NCAWGN channel can be achieved by input distributions that escape to infinity [21], [22]. This means that the behavior of the capacity for high SNR is unchanged if we assume that the input is constrained to be bounded away arbitrarily far from zero. Under this assumption we can select $\tilde{\alpha}(r) = \alpha(r) + \frac{1}{2}\left(\log\frac{4\pi}{e} - C_E\right)$. Although the condition $\tilde{\alpha}(r) \leq C_w(r),\ \ r \geq 0$ no longer holds, the functions $\tilde{\alpha}(r)$ and $C_w(r)$ have the same asymptotic behavior [18]. Therefore, for every $\epsilon > 0$ we can find a sufficiently large r_{min} such that $\tilde{\alpha}(r) - \epsilon \leq C_w(r),\ \ r \geq r_{min}$. This means that for every $\epsilon > 0$ choosing $\tilde{\alpha}(r) - \epsilon$ in place of $C_w(r)$ in (42) yields an optimization problem the solution of which upper bounds capacity for asymptotically high SNR. Since the capacity is a continuous function of the constraints, we can take the limit as $\epsilon \to 0$. Performing this substitution in (42) and solving once again yields the following bound for large SNR

$$C_{ub}^{\infty} = \frac{1}{2}\log\left(1 + \frac{P}{2\sigma^2}\right) - \frac{1}{2}\log 2.$$

This result coincides with Lapidoth [21] where this bound was shown to be asymptotically tight.

Our second approach is inspired by corollary 3.4 in [8, p. 142] which states the P average power constrained capacity of a discrete memoryless channel (DMC) with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and conditional probability matrix $W(y|x)$ by means of a dual expression, namely

$$C(P) = \min_{R\in\mathcal{P}(\mathcal{Y})}\ \min_{\gamma\geq 0}\ \max_{x\in\mathcal{X}} \left[D\left(W\left(\cdot|x\right)||R(\cdot)\right) + \gamma\left(P - x^2\right)\right] \tag{47}$$

where $\mathcal{P}(\mathcal{Y})$ denotes the set of all probability measures on $\mathcal{Y}$. In particular, for any output probability distribution $R(\cdot)$ capacity is bounded from above by

$$C(P) \leq \min_{\gamma\geq 0}\ \max_{x\in\mathcal{X}} \left[D\left(W\left(\cdot|x\right)||R\left(\cdot\right)\right) + \gamma\left(P - x^2\right)\right]. \tag{48}$$

Extension of (48) to channels with continuous alphabets is possible due to Lapidoth and Moser [22] who showed that for any probability measure $Q(\cdot)$ on $\mathcal{X}$ and any probability measure $R(\cdot)$ on $\mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are any separable metric spaces, there holds

$$I(Q\,;W) \leq \int D\left(W\left(\cdot|x\right)||R(\cdot)\right)\right)\,\mathrm{d}Q(x) \tag{49}$$

where $I(Q\,;W)$ is as defined in [22]. In Appendix E this inequality is extended to the case where an average power constraint is imposed on the input implying the validity of (48) for channels with continuous alphabets. Rewriting (48) in the notations of the NCAWGN channel we get the upper bound

$$C(P) \leq \min_{\gamma \geq 0} \max_{r \geq 0} \left[\int_0^\infty f(R\,|\,r) \log \frac{f(R\,|\,r)}{f_R(R)} \, dR + \gamma \left(P - r^2 \right) \right] \tag{50}$$

where $f_R(R)$ is any output density. Selecting in (50) an output density which is induced by a geometrically distributed input as in (37) yields the bound plotted in figure 3 and denoted geometric induced output (divergence based). An even better upper bound is obtained by using a parameterized output density of the form

$$f_R(R;\,\mu) = \frac{2\mu^\mu}{\Gamma(\mu)\,(P + 2\sigma^2)^\mu} R^{2\mu-1} e^{-\frac{\mu R^2}{P+2\sigma^2}}. \tag{51}$$

Substituting this density in (50) and optimizing the RHS over μ yields the bound in figure 3 denoted parametric induced output (divergence based). Numerical computation of this upper bound reveal that the behavior of this upper bound for large SNR is $\frac{1}{2}\log(1+SNR) - \frac{1}{2}\log 2$ as in the case of the first approach. In other words this bound is asymptotically tight.

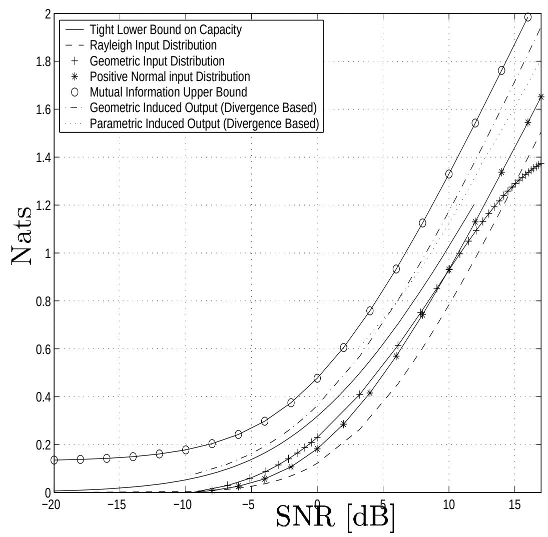

Figure 3: Upper and lower bounds on capacity

5 Extension To the Partially-Coherent Case

In this section we consider the PCAWGN channel where some phase tracking scheme is employed (such as a PLL). We assume that an ideal interleaver is present, rendering the residual phase estimation errors independent from one symbol interval to another. In this case we can consider the memoryless channel model $Y = Xe^{j\theta} + N$, where N is the circularly symmetric complex

