On the Distribution of the Peak-to-Average Power Ratio in OFDM Signals

Hideki Ochiai, Student Member, IEEE, and Hideki Imai, Fellow, IEEE

Abstract—The distribution of the peak-to-average power ratio (PAPR) in strictly band-limited orthogonal frequency-division multiplexing (OFDM) signals is studied. Assuming that the baseband OFDM signal is characterized as a band-limited complex Gaussian process, we first attempt to derive the exact distribution of the PAPR in the band-limited OFDM signals. Since this distribution cannot be expressed in a closed form, we further develop a simple closed-form approximation, based on the level-crossing rate analysis. Comparisons of the proposed distributions with those obtained by computer simulations show good agreement and convergence with an increase in the number of subcarriers.

Index Terms—Distribution, level-crossing rate, orthogonal frequency-division multiplexing (OFDM), peak-to-average power ratio (PAPR).

I. INTRODUCTION

Due to its robustness against multipath fading with certain implementation advantages over single-carrier systems, coded orthogonal frequency-division multiplexing (OFDM) has become a popular technique in various high-speed wireless data transmission systems [1], [2]. One of the major drawbacks of OFDM signals (or more generally, multicarrier signals) is its waveform with a very high peak-to-average power ratio (PAPR). OFDM signals therefore cause serious problems such as a severe power penalty at the transmitter, which is prohibitive for use in portable wireless systems where the terminals are powered by battery [3]. Recently, a number of PAPR reduction schemes have been proposed to alleviate this undesirable property of OFDM signals (see, e.g., [4]–[6]).

To evaluate the capability of PAPR reduction schemes or design systems involving nonlinear devices, it is necessary to understand the properties of the PAPR in OFDM signals. If all the subcarriers are modulated by phase-shift keying (PSK), the theoretical upper bound of the PAPR in OFDM signals with $N$ subcarriers is $N$. Thus, the probability of observing such a PAPR is $P_\text{max}^2/N = 2^{-N}$. Even if $N$ is as small as, say 32, and quadrature PSK (QPSK) modulation is employed ($M = 4$), this probability is only $8.7 \times 10^{-19}$. (If the symbol period of the OFDM system is 100 $\mu$s, the theoretically maximum PAPR will be observed statistically only once in 3.7 million years!) Therefore, the upper bound may not be meaningful for characterizing the PAPR of OFDM signals, and the statistical distribution of the PAPR should be taken into account.

Recently, the distribution of the PAPR of OFDM signals has been derived by several researchers [3]–[5], in conjunction with evaluation of the PAPR reduction schemes. However, simulation results have shown that the analysis found in [3]–[5], which we will refer to as a conventional analysis, considerably underestimates the distribution of the PAPR of the band-limited OFDM signals obtained by computer simulations. (In [5], therefore, an empirical modification has been proposed.) Thus, a more accurate expression with theoretical justification is needed.

In this paper, assuming the baseband OFDM signal as a band-limited complex Gaussian process and the peaks in the signals statistically uncorrelated, we derive the exact distribution of the PAPR in the band-limited OFDM signals, as well as a simple closed-form approximation based on the theory of the level-crossing rate analysis of the envelope process.

II. SYSTEM MODEL

We consider a simplified OFDM transmitter shown in Fig. 1. The effect of the guard interval required to alleviate the intersymbol interference is not considered for brevity. Complex data sequences of length $N$ are input to the inverse discrete Fourier transform (IDFT). Let $x(t)$ and $y(t)$ denote the real and imaginary parts of the output signal after the ideal low-pass filter.
A complex baseband signal, defined over the time interval \( t \in [0, T_s] \), can be expressed as

\[
\tilde{s}(t) = x(t) + jy(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} A_k e^{j2\pi(k-(N-1)/2)/T_s} t
\]  

(1)

where \( A_k \) is the complex data of the \( k \)th subcarrier and \( T_s \) is the OFDM symbol period. Without loss of generality, the \( A_k \) are assumed to be statistically independent, identically distributed (i.i.d.) random variables with zero mean and variance \( \sigma^2 = \mathbb{E}[|A_k|^2] \).

The PAPR of the baseband OFDM signals \( \mathcal{P} \) can be defined as

\[
\mathcal{P} \triangleq \frac{\max_{0 \leq t \leq T_s} |\tilde{s}(t)|^2}{P_{av}}
\]  

(2)

where \( P_{av} = \sigma^2 \) corresponds to the average power of the baseband OFDM signals. For mathematical convenience, we alternatively consider the crest factor (CF) \( C \), which is a square-root of the PAPR, i.e.,

\[
C \triangleq \sqrt{\mathcal{P}} = \max_{0 \leq t \leq T_s} \frac{|\tilde{s}(t)|}{\sqrt{P_{av}}} = \max_{0 \leq t \leq T_s} r(t)
\]  

(3)

where

\[
r(t) \triangleq \frac{|\tilde{s}(t)|}{\sqrt{P_{av}}} = \sqrt{x^2(t) + y^2(t)}/P_{av}
\]  

(4)

is the envelope of the complex baseband OFDM signal normalized by the average power. In the following, without loss of generality, we assume that the two-sided bandwidth of the signals \( x(t) \) and \( y(t) \), denoted by \( W \), is fixed, and the increase of the number of subcarriers results in the proportional increase of the symbol period \( T_s \), since \( T_s \approx N/W \) for large \( N \).

