The Capacity of the Noncoherent Channel*

GIULIO COLAVOLPE, RICCARDO RAHELI
Università di Parma, Dipartimento di Ingegneria dell'Informazione, Parco Area delle Scienze 181A, I-43100 Parma - Italy
{*colavolpe,raheli*}*@tlc.unipr.it*

Abstract. The capacity of a random-phase additive white Gaussian noise (AWGN) channel, referred to as *noncoherent channel*, is investigated in the case of a transmission of N information symbols. The non-Gaussianity of the capacity-achieving distribution is shown and a lower bound on the channel capacity is derived. For increasing values of the number of transmitted symbols N, the capacity of a noncoherent channel is shown to asymptotically approach that of a coherent channel, i.e., a known-phase AWGN channel. The asymptotical Gaussianity of the capacity-achieving distribution is also shown. Based on the derived lower bound, the inherent capacity loss of a noncoherent channel, as compared to a coherent one, may be considered very limited for all but very small values of N. This result may be viewed as the information theoretic counterpart of a similar conclusion derived by many authors with reference to the probability of detection error.

1 INTRODUCTION AND MOTIVATIONS

In bandpass transmissions, a coherent phase reference is normally not available at the receiver. Rare exceptions to this situation may be conceived, such as in cases where a strong pilot tone is used. Therefore, a noncoherent channel, i.e., an AWGN channel which introduces a random phase rotation, is a very general model for bandpass transmission channels. Two approaches are commonly adopted to solve the problem of the detection of a possibly encoded information sequence transmitted over an AWGN noncoherent channel.

The first approach is based on an approximate coherent detection. This approximation is based on the use of a phase synchronization scheme, which extracts a phase reference from the incoming signal, in conjunction with a detection scheme which is optimal under the assumption of perfect synchronization. We refer to these detection schemes as *pseudocoherent receivers*. Different synchronization schemes are possible, either based on a non data-aided, a decision-directed or a data-aided strategy, the latter if a known preamble is present. Many years of digital communications have shown that, when the phase rotation introduced by the channel is constant or slowly varying and the length of the transmission is sufficiently large, the performance of ideal coherent detection may be practically achieved.

The second approach is represented by noncoherent detection. In the technical literature, a growing interest has been recently shown towards improved noncoherent detection or decoding schemes. Two main classes of algorithms have been proposed. *Multiple symbol differential detection* (MSDD) [1], [2] is based on maximum likelihood detection of a block of information symbols based on a corresponding finite signal observation. *Noncoherent sequence detection* (NSD) [3], [4] starts from the optimal noncoherent maximum likelihood sequence detection strategy and introduces some approximations in order to realize simple suboptimal detection or decoding schemes based on the Viterbi algorithm. In both cases, for constant phase and sufficiently large receiver complexity, the performance is shown to approach that of ideal coherent detection, either by computer simulation or analytically in the case of coded M-ary phase-shift keying (M-PSK) systems [1]-[7].

The fact that the performance of both the approximate coherent and noncoherent strategies approaches that of optimal coherent detection, as long as the channel phase is constant or slowly varying and the transmission length is sufficiently large, intuitively suggests that the capacity of a noncoherent channel tends to that of a coherent channel for an increasing number of transmitted symbols.

The described noncoherent schemes, namely MSDD and NSD, use the channel in a different way. For MSDD, the receiver bases its decisions on the observation of a block of few symbols. As a consequence, the phase is required to be constant over the block and independent consecutive noncoherent channels are used. On the contrary, NSD is based on an approximation of the optimal nonco-

*This paper was presented in part at the *International Conference on Communications*, 1999. This work was supported by Ministero dell'Università e della Ricerca Scientifica e Tecnologica (MURST), Italy.

herent receiver that exploits a single noncoherent channel, on which all the transmitted symbols are sent. It is intuitive that NSD corresponds to a better use of the channel. In this paper, this will be justified by means of some considerations on the channel capacity.

In the widely used digital cellular systems with burst-mode transmission, a preamble is often use, in a pseudo-coherent scheme, for phase synchronization purposes also. Is this receiver task really necessary? What is the loss, if there is any, in terms of first principles, such as the channel capacity, consequent to the use of a noncoherent system for these applications?

In recent years, great attention has been devoted to turbo codes. As demonstrated in the case of coherent decoding, turbo codes achieve a performance limited by the channel capacity [8]. Therefore, the problem of the capacity of an AWGN noncoherent channel is important in order to investigate the performance limits of turbo codes in common bandpass applications. More recently, noncoherent soft-output iterative decoding schemes have also been proposed [9]-[12].

The capacity of an AWGN noncoherent channel has been calculated in the case of input symbols belonging to an M-PSK alphabet [13]. In this paper, we consider the case in which no constraint is given on the input symbols. We show that the distribution which maximizes the average mutual information, and thus achieving capacity, is composed of zero-mean and uncorrelated components, with uniformly distributed phases. We further show that, unlike the coherent case, the capacity-achieving distribution is not Gaussian. An asymptotically tight lower bound is provided. The computational complexity of this bound is independent of the number N of transmitted symbols. This bound shows that the capacity rapidly approaches that of an AWGN coherent channel. This asymptotic result is proved and it is shown that, when $N \to \infty$, the capacity-achieving distribution is Gaussian with independent and identically distributed zero-mean components.

In the next section, we describe the assumed channel model. Some important properties of the capacity-achieving distribution are derived in Section 3. A lower bound on the channel capacity is presented in Section 4. The asymptotic channel capacity and capacity-achieving distribution, when the number N of channel utilizations tends to infinity, are addressed in Section 5. Finally, conclusions are drawn in Section 6.

