Modeling and Analysis of Short Distance Sub-Terahertz Communication Channel via Mixture of Gamma Distribution

Kürşat Tekbıyık, *Graduate Student Member, IEEE,* Ali Rıza Ekti, *Senior Member, IEEE,* Güneş Karabulut Kurt, *Senior Member, IEEE,* Ali Görçin, *Senior Member, IEEE* Serhan Yarkan, *Senior Member, IEEE*

***Abstract*—With the recent developments on opening the terahertz (THz) spectrum for experimental purposes by the Federal Communications Commission, transceivers operating in the range of 0.1THz-10THz, which are known as THz bands, will enable ultra-high throughput wireless communications. However, actual implementation of the high-speed and high reliability THz band communication systems should start with providing extensive knowledge in regards to the propagation channel characteristics. Considering the huge bandwidth and the rapid changes in the characteristics of THz wireless channels, ray tracing and one-shot statistical modeling are not adequate to define an accurate channel model. In this work, we propose Gamma mixture based channel modeling for the THz band via the expectation-maximization (EM) algorithm. First, maximum likelihood estimation (MLE) is applied to characterize the Gamma mixture model parameters, and then EM algorithm is used to compute MLEs of the unknown parameters of the measurement data. The accuracy of the proposed model is investigated by using the Weighted relative mean difference (WMRD) error metrics, Kullback-–Leibler (KL)-divergence, and Kolmogorov-Smirnov (KS) test to show the difference between the proposed model and the actual probability density functions (PDFs) that are obtained via the designed test environment. To efficiently evaluate the performance of the proposed method in more realistic scenarios, all the analysis is done by examining measurement data from a measurement campaign in the 240GHz to 300GHz frequency range, using a well-isolated anechoic chamber. According to WMRD error metrics, KL-divergence, and KS test results, PDFs generated by the mixture of Gamma distributions fit to the actual histogram of the measurement data. It is shown that instead of taking pseudo-average characteristics of sub-bands in the wide band, using the mixture models allows for determining channel parameters more precisely.**

I. Introduction

The ability of wireless communication technology to meet consumer needs requires that next generation wireless networks' data rates reach terabits per second (Tbps) levels at a higher link density [1, 2]. Although free space optical (FSO) and millimeter wave (mmWave) communications are proposed for high data rates, the requirements of both systems, and especially a bandwidth of only 9GHz around 60GHz, are not expected to deliver Tbps for mobile and personal communication systems [3]. As there is no block wider than 10GHz below 100GHz [4], the researchers push the frequency limits towards terahertz (THz) band, which is in between 0.1THz - 10THz. Due to the flat frequency response and also the capabilities of the current state of the signal generators, most of the researches focus on the band between 200GHz - 300GHz.

To be able to fully discover the potential of a wireless communication system, the proper channel model must be used. Then, other parts of the system can be designed. Although the THz band will provide a way to achieve Tbps data rates, the THz band differs from the currently used bands in the channel characteristics that change rapidly and sharply across the spectrum [2, 4]. Therefore, all elements of the system should be re-examined and designed to develop a proper communication system. For example, the propagation channel is required to be analyzed on the aspects of materials in the medium, and the operating frequency. Wireless communication in wide band around 60GHz requires a channel model considering characteristics of sub-bands which are windows such that propagation characteristics can be assumed to be static throughout the window.

A. Related Works

In the studies on channel modeling, various approaches can be encountered for frequency, time and spatial analysis. It is worth saying that THz channels have many differences from lower frequency bands in terms of noise, propagation, and molecular absorption. Thus, channel modeling studies require to be reconsidered for THz band rather than employ the model proposed for lower bands. The wireless communication channel can be modeled by using deterministic or statistical methods. Although deterministic models like ray-tracing are most accurate if the detailed description of the given environment is properly and extensively fed into [5], their performance can be hindered even in the presence of a slightest change in the propagation environment. All parameters of

K. Tekbıyık and G.K. Kurt are with the Department of Electronics and Communications Engineering, İstanbul Technical University, İstanbul, Turkey. e-mail: {tekbiyik, gkurt}@itu.edu.tr

K. Tekbıyık, A. Görçin and A.R. Ekti are with the HISARLab at Informatics and Information Security Research Center (BİLGEM), TÜBİTAK, Kocaeli, Turkey. e-mail: {kursat.tekbiyik, ali.gorcin, aliriza.ekti}@tubitak.gov.tr

A.R. Ekti is with the Department of Electrical-Electronics Engineering, Balıkesir University, Balıkesir, Turkey. e-mail: arekti@balikesir.edu.tr