Note that in many OFDM systems with a large number of subcarriers, pulse shaping is not usually employed [2], since the power spectral density (PSD) of the band-limited OFDM signal is approximately rectangular. The amplitude of radio-frequency (RF) OFDM signals can be written as

\[
|s(t)| = |\Re \{ \tilde{s}(t) e^{j2\pi f_c t} \}|
\]  

(5)

where \( f_c \) is the carrier frequency of RF signals (see Fig. 1). Since \( f_c \gg 1/T_s \) in most wireless applications, the peak power of RF signals may be equivalent to that of the complex baseband signals. Therefore, in what follows, we only consider the PAPR of the baseband OFDM signals.

In order to proceed a theoretical analysis of OFDM signals, some assumptions may be necessary. In this paper, it is assumed that the baseband OFDM signal of (1) converges to a complex Gaussian random process for large \( N \). A stringent evaluation of the convergence of the process may be beyond the scope of this paper. However, one may at least confirm the pointwise convergence of the OFDM signal to Gaussian;\(^1 \) at any given time instant \( t_\ell \in [0, T_s] \), the two random variables \( x(t_\ell) \) and \( y(t_\ell) \) are expressed as

\[
x(t_\ell) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} A_k \cos \left( 2\pi \frac{k-(N-1)/2}{T_s} t_\ell + \arg A_k \right)
\]

\[
y(t_\ell) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} A_k \sin \left( 2\pi \frac{k-(N-1)/2}{T_s} t_\ell + \arg A_k \right)
\]  

(6)

and since the above random variables \( x(t_\ell) \) and \( y(t_\ell) \) are given by the sum of \( N \) i.i.d. random variables with zero mean and variance \( \mathbb{E}[|A_k|^2]/2 = \sigma^2/2 \), both \( x(t_\ell) \) and \( y(t_\ell) \) asymptotically become Gaussian with zero mean and variance \( \sigma^2/2 \) for large \( N \) by the central limit theorem.

With this assumption in mind, for any \( t_\ell \in [0, T_s] \), \( x(t_\ell) \) and \( y(t_\ell) \) are uncorrelated Gaussian random variables and the envelope \( r(t_\ell) \) becomes a Rayleigh random variable. In the following, the envelope process \( r(t) \) of a complex Gaussian process will be thus referred to as a Rayleigh process for simplicity.

III. CONVENTIONAL ANALYSIS OF THE PAPR OF OFDM SIGNALS

Consider the baseband OFDM signals of (1) sampled at the Nyquist rate. Let \( \tilde{s}_l \) denote the complex sample of \( s(lT_s/N) \), and let \( x_l \) and \( y_l \) be the corresponding samples such that \( \tilde{s}_l = x_l + jy_l \). These samples may be alternatively obtained by the \( N \)-point IDFT with the \( N \) complex inputs \( A_k \). Then, the corresponding amplitude of the envelope normalized by the average power is given by \( r_l \triangleq |\tilde{s}_l|/\sqrt{P_{av}} \). From (1), it may be straightforward to show that

\[
\mathbb{E}[\tilde{s}_l \tilde{s}_m^*] = \begin{cases} \sigma^2, & \text{for } l = m \\ 0, & \text{for } l \neq m \end{cases}
\]  

(7)

and thus the \( \tilde{s}_l \) are mutually uncorrelated\(^2 \). Consequently, it can be shown that

\[
\mathbb{E}[x_l x_m] = \mathbb{E}[y_l y_m] = \begin{cases} \sigma^2/2, & \text{for } l = m \\ 0, & \text{for } l \neq m \end{cases}
\]  

(8)

By assumption, the processes \( x(t) \) and \( y(t) \) are asymptotically Gaussian for large \( N \), and the uncorrelated samples \( x_l \) and \( y_l \) become independent Gaussian random variables, due to the fact that uncorrelated Gaussian random variables are statistically independent. This allows the \( \tilde{s}_l \) to be statistically independent and thus the \( r_l \) to be i.i.d. Rayleigh random variables of which the probability density function (pdf) is

\[
f_{r_l}(r) = 2r e^{-r^2}.
\]  

(9)

Therefore, the cumulative distribution function (cdf) of the CF can be given by

\[
F_{\mathcal{P}}(r) = \Pr(\max_{0 \leq l \leq N} r_l < r) = \Pr(r_0 < r) \cdot \Pr(r_1 < r) \cdots \cdot \Pr(r_{N-1} < r) = (1 - e^{-r^2})^N.
\]  

(10)

\(^1\)The pointwise convergence of the random variables \( x(t_\ell) \) does not imply that the finite-dimensional distributions of \( x(t) \) converge to those of a jointly Gaussian process.