Gaussian noise, X is the complex input, and θ is Tikhonov distributed according to $p(\theta) = \frac{e^{\rho \cos\theta}}{2\pi I_0(\rho)}$ for some positive ρ. We denote the channel output by $Y = Re^{j\psi}$ and channel input by $X = re^{j\varphi}$. The channel's conditional output density is

$$\begin{aligned} f_{R,\psi\,|\,r,\varphi}(R,\,\psi\,|\,r,\,\varphi) &= \int_{-\pi}^{\pi} f_{R,\psi\,|\,r,\varphi,\theta}(R,\,\psi\,|\,r,\,\varphi,\,\theta)\, p(\theta)\, d\theta \\ &= \frac{R}{2\pi\sigma^2} \exp\left\{-\frac{R^2+r^2}{2\sigma^2}\right\} \frac{I_0\,(\nu)}{I_0(\rho)} \end{aligned} \tag{52}$$

where $\nu = \sqrt{\frac{R^2r^2}{\sigma^4} + 2\rho\frac{Rr}{\sigma^2}\cos(\varphi-\psi) + \rho^2}$. It has been shown in [14]-[15] that the capacity achieving input distribution is circularly symmetric, i.e. the phase φ is independent of the amplitude r, and uniformly distributed over $[-\pi,\pi)$. Assuming such a characteristic of the input distribution, we are left with finding the amplitude distribution only. Thus, the capacity of the channel is given by

$$\sup_{\substack{F_r \\ \mathbb{E}[r^2]\le P}} \; I(R,\,\psi\,;\,r,\,\varphi\,:\,F_r). \tag{53}$$

where φ is uniformly distributed and independent of r. Using similar arguments to those which were already used in the NCAWGN channel, it can be shown [18] that necessary and sufficient conditions (Kuhn Tucker) for a distribution $F_0(r)$ to achieve the capacity are given by

$$\begin{aligned} \int_0^{\infty}\!\!\int_{-\pi}^{\pi} f(R,\,\psi\,|\,r,\,\varphi) \log\frac{f(R,\,\psi\,|\,r,\,\varphi)}{f(R,\,\psi\,:\,F_0)}\, dR\, d\psi &\le C + \gamma(r^2-P) \quad \forall r \\ \int_0^{\infty}\!\!\int_{-\pi}^{\pi} f(R,\,\psi\,|\,r,\,\varphi) \log\frac{f(R,\,\psi\,|\,r,\,\varphi)}{f(R,\,\psi\,:\,F_0)}\, dR\, d\psi &= C + \gamma(r^2-P) \quad \forall r \in E_0 \end{aligned} \tag{54}$$

where E_0 is the set of increase points of F_0. Continuing along the same path as in the NCAWGN channel it is possible to show [18] the analyticity of the function $g(z)$ defined by

$$g(z) = \int_0^{\infty}\!\!\int_{-\pi}^{\pi} f(R,\,\psi\,|\,z,\,\varphi) \log\frac{f(R,\,\psi\,|\,z,\,\varphi)}{f(R,\,\psi\,:\,F_0)}\, d\psi\, dR - C - \gamma(z^2-P). \tag{55}$$

By identical techniques to those employed for the NCAWGN channel it can be shown that E_0 does not contain an infinite set of mass points on any bounded interval, extending the result which was established for the NCAWGN channel. The proof is omitted for brevity, as it does not convey any new concepts and can be found in [18].

6 Summary and Discussion

We considered the capacity of the discrete-time NCAWGN memoryless channel under the average power constraint. Incorporating standard optimization techniques, it was shown that the capacity achieving distribution is discrete with an infinite number of mass points, and with only a finite number of them located over any bounded interval. Also, necessary and sufficient conditions for a capacity achieving distribution were obtained. This result was partly extended to the more general case where the unknown carrier phase manifests itself through the Tikhonov distribution, corresponding to the case where it is estimated by a PLL in conjunction to the employment of an ideal interleaver. Finally, some lower and upper bounds on channel capacity were derived, shedding some light on the behavior of the capacity for various values of SNR. These bounds could be used to examine the performance of coding schemes in communication systems which work under a regime where the carrier phase is rapidly varying with time.

A serious pitfall in computing the capacity achieving distribution for the NCAWGN channel is embodied in the fact that the optimal input distribution has infinitely many mass points. This fact makes it difficult for us to compute the capacity achieving distribution with standard numerical optimization tools. Any numerical optimization of the average mutual information over a finite collection of mass points is bound to be sub-optimal, and hence considered a lower bound on capacity. An unanswered question is the quantification of this sub-optimality. It would be of practical importance if one could quantify the trade off which exists between the number of optimized mass points and the relative distance from the actual capacity curve.

