Paper

Normalized Gaussian Approach to Statistical Modeling of OFDM Signals

Adam Rudziński

SKA Polska Sp. z o.o., Warsaw, Poland

Abstract—This article concerns modeling of statistical properties of OFDM signals with the help of "normalized Gaussian" model, proposed by Kotzer *et al.* In this paper there is provided an extended formulation of the model, supplemented by an expression for probability density, extending possible applications of the model in theoretical works. Numerical results for verification of the model are provided and a more accurate alternative is suggested.

Keywords—M-PSK, OFDM, sample value distribution, statistical model.

1. Introduction

In nowadays world there exists a demand for novel technical solutions introducing new qualities and functionalities into everyday life. The recent 10–20 years have led to an enormous technical progress, in particular in the field of electronics and telecommunications – global computer network or mobile telephony are good examples of developments which had a great impact on the way of living. Still, the pursuit of new technical solutions and higher performance continues.

The development in telecommunications was to large extent related to practical implementation of Orthogonal Frequency Division Multiplexing (OFDM) technique. It was first proposed and examined over 40 years ago [1]–[3], but it was impossible to use it without fast digital signal processing circuits available today. Currently this technique is used in, e.g., wireless LAN networks, terrestrial digital television and optical fibre communications, so the possibility of increasing data throughput depends largely on adaptation of devices for operation with larger numbers of subcarriers (increased bandwidth) and higher-order modulation schemes, i.e., with more densely populated constellations. A relevant example is the new DVB-T2 standard, including 256-QAM modulation, in contrast to the older DVB-T, limited to 64-QAM. Such advancement requires the understanding and proper modeling of phenomena occurring in signal transmission. In particular, this can be achieved by construction of analytic models, which explicitly reveal fundamental properties and relations in the system.

The conventional way of modeling statistical properties of the OFDM signal is to invoke the Central Limit Theorem and assume for the signal value x a zero-mean Gaussian probability distribution

$$G_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{1}$$

with variance σ^2 equal to the mean power of the signal. This approach usually leads to sufficiently accurate results, especially if a large number of subcarriers is in application. Still, this model is obviously an approximation, since it does not limit the signal value. It predicts non-zero probability even for nonphysically large x (in the so-called distribution tails). As such, it tends to become inaccurate in some cases, where suppressing of the distribution tails is of high importance, for example, in calculation of clipping noise when the clipping is small. Clearly, advanced modeling should be based on a model better resembling properties of the signal, but such models, if known at all, are very rarely found in the literature.

As an attempt to overcome the problem of unlimited values, Kotzer, Har-Nevo, Sodin and Litsyn have proposed another statistical model, which they have called "normalized Gaussian" model [4]. It is applicable to OFDM signal with subcarriers modulated according to a constant amplitude constellation (e.g. M-PSK). This model is based on such modification of the Gaussian distribution given by Eq. (1), which explicitly limits the values and assures preservation of energy in the whole OFDM symbol, resulting in inherent confinement of signal values within a finite range. Hence, it offers a potentially advantageous alternative to conventional Gaussian model. However, the original formulation of the normalized Gaussian model makes it suitable only for problems of specific kind, with limited usability for analytic derivations. These issues are addressed in this work, the aim of which is to extend or modify the formulation of the model and examine its most important properties.

This paper is organized as follows. Section 2 defines the notation and introduces the concept of the normalized Gaussian model according to formulation by Kotzer *et al.* [4]. Next, the model is reformulated in a more general and analytic way in Section 3. In Section 4 there is provided a numerical verification of the model. Section 5 proposes an improved model, obtained as a combination of the conventional and the normalized Gaussian models. An example of application and comparison of all these models is presented in Section 6, and Section 7 concludes the paper.

2. OFDM Signal and the Normalized Gaussian Model

In a modern communication device, transmitted data are mapped onto points ζ_m of a constellation, i.e. set

$$\Omega = \{\zeta_m \in \mathbb{C}; m = 1, \dots, \|\Omega\|\}, \qquad (2)$$

where $\|\Omega\|$ denotes the number of elements in Ω. A signal with Orthogonal Frequency Division Multiplexing (OFDM) is generated by Inverse Discrete Fourier Transformation (IDFT) of a vector $\boldsymbol{A} = \left[A_1 e^{j\phi_1} \dots A_K e^{j\phi_K}\right] \in \Omega^K$ containing K complex values representing spectral components of the signal (subcarriers), with amplitudes $A_k > 0$ and phases $\phi_k \in [0; 2\pi)$, where $k = 1, \dots, K$. The result of the transformation constitutes the time-domain representation of the signal. In general, this signal is complex, but for physical processing, e.g., for digital-to-analog conversion, it is split into two real signals, given by its real and imaginary parts, and each of them is processed at least in a part separately. Here it is assumed, that these two signals have the same statistical properties and further only the real part of the complex time-domain signal is considered. It can be represented as a sequence of N_S samples

$$x_i \equiv x(iT), \quad i = 0, 1, \dots, N_S - 1, \qquad (3)$$

of the continuous time-domain signal

$$x(t) = \sum_{k=1}^{K} A_k \cos(\omega_k t + \phi_k), \qquad (4)$$

where T is the sampling period and ω_k denotes angular frequency of the k-th subcarrier. The signal is divided into symbols of duration T_S (in practical applications extended by guard interval, which is not relevant here and therefore ignored), so to preserve orthogonality of subcarriers their angular frequencies are chosen as

$$\omega_k = k\omega_1, \quad \text{with } \omega_1 = \frac{2\pi}{T_S}. \qquad (5)$$