\(^2\)In [3], it is stated that the IDFT of statistically independent inputs will produce statistically independent outputs. This may not be necessarily true unless the inputs are Gaussian.
Simply changing the variable as \( \lambda = r^2 \) yields the cdf of the PAPR

\[
F_P(\lambda) = (1 - e^{-\lambda})^N. \tag{11}
\]

The distributions obtained by the conventional analysis, however, does not fit those of the PAPR of the band-limited OFDM signals obtained by computer simulations [3], [4], even for very large \( N \). The reason for this may be stated as follows. The conventional analysis yields the maximum sample of OFDM signals, but this does not necessarily correspond to the maximum peak of the band-limited OFDM signals, even though the obtained sample may be close to the maximum peak. As a result, the conventional analysis underestimates the distribution of the PAPR in the band-limited OFDM signals.

In [5], van Nee and de Wild gave an empirical approximation

\[
F_P(\lambda) \approx (1 - e^{-\lambda})^{\alpha N} \tag{12}
\]

where \( \alpha \) is a parameter determined by computer simulation to be 2.8. It should be noted that this approximation not only lacks theoretical justification but also yields some discrepancies with the simulation results for large \( N \), as is shown in Section V.

IV. ANALYSIS OF THE PEAK DISTRIBUTION OF OFDM SIGNALS

The mathematical analysis of the peaks in random signals dates back to the early work of Rice [7], [8] and Cartwright and Longuet-Higgins [9]. Rice [7], [8] carried out the mathematical analysis of peaks of Gaussian and Rayleigh processes, and the results were extended to the analysis of extreme peaks by Cartwright and Longuet-Higgins [9] for a Gaussian process (see, e.g., [10]).

Since both \( x(t) \) and \( y(t) \) of (6) are the sums of orthogonal sinusoids with random phases, they are wide-sense stationary. Thus, in the following, we assume that \( x(t) \) and \( y(t) \) are independent stationary band-limited Gaussian processes and consequently \( r(t) \) is a stationary band-limited Rayleigh process. We start with the derivation of the exact peak distribution of the band-limited Rayleigh process.

A. Exact Distribution of the CF

The probability that an arbitrary peak \( \rho \) in one OFDM symbol is above the level \( r \) may be given by

\[
Pr(\rho > r) = \frac{\text{the mean number of the peaks above } r}{\text{the mean number of total peaks}} = \frac{\bar{N}_p(r)}{\bar{N}_p(0)} \tag{13}
\]

where \( \bar{N}_p(\alpha) \) is the mean number of the peaks above the level \( \alpha \) in one OFDM symbol. Assuming that \( x(t) \) and \( y(t) \) are uncorrelated, stationary, and ideally band-limited Gaussian processes, the mean number of the peaks above the level \( \alpha \) in one OFDM symbol can be written as (see Appendix B)

\[
\bar{N}_p(\alpha) = \frac{4N}{\sqrt{15\pi}} \int_0^\infty u^2 \int_0^\infty c^{-\alpha u^2} \left\{ e^{-\frac{u^2}{2}} - \frac{\sqrt{\frac{3}{\pi}}}{2} (\alpha - 1) u \text{erfc} \left( \frac{\sqrt{3} u}{2} \right) \right\} du \tag{14}
\]

Suppose that there are \( N_p \) peaks in a sample of OFDM symbols, and let \( \rho_1, \rho_2, \ldots, \rho_{N_p} \) denote the corresponding successive peaks observed at time instants \( t_1, t_2, \ldots, t_{N_p} \), respectively, where \( t_1 < t_2 < \cdots < t_{N_p} \). Then, the CF in this sample symbol is given by

\[
C = \max_{1 \leq i \leq N_p} \rho_i. \tag{15}
\]

To proceed the analysis, the knowledge of statistical behavior of the peak process will be required. Obviously, peaks may not occur at predetermined time instants, and thus the time sequence \( t_1, t_2, \ldots, t_{N_p} \) may be considered as a point process. Let \( \gamma_i = t_i - t_{i-1} \) denote the time interval of the two successive peaks \( \rho_{i-1} \) and \( \rho_i \). Since the sequence of the peak intervals \( \gamma_i \) itself is a random process, statistical characterization of the \( \rho_i \) may require the knowledge of behavior of the \( \gamma_i \).