Another attempt to assess channel capacity here could be by first introducing an additional constraint, which causes the capacity achieving input to have a finite number of mass points. As was mentioned in section 1, imposing a peak constraint on the input guarantees in many other cases a capacity achieving distribution with a finite number of mass points. A finite number of mass points might even achieve the capacity of the NCAWGN channel if in addition to the average power constraint a fourth moment constraint is imposed on the input. Recent studies [13] have investigated cases where such a constraint does yield a finite mass point capacity achieving distribution. In any case, given that adding a constraint causes the channel to have a discrete and finite capacity achieving distribution, one can then relax this constraint asymptotically, and examine the capacity behavior which should tend to that of the NCAWGN channel under the average power constraint, which we have explored in our work. Such results can then be compared with ours.

A subsequent extension of part of our results [19], has recently been reported in [26] for the blockwise non-coherent channel. The cardinality of the capacity achieving discrete set is not yet known for this general case.

As to the partially coherent case, it should be noted that the infiniteness of the capacity achieving distribution was not established here. It should not be surprising if the same regime holds here as well, when one considers the limiting case were the loop SNR parameter ρ is taken to be very large (the coherent case) where the capacity achieving distribution being Gaussian definitely has an infinite number of mass points. A rigorous proof of this conjecture is called for.

The NCAWGN and PCAWGN channels are special cases of the general fading channel, where the complex fading variable assumes an amplitude of unity and where the phase assumes the uniform distribution in the NCAWGN case, and the Tikhonov distribution in the PCAWGN case. Both channels have been shown here to have discrete capacity achieving distributions. When considering these channels in the wider context of fading channels, it is desirable to be able to fully characterize the fading variable which causes the capacity achieving distribution to be discrete.

Appendix A
Weak* Continuity and Strict Concavity of Mutual Information

Weak* continuity of a functional f defined over the set of probability distributions is equivalent to $F_n(r) \stackrel{w^*}{\to} F(r) \implies f(F_n) \to f(F)$. We write the mutual information functional from (9) as the difference $I(R; r : F) = \tilde{h}_R(F) - \tilde{h}_{R|r}(F)$ where

$$\begin{aligned} \tilde{h}_R(F) &= -\int_0^\infty f(R : F) \log \frac{f(R : F)}{R} \, dR \\ \tilde{h}_{R|r}(F) &= -\int_0^\infty \int_0^\infty f(R \,|\, r) \log \frac{f(R \,|\, r)}{R} \, dR \, dF. \end{aligned} \tag{A.1}$$

We first show that $\tilde{h}_R(F)$ is weak* continuous. Let $F_n(r) \stackrel{w^*}{\to} F(r)$. We will show that $\tilde{h}_R(F_n) \to \tilde{h}_R(F)$.

$$\lim_{n\to\infty} \tilde{h}_R(F_n) = \lim_{n\to\infty} \left\{ -\int_0^\infty f(R : F_n) \log \frac{f(R : F_n)}{R} \, dR \right\} \tag{A.2}$$

$$= -\int_0^\infty R \lim_{n\to\infty} \left\{ \frac{f(R : F_n)}{R} \log \frac{f(R : F_n)}{R} \right\} dR \tag{A.3}$$

$$= -\int_0^\infty f(R : F) \log \frac{f(R : F)}{R} \, dR \tag{A.4}$$

$$= \tilde{h}_R(F). \tag{A.5}$$

Equations (A.2) and (A.5) are definitions. To establish (A.4) note that $\frac{f(R|r)}{R}$ where $f(R \,|\, r)$ is defined in (7) is a bounded continuous function of r. By definition of the weak* topology

$$\frac{f(R : F)}{R} = \int_0^\infty \frac{f(R \,|\, r)}{R} \, dF(r) \tag{A.6}$$

is a continuous function of F for every $R \geq 0$. Since $x \log x$ is continuous, it follows that $\frac{f(R:F)}{R} \log \frac{f(R:F)}{R}$ is continuous in F. By the definition of weak continuity there holds

$$\lim_{n\to\infty} \left\{ \frac{f(R:F_n)}{R} \log \frac{f(R:F_n)}{R} \right\} = \frac{f(R:F)}{R} \log \frac{f(R:F)}{R}. \tag{A.7}$$

To establish the interchange of integral and limit in (A.3) we use the Lebesgue dominated convergence theorem, and find an integrable function g such that

$$\left| \frac{f(R:F_n)}{R} \log \frac{f(R:F_n)}{R} \right| \leq g(R) \quad \forall n. \tag{A.8}$$

