The Capacity of Constant Envelope, Continuous Phase Signals over AWGN Channel under Carson’s Rule Bandwidth Constraint

Chun-Hsuan Kuo and Keith M. Chugg
Communication Sciences Institute
Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, California, 90089
{chunhsuk,chugg}@usc.edu

Abstract—In this paper, we derive an equivalent channel model for constant envelope, continuous phase (CECP) signals transmitted over the bandlimited additive white Gaussian noise (AWGN) channel. Using this equivalent channel model, it is proved that at high carrier-to-noise ratio (CNR), the capacity of coherent and non-coherent CECP-AWGN channel are the same. We then derive the main result of this paper, i.e., the capacity of the CECP-AWGN channel under Carson’s rule bandwidth constraint. For comparison, the symmetric information rate (SIR) of several commonly used bandwidth efficient continuous phase modulation (CPM) signals are estimated using the method described in [8]. We conclude that under Carson’s Rule bandwidth measure, the bandwidth efficiency of these CPM signals is still far below the CECP-AWGN capacity.

I. INTRODUCTION

Constant envelope, continuous phase (CECP) signals are widely used in communication systems due to their relatively low spectral sidelobes. Also, since CECP signals have constant envelope, the power amplifier can be operated in the saturation (nonlinear) region without causing much signal distortion, yielding better power efficiency. The continuous phase modulation (CPM) signal, such as Gaussian Minimum Shift Keying (GMSK, which is adapted in both GSM and DECT standards) and Continuous Phase Frequency Shift Keying (CPFSK), is a subclass of the CECP signals which are commonly used in practice. Although these modulation schemes are claimed to be bandwidth efficient, the limit of the bandwidth efficiency of the CECP-AWGN channel (or equivalently, the capacity of the CECP-AWGN channel with bandwidth constraint) is still unknown. In [1], the phase modulated Wiener process (WPM) is shown to approach the capacity of the additive white Gaussian noise (AWGN) channel with a Lorenzian power spectral density (PSD) constraint at high signal to noise ratio (SNR). However, with the constraint on the exact shape of the PSD, rather than on the bandwidth only, the result in [1] can only serve as a lower bound on the CECP-AWGN capacity with bandwidth constraint. In [2], a lower bound on the capacity of the AWGN channel with constant-envelope input constraint was developed using binary waveforms taking values of $\pm A$. The result does not fit into our framework since the signal being considered in [2] does not have continuous phase. Finally, the capacity of FM signal under RMS bandwidth constraint was derived in [3]. However, the results in [3] indicate that the rms bandwidth is not a suitable bandwidth measure for defining bandwidth efficiency.

A general CECP signal has the following form:

$$s(t) = \Re\{\sqrt{2}A e^{j\theta(t)} e^{j(2\pi f_c t + \varphi)}\}$$  \hspace{1cm} (1)

where $\theta(t)$ is the continuous message containing signal, and $f_c$ is the carrier frequency. For convenience, we assume $\theta(0) = 0$ without loss of generality. Since $\theta(t)$ is continuous, its time derivative can be defined (at least piecewise), and we have

$$s(t) = \sqrt{2}A \cos \left(2\pi f_c t + 2\pi \Delta \int_0^t m(\alpha) d\alpha + \varphi \right)$$  \hspace{1cm} (2)

where

$$m(t) \equiv \frac{\theta'(t)}{2\pi \Delta}$$  \hspace{1cm} (3)

We note that by setting $\theta(0) = 0$, there is a one to one correspondence between $m(t)$ and $\theta(t)$. Hence, we shall define $m(t)$ as the input message signal, and use equation (2) to represent the general CECP signal in the rest of our work.