Unfortunately, very little is known about the theoretical distribution of the exact peak intervals\(^3\) due to its difficulty of analysis. Thus, in order to simplify the analysis, we further make a heuristic assumption that the peaks are statistically mutually uncorrelated, which will be explained as follows. Consider the correlation time \( T_c \) of the random processes \( x(t) \) and \( y(t) \), which may be grossly defined by the reciprocal of the bandwidth of the signals, i.e., \( T_c \approx 1/W \approx T_K/N \), and suppose that the correlation of the stationary processes may rapidly diminish as the interval of observation exceeds \( T_c \). The mean number of peaks in one OFDM symbol period is numerically calculated from (14) as \( \bar{N}_p(0) \approx 0.64N \), which is fewer than \( N \). Thus, with the stationarity of the process in mind, successive peaks may have a temporal separation larger than \( T_c \) with high probability. Therefore, roughly speaking, the values of \( x(t) \) and \( y(t) \) at the sample time where one peak occurs are uncorrelated from the values of \( x(t) \) and \( y(t) \) at sample times where another peak occurs.

Consequently, since the values of \( x(t) \) and \( y(t) \) at the sample times of interest are uncorrelated and are jointly Gaussian, they are statistically independent. Thus, the distribution of the CF in the sample symbol is given by

\[
Pr(C < r) = Pr(\rho < r)^{N_p} \approx (1 - \Pr(\rho > r))^{N_p} \approx (1 - \Pr(\rho > r))^{0.64N}. \tag{16}
\]

Finally, the distribution can be determined by substituting (13) and (14) into (16). We refer to the above equation as an exact distribution. It should be noted, however, that the distribution may be theoretically accurate only if the following two assumptions hold: 1) the processes \( x(t) \) and \( y(t) \) are ideally band-limited Gaussian and 2) peaks are statistically uncorrelated.

Apparently, the exact distribution is not a convenient form to deal with, since one must numerically carry out the double integral of (14). Therefore, in the following, we develop a simpler approximation of the distribution which is asymptotically as accurate as the exact distribution for large \( N \).

\(^3\) For the distributions of the intervals of level-crossings, see, e.g., Rainal [11] and the references therein.
Fig. 2. Envelope of the baseband OFDM signal with several peaks. As the level $\tau$ increases, the number of the positive level-crossings ($\times$) approaches the number of the peaks ($\circ$) that are above $\tau$.

B. Level-Crossing Rate Approximation of the Peak Distribution

We start with deriving the peak distribution of the band-limited Rayleigh process based on the level-crossing rate approximation similar to [12] and [13], and then apply the result to the derivation of the distribution of the CF in OFDM signals.

For a band-limited Gaussian random process, the probability that an arbitrary peak $\rho$ is above the level $\tau$ may be approximated by the ratio of the level-crossings to the zero-crossings, i.e., [12] and [13]

$$\Pr(\rho > \tau) \approx \frac{\nu^+_c(\tau)}{\nu^+_c(0)}$$

(17)

where $\nu^+_c(\alpha)$ is the mean number of the positive (i.e., upward) crossings at the level $\alpha$. This approximation is valid because for a band-limited Gaussian process, the assumption holds with very high accuracy that each cycle (from one positive zero-crossing to the next) has a single positive peak [13]. This approach makes the derivation of the peak distribution possible without dealing with the peak distribution which involves quite complicated expressions.

However, this simplification cannot be applied to the envelope of OFDM signals in a straightforward manner, because the process of our primary interest is not Gaussian. Nevertheless, since the process is band-limited, a similar approach based on the level-crossings may be applicable.

Consider the only peaks that exceed a given threshold $\bar{\tau}$ well above zero. Then, a proper selection of $\bar{\tau}$ makes the following assumption valid: each positive crossing of the level $\bar{\tau}$ has a single positive peak that is above the level $\bar{\tau}$. This is illustrated in Fig. 2, where it is observed that as the reference level $\tau$ increases, the number of the level-crossings with positive slope approaches the number of the peaks above this reference level.

Therefore, the conditional probability of (13) given that a peak exceeds the level $\bar{\tau}$ can be approximated as

$$\Pr(\rho > \tau | \rho > \bar{\tau}) = \frac{\bar{\nu}_c^+(\tau)T_s}{\bar{\nu}_c^+(\bar{\tau})T_s} = \frac{\nu^+_c(\tau)}{\nu^+_c(\bar{\tau})},$$

for $\tau > \bar{\tau}$.

(18)

Since the level-crossing rate $\nu^+_c(\tau)$ of the band-limited OFDM signal is given by (see Appendix A)

$$\nu^+_c(\tau) = \sqrt{\frac{\pi}{3N}}T_s e^{-\tau^2}$$

(19)

(18) reduces to

$$\Pr(\rho > \tau | \rho > \bar{\tau}) \approx \frac{e^{-\tau^2}}{\bar{\nu}_c^+(\tau)}.$$  

(20)

Consequently, the conditional distribution of the peaks can be expressed as

$$\Pr(\rho < r | \rho > \bar{\tau}) \approx 1 - \frac{e^{-\tau^2}}{\bar{\nu}_c^+(\tau)}.$$  

(21)

C. Distribution of the CF Based on the Level-Crossing Approximation

For a given reference level $\bar{\tau}$, the cdf of the CF can be written as [14]

$$F_C(r) = (1 - F_C(\tau))F_C(r | C > \bar{\tau}) + F_C(\tau),$$  

for $r > \bar{\tau}$.