In [18] it is shown that the function $\frac{f(R:F_n)}{R}$ is bounded for all F_n by the function $\tilde{g}(R)$ defined by

$$\tilde{g}(R) = \begin{cases} \frac{1}{\sigma^2} e^{-\frac{R^2}{2\sigma^2}} & \text{if } R \leq \sqrt{2\sigma^2} \\ \frac{M}{R^{2+\delta}} & \text{if } R > \sqrt{2\sigma^2} \end{cases} \tag{A.9}$$

for some $M > 0$ and $\delta \in (0,\, 1)$. Thus, it follows that

$$\left| \frac{f(R:F_n)}{R} \log \frac{f(R:F_n)}{R} \right| \quad \leq \quad g(R)$$

where g is given by

$$g(R) = \begin{cases} \frac{1}{\sigma^2} e^{-\frac{R^2}{2\sigma^2}} \left| \log \sigma^2 + \frac{R^2}{2\sigma^2} \right| & \text{if } R \leq \sqrt{2\sigma^2} \\ \frac{M}{R^{2+\delta}} \left| \log M - (2+\delta) \log R \right| & \text{if } R > \sqrt{2\sigma^2} \end{cases} \tag{A.10}$$

which is piecewise continuous, bounded, and integrable. We have shown so far that $\tilde{h}_R(F)$ is weak* continuous. Next, we show that $\tilde{h}_{R|r}(F)$ is weak* continuous in F. Note from (A.1) that the innermost integral can be expressed in terms of Wyner's capacity of polyphase coding [41]. The Wyner capacity $C_w(r)$ expresses the capacity of an AWGN channel where a polyphase coding scheme is employed as function of the (fixed) input amplitude r. This capacity is given by [41]

$$C_w(r) = -\int_0^\infty f(R\,|\,r) \log \frac{f(R\,|\,r)}{R}\, dR - \log \sigma^2 e. \tag{A.11}$$

The function $\tilde{h}_{R|r}(F)$ can now be written as

$$\tilde{h}_{R|r}(F) = \int_0^\infty \left[C_w(r) + \log \sigma^2 e \right] \, dF(r).$$

Let $F_n(r) \xrightarrow{w^*} F(r)$. We must show that

$$\int_0^\infty C_w(r)\, dF_n(r) \to \int_0^\infty C_w(r)\, dF(r).$$

This follows if the function

$$g(r) = \begin{cases} 0 & \text{if } r < 0 \\ C_w(r) & \text{if } r \geq 0 \end{cases} \tag{A.12}$$

is uniformly integrable in F_n [23, Sec. 11.4], that is, if g is continuous, if $\int |g|\, dF_n < \infty$, and if

$$\lim_{b \to \infty} \int_b^\infty C_w(r)\, dF_n(r) = 0 \tag{A.13}$$

uniformly in n. The function g can be shown to be an analytic function of r on the right half of the complex plane (see Appendix B). In particular, it is continuous in r for $r \geq 0$. Furthermore, from [41] it is known that for small values of r, $C_w(r) = \frac{r^2}{2\sigma^2} + O(r^4)$, and for large values of r, $C_w(r) = \frac{1}{2} \log \frac{2\pi r^2}{\sigma^2 e} + \epsilon(r)$, where $\epsilon(r) \overset{r\to\infty}{\longrightarrow} 0$. From these two assertions it follows that there exists some positive number K such that $|g| < \frac{r^2}{2\sigma^2} + K$ for all r. Therefore, there holds

$$\int_0^\infty |g|\, dF_n \leq \int_0^\infty \left[\frac{r^2}{2\sigma^2} + K \right] dF_n = \frac{P}{2\sigma^2} + K \tag{A.14}$$

for every n. To verify (A.13) note that for similar arguments there must exist a number $a > 0$ such that $C_w(r) < ar$ for $r \geq 0$. Therefore, for $b \geq 1$

$$\begin{aligned} \int_b^\infty C_w(r)\, dF_n(r) &\leq \int_b^\infty ar\, dF_n(r) \\ &\leq \frac{a}{b} \int_b^\infty r^2\, dF_n(r) \\ &\leq \frac{aP}{b} \end{aligned} \tag{A.15}$$

which is a bound independent of n that converges to 0 as $b \to \infty$. Hence, g is uniformly integrable in F_n, and $\tilde{h}_{R|r}(F)$ is a weak* continuous function of F. Being the difference of two weak* continuous functions, the mutual information $I(R;\, r\, :\, F)$ is consequently a weak* continuous function of F.

That mutual information is concave is a known fact (see for example [8, p. 50]). It remains to be shown that the concavity is strict, i.e. that for any distribution functions $F_1, F_2 \in \mathcal{F}$ and for any $\theta \in (0,\, 1)$ there holds

$$I(R;\, r\, :\, (1-\theta)F_1 + \theta F_2) = (1-\theta) I(R;\, r\, :\, F_1) + \theta I(R;\, r\, :\, F_2) \tag{A.16}$$

if and only if $F_1 = F_2$. The "if" part is trivial. For the "only if" part consider the decomposition of the mutual information in the beginning of this Appendix. Note that $\tilde{h}_{R|r}(F)$ is linear in F. To show that $\tilde{h}_R(F)$ is strictly concave let $F_\theta = (1-\theta)F_1 + \theta F_2$. By (10) the corresponding output PDF is $f(R : F_\theta) = (1-\theta)f(R : F_1) + \theta f(R : F_2)$. Now, evaluate the difference

$$\begin{aligned}
&\tilde{h}_R(F_\theta) - (1-\theta)\tilde{h}_R(F_1) - \theta\tilde{h}_R(F_2) = \\
&= (1-\theta)D(f(R : F_1)||f(R : F_\theta)) + \theta D(f(R : F_2)||f(R : F_\theta)). \qquad \text{(A.17)}
\end{aligned}$$