In general, for consideration of statistical properties of the signal, amplitudes A_k are represented by a random variable A with the probability distribution

$$f_A(A) = \frac{1}{\|\Omega\|} \sum_{m \in \Omega} \delta(A - |\zeta_m|), \qquad (6)$$

and phases ϕ_k are treated as independent random variables, distributed uniformly on a subset of the interval $[0; 2\pi)$. The mean power of the OFDM symbol defined by Eqs. (3) or (4) is equal to

$$\sigma^2 = K\frac{\langle A^2 \rangle}{2}. \qquad (7)$$

Here, for consideration of only M-PSK constellations, it is assumed that each $|\zeta_m| = \frac{1}{K}$. With this choice the signal is scaled to obtain values from the range $[-1; 1]$ and its mean power becomes $\sigma^2 = \frac{1}{2K}$.

The limitation of signal values is violated in the conventional model based on the Central Limit Theorem and Gaussian probability distribution, what, as mentioned above, may cause this model to be inadequate for modeling of certain phenomena or in some cases. The problem of non-physically large values is alleviated by proposition of Kotzer *et al.*, who have suggested to introduce normalization of the Gaussian-distributed estimate, which would explicitly impose the preservation of energy of the whole OFDM symbol [4]. According to this approach, assuming $K = N_S$, the absolute value of the i-th sample of the complex signal obtained by the IDFT is a random variable defined as

$$r_i = \frac{|g_i|}{\sqrt{\sum_{k=0}^{N_S-1} |g_k|^2}}, \quad i = 0, 1, \dots, N_S - 1, \qquad (8)$$

where g_k are independent and identically distributed random variables with complex Gaussian distribution, $g_k \sim \mathscr{CN}(0, \sigma^2)$. This way, for each sample the value of r_i remains within the range from 0 to 1. Such formulated model has been called the normalized Gaussian model. It is applicable to OFDM signals obtained for subcarriers modulated according to a constant amplitude modulation scheme, e.g. M-PSK.

In [4] it has been shown, that the normalized Gaussian model reproduces the properties of the signal very well. Thus, it may be advantageous to use this model instead of the conventional Gaussian-based. However, the model as formulated by Kotzer *et al.* in [4] relates to absolute values of complex signal samples, therefore is applicable only to severely limited range of problems, not including, for example, digital-to-analog conversion, performed separately for real and imaginary parts. It is not defined in terms of probability distribution for r_i, hence impractical for analytic derivations. The model assumes the number of samples to be equal to the number of subcarriers and as such it concerns only signal at the Nyquist limit for the highest frequency subcarrier, i.e. without any oversampling, always present in practical applications. Therefore, for wider application, the model needs to be generalized to describe signed values of samples, it should be reformulated in terms of probability density function and apply to the cases in which the signal is oversampled.

3. Modification of the Normalized Gaussian Model

The absolute value of a complex OFDM signal (i.e. IDFT output) does not depend on the phases of subcarriers per se, but on their differences. Thus, one of the phases can be considered to be the global phase of the signal. Hence, variable $r_i e^{j\phi_i}$ can be identified with the complex value of the i-th sample. However, the random phase ϕ_i can be also identified with the phase of the complex variable g_i, and therefore the variable corresponding to one of the real-valued signals, that can be fed to a digital-to-analog converter or other processing device, is

$$\mathrm{Re}\left\{r_i e^{j\phi_i}\right\} = \frac{\mathrm{Re}\{g_i\}}{\sqrt{\sum_{k=0}^{N_S-1}|g_k|^2}}. \quad (9)$$

The Eq. (9) can be used for generation of random vectors of samples for simulation, since it imposes the correct mean signal power for a group of samples by using a common normalization factor, explicitly related to individual samples of the whole vector. Obviously, $|g_k|^2 = \mathrm{Re}\{g_k\}^2 + \mathrm{Im}\{g_k\}^2$ and both real and imaginary part of g_k can be considered to be independent random variables with real-valued normal distribution, $\mathrm{Re}\{g_k\}, \mathrm{Im}\{g_k\} \sim \mathcal{N}(0,\sigma^2)$. Therefore, for each sample, the sum in the nominator comprises: the square of the variable $\mathrm{Re}\{g_i\}^2$, and other $2N_S - 1$ independent random variables with the same Gaussian distribution. Because statistical properties are the same for each sample, at least to large extent the properties of the signal can be characterized by properties of a single sample. Focusing on just a single sample one can treat all the samples as statistically independent, and then the sum in nominator of Eq. (9) can be expressed as