(22)

For large $N$, $\bar{\tau}$ can be chosen such that the highest peak always occurs above the level $\bar{\tau}$, or, in other words, the probability that all the peaks are less than $\bar{\tau}$ becomes negligible, i.e.,

$$F_C(\bar{\tau}) = \Pr(C < \bar{\tau}) \approx 0.$$  

(23)

Since the CF tends to increase with an increase in $N$, the approximation error in (23) becomes less significant as $N$ increases, provided that the level $\bar{\tau}$ is not too high. In what follows, we suppose that $\bar{\tau}$ is chosen such that (23) holds.

Following the derivation of the exact distribution, we assume that the peaks above $\bar{\tau}$ are conditionally independent given that a sample of the envelope process is higher than the reference level $\bar{\tau}$. It should be noted that this assumption of conditional independence is more convincing than the one in the case of the exact distribution, since the number of the peaks to be dealt with is now fewer and thus the expected temporal separation among the peaks of interest becomes greater as the level $\bar{\tau}$ increases. Consequently, the conditional cdf of the CF is given by

$$F_C(r | C > \bar{\tau}) = \Pr(\rho < r | \rho > \bar{\tau}) \bar{\nu}_c^+(\tau),$$

(24)

where $\bar{\nu}_c^+(\tau)$ is the mean number of the peaks above $\bar{\tau}$, which can be approximated for high $\bar{\tau}$ by

$$\bar{\nu}_c^+(\tau) \approx \nu^+_c(\tau)T_s = \sqrt{\frac{\pi}{3N}}T_s e^{-\tau^2}.$$  

(25)

Finally, from (21)–(25), the following simple expression for $F_C(r)$ can be obtained:

$$F_C(r) \approx F_C(r | C > \bar{\tau}) = \begin{cases} 
\sqrt{\frac{\pi}{3N}}T_s e^{-\tau^2}, & \text{for } r > \bar{\tau} \\
0, & \text{for } r \leq \bar{\tau}, 
\end{cases}$$

(26)

The corresponding distribution of the PAPR is obtained by simply changing the variable as $\lambda = r^2$. 
There is one parameter $\varphi$ left to be determined in the above equation. Since the probability of (20) must be less than or equal to 1 for $\forall r \geq \varphi$, the function $f(x) \triangleq xe^{-x^2}$ must be monotonically decreasing for $\forall x \geq \varphi$. This constraint provides a lower bound of the reference level as $\varphi \geq (1/\sqrt{2}) = 0.71$.

The effect of $\varphi$ on the accuracy of the distribution will be numerically evaluated in the next section.

V. NUMERICAL RESULTS

Computer simulations were performed to obtain the cdf of the band-limited OFDM signals. In the subsequent simulations, 1 000 000 OFDM symbols were randomly generated. The continuous signal after the ideal LPF was approximated by at least 16-time oversampling.\(^4\)

A. Effect of the Reference Level $\varphi$

We examine the effect of the reference level $\varphi$. This value may be determined by comparing the resultant distribution with either the exact distribution obtained by numerical integration or the distribution obtained by simulations. Now, we choose $\varphi$ such that the cumulative distribution of the CF of (26) agrees with the simulation results at a certain target point. In the following, as an example, the value of $r$ at $F_C(r) = 0.001$ will be chosen as a target.

In Fig. 3, (26) is plotted as a function of $\varphi$ for $N = 100, 500, 1000$, along with the corresponding simulation results at $F_C(r) = 0.001$ where the subcarriers are modulated by QPSK or 16QAM (rectangular). Also shown in this figure is the value $r$ of the exact distribution (16) at $F_C(r) = 0.001$.

The distribution based on the proposed approximation seems to agree with the simulation result of QPSK for $\varphi$ around 1.7–1.8. It appears that $\varphi = \sqrt{\pi} \approx 1.77$, which is twice the mean of the Rayleigh process $\gamma(t)$, shows a good agreement. Therefore, we choose $\varphi = \sqrt{\pi}$ for QPSK, as plotted in Fig. 3. For 16QAM, however, slightly lower $\varphi$ will be recommended.

Note that in Fig. 3 the discrepancies are observed among QPSK, 16QAM, and the exact distributions particularly for $N = 100$. This may be due to the fact that the Gaussian approximation of the processes may not be accurate enough, since the input data $A_k$ is not complex Gaussian but has a constraint on its signal constellation in practice. As the signal constellation of $A_k$ approaches two-dimensional Gaussian (or contains higher entropy under a unit power constraint on $A_k$), the convergence of the processes to Gaussian may be faster than the PSK-type constellation. However, for large $N$, the constellation of each subcarrier becomes irrelevant to the resultant distribution of the CF.