2 CHANNEL MODEL

The input to the channel is modeled as a vector $\mathbf{x} \triangleq (x_1, x_2, \ldots, x_N)^T$ of N complex symbols and the output is a vector $\mathbf{y} \triangleq (y_1, y_2, \ldots, y_N)^T$, whose components may be expressed as

$$y_k = x_k e^{j\theta} + w_k \tag{1}$$

where θ is a phase shift introduced by the channel, modeled as a continuous random variable with uniform distribution, and w_k are independent and identically distributed (i.i.d.), zero-mean, complex Gaussian random variables with independent real and imaginary components, each with variance σ^2. In the following, these random variables are collected in a vector $\mathbf{w} \triangleq (w_1, w_2, \ldots, w_N)^T$.

Each component of vector $\mathbf{x}$ may be also expressed in polar coordinates as $x_i = r_i e^{j\phi_i}$ and vectors $\mathbf{r} \triangleq (r_1, r_2, \ldots, r_N)^T$ and $\boldsymbol{\phi} \triangleq (\phi_1, \phi_2, \ldots, \phi_N)^T$ defined accordingly. The signal-to-noise ratio (SNR) per information symbol is defined as

$$\gamma \triangleq \frac{E\{||\mathbf{x}||^2\}}{E\{||\mathbf{w}||^2\}} = \frac{E\{||\mathbf{x}||^2\}}{2N\sigma^2} \tag{2}$$

where $||\mathbf{x}||$ denotes the Euclidean norm of $\mathbf{x}$.

3 CAPACITY-ACHIEVING DISTRIBUTION

The capacity-achieving distribution of vector $\mathbf{x}$ is characterized by a probability density function $p(\mathbf{x})$ which maximizes the *average mutual information* (AMI), defined as

$$I_{nc} \triangleq I(\mathbf{x};\mathbf{y}) = E\left\{\log_2 \frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\right\} \tag{3}$$

in which the conditional probability density function $p(\mathbf{y}|\mathbf{x})$ may be expressed as [12], [14]

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^N} \exp\left\{-\frac{||\mathbf{x}||^2}{2\sigma^2} - \frac{||\mathbf{y}||^2}{2\sigma^2}\right\} \mathrm{I}_0\left(\frac{|\mathbf{y}^T\mathbf{x}^*|}{\sigma^2}\right) \tag{4}$$

where I_0 is the zero-th order modified Bessel function of the first kind. The capacity per channel use of this AWGN noncoherent channel is defined as[1]

$$C_{nc} \triangleq \frac{\max\{I_{nc}\}}{N}\,. \tag{5}$$

In the following we denote by C_c the capacity of the same channel in the case of coherent transmission, i.e., when the receiver perfectly knows the phase θ. As well known, this capacity is $C_c = \log_2(1+\gamma)$ [15].

We now prove some important properties of the capacity-achieving distribution.

Theorem 1 *For a transmission of N symbols over an AWGN noncoherent channel, the AMI is maximum when the random vector $\boldsymbol{\phi}$ is independent of $\mathbf{r}$ and characterized by independent uniformly distributed (i.u.d.) components.*

[1] In the case of MSDD with block length of N symbols, the capacity of the used noncoherent channel is $C_{nc} = \frac{\max\{I_{nc}\}}{N-1}$, due to the overlap of one symbol [13].

$$\int_{\mathbb{C}^N}\left[\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\log_2\frac{p(\mathbf{y}|\mathbf{x})}{\int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\,d\mathbf{x}}\,d\mathbf{y}+\lambda p(\mathbf{x})+\mu||\mathbf{x}||^2 p(\mathbf{x})\right]d\mathbf{x}$$

Proof. As in [13], we consider the *virtual channel*, obtained from our channel considering the phase θ as an input, and whose AMI is

$$I_v \triangleq I(\theta,\mathbf{x};\mathbf{y}) = I(\theta,\boldsymbol{\phi},\mathbf{r};\mathbf{y}) = I(\boldsymbol{\theta},\mathbf{r};\mathbf{y}) \tag{6}$$

having defined $\boldsymbol{\theta} \triangleq (\theta_1,\theta_2,\ldots,\theta_N)^T = (\phi_1+\theta,\phi_2+\theta,\ldots,\phi_N+\theta)^T$, whose distribution is determined by that of $\boldsymbol{\phi}$ (because the distribution of θ is known). The AMI of the virtual channel may be related to the AMI of the AWGN channel in the case of noncoherent transmission as [13]

$$I_{nc} = I_v - I(\theta;\mathbf{y}|\mathbf{x}) = I_v - I(\theta;\mathbf{y}|\mathbf{r},\boldsymbol{\phi})\,. \tag{7}$$

The term $I(\theta;\mathbf{y}|\mathbf{r},\boldsymbol{\phi})$ is the AMI of the diversity channel in which θ is the input, $\mathbf{y}$ the output, and the receiver perfectly knows the channel state information (CSI), represented by $\mathbf{x}$. In general, this AMI depends on the joint distribution of $\mathbf{r}$ and $\boldsymbol{\phi}$, i.e., on the two factors $p(\mathbf{r})$ and $p(\boldsymbol{\phi}|\mathbf{r})$. However, a more in depth analysis reveals that this AMI is independent of $p(\boldsymbol{\phi}|\mathbf{r})$. The reason is intuitive due to the fact that the receiver knows perfectly the phases ϕ_k, whose realizations do not affect the receiver performance. This result may also be proved in the following more rigorous manner. In fact, being the phases ϕ_k perfectly known by the receiver, it is possible to transform the vector $\mathbf{y}$ using the diagonal unitary matrix

$$diag(e^{-j\phi_1}, e^{-j\phi_2},\ldots,e^{-j\phi_N})\,. \tag{8}$$

Since this transformation is reversible, it does not modify the AMI (data processing theorem [16]). Furthermore, this transformation does not change the statistics of the noise vector. Hence, the AMI $I(\theta;\mathbf{y}|\mathbf{r},\boldsymbol{\phi})$ is independent of $p(\boldsymbol{\phi}|\mathbf{r})$.