A. Görçin is with the Department of Electronics and Communications Engineering, Yıldız Technical University, İstanbul, Turkey. e-mail: agorcin@yildiz.edu.tr

S. Yarkan is with the Department of Electrical and Electronics Engineering, İstanbul Commerce University, İstanbul, Turkey. e-mail: syarkan@ticaret.edu.tr

the propagation environment are required by these models. As a result, considering that even molecular changes affect the propagation characteristics of THz waves, it can be said that the ray-tracing method might not adequately model THz channels. Another reason that makes this method complex and computationally cumbersome is the exponential increase in the complexity of the method as the size of the medium to be modeled increases. On the contrary, temporospatial characteristics of wireless channels of data centers are investigated in [6, 7]. Furthermore, multi-dimensional parameters of kiosk's wireless channels are modelled for each type of THz rays in [8]. The statistical approaches use the average of the environmental effects, unlike a deterministic model (e.g., ray-tracing). Some stochastic models have been recently proposed in [9-11]. Our previous work [12] proposes a two-slope path loss model for short-range THz communication links. In [13], the statistical channel parameters such as delay-spread, cluster delays and cluster powers are obtained throughout extensive measurements at 140GHz. It should be noted that that ray-tracing and stochastic methods are not antipodes; but, they are complement of each other to describe a channel more accurate. For example, [14] proposes a hybrid channel model by combining statistical approaches and ray-tracing methods for sub-THz band.

Another important consideration in channel modeling is the careful selection of signal processing methods to be used for modeling the wide-band channel. To set an example, in [15], the frequency sweeping method, which is not safe due to artifacts created when the post-processing of the smaller chunks of bandwidth, is employed to model the spectrum between 260GHz and 400GHz. Another problem in channel modeling is to make the assumption that the derived impulse response has a linear phase. This assumption implies that the impulse response is symmetrical to line-of-sight (LOS) propagation delay. However, the real physical environments do not allow this phenomenon because it contradicts causality. Therefore, Kazuhiro *et. al.* propose a causal channel model for THz band [16].

Multi-input multi-output (MIMO) can provide coverage improvements in addition to capacity enhancements for THz communications; thus, channel models for 2×2 MIMO systems are investigated in [17, 18]. The results indicate that MIMO systems can achieve high data rates. In [19], Doppler shift caused by airflow turbulance in data-center is measured for band between 300GHz and 320GHz. Besides Doppler shift measurements, it presents that channel amplitude gains for 4×4 MIMO follow m-Nakagami distribution. Furthermore, the cluster shadowing gain in a data-center is Gaussian distributed for that band [20]. Also, by using graphene-based MIMO system, the spectral efficiency can be enhanced. Massive MIMO systems benefits most from the ultra small antenna sizes at THz frequencies; therefore, massive MIMO antenna structures for these bands are researched in [21-23]. These inquiries show that the capabilities of the THz communication can be advanced by utilizing nano antenna structures and massive MIMO systems. Besides these works, some studies focus on the application specific aspects of these bands; in [24], indoor channel measurements are conducted for 300GHz. Also, in [25], the behavior of the digital communication schemes are analyzed for the same band.

Studies up to this point assume that the THz band of interest has a single statistical distribution. In this case, it can be concluded that channel modeling with a single probability density function (PDF) is not sufficient, considering the presence of windows that behave differently in the THz band due to the effect of molecules in the medium. The THz band contains changes across the spectrum, so it may not be sufficient to express this extremely wide-band with a single statistical model. For example, suppose that the three sub-bands behave differently from each other as demonstrated in Fig. 1. Hence, the use of mixtures to add the characteristics of each sub-band into the model provides better convergence to the actual histogram. It is worth noting that an arbitrary PDF can be modeled by utilizing Gamma mixtures [26, 27]. Therefore, in this study, we investigate the channel modeling with Gamma mixtures for short-range sub-THz channels.

Although mixture models have been used in many different fields, in this study, we are confined to mentioning only the studies on wireless communication channel models. In [28], it is stated that mixture Gamma is able to model $\alpha-\kappa-\mu$ shadowed fading channels, even though they consist of intractable statistical properties. [29] and [30] employ the mixture of Gaussian distributions to construct a generalized shadowing model, by adopting expectation-maximization (EM) algorithm to find the mixture parameters. The error probability and ergodic capacity can be analyzed by using Gamma mixtures for diversity reception schemes over generalized-K fading channels [31]. Moreover, the physical layer security analysis can be performed by utilizing mixture models in generalized-K fading channels [32]. [33], which is one of the most important studies in this research area, proposes a mixture of Gamma distributions for the signal-to-noise ratio (SNR) of fading channels; thereby, it allows to derive the average channel capacity, the outage probability, and the symbol error rate.