Note that for high $r \gg \varphi$, the distribution of the CF becomes less sensitive to the value of the reference $\varphi$, as is shown in the following.

B. Cumulative Distribution Function of the CF

Fig. 4 compares the cdfs of the CF of (26) with $\varphi = \sqrt{\pi}$ and the exact distribution of (16) along with simulation results for $N = 100, 500, 1000$.

\(^4\)In this paper, an $n$-time oversampling refers to the ideal interpolation sampled at $n$ times the Nyquist rate.

Fig. 3. Threshold crest factors at $F_C(r) = 0.001$ of the proposed distributions as a function of the reference level $\varphi$, along with the corresponding simulation results.

Fig. 4. Comparison of the proposed distributions of the crest factor with simulation results for $N = 100, 500, 1000$.

For QPSK, however, slightly lower $\varphi$ will be recommended. Note that in Fig. 3 the discrepancies are observed among QPSK, 16QAM, and the exact distributions particularly for $N = 100$. This may be due to the fact that the Gaussian approximation of the processes may not be accurate enough, since the input data $A_k$ is not complex Gaussian but has a constraint on its signal constellation in practice. As the signal constellation of $A_k$ approaches two-dimensional Gaussian (or contains higher entropy under a unit power constraint on $A_k$), the convergence of the processes to Gaussian may be faster than the PSK-type constellation. However, for large $N$, the constellation of each subcarrier becomes irrelevant to the resultant distribution of the CF.

Note that for high $r \gg \varphi$, the distribution of the CF becomes less sensitive to the value of the reference $\varphi$, as is shown in the following.

C. Asymptotic Form of the Distribution for High $r$

In practice, $F_C(r)$ with high value of $r$, or the complementary cdf $1 - F_C(r) = \Pr(C > r)$ is of particular interest. From Fig. 4, it is observed that as $r$ increases, both the approximate and exact distributions approach the simulation results. The distribution appears to become less sensitive to the value of $\varphi$ as $F_C(r)$ approaches 1.

Indeed, for high $r$, the conditional probability that the peak exceeds $r$ may be very small, i.e., $\Pr(\rho > r | \rho > \varphi)$ of (20) approaches 0. In this case, (26) can be further simplified by using
a limiting form of the exponential function (see, e.g., [10, p. 180]) as
\[ F_C(r) \approx \exp\left(-\sqrt{\frac{2}{3}} N r e^{-r^2/2}\right), \quad \text{for } \Pr(r > r | r > \bar{r}) \to 0 \]
which does not depend on \( \bar{r} \).

We note that all the three complementary cdfs obtained from (16), (26), and (27) become quite indiscernible for large \( N \) and relatively high values \( r \) of interest.

D. Comparison of the Accuracy

We evaluate the accuracy of the approximation with respect to the simulation results and also the empirical approximation of [5], i.e., (12). In order to closely examine the accuracy of the distribution for high \( r \), the complementary cdf is calculated.

The results are shown in Fig. 5 for \( N = 100, 500, \) and 1000, with simulation results employing 16QAM modulation. As an approximation model, only the asymptotic form of the cdf (27) is shown for all \( N \). As we noted, however, both the other distributions (16) and (26) look exactly the same. As expected, the simulation and proposed approximation results converge as \( N \) increases, due to the improvement of the accuracy of the Gaussian approximation. It should be noted that there are noticeable discrepancies between the distributions of (12) and those of simulations for large \( N \).

The accuracy of the results may also suggest that the assumptions of Gaussian process approximation and statistical independence of the peaks made in this paper are reasonable for large \( N \).

E. Complementary CDF of OFDM Signals with a Large Number of Subcarriers

For large \( N \), say \( N > 5000 \), even computer simulations become difficult to perform in order to obtain accurate distributions of the PAPR. Thus, in Fig. 6, we show the complementary cdf using (27) for \( N \) from 100 to 100 000. For practical interest, PAPR in decibels is used as a measure.

It is observed that even if \( N \) is as large as 100 000, the probability of the PAPR exceeding 15 dB is only about 10\(^{-8}\), which may be negligibly low for many applications. Although the PAPR of OFDM signals is in general much higher than that of single-carrier signals, the results show that the PAPR of OFDM signals is practically not so sensitive to the increase in the number of subcarriers.

VI. CONCLUSION

Assuming that the baseband OFDM signal is characterized as a complex Gaussian process and the peaks in the envelope process are statistically uncorrelated, we have derived the distributions of the PAPR in the band-limited OFDM signals based on the level-crossing rate approximation of the peak distribution along with the exact distribution. By appropriately adjusting the reference level, the resultant approximation of the PAPR distribution can be made accurate for a relatively large number of subcarriers. If the range of the PAPR of interest is high, the distribution can be further simplified without loss of the accuracy. Even though our analysis is based on the aforementioned assumptions, rigorous comparisons by computer simulation may suggest that these heuristic assumptions are reasonable.