$$\sum_{k=0}^{N_S-1}|g_k|^2 = \mathrm{Re}\{g_i\}^2 + \sigma^2 Q_{2N_S-1}, \quad (10)$$

where $Q_n \in [0;\infty)$ is defined as sum of n squares of independent and identically distributed random variables with standard normal distribution $\mathcal{N}(0,1)$. It is well known that such Q_n is distributed according to the chi-squared distribution with n degrees of freedom, $Q_n \sim \chi_n^2$, with the probability density function

$$\chi_n^2(q) = \frac{q^{\frac{n}{2}-1}e^{-\frac{q}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}\theta(q), \quad (11)$$

where $\Gamma(x)$ denotes the gamma function and $\theta(q)$ is the Heaviside step function. The random variable defined by Eq. (9) can be further subjected to the following:

- substitution $v = \mathrm{Re}\{g_i\}/\sigma \sim \mathcal{N}(0,1)$,
- relating the mean power to the number of subcarriers K, by using in sum (10) a random variable Q_{2K-1} instead of Q_{2N_S-1}.

This way, each sample of considered signal is represented by random variable

$$\xi = \frac{v}{\sqrt{v^2 + Q_{2K-1}}} \in [-1;1]. \quad (12)$$

Because the random variables v and Q_{2K-1} are independent, the probability density function $f_\xi(\xi)$ for variable ξ can be calculated using the formula (with $q \equiv Q_{2K-1}$ for brevity):

$$f_\xi(\xi) = \int_0^\infty \mathrm{d}q\, \chi_{2K-1}^2(q)\, G_1(v(q,\xi))\left|\frac{\partial v}{\partial \xi}\right|. \quad (13)$$

The variable v in function of q and ξ is given by

$$v(q,\xi) = \xi\sqrt{\frac{q}{1-\xi^2}}. \quad (14)$$

Thus, remembering that

$$\int_0^\infty \mathrm{d}q\, q^n e^{-aq} = \frac{\Gamma(n+1)}{a^{n+1}}, \quad \text{where } a > 0,\ n > -1, \quad (15)$$

$\Gamma(n+1) = n\Gamma(n)$ and $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, one obtains:

$$f_\xi(\xi) = \frac{1}{2}\binom{K-1}{\frac{1}{2}}\left(1-\xi^2\right)^{K-\frac{3}{2}}\theta(1-|\xi|), \quad (16)$$

with the generalized binomial for $\alpha,\beta \in \mathbb{R}$ defined as

$$\binom{\alpha}{\beta} = \frac{\Gamma(\alpha+1)}{\Gamma(\beta+1)\Gamma(\alpha-\beta+1)}. \quad (17)$$

To check that this distribution is in fact normalized to 1 it is convenient to change the integration variable to $u = \xi^2$ and make use of the identity

$$\int_0^1 \mathrm{d}u\, u^\alpha (1-u)^\beta = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}. \quad (18)$$

Similarly, one can easily calculate various absolute moments of this distribution

$$\langle|\xi|^n\rangle = \frac{1}{\sqrt{\pi}}\frac{\Gamma(K)\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(K+\frac{n}{2}\right)}, \quad (19)$$

in particular the mean square value $\langle\xi^2\rangle = \frac{1}{2K}$, reproducing correctly the mean power of the signal. It is interesting to

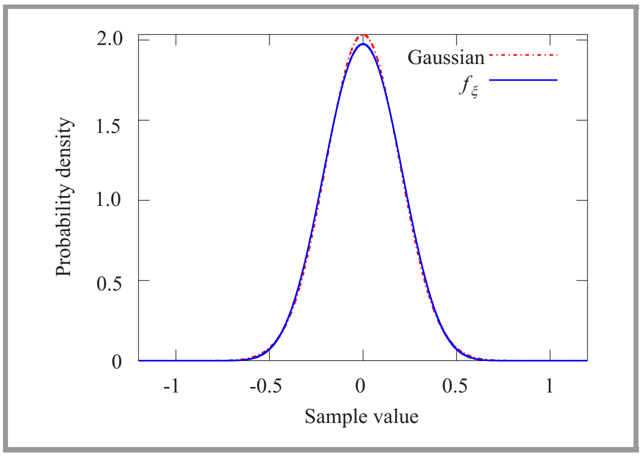

Fig. 1. Comparison of Gaussian probability distribution and the distribution f_ξ, for $K = 13$.

note, that the expression $f_\xi(\xi)$ for one subcarrier, $K = 1$, leads to the arcsine distribution, what is the correct result for a single harmonic function (sine or cosine) of a uniformly distributed variable.

The plot of the distribution f_ξ in comparison to Gaussian is presented in Fig. 1. Both distributions are very similar in the central portion. The Gaussian distribution has a little higher value near 0, on the slopes is a little below f_ξ and again becomes higher at the tails. The most important difference is that the distribution f_ξ has a compact support, and is zero outside the range of allowed values of the signal. It will be seen on the next plots, which are drawn in the logarithmic scale.