Unlike linear modulations such as PSK and QAM, the spectral property of the CECP signal generally depends on the complete statistical description, not just the second moment description, of the message signal $m(t)$ [4]. This makes the bandwidth computation of the CECP signal difficult for a general message signal $m(t)$, and numerical evaluation or simulation is usually required. However, by noting that (2) is exactly the form of frequency modulation (FM), we can use the well known Carson’s Rule as a bandwidth measure. Carson’s Rule, which is defined for FM signals, depends only on the second moment description of the unmodulated signal, and despite its simplicity, it gives surprisingly satisfying results. In fact, it is used by NTIA for federal radio frequency management [7]. The one-sided Carson’s Rule bandwidth of $s(t)$ in (2) is defined as

$$B_c \equiv 2(\beta + 1)f_m$$  \hspace{1cm} (4)

Where

$$\beta \equiv \frac{\Delta \sqrt{P_m}}{f_m}$$  \hspace{1cm} (5)
Here, \( f_m \) is the largest “significant” frequency of \( m(t) \), and \( P_m \) is the power of \( m(t) \). The parameter \( \beta \) is often referred to as the effective FM modulation index.

This paper is organized as follows. In Section II, we will derive an equivalent channel model for a CECP-AWGN channel. In Section III, we use this equivalent channel model to prove that at high CNR, the capacity of coherent (i.e. \( \varphi \) known at the receiver) and non-coherent (i.e. \( \varphi \) unknown at the receiver) CECP-AWGN channel are the same. In Section IV, we use Carson’s Rule bandwidth as a constraint to derive the capacity of CECP-AWGN channel. Different measures of the largest significant frequency (i.e. \( f_m \)) and its effects on capacity will be considered in detail. Finally, in Section V, we compare the CECP-AWGN channel capacity with the symmetry information rate (SIR) of several commonly used bandwidth efficient CPM signals.

II. EQUIVALENT MODEL

In this section, we assume coherent detection, i.e. \( \varphi \) is known at the receiver. This assumption enables us to establish the complex baseband CECP-AWGN channel model shown in Figure 1. Consider the equivalent complex baseband CECP signal (see (2)) given by

\[
\tilde{s}(t) = A \exp\{j2\pi\Delta \int_0^t m(\alpha) \, d\alpha\} 
\]  (6)

This CECP signal is transmitted over a bandlimited AWGN channel. The equivalent complex baseband AWGN process can be described as

\[
\tilde{n}(t) = n_c(t) + jn_s(t) 
\]  (7)

where \( n_c(t) \) and \( n_s(t) \) are mutually independent Gaussian processes both having power spectral density

\[
S_{n_c}(f) = S_{n_s}(f) = \left\{ \begin{array}{ll} \frac{N_0}{2} & \text{if } |f| < W, \\ 0 & \text{otherwise.} \end{array} \right. 
\]  (8)

Now we make the following assumption:

Assumption 1: \( \tilde{s}(t) \) has one-sided bandwidth \( B \ll W \ll f_s \) such that it passes through the channel without distortion. Furthermore, the phase process \( \arg(\tilde{s}(t)) \) is roughly constant during the coherence time of the noise processes \( n_c(t) \) and \( n_s(t) \).

This assumption enables us to write the complex baseband output of the channel as (see [4])

\[
\tilde{r}(t) = \alpha(t) \exp\{j[2\pi\Delta \int_0^t m(\alpha) \, d\alpha + v(t)]\} 
\]  (9)

where

\[
\alpha(t) = \sqrt{(A + n_c(t))^2 + n_s^2(t)} 
\]  (10)

and

\[
v(t) = \arg ((A + n_c(t)) + jn_s(t)) 
- 2\pi \int_0^t \text{sgn}[n_s(\tau)]U[-A - n_c(\tau)]\delta[n_s(\tau)]d\tau 
\]  (11)

Here, \( U(\cdot) \) is the unit step function, and \( \delta(\cdot) \) is the Dirac-delta function. The second term in (11) compensates for the branch cut of the \( \arg \) function, such that \( v(t) \) is a continuous function ranging from \( -\infty \) to \( \infty \). This term also contributes to the “click noise” at the output of the FM discriminator [5]. Using equations (9)-(11), the following lemma is obtained.

Lemma 1: There exists a one to one mapping between \( \tilde{r}(t) \) and \( [\alpha(t), v(0), 2\pi\Delta m(t) + v(t)] \).

This gives us the equivalent model shown in Figure 1. The proof of Lemma 1 is omitted for brevity.