Finally, in spite of the fact that the theoretical upper bound of the PAPR in OFDM signals is proportional to the number of subcarriers, it has been shown that the statistical distribution of the PAPR of OFDM signals is not so sensitive to the increase in the number of subcarriers.

APPENDIX A

LEVEL-CROSSING RATE OF THE BAND-LIMITED OFDM SIGNALS

In this appendix, we derive the level-crossing rate of the band-limited OFDM signals. First, the joint pdf of the envelope \( r(t) \) and its time-derivative \( y(t) \) is derived. The envelope \( r(t) \) of (4) is given by

\[ r(t) = \sqrt{\frac{x^2(t) + y^2(t)}{2\sigma^2_x}} \]  \hspace{1cm} (A.1)
where \( x(t) \) and \( y(t) \) are statistically independent stationary Gaussian random processes with zero mean and variance \( \sigma^2_x \).

Since the derivative operation is linear, the derivatives of \( x(t) \) and \( y(t) \) are also Gaussian random processes with zero mean and variance \( \sigma^2_x \).

Let \( x, \dot{x}, y, \dot{y} \) denote the samples of the Gaussian processes \( x(t), \dot{x}(t), y(t), \dot{y}(t) \), respectively, at the same time instant. The joint pdf of \( x, \dot{x}, y, \dot{y} \) is written as

\[
 f_{x,y}(X) = \frac{1}{\sqrt{(2\pi)^n|R|}} \exp\left[-\frac{1}{2}XR^{-1}X'\right] \tag{A.2}
\]

where \( X = [x, \dot{x}, y, \dot{y}] \), \( n = 4 \), \( R \) is the covariance matrix, and \( |R| \) is the determinant of \( R \). Since \( x(t) \) and \( y(t) \) are uncorrelated and \( E[x^2] = (1/2)(d/dt)E[\dot{x}^2] = 0 \), the covariance matrix is given by

\[
 R = \begin{bmatrix}
 \sigma_x^2 & 0 & 0 & 0 \\
 0 & \sigma_x^2 & 0 & 0 \\
 0 & 0 & \sigma_y^2 & 0 \\
 0 & 0 & 0 & \sigma_y^2
\end{bmatrix}. \tag{A.3}
\]

By changing the variables to polar coordinates as

\[
 x = \sqrt{2\sigma_x^2}r \cos \theta \\
 y = \sqrt{2\sigma_y^2}r \sin \theta,
\]

the joint pdf of \( r, \theta, \dot{r}, \dot{\theta} \), is given by

\[
 f_{r,\theta}(r, \theta) = \frac{r^2}{\pi K} \exp\left[-\frac{r^2}{2} - \frac{1}{K} (r^2 + \theta^2)\right] \tag{A.5}
\]

where

\[
 K \triangleq \frac{\sigma_y^2}{\sigma_x^2}. \tag{A.6}
\]

Integrating \( \theta \) from 0 to \( 2\pi \) and \( \dot{\theta} \) from \(-\pi \) to \( \pi \), we obtain the joint pdf of \( r \) and \( \dot{r} \)

\[
 f_r(r, \dot{r}) = 2r e^{-r^2} \frac{1}{\sqrt{\pi K}} e^{-\dot{r}^2}. \tag{A.7}
\]

Now we determine \( K \) for the band-limited OFDM signals. Let \( R_x(\tau) \) and \( S_x(f) \) be the autocorrelation function and the PSD of \( x(t) \), respectively. Then, the variances \( \sigma_x^2 \) and \( \sigma_y^2 \) can be written as [10], [14]

\[
 \sigma_x^2 = R_x(0) = \int_{-\infty}^{\infty} S_x(f) df \tag{A.8}
\]

and

\[
 \sigma_y^2 = R_y(0) = \int_{-\infty}^{\infty} (2\pi f)^2 S_x(f) df. \tag{A.9}
\]

Since the PSD of the band-limited OFDM signals is nearly rectangular, \( S_x(f) \) can be assumed constant over the entire bandwidth, i.e.,

\[
 S_x(f) \approx 0, \quad \text{for } |f| > W/2.
\]

Substituting this into \( A.9 \) yields

\[
 \sigma_y^2 = \frac{\pi^2}{3} W^2 \sigma_x^2. \tag{A.10}
\]

Consequently, from \( A.6 \) we have

\[
 K = \frac{\pi^2}{3} W^2. \tag{A.12}
\]

The level-crossing rate, which is the mean number of positive crossings of the process \( r(t) \) at the level \( r = a \) per unit time, is given by [7, eq. (3.3-5)] or [8, eq. (4.1)]

\[
 \nu^c_r(a) = \int_0^\infty \dot{r} f_r(a, \dot{r}) d\dot{r}. \tag{A.13}
\]

Substituting \( \dot{r}(r, \dot{r}) \) of \( A.7 \) and carrying out the integration, we have

\[
 \nu^c_r(a) = \sqrt{\frac{K}{\pi}} a e^{-a^2}. \tag{A.14}
\]