4. Numerical Verification

The correctness of the derived probability distribution f_ξ has been verified by numerical simulation performed in Octave [7]. Each numerical estimate has been obtained as a normalized histogram of values of samples for 1 million randomly generated discrete OFDM symbols with specified constellation, number of subcarriers K and number of samples per symbol N_S. For calculation of the histogram, full range of signal samples (i.e. $[-1;1]$) has been divided into 101 "bins" of equal width, but in the presented plots, this range has been narrowed down to the range of values observed in the simulation (apparently, the other values occur with probabilities too small to have been observed).

First, it was checked that the expression (16) for f_ξ correctly reproduces the probabilities of observation of sample values $\mathrm{Re}\left\{r_i e^{j\phi_i}\right\}$ defined by Eq. (9). Exemplary comparisons of the derived and numerical results are shown in Fig. 2. The plots in the figure illustrate that the expression (16) in fact defines the correct probability distribution for the normalized Gaussian model.

Next, the derived probability distribution f_ξ has been compared with numerical results for various OFDM signals and with conventional Gaussian distribution $G_{1/\sqrt{2K}}$ with the same variance. For each case, discrete symbols according to expressions (3) and (4) have been generated with the help of IDFT. The results are presented in Fig. 3 for QPSK constellation and in Fig. 4 for other constant amplitude constellations (8-PSK and 32-PSK). As it can be seen, both analytic models give similar results, in general reproducing the numerical results well. However, neither of them predicts, that the signal samples tend to group at particular values, especially for lower order constellations with a low number of subcarriers K. This effect becomes smaller with increased oversampling (the ratio N_S/K) or the order of constellation. It can be alleviated by introduction of a random global phase shift in constellation for each symbol separately, what causes "smoothing" of the numerical results – for example, see the plot shown in Fig. 5.

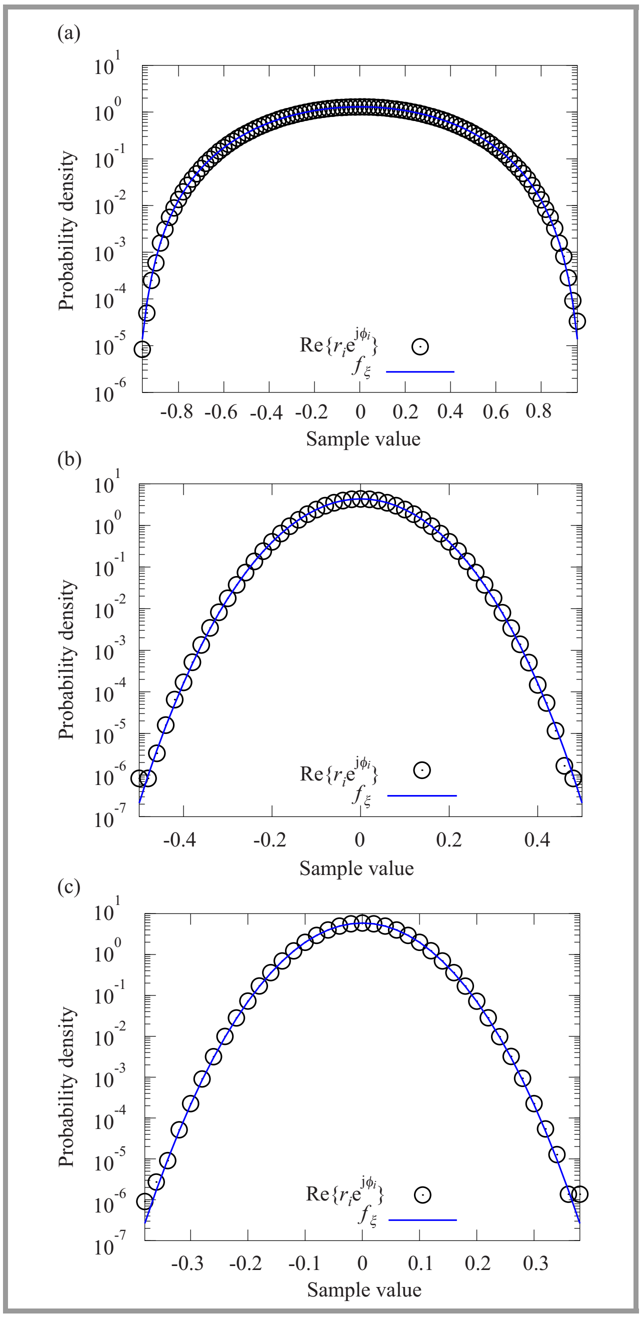

Fig. 2. Comparison of numerically estimated probability distribution of $\mathrm{Re}\left\{r_i e^{j\phi_i}\right\}$ and the derived expression f_ξ for OFDM symbols: (a) $K = 6$ and $N_S = 16$, (b) $K = 60$ and $N_S = 256$, (c) $K = 110$ and $N_S = 256$.