B. Contributions

In this study, we utilize mixture models to investigate channel models for sub-THz band between 240GHz and 300GHz. With inspiration from the studies such as [34, 35], which adopt the mixture models to characterize the wireless propagation channel, we propose mixture models which are suitable for the nature of the THz band to model the distribution of the received power for the sub-THz band between 240GHz and 300GHz. Based on the intrinsic propagation characteristics of the wide-band THz communication channel, a channel model based on a single distribution can not provide an adequate representation. As the THz band allows for very broadband communication and there is a significant change [12] in channel characteristics throughout this wide band. The contributions of this study can be categorized under three main points:

- For the THz band, measurement based channel model study is performed. Using measurement data, it is shown that Gamma mixtures can be used effectively in channel

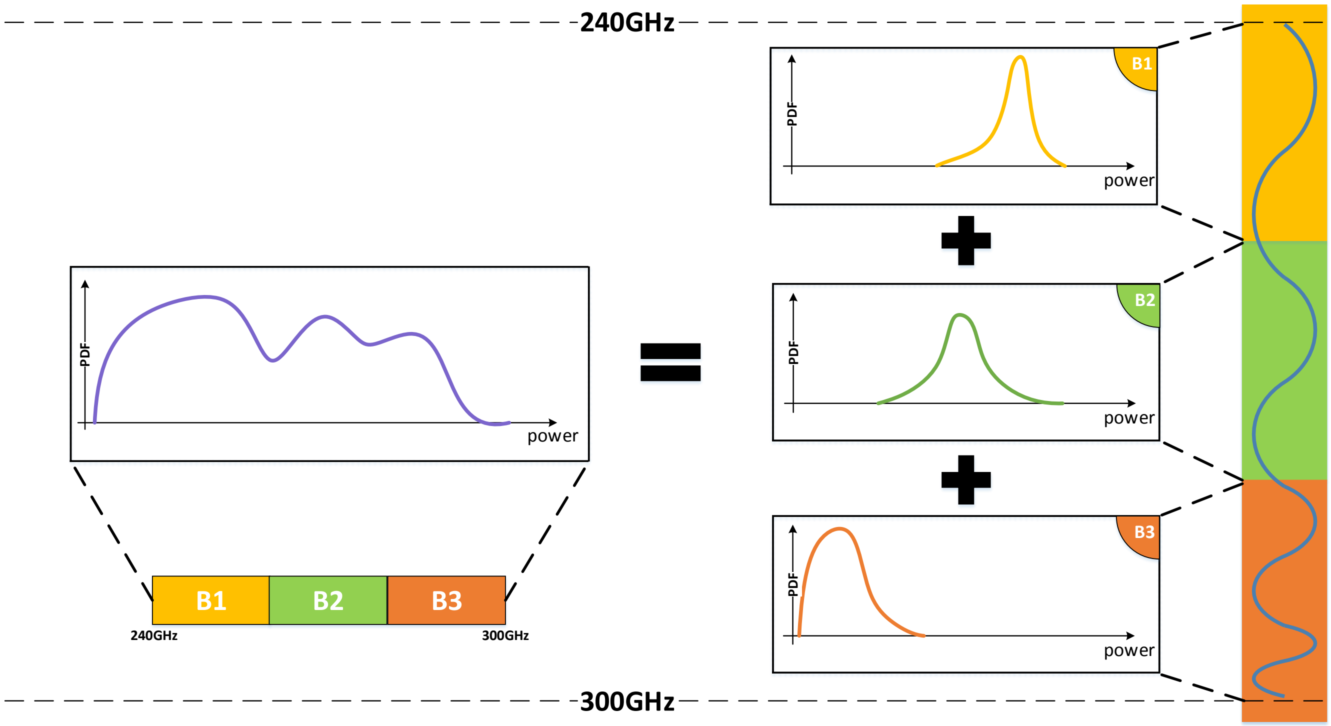


Fig. 1. Instead of using a single distribution to model the received power characteristics, mixture model is able to give information about each sub-band characteristics. It can be said that the wide-band signal covers the characteristics of each sub-band signal propagates in the bands B1, B2, and B3.

modeling for THz band. Thus, the characteristics of the channel can be expressed in a realistic manner.