Obviously, the AMI of the virtual channel is [15]

$$I_v = I(\boldsymbol{\theta},\mathbf{r};\mathbf{y}) = H(\mathbf{y}) - H(\mathbf{y}|\boldsymbol{\theta},\mathbf{r}) \tag{9}$$

where $H(\mathbf{y})$ and $H(\mathbf{y}|\boldsymbol{\theta},\mathbf{r})$ are the entropy and the conditional entropy of $\mathbf{y}$, respectively. This conditional entropy is [15]-[19]

$$H(\mathbf{y}|\boldsymbol{\theta},\mathbf{r}) = N\log_2 2\pi e\sigma^2 \tag{10}$$

independently of the joint distribution of $\boldsymbol{\phi}$ and $\mathbf{r}$. With a change to polar coordinates, $y_i = v_i e^{j\psi_i}$, and letting $\mathbf{v} \triangleq (v_1,v_2,\ldots,v_N)$, $\boldsymbol{\psi} \triangleq (\psi_1,\psi_2,\ldots,\psi_N)$, it is possible to express

$$\begin{aligned} H(\mathbf{y}) &= H(\mathbf{v},\boldsymbol{\psi}) \\ &\quad + \int_0^{+\infty}\!\!\cdots\int_0^{+\infty} p(v_1,\ldots,v_N)\log_2\prod_{i=1}^N v_i\ dv_1\ldots dv_N \\ &= H(\mathbf{v},\boldsymbol{\psi}) + \sum_{i=1}^N\int_0^{+\infty} p(v_i)\log_2 v_i\,dv_i \end{aligned} \tag{11}$$

which is a multidimensional version of (33) in [17]. Noting that $H(\mathbf{v},\boldsymbol{\psi}) \le H(\mathbf{v}) + H(\boldsymbol{\psi})$, with equality if and only if $\mathbf{v}$ and $\boldsymbol{\psi}$ are independent, we have

$$\begin{aligned} I_{nc} \ \le\ & H(\mathbf{v}) + H(\boldsymbol{\psi}) + \sum_{i=1}^N\int_0^{+\infty} p(v_i)\log_2 v_i\,dv_i \\ & -N\log_2 2\pi e\sigma^2 - I(\theta;\mathbf{y}|\mathbf{r},\boldsymbol{\phi})\,. \end{aligned} \tag{12}$$

The first and third terms depend on the probability density function $p(\mathbf{v})$ only, which is independent of the conditional distribution $p(\boldsymbol{\phi}|\mathbf{r})$, as shown in Appendix A, Lemma 1. The term $H(\boldsymbol{\psi})$ is maximized when $\boldsymbol{\psi}$ has i.u.d. components [15]. As shown in Appendix A, Lemma 2, $\boldsymbol{\psi}$ has i.u.d. components if $\boldsymbol{\theta}$ (and hence $\boldsymbol{\phi}$) has i.u.d. components, independently of the distribution of $\mathbf{r}$. In this case, (12) holds with equality. Therefore, from (12) we may conclude that I_{nc} is maximum when vectors $\mathbf{r}$ and $\boldsymbol{\phi}$ are independent and $\boldsymbol{\phi}$ has i.u.d. components. This distribution of $\boldsymbol{\phi}$ achieves capacity and the corresponding components of $\mathbf{x}$ have circular symmetry. □

Corollary 1 *The capacity-achieving distribution of* $\mathbf{x}$ *is characterized by zero mean and uncorrelated components.*

Proof. The proof is straightforward considering the independence of $\mathbf{r}$ and $\boldsymbol{\phi}$ and the circular symmetry of the components of $\mathbf{x}$. □

This result will be used in the proof of the following theorem.

Theorem 2 *For a transmission of* N *symbols over an AWGN noncoherent channel, the capacity-achieving distribution is not Gaussian.*

Proof. We begin by searching for the probability density function $p(\mathbf{x})$ which maximizes the AMI I_{nc} under a constant power constraint

$$\int_{\mathbb{C}^N} ||\mathbf{x}||^2 p(\mathbf{x})d\mathbf{x} = K \tag{13}$$

where $\mathbb{C}$ denotes the set of complex numbers and K is a constant. The constraint $\int_{\mathbb{C}^N} p(\mathbf{x})d\mathbf{x} = 1$ is also necessary due to the fact that $p(\mathbf{x})$ is a probability density function. Using Lagrange multipliers, we have to maximize the expression at the top of this page. By means of variational techniques, after some tedious but straightforward manipulations we obtain that the capacity-achieving distribution must satisfy the following condition

$$\begin{aligned} \int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})\log_2 p(\mathbf{y})\,d\mathbf{y} &= \lambda - \log_2 e + \mu||\mathbf{x}||^2 \\ &\quad + \int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})\log_2 p(\mathbf{y}|\mathbf{x})\,d\mathbf{y} \end{aligned} \tag{14}$$

or equivalently[2]

$$\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x}) \log_2 \frac{p(\mathbf{y})}{p(\mathbf{y}|\mathbf{x})}\, d\mathbf{y} = \lambda - \log_2 e + \mu||\mathbf{x}||^2 \,. \quad (15)$$