5. Geometrically Averaged Model

From the plots presented in Section 4 it is apparent that the conventional Gaussian model overestimates the tail distribution, while the distribution f_ξ underestimates it. Therefore, an obvious attempt to increase the accuracy of the modeling is to use the geometrical average of the probability distributions in these two models, i.e.

$$h_\xi(\xi) = N_h H_\xi(\xi), \tag{20}$$

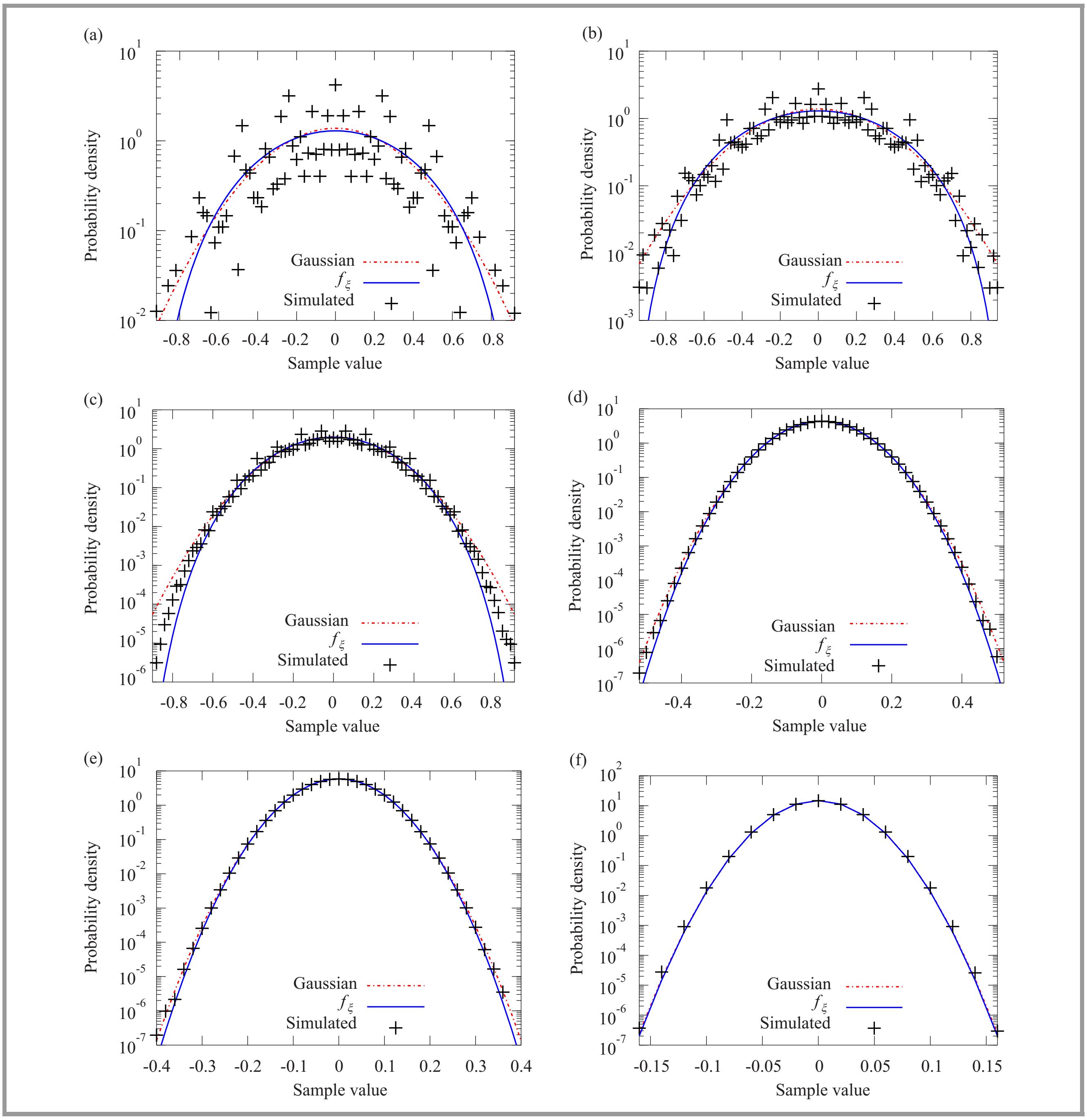

Fig. 3. Exemplary results for verification of distribution f_ξ for QPSK constellation: (a) $K = 6$ and $N_S = 16$, (b) $K = 6$ and $N_S = 32$, (c) $K = 13$ and $N_S = 32$, (d) $K = 60$ and $N_S = 256$, (e) $K = 110$ and $N_S = 256$, (f) $K = 700$ and $N_S = 2048$.

where dependence on ξ is given by

$$H_\xi(\xi) = \sqrt{G_{1/\sqrt{2K}}(\xi)\, f_\xi(\xi)}, \tag{21}$$

and the normalization constant

$$N_h = \left(\int_{-1}^{1} \mathrm{d}\xi\, H_\xi(\xi)\right)^{-1}. \tag{22}$$

The comparison of such defined probability distribution with numerical results presented above is shown in Fig. 6. It can be seen, that both results fit each other very well, thus the distribution h_ξ could be used for better accuracy. However, it seems that analytic expression for h_ξ is rather complicated and it will not be derived here any further.