- Weighted relative mean difference (WMRD), Kolmogorov-Smirnov (KS) and Kullback-–Leibler (KL)-divergence approaches are studied to investigate how well Gamma mixture models fit into measurement data.
- Moreover, considering that the measurement data used in this study is a very valuable source of information and the necessity of making serious investments to reach such data, it is offered as a public dataset [36]. We believe that the sharing of this measurement data will foster new studies.

C. Organization of the Paper

The rest of this manuscript is organized as follows. Section II details the signal model and gives mathematical preliminaries. The measurement setup is introduced in Section III. In Section IV, Gamma mixture modeling results are given and discussed. Finally, Section V concludes the study.

II. Background

A. Signal Model

The received signal is represented as:

$$r(t) = \mathrm{Re}\{[x_I(t) + jx_Q(t)]e^{j2\pi f_c t}\}, \tag{1}$$

where j denotes the unit imaginary number and $\mathrm{Re}\{\cdot\}$ is the real part of the complex number. $x_I(t)$ and $x_Q(t)$ are in-phase and quadrature (I/Q) parts of the complex baseband signal. f_c stands for the carrier frequency of the signal.

The multipath channel at passband with different delays and attenuation levels can be given as:

$$h(t) = \sum_{l=0}^{L-1} a_l \delta(t - t_l), \tag{2}$$

where L is the number of multipath components. a_l and t_l denote the attenuation and delay factors for the lth path, respectively. The complex baseband representation of Eq. (2) is

$$h(t) = \sum_{l=0}^{L-1} a_l \delta(t - t_l) e^{-j2\pi f_c t_l}. \tag{3}$$

If the channel consists of only LOS component, L in Eq. (3) is equal to 1. Then, LOS channel is given as:

$$h(t) = a_0 \delta(t - t_0) e^{-j2\pi f_c t_0}, \tag{4}$$

where a_0 and $2\pi f_c t_0$ denote amplitude and phase of channel, respectively. t_0 is propagation delay given with

$$t_0 = \frac{d}{c}, \tag{5}$$

where d is the distance between transmitter and receiver and c is the speed of light.

Anechoic chambers, as used in our measurements, do not allow non-line-of-sight (NLOS) propagation. The losses are limited to antenna misalignment, imperfections created by hardware, and path loss. Thus, the signal model can be reduced to a direct path which is comprised of distant dependent path loss and antenna misalignment. The contribution of path loss to the channel amplitude a_0 is given as:

$$P_{RX} = P_{TX} - 10n\log(d) + M. \tag{6}$$

The received power P_{RX} including antenna gain considering misalignment, M, is calculated as the difference between transmitted power P_{TX} and path loss with exponent n.

B. Gamma Distribution

The Gamma function, $\Gamma(a)$, is defined as [37]:

$$\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx,\ a > 0. \tag{7}$$

By using integration by parts, $\Gamma(a) = (a-1)!$ when a is a positive integer. Consider the random variable G which is a mixture of m Gamma distributions and defined as:

$$f_G(x) = \sum_{l=1}^{m} \rho_l f_l(x;\alpha_l,\beta_l),\ l = 1,2,\ldots,m,\quad x > 0,\quad \rho_l > 0 \tag{8}$$

where $f_l(x;\alpha_l,\beta_l) = \frac{1}{\beta_l^{\alpha_l}\Gamma(\alpha_l)}x^{\alpha_l-1}e^{-x/\beta_l}$; $\alpha_l > 0$ and $\beta_l > 0$ are the shape and scale parameters of the lth component of the mixture distribution; ρ_l denotes mixture proportions or weights that satisfy the conditions (a) $0 < \rho_l < 1$, $\forall l = 1,2,\ldots,m$ and (b) $\sum_{l=1}^{m}\rho_l = 1$. Please note that, m denotes the number of components in the mixture. The main reasons for using a mixture of Gamma distributions in the paper are: (i) the tractability of its cumulative distribution function (CDF) and moment generating function (MGF), (ii) giving an approximation for small-scale fading channels [33], and (iii) high accuracy by properly adjusting parameters.

C. Maximum Likelihood Estimation

The maximum likelihood estimation (MLE) technique is provided to obtain the parameters of the gamma mixture from the actual channel PDF. Let assume that $X_1,\cdots,X_n$ are random variables with Gamma distribution (with unknown parameters $\alpha > 0$ and $\beta > 0$). The likelihood function is given as:

$$L(x;\alpha,\beta) = \prod_{i=1}^{n}\frac{x_i^{\alpha-1}e^{-\frac{x_i}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}$$
$$= \left\{\prod_{i=1}^{n}x_i\right\}^{-1}\left\{\prod_{i=1}^{n}x_i\right\}^{a} e^{-\frac{\sum_{i=1}^{n}x_i}{\beta}}\beta^{-n\alpha}\Gamma^{-n}(\alpha)$$