III. MUTUAL INFORMATION RATE AT HIGH CNR

Consider the equivalent channel model in the previous section. We first examine the statistical property of the noise process \( v'(t) \) at high CNR. When the CNR is high, we can approximate \( v'(t) \) by its first order approximation [4]. Hence, we have

\[
v'(t) \approx \frac{n'_s(t)}{A} 
\]  (12)

and the PSD of \( v'(t) \) becomes

\[
S_{v'}(f) = \frac{2\pi^2 N_0}{A^2} f^2 
\]  (13)

Similarly, we also have the following first order approximations:

\[
\alpha(t) \approx A + n_c(t) \]  (14)

and

\[
v(0) \approx \frac{n_s(0)}{A} 
\]  (15)

The mutual information rate between random processes \( x(t) \) and \( y(t) \) is defined as

\[
I_R[x(t); y(t)] \equiv \lim_{T \to \infty} \frac{1}{T} I(X^N; Y^N) 
\]  (16)

where \( X^N \) and \( Y^N \) are the orthogonal expansion vectors of \( x(t) \) and \( y(t) \) over the time interval \([0, T]\) respectively. Using the standard information theory argument, we have the following 2 Lemmas.

Lemma 2: \( I_R[m(t); \alpha(t) \mid z(t), v(0)] = 0 \) under the first order approximation, where \( z(t) \equiv 2\pi\Delta m(t) + v(t) \).

Lemma 3: \( I_R[m(t); v(0) \mid z(t)] = 0 \) under the first order approximation.

Again, the proofs of the above two lemmas are omitted for brevity. By lemma 2 and 3, together with the chain rule of mutual information, we have the following theorem.

Theorem 1: \( I_R[m(t); \alpha(t), z(t), v(0)] = I_R[m(t); z(t)] \) under the first order approximation.
This theorem allows us to establish a simpler equivalent CECP-AWGN channel model at high CNR when capacity is concerned, as shown in Figure 2. Note that even though we start our derivation by assuming coherent detection (φ known to the receiver), z(t) is exactly the output of the non-coherent FM receiver (i.e. the hardlimiter-discriminator receiver), which could be obtained without knowing φ. Hence, we conclude the following corollary:

**Corollary 1:** The capacity of coherent and non-coherent CECP-AWGN channel are the same under the first order approximation.

**IV. THE CAPACITY OF CECP-AWGN CHANNEL UNDER CARSON’S RULE BANDWIDTH CONSTRAINT**

Using the equivalent channel model derived in the section III, the capacity of the CECP-AWGN channel under Carson’s Rule bandwidth constraint at high CNR is given by

$$C = \lim_{T \to \infty} \frac{1}{T} \left[ \sup I(M^N; Z^N) \right]$$

where $M^N$ and $Z^N$ are the orthogonal expansion vectors of $m(t)$ and $z(t)$ over the time interval $[0, T]$ respectively, and the supremum is over all input probability distributions satisfying the following constraint

$$B_c = \Delta \sqrt{P_m} + f_m \leq B$$

The capacity $C$ has units of (bits/s), and will be given as a function of $B$ and signal to noise ratio $\rho$ defined as

$$\rho \equiv \frac{A^2}{2N_0B}$$

We summarize our result below. The detailed derivations can be found in the appendices.

1) $f_m :$ Strict Bandwidth. Then we have

$$C = B \max_{0 \leq u \leq x'} \frac{u(\ln(\frac{2\rho(1-u)^2}{u^2} + \frac{1}{2}) + 2)}{\ln 2}$$

where $x'$ is the unique real solution to the following equation

$$u^3 - 3\rho(1-u)^2 = 0$$

2) $f_m : (100x_o)\%$ Bandwidth. Then we have

$$C \equiv B \max_{0 \leq k \leq K_o} \frac{q(\rho, k)}{\ln 2} \left[ 2f(k, x_o) + \ln(1 + \frac{1}{3k^2f^2(k, x_o)} - f(k, x_o)) \right]$$

where

$$0 \leq K_o \equiv 2 \cos \left( \frac{\pi + \cos^{-1}(x_o)}{3} \right) \leq 1$$

$$f(k, x_o) \equiv \left\{ \begin{array}{ll} \frac{1}{2}[1 + 2\cos\left(\frac{\pi - \gamma(k, x_o)}{4}\right)] & k^3 \geq 2(1-x_o) \\ \frac{1}{2}[1 + \beta(k, x_o) + \frac{1}{3\gamma(k, x_o)}] & k^3 \leq 2(1-x_o) \end{array} \right.$$}