6. Clipping Noise

A commonly known problem in which tails of the probability distribution used for modeling of the signal

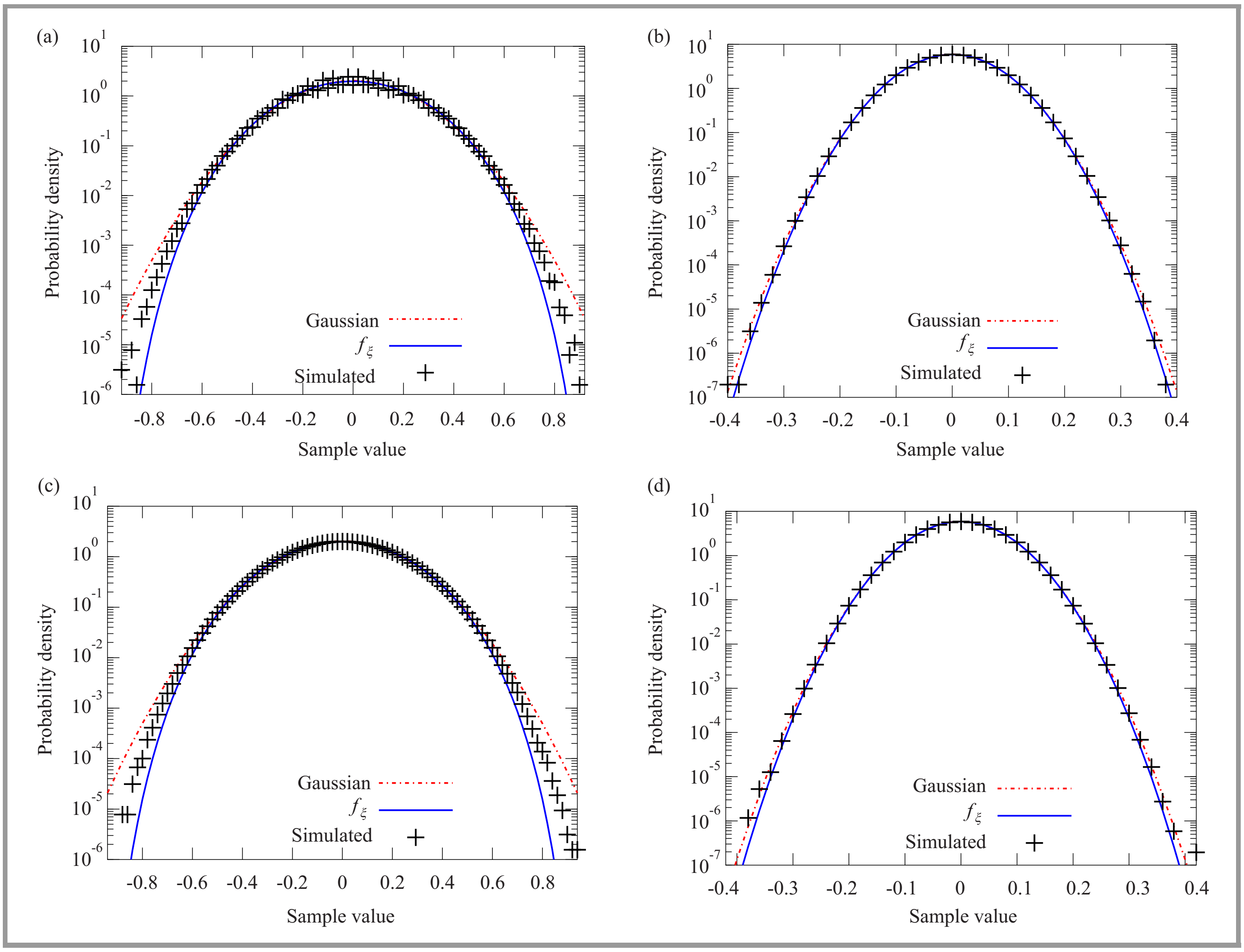

Fig. 4. Exemplary results for verification of distribution f_ξ for M-PSK constellations: (a) 8-PSK, $K = 13$, $N_S = 32$, (b) 8-PSK, $K = 110$, $N_S = 256$, (c) 32-PSK, $K = 13$, $N_S = 32$, (d) 32-PSK, $K = 110$, $N_S = 256$.