The uninformative factor, $\left\{\prod_{i=1}^{n}x_i\right\}^{-1}$, is discarded

$$= \left\{\prod_{i=1}^{n}x_i\right\}^{a} e^{-\frac{\sum_{i=1}^{n}x_i}{\beta}}\beta^{-n\alpha}\Gamma^{-n}(\alpha)$$
$$= \beta^{-n\alpha}\Gamma^{-n}(\alpha)\left\{\prod_{i=1}^{n}x_i\right\}^{a} e^{-\frac{\sum_{i=1}^{n}x_i}{\beta}} \tag{9}$$

The corresponding log likelihood function of Eq. (9) leads to:

$$\ln(L) = -n\alpha\ln(\beta) - n\ln(\Gamma(a)) + \alpha\sum_{i=1}^{n}\ln(x_i) - \sum_{i=1}^{n}\frac{x_i}{\beta}. \tag{10}$$

Maximum likelihood estimates can be found for α and β by taking partial derivatives of Eq. (10) with respect to α and β, then we obtain:

$$\frac{\partial\ln(L)}{\partial\alpha} = -n\ln(\beta) - n\frac{\partial\ln(\Gamma(a))}{\partial\alpha} + \sum_{i=1}^{n}\ln(x_i)$$
$$\frac{\partial\ln(L)}{\partial\beta} = -n\alpha\frac{1}{\beta} + \sum_{i=1}^{n}\frac{x_i}{\beta^2} \tag{11}$$

Because of the diGamma and logarithm functions in Eq. (11), a closed-form solution could not be provided [37]. Numerical methods such as Newton-Raphson can be applied to find the values for α and β which is not the scope of this study.

D. Expectation Maximization

We have a training set $\mathbf{r} = (r_1, r_2, \cdots, r_m)$ consisting of m independent observations captured by considering each measurement data at different transmitter-receiver separation distances such as d=20cm, 30cm, 40cm, 60cm, 80cm. Our goal is to fit the Gamma distribution parameters by utilizing the EM algorithm. EM algorithm, which is a machine learning technique [38], provides a simplification to MLE problems, which are mostly seen in mixture models [35]. The EM algorithm consists of two steps, namely, the expectation (E)-step and the maximization (M)-step. The reader is referred to [39] for more detailed explanations about the EM algorithm.

The EM algorithm requires number of mixtures as a priori. Initially, the parameters are randomly chosen for the mixture model parameters $\theta_{1:M} = (\theta_1,\cdots,\theta_M)$. Then, the parameters are updated in each iteration until the convergence criteria hold. E-step calculates membership coefficients for all data point ($i = 1,\cdots,L$) and mixture components ($k = 1,\cdots,M$) by utilizing the current parameters $\theta_{1:M}$ [35, 40]

$$\phi_{ik} = \frac{\pi_k p_k(x_i|\theta_k)}{\sum_{k=1}^{M}\pi_k p_k(x_i|\theta_k)}, \tag{12}$$

where x_i is the data in the kth mixture; π_k denotes the mixing proportion. It is obvious that $\sum_{k=1}^{M}\phi_{ik} = 1$. Then, the parameter values and the mixing proportions for each mixture components are updated to maximize the likelihood probability in the M-step. In the M-step, the membership coefficients calculated in E-step are used to find parameters and mixing proportions as:

$$\pi_k^{new} = \frac{\sum_{i=1}^{L}\phi_{ik}}{L}$$
$$\mathbb{E}[X_k]^{new} = \frac{\sum_{i=1}^{L}\phi_{ik}x_i}{\sum_{i=1}^{L}\phi_{ik}} = \alpha\beta$$
$$\mathrm{Var}[X_k]^{new} = \frac{\sum_{i=1}^{L}\phi_{ik}(x_i - \mathbb{E}[X_k]^{new})^2}{\sum_{i=1}^{L}\phi_{ik}} = \alpha\beta^2. \tag{13}$$

The parameters (α,β) for each Gamma mixture can be found by using Eq. (13).

E. Error Metrics

In this subsection, we provide an overview of the possible error metrics to determine the goodness-of-fit for the proposed model.

1) Weighted Mean Relative Difference: The proposed models are quantified by using WMRD, which gives a measurement for the difference between the model and actual PDFs. It is defined as [35]:

$$\mathrm{WMRD} = \frac{\sum_{\rho} |y_\rho - \hat{y}_\rho|}{\sum_{\rho}(y_\rho + \hat{y}_\rho) \times 0.5}, \tag{14}$$

where ρ represents the received power and y_ρ is the number of ρ value observations in the received power set. As well as, $\hat{y}_\rho$ is related to the estimated model.