The RHS of (A.17) is identically zero if and only if $f(R : F_1) = f(R : F_2)$. It remains to be shown that $f(R : F_1) = f(R : F_2)$ implies $F_1 = F_2$, or equivalently that

$$\int_0^\infty \frac{R}{\sigma^2} e^{-\frac{R^2+r^2}{2\sigma^2}} I_0\left(\frac{rR}{\sigma^2}\right) d(F_1 - F_2) = 0 \Rightarrow F_1 - F_2 = 0.$$

However, this fact was shown to hold in [34]. Therefore, $\tilde{h}_R(F)$ is strictly concave. Since the mutual information functional is a sum of a linear function and a strictly concave function, it is therefore strictly concave.

Appendix B
Analyticity Of $g(z)$ for the NCAWGN Channel

In this Appendix we show that the function $g(z)$

$$g(z) = \int_0^\infty f(R\,|\,z) \log \frac{f(R\,|\,z)}{f(R : F_0)}\, dR - C - \gamma(z^2 - P), \quad \mathfrak{Re}(z) > 0 \qquad \text{(B.1)}$$

is analytic throughout the domain $D = \{z \in \mathbb{C} : \mathfrak{Re}(z) > 0\}$. It suffices to show that the first term of $g(z)$ is analytic, because every polynomial is analytic in the whole complex plane. Denote the first term of $g(z)$ by

$$\begin{aligned}
w(z) &= \int_0^\infty f(R\,|\,z) \log \frac{f(R\,|\,z)}{f(R : F_0)}\, dR \\
&= \int_0^\infty h_1(R,\, z)\, dR - \int_0^\infty h_2(R,\, z)\, dR \qquad \text{(B.2)}
\end{aligned}$$

where

$$\begin{aligned}
h_1(R,\, z) &= \frac{R}{\sigma^2}\left[-\log\sigma^2 - \frac{R^2+z^2}{2\sigma^2} + \log I_0\left(\frac{zR}{\sigma^2}\right)\right] \exp\left\{-\frac{R^2+z^2}{2\sigma^2}\right\} I_0\left(\frac{zR}{\sigma^2}\right) \\
h_2(R,\, z) &= \frac{R}{\sigma^2} \exp\left\{-\frac{R^2+z^2}{2\sigma^2}\right\} I_0\left(\frac{zR}{\sigma^2}\right) \log \frac{f(R : F_0)}{R}.
\end{aligned}$$

We use the differentiation lemma [20, p. 392] to prove the analyticity of the two terms in (B.2). Consider the first term in (B.2). h_1 is continuous on $\{z \in \mathbb{C} : \mathfrak{Re}(z) > 0\} \times [0, \infty)$ and analytic for each R as a combination of continuous and analytic functions (the principal branch of the logarithm taken). Now, since every compact set in the complex plane is closed and bounded, it suffices to show that the first term in (B.2) is uniformly convergent for any rectangle $\{z \in \mathbb{C} : a \leq \mathfrak{Re}(z) \leq b, -b \leq \mathfrak{Im}(z) \leq b\}$. Any other compact set in the right half of the complex plane is contained in some rectangle of that form. Let K be such a compact set for some positive a and b, and let z be an element of K. Evaluation of the first term in (B.2) yields

$$\begin{aligned}
&\left|\int_0^\infty h_1(R, z)\, dR\right| \leq \\
&\leq \frac{1}{\sigma^2} e^{\frac{b^2}{2\sigma^2}} \int_0^\infty R \left| -\log \sigma^2 - \frac{R^2 + z^2}{2\sigma^2} + \log I_0\left(\frac{zR}{\sigma^2}\right)\right| e^{-\frac{R^2}{2\sigma^2}} \left| I_0\left(\frac{zR}{\sigma^2}\right)\right| dR \\
&\leq \frac{1}{\sigma^2} e^{\frac{b^2}{2\sigma^2}} \int_0^\infty R \left[|\log \sigma^2| + \frac{R^2}{2\sigma^2} + \frac{a^2 + b^2}{2\sigma^2} + B \right] e^{-\frac{R^2}{2\sigma^2}} I_0\left(\frac{Rb}{\sigma^2}\right) dR \qquad \text{(B.3)} \\
&< \infty
\end{aligned}$$

where (B.3) comes by the fact that $|I_0(z)| \leq I_0(\mathfrak{Re}(z))$ for $z \in \mathbb{C}$ and by the fact that $\left|\log I_0\left(\frac{zR}{\sigma^2}\right)\right|$ can be bounded by B for any $z \in K$ as a continuous function over K. We conclude that the first term in (B.2) is uniformly convergent for all $z \in K$. Hence, by the differentiation lemma, it is analytic for $\mathfrak{Re}(z) > 0$.