Using (4), we have

$$\begin{aligned}
&\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x}) \log_2 p(\mathbf{y}|\mathbf{x})\, d\mathbf{y} \\
&\quad = -N \log_2 2\pi\sigma^2 - \log_2 e \frac{||\mathbf{x}||^2}{2\sigma^2} \\
&\qquad - \frac{\log_2 e}{2\sigma^2}(2N\sigma^2 + ||\mathbf{x}||^2) \\
&\qquad + \int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x}) \log_2 \mathrm{I}_0\left(|\mathbf{y}^T\mathbf{x}^*|\right) d\mathbf{y} \qquad (16)
\end{aligned}$$

where the following relation has been used

$$\begin{aligned}
&\int_{\mathbb{C}^N} ||\mathbf{y}||^2 p(\mathbf{y}|\mathbf{x})\, d\mathbf{y} \\
&\quad = \frac{1}{2\pi}\int_0^{2\pi}\int_{\mathbb{C}^N} ||\mathbf{y}||^2 p(\mathbf{y}|\mathbf{x},\theta)\, d\mathbf{y}d\theta \\
&\quad = (2N\sigma^2 + ||\mathbf{x}||^2)\,. \qquad (17)
\end{aligned}$$

Therefore, (14) becomes

$$\begin{aligned}
\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x}) \log_2 p(\mathbf{y})\, d\mathbf{y} &= \lambda' + \mu'||\mathbf{x}||^2 \\
&+ \int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x}) \log_2 \mathrm{I}_0\left(|\mathbf{y}^T\mathbf{x}^*|\right) d\mathbf{y}\,. \qquad (18)
\end{aligned}$$

where

$$\lambda' \triangleq \lambda - \log_2 e - N\log_2 2\pi e\sigma^2 \qquad (19)$$

$$\mu' \triangleq \mu - \frac{\log_2 e}{\sigma^2}\,. \qquad (20)$$

We now consider the left hand side of (18). By contradiction, if $\mathbf{x}$ were Gaussian, it would have independent components (being uncorrelated, see corollary 1), and its probability density function would be

$$p(\mathbf{x}) = \frac{1}{(2\pi\sigma_x^2)^N}\exp\left\{-\frac{||\mathbf{x}||^2}{2\sigma_x^2}\right\}\,. \qquad (21)$$

The distribution of $\mathbf{y}$ would be obviously Gaussian with independent components as well

$$p(\mathbf{y}) = \frac{1}{(2\pi\sigma_y^2)^N}\exp\left\{-\frac{||\mathbf{y}||^2}{2\sigma_y^2}\right\} \qquad (22)$$

where $\sigma_y^2 = \sigma_x^2 + \sigma^2$. As a consequence, we would have

$$\begin{aligned}
&\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x}) \log_2 p(\mathbf{y})\, d\mathbf{y} \\
&\quad = -N\log_2(2\pi\sigma_y^2) - \frac{\log_2 e}{2\sigma_y^2}\int_{\mathbb{C}^N} ||\mathbf{y}||^2 p(\mathbf{y}|\mathbf{x})\, d\mathbf{y} \\
&\quad = -N\log_2(2\pi\sigma_y^2) - \frac{\log_2 e}{2\sigma_y^2}(2N\sigma^2 + ||\mathbf{x}||^2) \qquad (23)
\end{aligned}$$

which clearly exhibits a quadratic dependence on $\mathbf{x}$. On the contrary, the right hand side of (18) is not a quadratic form due to presence of the term $\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x}) \log_2 \mathrm{I}_0\left(|\mathbf{y}^T\mathbf{x}^*|\right) d\mathbf{y}$. Therefore, $\mathbf{x}$ cannot have a Gaussian distribution. □

The problem of finding the distribution of $\mathbf{r}$ which maximizes the AMI, presents some analytical difficulties which could not be overcome. In particular, note that the maximizing distribution depends on N. For this reason, we now consider an asymptotically tight, for $N \to \infty$, lower bound on channel capacity.

4 Lower bound on channel capacity

We now provide a lower bound on channel capacity which is tight for large values of N, as proved in the next section. This bound is the AMI I_{nc} for a specific probability density function $p(\mathbf{x})$.

We assume that the vector $\mathbf{x}$ is composed of i.i.d., zero-mean, complex Gaussian random variables with independent real and imaginary components, each with variance σ_x^2. Therefore, the probability density function of $\mathbf{x}$ is given by (21) and the SNR (2) is

$$\gamma = \frac{\sigma_x^2}{\sigma^2}\,. \qquad (24)$$

Under this assumption on $p(\mathbf{x})$, the probability density function of $\mathbf{y}$ is obviously given by (22). In this case, the AMI per channel use is

$$\begin{aligned}
\frac{I_{nc}}{N} &= \frac{1}{N}\int_{\mathbb{C}^N}\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \log_2 \frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}\, d\mathbf{x}d\mathbf{y} \\
&= -\frac{1}{N}\int_{\mathbb{C}^N} p(\mathbf{y}) \log_2 p(\mathbf{y})\, d\mathbf{y} \\
&\quad + \frac{1}{N}\int_{\mathbb{C}^N}\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \log_2 p(\mathbf{y}|\mathbf{x})\, d\mathbf{x}d\mathbf{y} \\
&= \log_2 2\pi e\sigma_y^2 \\
&\quad + \frac{1}{N}\int_{\mathbb{C}^N}\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \log_2 p(\mathbf{y}|\mathbf{x})\, d\mathbf{x}d\mathbf{y}\,. \qquad (25)
\end{aligned}$$