2) Kolmogorov-Smirnov Test: KS test is a non-parametric goodness-of-fit test, namely it does not make an assumption of any distribution. In addition to vector norm based error technique, the KS test is employed as goodness-of-fit test with the confidence level $p = 0.05$ to compare the actual PDF with the estimated mixture models.

3) Kullback—Leibler Divergence: KL distance or divergence is interpreted as the distance between the actual probability distribution, P_{act} and the estimated probability distribution, P_{est}. Let $P_{act} = \{p_1, p_2, \cdots, p_n\}$ and $P_{est} = \{q_1, q_2, \cdots, q_n\}$, then KL-divergence is defined as

$$D_{\mathrm{KL}}(P_{act} \| P_{est}) = -\sum_{x \in \mathcal{X}} P_{act}(x) \log \left(\frac{P_{est}(x)}{P_{act}(x)} \right). \tag{15}$$

In this paper, KL-divergence is utilized to compare the actual distribution and the estimated models via the EM algorithm. KL-divergence gets a higher value when two distributions have less similarities.

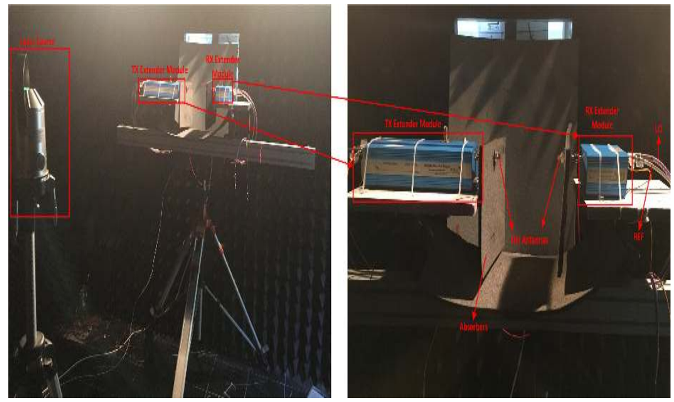


Fig. 2. Measurement setup is prepared in the anechoic chamber to suppress possible reflections and guarantee LOS conditions. Laser source is used to eliminate any misalignment.

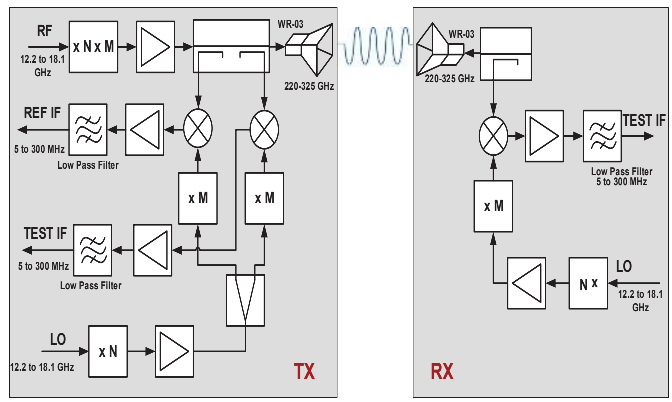


Fig. 3. Block diagram for the measurement setup, which uses bottom-up approach to generate THz signals.

III. Channel Measurement Campaign and Data Processing

The measurement setup exihibited in Fig. 2 is allocated in one of the anechoic chambers of Turkish Science Foundation [41] with the dimensions of $7m \times 4m \times 3m$ to make sure that the LOS components of the transmissions are observed and received properly.

The setup is comprised of four main hardware components: (i) a performance network analyzer (PNA) vector network analyzer (VNA), which is coded as E8361A, (ii) extender modules for millimeter wave propagation, i.e., V03VNA2-A and V03VNA2-T/R-T coded devices from Oleson Microwave Labs (OML), and (iii) N5260 coded controller for extenders from the same company. VNA can analyze signals up to 67GHz, therefore, extender modules are utilized to be able to cover the 220GHz to 325GHz bands. The V03VNA2-T/R-A has 18 multipliers that can modulate signals on the 10GHz to 20GHz range up to the 300GHz region. On the contrary, the transmitted signal from the wireless channel is down-converted via V03VNA2-T by using the same number of mixers and the resultant signal is at the intermediate frequency (IF) of 5MHz to 300MHz. Following the down-conversion, the IF signal is provided as input to the VNA. The channel characteristics analysis is conducted considering the difference between the characteristics of transmitted and received signals. Please also note that we utilize a laser level tool to ensure that both transmitter and receiver are perfectly aligned for LOS transmission. The block diagram of this process is depicted in Fig. 3.