We now show that the second term in (B.2) is also analytic. The function $h_2(R, z)$ is continuous on $\{z \in \mathbb{C} : \mathfrak{Re}(z) > 0\} \times [0, \infty)$ because the function $f(R : F)$ is continuous for all $R \geq 0$ [18]. Moreover, it is also an analytic function of z for each R, being a combination of analytic functions. Thus, it is left for us to establish that the function $h_2(R, z)$ is uniformly convergent for all $z \in K$. It can be shown [18] that the function $\log \frac{f(R:F)}{R}$ is bounded by

$$\log \frac{f(R : F)}{R} \leq \max\left\{ \left| \frac{R^2}{2\sigma^2} - \log \frac{\beta_F}{\sigma^2} \right|, \left|\log \sigma^2\right| \right\}. \qquad \text{(B.4)}$$

where $\beta_F = \int_0^\infty e^{-\frac{r^2}{2\sigma^2}} dF(r)$ is a constant belonging to $(0, 1]$. Now, we evaluate the second term in (B.2)

$$\begin{aligned}
&\left|\int_0^\infty h_2(R, z)\, dR\right| \leq \\
&\leq \frac{1}{\sigma^2} e^{\frac{b^2}{2\sigma^2}} \int_0^\infty R e^{-\frac{R^2}{2\sigma^2}} \left| I_0\left(\frac{zR}{\sigma^2}\right)\right| \left|\log \frac{f(R : F_0)}{R}\right| dR \\
&\leq \frac{1}{\sigma^2} e^{\frac{b^2}{2\sigma^2}} \int_0^\infty R e^{-\frac{R^2}{2\sigma^2}} I_0\left(\frac{Rb}{\sigma^2}\right) \max\left\{ \left| \frac{R^2}{2\sigma^2} - \log \frac{\beta_F}{\sigma^2} \right|, \left|\log \sigma^2\right| \right\} dR < \infty
\end{aligned}$$

which implies the uniform convergence of the second term in (B.2). To conclude, we have shown that the conditions of the differentiation lemma hold also for the second term in (B.2). This establishes the analyticity of $w(z)$, and therefore, the analyticity of $g(z)$.

Appendix C
Weak Differentiability of $I(R\,;\,r\,:\,F) - \gamma g(F)$

We define the marginal mutual information

$$i(r\,;\,F) = \int_0^\infty f(R\,|\,r)\log\frac{f(R\,|\,r)}{f(R\,:\,F)}\,dR \tag{C.1}$$

denote $F_\theta = (1-\theta)F_0 + \theta F$, and compute the difference

$$\begin{aligned} & I(R\,;\,r\,:\,F_\theta) - I(R\,;\,r\,:\,F_0) = \qquad\qquad (C.2)\\ &= \int_0^\infty\!\!\int_0^\infty f(R\,|\,r)\log\frac{f(R\,:\,F_0)}{f(R\,:\,F_\theta)}\,dR\,dF_0(r) + \theta\int_0^\infty\!\!\int_0^\infty i(r\,;\,F_\theta)\,dR\,dF(r) + \\ &\quad -\theta\int_0^\infty\!\!\int_0^\infty i(r\,;\,F_0)\,dR\,dF_0(r). \end{aligned}$$

The weak derivative is given by [35]

$$\begin{aligned} I'_{F_0}(R\,;\,r\,:\,F) &= \lim_{\theta\downarrow 0}\frac{I(R\,;\,r\,:\,F_\theta) - I(R\,;\,r\,:\,F_0)}{\theta} \\ &= \int_0^\infty i(r\,;\,F_0)\,dF(r) - I(R\,;\,r\,:\,F_0). \end{aligned}$$

For the weak derivative of g we have

$$g'_{F_0}(F) = g(F) - g(F_0). \tag{C.3}$$

Since these two derivatives are valid for all F_0 and all F in $\mathcal{F}$ it follows that the functional $I(R\,;\,r\,:\,F) - \gamma g(F)$ is weakly differentiable.

Appendix D
Conditions for the Hankel Transform

The continuity of $f(R\,:\,F_0)$ implies the continuity of $\tilde{f}(x\,:\,F_0)$, and thus we have that the function $e^{-\frac{1}{2}\sigma^2x^2}\log\frac{\tilde{f}(x\,:\,F_0)}{\sigma^2x}$ is continuous for $x \geq 0$. Next, we show that the function $\sqrt{x}e^{-\frac{1}{2}\sigma^2x^2}\log\frac{\tilde{f}(x\,:\,F_0)}{\sigma^2x}$

is absolutely integrable.

$$\begin{aligned}
&\int_0^\infty \left| \sqrt{x} e^{-\frac{1}{2}\sigma^2 x^2} \log \frac{\tilde{f}(x\,:\,F_0)}{\sigma^2 x} \right| dx = \\
&= \int_0^\infty \sqrt{x} e^{-\frac{1}{2}\sigma^2 x^2} \left| \log \frac{\tilde{f}(x\,:\,F_0)}{\sigma^2 x} \right| dx \\
&\leq \int_0^\infty \sqrt{x} e^{-\frac{1}{2}\sigma^2 x^2} \max \left\{ \left| \frac{1}{2}\sigma^2 x^2 - \log \frac{\beta_{F_0}}{\sigma^2} \right|, \left| \log \sigma^2 \right| \right\} dx < \infty
\end{aligned}$$

where in the last inequality we have used the results of Appendix B, namely, (B.4).