Using (4) and (17), we have

$$\begin{aligned}
\frac{I_{nc}}{N} &= \log_2 2\pi e\sigma_y^2 - \log_2(2\pi\sigma^2) \\
&\quad - \frac{\log_2 e}{2\sigma^2 N}\int_{\mathbb{C}^N} ||\mathbf{x}||^2 p(\mathbf{x})\, d\mathbf{x} \\
&\quad - \frac{\log_2 e}{2\sigma^2 N}\int_{\mathbb{C}^N}\int_{\mathbb{C}^N} ||\mathbf{y}||^2 p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\, d\mathbf{x}d\mathbf{y} \\
&\quad + \frac{1}{N}\int_{\mathbb{C}^N}\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \log_2 \mathrm{I}_0\left(\frac{|\mathbf{y}^T\mathbf{x}^*|}{\sigma^2}\right) d\mathbf{x}d\mathbf{y} \\
&= \log_2(1+\gamma) - 2\gamma\log_2 e \\
&\quad + \frac{1}{N}\int_{\mathbb{C}^N}\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \log_2 \mathrm{I}_0\left(\frac{|\mathbf{y}^T\mathbf{x}^*|}{\sigma^2}\right) d\mathbf{x}d\mathbf{y}
\end{aligned}$$

[2] Eq. (15) may be interpreted as the continuous-distribution constrained-power counterpart of theorem 5.4.3 of [18].

$$= C_c - 2\gamma \log_2 e + \frac{1}{N}\int_{\mathbb{C}^N}\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \log_2 \mathrm{I}_0\left(\frac{|\mathbf{y}^T\mathbf{x}^*|}{\sigma^2}\right) d\mathbf{x}d\mathbf{y}\,. \tag{26}$$

Hence, a lower bound on the capacity of the considered noncoherent channel is

$$C_{nc} \geq C_c - 2\gamma \log_2 e + \frac{1}{N}\int_{\mathbb{C}^N}\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \log_2 \mathrm{I}_0\left(\frac{|\mathbf{y}^T\mathbf{x}^*|}{\sigma^2}\right) d\mathbf{x}d\mathbf{y}\,. \tag{27}$$

The integral in (27) may be numerically calculated. In fact, the argument of the integral depends on $||\mathbf{x}||$, $||\mathbf{y}||$ and $|\mathbf{y}^T\mathbf{x}^*|$, only. Computing first the integral with respect to $\mathbf{x}$, the value of $\mathbf{y}$ is fixed and we may rotate the coordinate system in such a way that the first axis has the same direction of $\mathbf{y}$. Denoting by $\mathbf{x}'$ and $\mathbf{y}'$ the rotated vectors, we have $|\mathbf{y}^T\mathbf{x}^*| = |\mathbf{y}'^T\mathbf{x}'^*| = |y_1'||x_1'| = ||\mathbf{y}'|||x_1'| = ||\mathbf{y}|||x_1'|$. Therefore, after this rotation the integrand depends on $||\mathbf{x}||$, $||\mathbf{y}||$ and $|x_1'|$. Using the result of appendix B, equations (4) and (21), and another change to polar coordinates for the integral with respect to $\mathbf{y}$, we may express the last term in (27) as shown in equation (28) at the top of the next page, where $\beta \stackrel{\triangle}{=} 1 + \frac{1}{\gamma}$. This integral may be calculated using the Gauss-Hermite quadrature formula [19] and its complexity is independent of N.

In Figure 1, the proposed lower bound on channel capacity is shown for various values of N. For comparison, the coherent channel capacity is also shown. It may be observed that, for increasing values of N, the lower bound rapidly approaches the capacity of a coherent AWGN channel. This result was also found in [13] in the case of M-ary PSK input symbols and is in agreement with the experience which says that the limit represented by the performance of an ideal coherent receiver may be reached, for constant or slowly varying channel phase and continuous transmissions, with practical pseudocoherent or noncoherent schemes. Moreover, since for $N \geq 10$ the capacity loss is negligible, a burst-mode transmission, even if it is theoretically limited in capacity, it is not so in practice.

5 Asymptotic behavior

We now analyze the asymptotic behavior for $N \to \infty$ of the capacity of the considered noncoherent channel using some arguments similar to those in [13]. We prove the following theorem.

Theorem 3 *When $N \to \infty$, the capacity of an AWGN noncoherent channel approaches the capacity of a coherent channel. The capacity-achieving distribution is asymptotically Gaussian with i.i.d. components.*

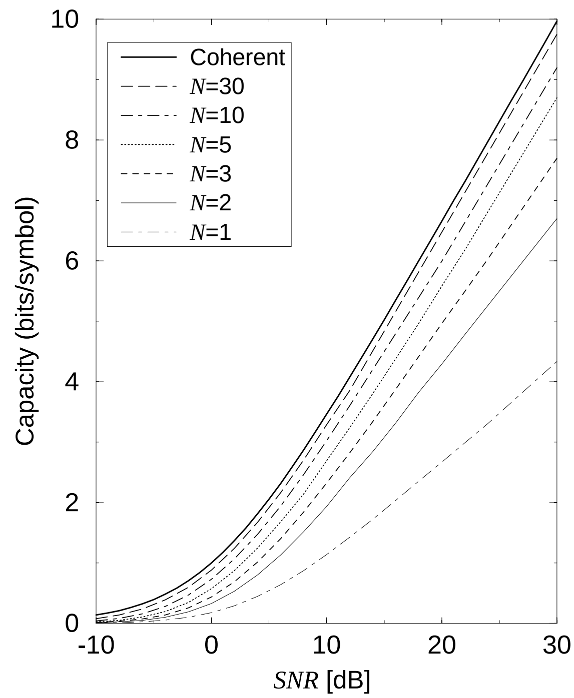

Figure 1: *Lower bound on channel capacity for different values of N.*

Proof. The AMI of the virtual channel may be expressed as

$$I_v = I(\theta, \mathbf{x}; \mathbf{y}) = I(\theta; \mathbf{y}) + I(\mathbf{x}; \mathbf{y}|\theta) = I(\theta; \mathbf{y}) + I_c \tag{29}$$

where I_c is the mutual information of the AWGN coherent channel and $I(\theta; \mathbf{y})$ is the AMI of the diversity channel in which θ is the input, $\mathbf{y}$ the output, and the receiver operates without knowledge of CSI. Substituting (29) in (7), one has

$$I_{nc} = I_c + I(\theta; \mathbf{y}) - I(\theta; \mathbf{y}|\mathbf{x})\,. \tag{30}$$