When the hardware characteristics are considered, it is seen that the typical source match at the output is 9dB for balanced multipliers, which are connected to the WR-10 band extension multiplier chains. Each chain contains WR-03 wave-guide output interfaces. The signal generated can be a continuous wave (CW) or frequency sweeping signal. The level of RF power for the local oscillator (LO) to run OML modules should be in $+10$dBm range. The extender characteristics are as follows; the phase stability is ± 0.4dB in the range of $\pm 8°$, typical dynamic range is 75dB with the minimum of 60dB.

At the earlier stages of the measurement campaign, we realized relatively small impairments in terms of phase and magnitude stability are observed at the range of 240GHz to 300GHz frequencies. Thus, we decided to utilize these bands to be able to achieve the best results for our purpose in this study. Scattering parameters (s-parameters) are utilized to understand the channel transfer function of these bands and for the modelling of the wireless channel, first a calibration procedure is executed. In this context, a direct connection is

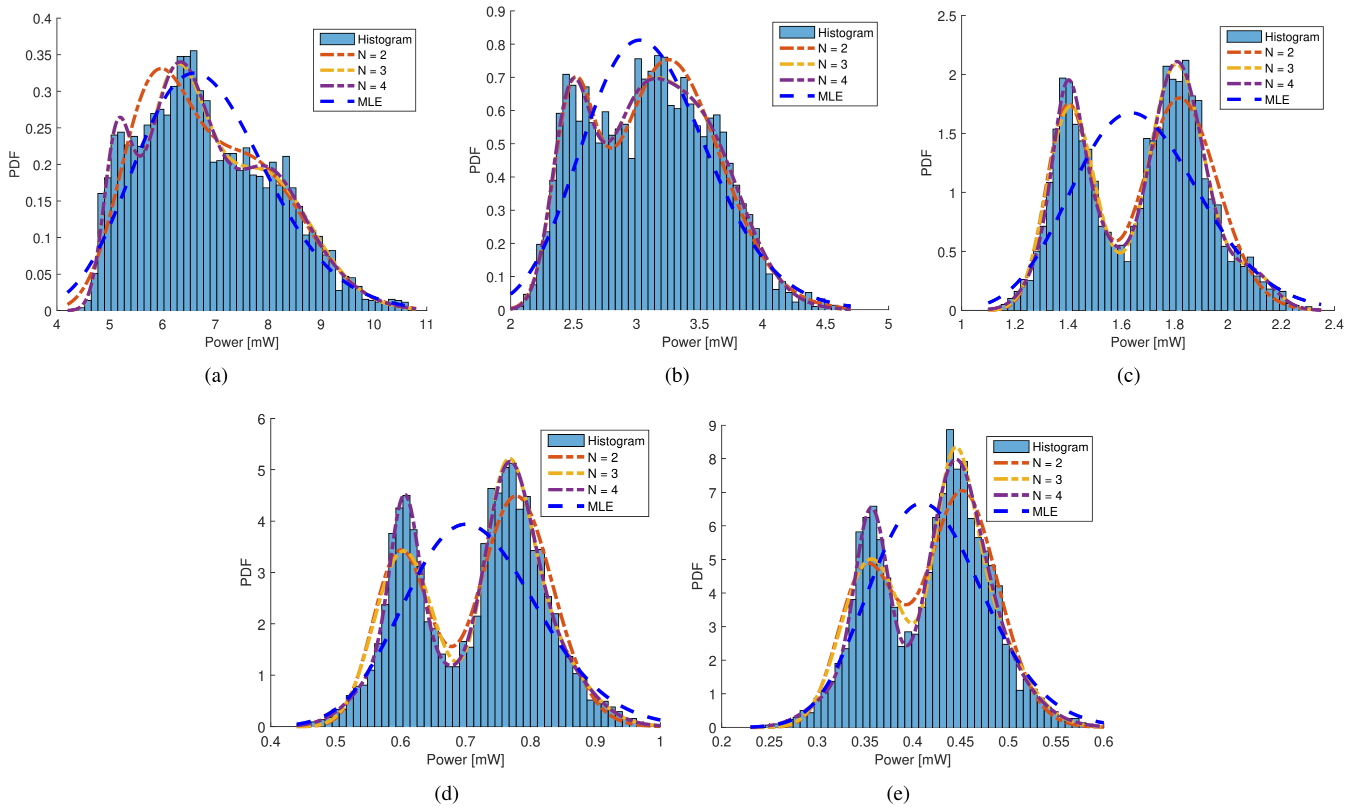


Fig. 4. Mixture models fit much better to the measurement data for (a) 20cm, (b) 30cm, (c) 40cm, (d) 60cm, and (e) 80cm compared to MLE.

established between the transmitting and receiving wave-guide ports of the extenders. Following this step, calibration data is saved inside the measurement devices of the setup and it is converted to the form of complex S_{21} parameters of each point of measurement. The measurement system also includes cables and connectors whic are also separately calibrated to eliminate the impairments.