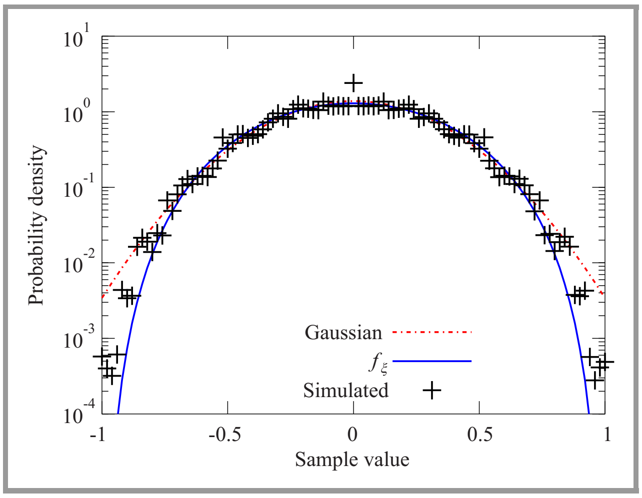

Fig. 5. Exemplary result of simulation with random global phase shift for QPSK, $K = 6$, $N_S = 16$.

have a significant impact is the calculation of clipping noise [5], [6]. This noise usually is considered for digital-to-analog or analog-to-digital converters and amplifiers. Clipping occurs when a device limits the absolute value of the output signal to a certain level A_{clip}. It is convenient to define this level in terms of a clipping factor

$$\gamma = \frac{A_{\text{clip}}}{\sigma}, \tag{23}$$

so that the signal is constrained to $\pm\gamma\sigma$. The clipped portion of the signal is a distortion contributing to the total noise. Following the simplest derivation signal with uncorrelated time-domain values or samples [5], the power of the clipping noise is

$$\sigma_c^2(\gamma) = 2\int_{\gamma\sigma}^{\infty} \mathrm{d}\xi\, (\xi - \gamma\sigma)^2\, p(\xi)\,, \tag{24}$$

where $p(\xi)$ denotes the probability distribution assumed for the signal value ξ. Obviously, if the signal is already limited below $\gamma\sigma$ no clipping occurs and any model allowing the signal to have arbitrarily large values (like the conventional Gaussian model) will lead to incorrect results. Assuming the Gaussian model, i.e $p(\xi) = G_\sigma(\xi)$,

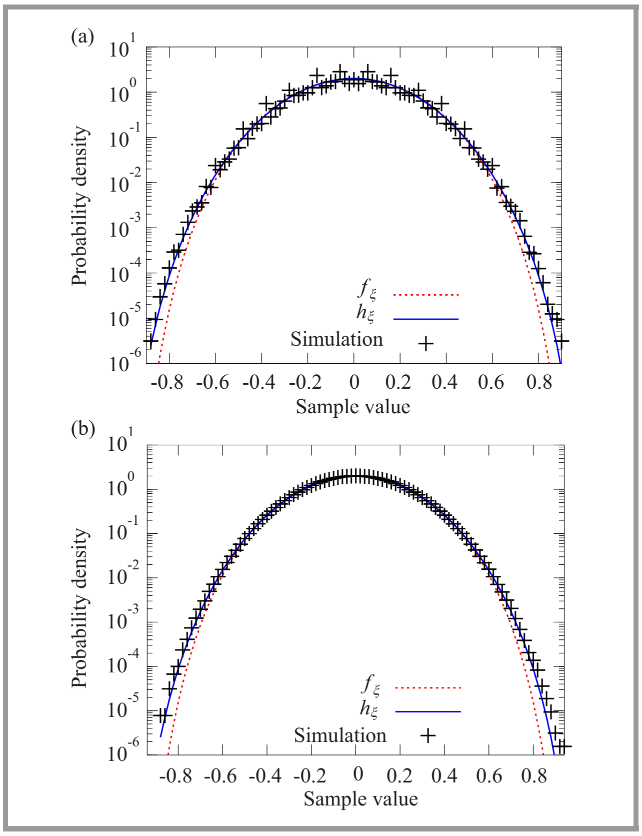

Fig. 6. Exemplary results for verification of distribution h_ξ for M-PSK constellations: (a) QPSK, $K = 13$, $N_S = 32$, (b) 32-PSK, $K = 13$, $N_S = 32$.

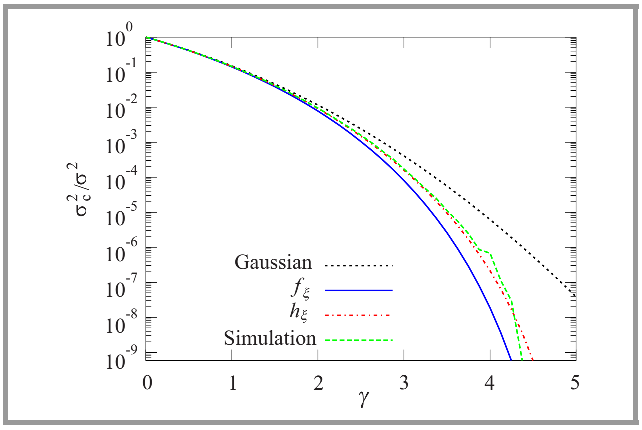

Fig. 7. Clipping noise to signal power ratio obtained with various signal models in comparison to results of simulation for QPSK, $K = 13$ and $N_S = 32$.