$$\gamma(k, x_o) \equiv \arg \left( \frac{k^3 - 4(1-x_o)}{k^3} + j\sqrt{8(\frac{1-x_o}{k^3}) - 16\left(\frac{1-x_o}{k^3}\right)^2} \right)$$

$$\beta(k, x_o) \equiv \left( \frac{-k^3 + 4(1-x_o)}{k^3} - \sqrt{-8(\frac{1-x_o}{k^3}) + 16\left(\frac{1-x_o}{k^3}\right)^2} \right)^{1/2}$$

$$q(\rho, k) \equiv \rho k^2 + \left( -3 + \rho k^3 + \sqrt{9 - 4\rho k^3} \right) \left( \frac{9}{2\rho} - k^3 + 3\sqrt{9 - 4\rho k^3} \right)^{1/2}$$

$$- \left( -3 + 2k^3 + \sqrt{9 - 4\rho k^3} \right) \left( \frac{9}{2\rho} - k^3 + 3\sqrt{9 - 4\rho k^3} \right)^{1/2} 2k^2$$

The bandwidth efficiency $R (\equiv C/2B)$ of the CECP-AWGN channel under Carson’s Rule bandwidth measure is plotted in Figure 3. For comparison, we also plot the bandwidth efficiency of the bandlimited AWGN channel without a modulation constraint, which is given by

$$R = \log_2 \left( 1 + \frac{E_b}{N_0} \right)$$
First, we note that due to our assumption, the result is valid only at high CNR, which corresponds to high $E_b/N_0$. Second, we can see that there is a large degradation from the bandlimited AWGN channel capacity to the CECP-AWGN channel capacity. For example, under the Carson’s Rule bandwidth measure where $m_f$ is defined as the 99% bandwidth of $m(t)$, in order to achieve the bandwidth efficiency of 5 bit/s/Hz, an additional 15dB of $E_b/N_0$ is needed when compared to the capacity of bandlimited AWGN channel. On the other hand, modulations required to approach the capacity of bandlimited AWGN channel at 5 bit/s/Hz (e.g. 64 QAM) may require significant amplifier back-off.

V. THE SIR OF CPM SIGNALS

The SIR is defined as the mutual information rate between the channel input and output when the channel input is a sequence of independent, identically distributed random variables with uniform distribution. Here, we will use the simulation based method described in [8] to evaluate the SIR of the modulation constrained CPM-AWGN channel. In general, the SIR is not the capacity of the CPM-AWGN channel. However, it is a lower bound of the channel capacity, and can be viewed as a practical limit for the channel.

Table I list the parameters of several commonly used CPM signals\(^1\). Their corresponding SIR are shown in Figure 4, together with the CECP-AWGN channel capacity derived in the last section. We notice that the Carson’s Rule bandwidth of MSK is significantly larger than that of the GMSK. This is due to the fact that MSK has a much slower sidelobe rolloff when compared to the GMSK signals. We can also see that although the GMSK signals are more bandwidth efficient than MSK, their bandwidth efficiencies are still far below the CECP-AWGN capacity at high $E_b/N_0$. Hence, at high CNR, the bandwidth efficiency of the CECP signal could still be improved via proper signal design.

VI. CONCLUSION

We have derived an equivalent channel model for CECP signals transmitted over the bandlimited AWGN channel. Using this equivalent channel model, it was proved that at high CNR, the capacity of coherent and non-coherent CECP-AWGN channel are the same. We then derived the capacity of the CECP-AWGN channel under Carson’s rule bandwidth constraint with different definitions of $m_f$. For comparison, the SIR of several commonly used bandwidth efficient CPM signals are also evaluated. We conclude that, under Carson’s Rule bandwidth measure, even though the the CECP-AWGN capacity suffers a large degradation from the capacity of the bandlimited AWGN channel without a modulation constraint, the bandwidth efficiency of the most commonly used CPM signals is still far below the CECP-AWGN capacity at high CNR. Hence, significant improvement on the bandwidth efficiency of the CECP signal at high CNR is very possible.