As in the previous section, considering a $p(\mathbf{x})$ given by (21), we have

$$C_{nc} \geq \frac{I_{nc}}{N} = \frac{I_c}{N} + \frac{I(\theta;\mathbf{y})}{N} - \frac{I(\theta;\mathbf{y}|\mathbf{x})}{N} = C_c + \frac{I(\theta;\mathbf{y})}{N} - \frac{I(\theta;\mathbf{y}|\mathbf{x})}{N}\,. \tag{31}$$

For this choice of $p(\mathbf{x})$, the random vector $\mathbf{y}$ is independent of θ and then $I(\theta; \mathbf{y}) = 0$, obtaining

$$C_{nc} \geq C_c - \frac{I(\theta;\mathbf{y}|\mathbf{x})}{N}\,. \tag{32}$$

Let us now consider the property of the capacity-achieving distribution, shown in theorem 1, of having a phase vector $\boldsymbol{\phi}$ with i.u.d. components. This property allows us to conclude that $I(\theta; \mathbf{y}) = 0$ even for the capacity-achieving distribution. As a consequence, being the AMI

$$\frac{1}{N}\int_{\mathbb{C}^N}\int_{\mathbb{C}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\log_2 \mathrm{I}_0\left(\frac{|\mathbf{y}^T\mathbf{x}^*|}{\sigma^2}\right) d\mathbf{x}d\mathbf{y} = \frac{8}{N\Gamma(N-1)\Gamma(N)(\gamma+1)^N}$$
$$\cdot\int_0^\infty\int_0^\infty\int_0^{\pi/2} e^{-\rho_1^2-\rho_2^2}\rho_1^2\rho_2^2(\rho_1\rho_2\sin\alpha)^{2N-3}\cos\alpha\, \mathrm{I}_0\!\left(\frac{2}{\sqrt{\beta}}\rho_1\rho_2\cos\alpha\right)\log_2 \mathrm{I}_0\!\left(\frac{2}{\sqrt{\beta}}\rho_1\rho_2\cos\alpha\right) d\rho_1 d\rho_2 d\alpha \quad (28)$$

$I(\theta;\mathbf{y}|\mathbf{x})$ non negative and noting that $I_c/N \le C_c$, from (30) we obtain $C_{nc} \le C_c$, as intuitively expected. Using (32), we finally have

$$C_c - \frac{I(\theta;\mathbf{y}|\mathbf{x})}{N} \le C_{nc} \le C_c\,. \quad (33)$$

We now consider the following diversity channel

$$\mathbf{y} = z\mathbf{x} + \mathbf{w} \quad (34)$$

where z is the input and $\mathbf{y}$ the output. If the probability density function of $\mathbf{x}$ has the expression given by (21), this channel is a diversity Rayleigh fading channel. The capacity C of this channel is [20]

$$C \le \log_2(1+N\gamma)\,. \quad (35)$$

Therefore, considering the special case $z = e^{j\theta}$, we have

$$I(\theta;\mathbf{y}|\mathbf{x}) \le \log_2(1+N\gamma) \quad (36)$$

and, from (33)

$$C_c - \frac{\log_2(1+N\gamma)}{N} \le C_{nc} \le C_c\,. \quad (37)$$

When $N\to\infty$, the lower bound, and then C_{nc}, tends to the limit represented by the capacity C_c of a coherent channel.

Since the lower bound tends to C_{nc} and is calculated assuming $\mathbf{x}$ Gaussian with i.i.d. components, when $N\to\infty$ this distribution achieves capacity. □

6 Conclusions

In this paper, the capacity of an AWGN channel in the case of a noncoherent transmission of N symbols has been considered. It has been proved that the capacity-achieving distribution is not composed of independent and identically distributed zero-mean Gaussian components. An asymptotically tight, for $N\to\infty$, lower bound has been provided. This bound shows that the capacity rapidly approaches that of the AWGN channel in the case of a coherent transmission. For $N \ge 10$, this limit is practically reached. Finally, this asymptotic result has been proved and it has been shown that, when $N\to\infty$, the capacity-achieving distribution is Gaussian with independent and identically distributed zero-mean components.

The practical equivalence between coherent and noncoherent channels, in the case of a constant or slowly varying phase rotation and continuous transmissions, demonstrated by many years of experience in digital communications has been theoretically proved. Besides noncoherent detection schemes, this result applies to commonly used pseudocoherent detection schemes as well. It has been also shown that, in the case of burst-mode transmissions, for practical burst lengths, in principle a preamble is not necessary for phase synchronization purposes.

Appendix A

In this appendix, we show two important lemmas used in the proof of theorem 1. Expressing the Gaussian probability density function $p(\mathbf{y}|\mathbf{x},\theta)$ in polar coordinates, we have

$$p(\mathbf{v},\boldsymbol{\psi}|\mathbf{r},\boldsymbol{\theta}) = \prod_{i=1}^{N}\frac{v_i}{2\pi\sigma^2}\exp\left\{-\frac{v_i^2+r_i^2-2v_ir_i\cos(\psi_i-\theta_i)}{2\sigma^2}\right\}\,. \quad (38)$$