Appendix E
Upper Bound For Continuous Alphabets

Begin with (49) and add the power constraint term on both sides to obtain

$$\begin{aligned}
I(Q\,;W) + \gamma(P - E_Q x^2) &\leq \int D\left(W\left(\cdot|x\right)||R(\cdot)\right)\, \mathrm{d}Q(x) + \gamma(P - E_Q x^2) \\
&= \int \left[D\left(W\left(\cdot|x\right)||R(\cdot)\right) + \gamma(P - x^2) \right] \mathrm{d}Q(x) \\
&\leq \max_{x \in \mathcal{X}} \left[D\left(W\left(\cdot|x\right)||R(\cdot)\right) + \gamma(P - x^2) \right] \qquad \text{(E.1)}
\end{aligned}$$

Now, since (E.1) holds for any Q, we can take the maximum of the LHS with respect to Q and obtain

$$\max_Q \left[I(Q\,;W) + \gamma(P - E_Q x^2) \right] \leq \max_{x \in \mathcal{X}} \left[D\left(W\left(\cdot|x\right)||R(\cdot)\right) + \gamma(P - x^2) \right]. \qquad \text{(E.2)}$$

Since the last inequality holds for any γ, we can minimize both sides with respect to γ

$$\min_{\gamma \geq 0} \max_Q \left[I(Q\,;W) + \gamma(P - E_Q x^2) \right] \leq \min_{\gamma \geq 0} \max_{x \in \mathcal{X}} \left[D\left(W\left(\cdot|x\right)||R(\cdot)\right) + \gamma(P - x^2) \right]$$

where the LHS is the capacity of the channel.

References

[1] I. C. Abou-Faycal, M. D. Trott and S. Shamai (Shitz) "The Capacity of Discrete-Time Memoryless Ralyeigh Fading Channels," *IEEE Trans. Inform. Theory,* vol. 47, no 4, pp. 1290-1301, May 2001.

[2] N. M. Blachman, "A Comparison of the Informational Capacities of Amplitude- and Phase-Modulation Communication Systems," *Proc. IRE*, vol. 41, pp. 748-759, June 1953.

[3] B. J. Belzer, A. D. Risley, P. Hou and T. R. Fischer "Capacity of AM-PSK on Partially Coherent Fading Channels," *IEEE Trans. Commun.*, Vol. 50, No. 3, pp. 369-373, March 2002.

[4] G. F. Carrier, M. Krook and C. E. Pearson, *Functions of a Complex Variable: Theory and Technique,* McGraw-Hill, 1966.

[5] G. Covalope, G. Ferrar and R. Raheli "Non-coherent Iterative (Turbo) Decoding," *IEEE Trans. Commun.*, Vol. 48, No. 9, pp. 1488-1498, Sep. 2000.

[6] G. Covalope and R. Raheli "The Capacity of Non-coherent Channels," *ETT (European Transactions on Telecommunications*), Vol. 12, No. 4, pp. 289-296, July/August 2001.

[7] G. Covalope and R. Raheli "Non-coherent Sequence Detection," *IEEE Trans. Commun.*, Vol. 47, No. 9, pp. 1376-1385, Sep. 1999.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Akadémiai Kiadó, Budapest (third impression).

[9] A. Das, "Capacity-Achieving Distributions for Non-Gaussian Additive Noise Channels," in *Proc. 2000 International Symposium on Information Theory*, Sorrento, Italy, June 25-30, p. 432.

[10] D. Divasalar and M. K. Simon, "Multiple-Symbol Differential Detection of MPSK," *IEEE Trans. Commun.,* vol. 38, pp. 300-308, Mar. 1990.

[11] D. Divasalar and M. K. Simon, "Maximum-Likelihood Differential Detection of Uncoded and Trellis Coded Amplitude Phase Modulation over AWGN and Fading Channels - Metrics and Performance," *IEEE Trans. Commun.,* vol. 42, pp. 76-89, Jan. 1994.

[12] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Willey, 1968.

[13] M. C. Gursoy, H. V. Poor and S. Verdú, "Capacity of Discrete Memoryless Channels with Second and Fourth Moment Constraints," Preprint of Princeton Technical Report.

[14] P. Hou, B. Belzer and T. R. Fischer, "Shaping Gain of Partially Coherent Additive White Gaussian Noise Channel," *Proc. 2000 Conference on Information Sciences and Systems,* pp. wp6-17-wp6-22.

[15] P. Hou, B. Belzer and T. R. Fischer, "Shaping Gain of the Partially Coherent Additive White Gaussian Noise Channel," *IEEE Communications Letters,* vol. 6, No. 5, pp. 175-177, May 2002.