\(^1\)The Carson’s Rule bandwidth in Table I and Figure 4 is given in equation (4), where $f_m$ is defined as the 99% bandwidth of the unmodulated signal.

### TABLE I

THE PARAMETERS OF SOME COMMONLY USED CPM SIGNALS.

<table>
<thead>
<tr>
<th>CPM Type</th>
<th>MSK</th>
<th>GMSK</th>
<th>GMSK</th>
<th>GMSK</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$h$</td>
<td>0.5</td>
<td>0.5</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>$L$</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>N/A</td>
<td>GSM</td>
<td>N/A</td>
<td>DECT</td>
</tr>
<tr>
<td>Standards</td>
<td>N/A</td>
<td>GSM</td>
<td>N/A</td>
<td>DECT</td>
</tr>
</tbody>
</table>

APPENDIX A

We first consider the set $V$ consisting of all the processes having average power $P_m$. Using the equivalent CECP-AWGN channel model in Figure 2 and the well known water-filling strategy[6], we have

\[ C_v(P_m) \equiv \lim_{T \to \infty} \frac{1}{T} \sup_{m(t) \in V} I(M^N; Z^N) \]

\[ = \frac{1}{2} \int_{-\alpha_o}^{\alpha_o} \log_2 \left( \frac{\pi^2 N_0 z^2}{4 z^2 - N_0} \right) df \]

\[ = \frac{2}{\ln 2} \alpha_o \tag{29} \]

where

\[ \alpha_o \equiv \left( \frac{3A^2 \Delta^2 P_m}{2N_0} \right)^{\frac{1}{2}} \tag{30} \]

For our convenience, we define the $C_v(P_m)$ achieving process to be $m^*(t)$. The power spectrum of $m^*(t)$ after water-filling is shown in Figure 5(a). Note that the one-sided strict bandwidth of $m^*(t)$ is $\alpha_o$. Now consider the set $S \subseteq V$ consisting of all the processes having average power $P_m$ and strict bandwidth $f_m$. Using (29) and water-filling, we have

\[ C_s(P_m, f_m) \equiv \lim_{T \to \infty} \frac{1}{T} \sup_{m(t) \in S} I(M^N; Z^N) \leq \frac{2}{\ln 2} \alpha_o \tag{31} \]
Let us define $k = f_m / \alpha_o$, and represent $C_s(P_m, f_m)$ as a function of $\alpha_o$ and $k$. We know that

$$C'_s(\alpha_o, k) \equiv C_s\left(\frac{2N_q\alpha_o^3}{3Q^2\Delta^2}, \alpha_o k\right) \leq \frac{2}{\ln 2} \alpha_o$$

and the equality holds when $k = K_o$. If $k \leq K_o$, then by water-filling, we have (after some computation)

$$C'_s(\alpha_o, k) = \frac{\alpha_o k}{\ln 2} \left[2f(k, x_o) + \ln(1 + \frac{2}{3}(\kappa^2f^2(k, x_o) - f(k, x_o)))\right]$$

where $f(k, x_o)$ is defined in (24). Since we need

$$\Delta \sqrt{P_m} + f_m \leq B$$

after some computation, we have

$$\alpha_o \leq Bq(\rho, k)$$

where $q(\rho, k)$ is defined in (27). Note that $q(\rho, k)$ is a monotonic decreasing function of $k$, and $C'_s(\alpha_o, k)$ is a non-decreasing function of $\alpha_o$. We conjecture that $C'_s(\alpha_o, k)$ is also a non-decreasing function of $k$ when $k \leq K_o$, which is strongly supported by numerical results. Accepting this conjecture, we have

$$C = \max_{0 \leq \alpha_o \leq Bq(\rho, k)} C'_s(\alpha_o, k)$$

$$= B \max_{0 \leq k \leq K_o} \frac{q(\rho, k)k}{\ln 2} \left[2f(k, x_o) + \ln(1 + \frac{2}{3}(\kappa^2f^2(k, x_o) - f(k, x_o)))\right]$$

Note that given $x_o$, the capacity is again an explicit function of $B$ and $\rho$, and the bandwidth efficiency $R = C/2B$ is an explicit function of $\rho$ only.

**REFERENCES**