Substituting \( K \) of \( A.12 \) and noticing that \( W \approx (N/T_s) \), we obtain the level-crossing rate of \( r(t) \)

\[
 \nu^c_r(a) = \sqrt{\frac{1}{3 N T_s}} a e^{-a^2}. \tag{A.15}
\]

### Appendix B

**Number of Peaks in the Band-Limited OFDM Signals**

Since the derivatives of \( \dot{x}(t) \) and \( \dot{y}(t) \) are also Gaussian with zero mean, variance \( \sigma_{\dot{x}}^2 \triangleq E[\dot{x}^2(t)] \), \( E[\dot{x}(t)\dot{y}(t)] = (1/2)(d/dt)E[\dot{x}^2(t)] = 0 \), and \( E[x(t)\dot{y}(t)] = (d/dt)E[x(t)\dot{x}(t)] = E[\dot{x}^2(t)] \), the joint pdf of jointly Gaussian random variables \( x, \dot{x}, y, \dot{y}, \dot{y}, \dot{y} \), which are the samples of the Gaussian processes \( x(t), \dot{x}(t), y(t), \dot{y}(t), \dot{y}(t), \dot{y}(t) \) at the same time instant, is given by \( A.2 \) with \( X = [x, \dot{x}, y, \dot{y}, \dot{y}, \dot{y}] \), \( n = 6 \), and

\[
 R = \begin{bmatrix}
 \sigma_x^2 & 0 & 0 & 0 & 0 & 0 \\
 0 & \sigma_x^2 & 0 & 0 & 0 & 0 \\
 0 & 0 & \sigma_y^2 & 0 & 0 & 0 \\
 0 & 0 & 0 & \sigma_y^2 & 0 & 0 \\
 0 & 0 & 0 & 0 & \sigma_y^2 & 0 \\
 0 & 0 & 0 & 0 & 0 & \sigma_y^2
\end{bmatrix}. \tag{B.1}
\]

Changing the variables as in \( A.4 \) and integrating out the variables \( \theta, \dot{\theta}, \dot{\theta} \), we obtain the joint pdf of the envelope

\[
 f_r(r, \theta, \dot{r}) = \int_{-\infty}^{\infty} 2^{M} 2^{M} (K\pi)^{\frac{1}{2}} \exp\left[-\frac{1}{2} (r^2 + \theta^2)\right] \times \exp\left[-\frac{1}{2} (1 + M + (1 - 2M)\phi^2 + M\phi^4) \right] d\phi
\]

\[
 = \frac{1}{\sqrt{\pi K}} \exp\left[-\frac{1}{2} (r^2 + \frac{M}{K} \phi^2)\right] d\phi
\]

where

\[
 M \triangleq \frac{\sigma_{\phi}^2}{\sigma_x^2 - \sigma_y^2}. \tag{B.3}
\]

Note that in \( B.2 \), we have changed the variable \( \phi \triangleq \sqrt{K} \theta \) and left the integral of \( \phi \) for mathematical convenience (see the recent work by Blachman [15]).
The rate density of peaks, which is the mean number of peaks of the process \( r(t) \) in the region \( u \leq r \leq u + du \) per unit time, is given by [7, eq. (3.8-1)]

\[
\nu_p(u) \, du = du \int_{-\infty}^{0} \frac{-i}{\pi} f_s(u, 0, \nu) \, d\nu. \tag{B.4}
\]

Substituting \( f_s(r, \nu, \nu') \) of (B.2) and after some algebra, (B.4) can be rewritten as

\[
\nu_p(u) \, du = du \frac{2}{\pi} \sqrt{\frac{K}{M}} \int_{0}^{\infty} e^{-\left(\nu^2+1\right)u^2} \left\{ e^{-\left(M\nu^2-1\right)u^2} - \sqrt{\pi M (\nu^2 - 1)} \text{erfc} \left( \sqrt{M (\nu^2 - 1)} u \right) \right\} \, d\nu, \tag{B.5}
\]

With

\[
\sigma_x^2 = R_x(0) = \int_{-\infty}^{\infty} (2\pi f)^4 S_x(f) \, df = \frac{\pi^4}{3} W^4 \sigma_z^2, \tag{B.6}
\]

and from (A.11) and (B.3), we obtain \( M = 5/4 \).

Consequently, the mean number of peaks exceeding the level \( \alpha \) in one OFDM symbol can be given by

\[
\mathcal{N}_p(\alpha) = T_s \int_{\alpha}^{\infty} \nu_p(u) \, du, \tag{B.7}
\]

which leads to (14). The above integration may not be expressed in a simple closed form and needs to be carried out numerically.

ACKNOWLEDGMENT

The authors are indebted to the anonymous reviewers for their careful reading of the manuscript and detailed comments and suggestions that significantly improved the quality of this paper.

REFERENCES