[16] S. Hranilovic and F. R. Kschischang, "Capacity Bounds for Power- and Band-Limited Optical Intensity Channels Corrupted by Gaussian Noise," Submitted to the IEEE Transactions on Information Theory, May 2002.

[17] Z. Jelonek, *A Comparison of Transmission Systems*, Presented at the Symposium on Applications of Communication Theory, Electrical Engineering Dept., Imperial College, London, September 1952 (published by: Messrs. Butterworths Scientific Press.).

[18] M. Katz, *On the Capacity-Achieving Distribution of the Discrete-Time Non-Coherent and Partially-Coherent Additive White Gaussian Noise Channels,* M.Sc. Thesis, Technion - Israel Institute of Technology, Haifa, June 2002.

[19] M. Katz and S. Shamai (Shitz), "On the Capacity-Achieving Distribution of the Discrete-Time Non-Coherent Additive White Gaussian Noise Channel," in *Proc. 2002 International Symposium on Information Theory*, Lausanne, Switzerland, June 30 - July 5, p. 165.

[20] S. Lang, *Complex Analysis*, 3rd ed. New York: Springer-Verlag, 1993.

[21] A. Lapidoth, "Capacity Bounds Via Duality: A Phase Noise Example," in *Proc. 2nd Asian-European Workshop on Information Theory*, Breisach, Germany, June 26-29, 2002, pp. 58-61.

[22] A. Lapidoth and S. M. Moser, "Convex-Programming Bounds on the Capacity of Flat Fading Channels," in *Proc. 2001 International Symposium on Information Theory*, Washington, DC, June 24-29, p. 53.

[23] M. Loève, *Probability Theory,* New-York: Van Nostrand, 1963.

[24] D. G. Luenberger, *Optimization by Vector Space Methods,* New York: Wiley, 1969.

[25] I. D. Marsland and P. T. Mathiopoulos "On the Performance of Iterative Non-coherent Detection of Coded M-PSK Signals," *IEEE Trans. Commun.*, Vol. 48, No. 4, pp. 588-596, Apr. 2000.

[26] R. Nuriyev and A. Anastasopoulos, "Capacity and Coding for the Noncoherent Block-independent AWGN Channel," Submitted to *ICC 2003*, Aug. 2002.

[27] W. Oettli, "Capacity-Achieving Input Distributions for Some Amplitude-Limited Channels with Additive Noise," *IEEE Trans. Inform. Theory*, vol. IT-20, no. 3, pp. 372-374, May 1974.

[28] R. Palanski, "On the Capacity Achieving Distributions of Some Fading Channels," Submitted to the *40th Annual Allerton Conference on Communication, Control, and Computing 2002.*

[29] M. Peleg and S. Shamai (Shitz), "On the Capacity of the Blockwise Incoherent MPSK Channel," *IEEE Trans. Inform. Theory,* vol. 46, no. 5, pp. 603-609, May 1998.

[30] M. Peleg and S. Shamai (Shitz), "Iterative Decoding of Coded and Interleaved Non-coherent Multiple Symbol Detected DPSK," *Electron. Lett.,* vol. 33, no. 12, pp. 1018-1020, June 1997.

[31] M. Peleg, S. Shamai (Shitz) and S. Galán "Iterative Decoding for Coded Non-Coherent MPSK Communications over Phase-Noisy AWGN Channel," *IEE Proc.-Commun.,* vol. 147, no. 2, pp. 87-95, Apr. 2000.

[32] A. D. Poularikas, *The Transforms and Applications Handbook,* IEEE Press, 2000.

[33] A. Risley, B. Belzer and Y. Zhu, "Turbo Trellis Coded Modulation on Partially Coherent Fading Channels," Submitted to *ISIT 2000*, Sep. 1999.

[34] S. Shamai (Shitz) and I. Bar-David "The Capacity of Average and Peak-Power-Limited Quadrature Gaussian Channels," *IEEE Trans. Inform. Theory,* vol. 41, no. 4, pp. 1060-1071, July 1995.

[35] J. G. Smith, "The Information Capacity of Amplitude and Variance Constrained Scalar Gaussian Channels," *Inform. Contr.,* vol. 18, pp. 203-219, 1971.

[36] I. N. Sneddon, *The Use of Integral Transforms.* McGraw-Hill, 1972.

[37] G. Taricco and M. Elia, "Capacity of Fading Channel With No Side Information," Electron. Lett., vol. 33, no. 16, pp. 1368-1370, July 31, 1997.

[38] E. Telatar, "Envelope Detected Complex Gaussian Noise Channel," Personal Communication, Bell Laboratories, Lucent Technologies, Murray-Hill, NJ, USA, January 21, 1999.

[39] S. Verdú, "Spectral Efficiency in the Wideband Regime," *IEEE Trans. Inform. Theory,* vol. 48, no. 6, pp. 1319-1343, June 2002.

[40] G. N. Watson, *A Treatise on the Theory of Bessel Functions,* 2nd ed. Cambridge university press, 1952.

[41] A. D. Wyner, "Bounds on Communication with Polyphase Coding," *Bell Syst. Tech. J.*, pp. 523-559, Apr. 1966.