The marginal (conditional) probability density functions are easily obtained

$$p(\mathbf{v}|\mathbf{r},\boldsymbol{\theta}) = p(\mathbf{v}|\mathbf{r}) = \prod_{i=1}^{N}\frac{v_i}{\sigma^2}\exp\left\{-\frac{v_i^2+r_i^2}{2\sigma^2}\right\}\mathrm{I}_0\left(\frac{v_ir_i}{\sigma^2}\right) \quad (39)$$

$$p(\boldsymbol{\psi}|\mathbf{r},\boldsymbol{\theta}) = \prod_{i=1}^{N}\left\{\frac{1}{2\pi}\exp\left[-\frac{r_i^2}{2\sigma^2}\right] + \frac{r_i\cos(\psi_i-\theta_i)}{\sqrt{2\pi}\sigma}\exp\left[-\frac{r_i^2\sin^2(\psi_i-\theta_i)}{2\sigma^2}\right]\cdot Q\left(-\frac{r_i\cos(\psi_i-\theta_i)}{\sigma}\right)\right\} \quad (40)$$

where $Q(x)$ is the Gaussian Q function defined as

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}}\int_x^{+\infty}\exp\left\{-\frac{t^2}{2}\right\}dt\,. \quad (41)$$

Lemma 1 *The probability density function $p(\mathbf{v})$ is independent of the conditional distribution $p(\boldsymbol{\phi}|\mathbf{r})$.*

Proof. We begin by noting that

$$p(\mathbf{v}|\mathbf{r}) = \int_{\Theta^N} p(\mathbf{v}|\mathbf{r},\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{r})\,d\boldsymbol{\theta} \quad (42)$$

where $\Theta \triangleq [0,2\pi)$. In general, $p(\mathbf{v}|\mathbf{r})$ would depend on $p(\boldsymbol{\theta}|\mathbf{r})$. However, (39) shows that $p(\mathbf{v}|\mathbf{r})$ is independent of

$p(\boldsymbol{\theta}|\mathbf{r})$. As a consequence, it is independent of $p(\boldsymbol{\phi}|\mathbf{r})$. Of course, this property holds for $p(\mathbf{v})$ also. □

Lemma 2 *If vector $\boldsymbol{\theta}$ has i.u.d. elements, $\boldsymbol{\psi}$ has also i.u.d. components, independently of the distribution of $\mathbf{r}$. Furthermore, $\mathbf{v}$ and $\boldsymbol{\psi}$ are independent.*

Proof. From (40), we obtain

$$p(\boldsymbol{\psi}|\mathbf{r}) = \int_{\Theta^N} p(\boldsymbol{\psi}|\mathbf{r},\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{r})\, d\boldsymbol{\theta}\,. \tag{43}$$

Assuming $p(\boldsymbol{\theta}|\mathbf{r}) = \prod_{i=1}^N p(\theta_i) = \left(\frac{1}{2\pi}\right)^N = p(\boldsymbol{\theta})$, for $0 \leq \theta_i < 2\pi$, due to the dependence on $\psi_i - \theta_i$ of the integrand in (43), the integrals with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are equal. We then obtain $p(\boldsymbol{\psi}|\mathbf{r}) = \prod_{i=1}^N p(\psi_i) = \left(\frac{1}{2\pi}\right)^N = p(\boldsymbol{\psi})$, for $0 \leq \psi_i < 2\pi$.

From (38) and (39), if $p(\boldsymbol{\theta}) = p(\boldsymbol{\psi}) = \left(\frac{1}{2\pi}\right)^N$, it is straightforward to verify that $p(\mathbf{v},\boldsymbol{\psi}|\mathbf{r}) = p(\mathbf{v}|\mathbf{r})p(\boldsymbol{\psi})$. As a consequence, averaging with respect to $\mathbf{r}$, we have $p(\mathbf{v},\boldsymbol{\psi}) = p(\mathbf{v})p(\boldsymbol{\psi})$. □

APPENDIX B

Let $\mathbf{x} \triangleq (x_1, x_2, \ldots, x_N)^T$ be a vector of N complex variables and denote its norm by $||\mathbf{x}||$. We now show that the integral over $\mathbb{C}^N$ of a generic function $f(|x_1|, ||\mathbf{x}||)$ may be expressed as

$$\int_{\mathbb{C}^N} f(|x_1|, ||\mathbf{x}||)\, d\mathbf{x} = \frac{4\pi^N}{\Gamma(N-1)} \cdot \int_0^\infty \int_0^{\pi/2} \rho^{2N-1}(\sin\alpha)^{2N-3}\cos\alpha\, f(\rho\cos\alpha, \rho)\, d\rho\, d\alpha\,. \tag{44}$$

In fact, being $||\mathbf{x}||$ non-negative, by definition of the Dirac delta function $\delta(x)$ [21], we have

$$f(|x_1|, ||\mathbf{x}||) = \int_0^\infty \delta(\rho - ||\mathbf{x}||) f(|x_1|, \rho)\, d\rho\,. \tag{45}$$

Defining $\tilde{\mathbf{x}} \triangleq (x_2, \ldots, x_N)^T$, we may express

$$\begin{aligned}
&\int_{\mathbb{C}^N} f(|x_1|, ||\mathbf{x}||)\, d\mathbf{x} \\
&= \int_0^\infty \left\{ \int_{\mathbb{C}} \left[\int_{\mathbb{C}^{N-1}} \delta(\rho - ||\mathbf{x}||)\, d\tilde{\mathbf{x}} \right] f(|x_1|, \rho)\, dx_1 \right\} d\rho \\
&= \int_0^\infty \left\{ \int_{\mathbb{C}} g(|x_1|, \rho) f(|x_1|, \rho)\, dx_1 \right\} d\rho
\end{aligned} \tag{46}$$

having defined

$$\begin{aligned}
g(|x_1|, \rho) &\triangleq \int_{\mathbb{C}^{N-1}} \delta(\rho - ||\mathbf{x}||)\, d\tilde{\mathbf{x}} \\
&= 2\rho \int_{\mathbb{C}^{N-1}} \delta(\rho^2 - ||\mathbf{x}||^2)\, d\tilde{\mathbf{x}}
\end{aligned} \tag{47}$$