the clipping noise to signal power ratio is given by the following expression:

$$\left(\frac{\sigma_c^2(\gamma)}{\sigma^2}\right)_G = \left(1+\gamma^2\right)\operatorname{erfc}\left(\frac{\gamma}{\sqrt{2}}\right) - \sqrt{\frac{2\gamma^2}{\pi}}\,\mathrm{e}^{-\frac{\gamma^2}{2}}, \quad (25)$$

wherein the complementary error function

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty \mathrm{d}t\,\mathrm{e}^{-t^2}. \quad (26)$$

Similar results for $p(\xi) = f_\xi(\xi)$ and $p(\xi) = h_\xi(\xi)$ are here calculated numerically and a comparison of these three results with simulation is presented in Fig. 7. In addition to being time-consuming, the simulation becomes inaccurate for higher values of γ, since the probability of observing sample value high enough to be clipped becomes too small for occurrence of clipping within a given number of random signal symbols (here 100,000). In particular, this relates to clipping at value of $\gamma \approx 4$, which is of practical interest. It can be seen, that, as could be predicted from previous results presented here, the Gaussian model overestimates the clipping noise, at $\gamma \approx 4$ significantly, while the normalized Gaussian model, based on expression for f_ξ, underestimates it, and using h_ξ, the geometrical average of the Gaussian distribution and f_ξ, reproduces the clipping noise very well. This illustrates well an advantage of using proper analytic models, allowing to obtain accurate results very quickly, except of providing a much better insight into the physical phenomenon.

7. Conclusions

In this paper there is provided an extension, with slight modification, of a statistical model of OFDM signal, so-called normalized Gaussian model, originally proposed by Kotzer *et al.* [4]. Treating the signal sample values as independent random variables allows to derive their probability distribution f_ξ given by expression (16). It has been shown, that this expression correctly reproduces statistical properties of the original model. The predictions of the extended model have been compared to results of numerical simulations and on their basis a more accurate model, combining both conventional Gaussian model and the normalized Gaussian model by means of a geometrical average, was proposed. As an example of application, the models were used for calculation of clipping noise.

The probability distribution f_ξ turns out to be accurate to the same degree as the conventionally used Gaussian distribution, with the difference, that while Gaussian overestimates the probability of occurrence of the higher values, the distribution f_ξ tends to underestimate it. Being expressed by a relatively simple function, distribution f_ξ allows to calculate and obtain closed-form results for numerous parameters of the signal, thus it may be particularly useful for theoretical research. The proposed here “geometrically averaged” distribution h_ξ, although providing the most accurate results of all three, seems to be more complicated and might lead to more cumbersome expressions. However, even such formulation as presented offers the benefit of quickly obtained results by simple numerical calculation.

Concluding, the expressions obtained in this paper provide alternatives to conventional Gaussian model and might contribute to theoretical characterization and development of telecommunication devices, especially those using susceptible to noise higher-order modulation schemes.

References

[1] R. W. Chang, "High-speed multichannel data transmission with bandlimited orthogonal signals", *Bell Sys. Tech. J.*, vol. 45, pp. 1775–1796, 1966.

[2] B. R. Saltzberg, "Performance of an efficient parallel data transmission system", *IEEE T. Commun. Technol.*, vol. 15, pp. 805–811, 1967.

[3] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform", *IEEE T. Commun. Technol.*, vol. 19, pp. 628–634, 1971.

[4] I. Kotzer, S. Har-Nevo, S. Sodin, and S. Litsyn, "A model for OFDM signals with applications", in *Proc. 18th Eur. Wirel. Conf.,* Poznań, Poland, 2012.

[5] D. J. G. Mestdagh, P. M. P. Spruyt, and B. Biran, "Effect of amplitude clipping in DMT-ADSL transceivers", *Electron. Lett.*, vol. 29, pp. 1354–1355, 1993.

[6] R. Gross and D. Veeneman, "SNR and spectral properties for a clipped DMT ADSL signal", in *Proc. 1994 IEEE ICC Conf. Rec.*, vol. 2, New Orleans, USA, 1994, pp. 843–847.

[7] Octave [Online]. Available: http://octave.org/

Adam Rudziński received the M.Sc. and Ph.D. degrees in Electronics from Warsaw University of Technology, Poland, in 2004 and 2010, respectively, and M.Sc. degree in Theoretical Physics from Warsaw University, Poland, in 2009. His scientific interests include: modeling of electronic devices and signals, construction of electronic devices , interaction of electromagnetic waves with matter, quantum optics. Currently, he is employed at SKA Polska Sp. z o.o. as an electronics designer.

E-mail: a.rudzinski@ska-polska.pl
SKA Polska Sp. z o.o.
Al. Jerozolimskie 125/127, lok. 406
02-017 Warsaw, Poland