where the last equality can be easily shown. The function $g(|x_1|, \rho)$ may be computed considering that

$$\begin{aligned}
g(|x_1|, \rho) &= 2\rho \int_{\mathbb{C}^{N-1}} \delta(\rho^2 - ||\mathbf{x}||^2)\, d\tilde{\mathbf{x}} \\
&= 2\rho \int_{\mathbb{C}^{N-1}} \delta(\rho^2 - ||\tilde{\mathbf{x}}||^2 - |x_1|^2)\, d\tilde{\mathbf{x}}
\end{aligned} \tag{48}$$

and, with a change to polar coordinates

$$\begin{aligned}
&g(|x_1|, \rho) \\
&= 2\rho \frac{2\pi^{N-1}}{\Gamma(N-1)} \int_0^\infty r^{2N-3}\delta(\rho^2 - r^2 - |x_1|^2)\, dr \\
&= \rho \frac{2\pi^{N-1}}{\Gamma(N-1)} \int_0^\infty \lambda^{N-2}\delta(\rho^2 - \lambda - |x_1|^2)\, d\lambda \\
&= \rho \frac{2\pi^{N-1}}{\Gamma(N-1)} (\rho^2 - |x_1|^2)^{N-2} u(\rho^2 - |x_1|^2)
\end{aligned} \tag{49}$$

where $u(x)$ is the unit step function. Substituting (49) in (46) and integrating with respect to the variable x_1 in polar coordinates, we obtain

$$\int_{\mathbb{C}^N} f(|x_1|, ||\mathbf{x}||)\, d\mathbf{x} = \frac{4\pi^N}{\Gamma(N-1)} \cdot \int_0^\infty \int_0^\infty \rho r(\rho^2 - r^2)^{N-2} f(r, \rho) u(\rho - r)\, dr\, d\rho\,. \tag{50}$$

Finally, by the change of variables $(\rho, r) = (\rho, \rho\cos\alpha)$, we obtain (44).

Manuscript received on July 9, 1999.

REFERENCES

[1] D. Divsalar and M. K. Simon. Multiple-symbol differential detection of MPSK. *IEEE Transactions on Communications,* Vol. 38, No. 3, pages 300–308, March 1990.

[2] D. Divsalar 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 Transactions on Communications,* Vol. 42, No. 1, pages 76–89, January 1994.

[3] G. Colavolpe and R. Raheli. Noncoherent sequence detection. *IEEE Transactions on Communications,* Vol. 47, No. 9, pages 1376–1385, September 1999.

[4] G. Colavolpe and R. Raheli. Noncoherent sequence detection of continuous phase modulations. *IEEE Transactions on Communications,* Vol. 47, No. 9, pages 1303–1307, September 1999.

[5] Y. Kofman, E. Zehavi, and S. Shamai. nd-convolutional codes-part I: performance analysis. *IEEE Transactions on Information Theory,* Vol. 43, No. 2, pages 558–575, March 1997.

[6] Y. Kofman, E. Zehavi, and S. Shamai. nd-convolutional codes-part II: structural analysis. *IEEE Transactions on Information Theory,* Vol. 43, No. 2, pages 576–589, March 1997.

[7] G. Colavolpe and R. Raheli. Theoretical analysis and performance limits of noncoherent sequence detection of coded PSK. *IEEE Transactions on Information Theory,* Vol. 46, No. 4, pages 1483–1494, July 2000.

[8] C. Berrou and A. Glavieux. Near optimum error correcting coding and decoding: turbo-codes. *IEEE Transactions on Communications,* Vol. 44, No. 10, pages 1261–1271, October 1996.

[9] M. Peleg and S. Shamai (Shitz). Iterative decoding of coded and interleaved noncoherent multiple symbol detected DPSK. *IEE Electronics Letters,* Vol. 33, No. 12, pages 1018–1020, 5th June 1997.

[10] M. Peleg, S. Shamai (Shitz), and S. Galán. On iterative decoding for coded noncoherent MPSK communications over block-noncoherent AWGN channel. In *IEEE International Conference on Telecommunications,* pages 109–114, 1998.

[11] I. D. Marsland and P. T. Mathiopoulos. On the performance of iterative noncoherent detection of coded M-PSK signals. *IEEE Transactions on Communications,* Vol. 48, No. 4, pages 588–596, April 2000.

[12] G. Colavolpe, G. Ferrari, and R. Raheli. Noncoherent iterative (turbo) decoding. *IEEE Transactions on Communications,* Vol. 48, No. 9, pages 1488–1498, September 2000.

[13] M. Peleg and S. Shamai (Shitz). On the capacity of the blockwise incoherent MPSK channel. *IEEE Transactions on Communications,* Vol. 46, No. 5, pages 603–609, May 1998.

[14] J. Proakis. *Digital communications,* 3rd ed. McGraw-Hill, New York, U.S.A., 1995.

[15] C. E. Shannon and W. Weaver. *The mathematical theory of communication*. University of Illinois Press, Urbana, U.S.A., 1963.

[16] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, New York, U.S.A., 1991.

[17] A. D. Wyner. Bounds on communication with polyphase coding. *Bell System Technical Journal,* Vol. XLV, pages 523–559, April 1966.

[18] R. E. Blahut. *Principles and practice of information theory.* Addison-Wesley, New-York, U.S.A., 1987.

[19] S. Benedetto, E. Biglieri and V. Castellani. *Digital transmission theory.* Prentice-Hall, Englewood Cliffs, U.S.A., 1987.

[20] W. C. Y. Lee. Estimate of channel capacity in Rayleigh fading environment. *IEEE Transactions on Vehicular Technology,* Vol. 39, No. 3, pages 187–189, August 1990.

[21] A. Papoulis. *The Fourier integral and its applications*. McGraw-Hill, New York, U.S.A., 1987.