Two identical horn antennas with 24.8dBi gain at their center frequencies are connected at both transmitting and receiving ends of the measurement system. Therefore, this setup covered 60GHz band between 240GHz to 300GHz and recordings are done over 4096 points utilizing an IF bandwidth (BW) of 100Hz. Such process led to the improvement of observed dynamic range and reduction of the noise floor. Eventually 14.648MHz became the spectrum resolution available. Each set of captured I/Q samples are transferred into a laptop computer. Necessary conversions are applied and all the analyses are done on MATLAB R2018b software to carry out the baseband operations for each transmitter-receiver separation distance given in Table I.

IV. Gamma Mixture Model for Terahertz Wireless Channels

In this section, Gamma mixture models are employed to model received power distribution for five measurement described in Section III. The received power, P_{rx}, is calculated in the linear scale as

$$P_{rx} = |S_{21}|^2 P_{tx}, \tag{16}$$

where P_{tx} is the transmitted signal power and constant during the transmission time. $|S_{21}|$ denotes the amplitude response of the propagation channel. It is known that the instantaneous SNR for a signal with bandwidth of W is defined as:

$$\gamma = \frac{P_{rx}}{WN_0} \tag{17}$$

under the additive white Gaussian noise (AWGN) with power spectral density $N_0/2$. Therefore, SNR is related to the fading channel parameters, as well as the received power. By utilizing the instantaneous SNR, it is possible to derive the channel outage probability and the channel capacity [42].

A. Gamma Mixture Model Results

In order to model the received signal power, both MLE and the EM algorithm are used. EM algorithm enables to determine the parameters of the estimated mixture components for the measurements. Based on the aforementioned points stated in Section II-B, the Gamma distributions are utilized because of the facts that its MGF is tractable and there is an approximation for small-scale fading channels.

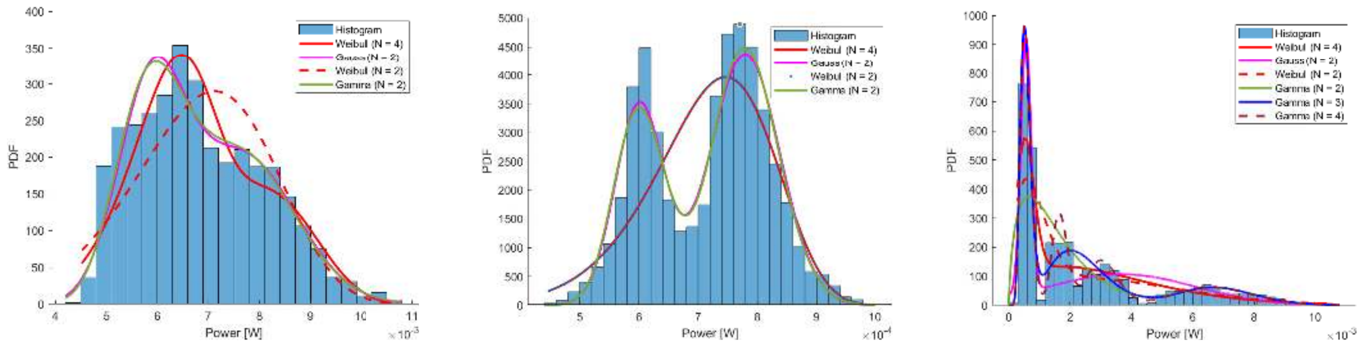
In Fig. 4, it can be clearly seen that MLE estimation is not a good fit for the measured histogram; however, the mixture models fit better. For example, it can be said that three mixtures of Gamma distribution are sufficient to estimate the actual histogram for the distance of 20cm. However, MLE gets hampered to fit since it assumes that there is no serious change in the channel behavior through the transmission band due to the characteristics of the molecules in the propagation

TABLE I
Error metrics for PDF estimations at distinct distances.

Distance	Mixture	Parameters π	α	β	WMRD ($\times 10^{-2}$)	KL Divergence	KS Test ($p = 0.05$)
20cm	MLE	1.00	30.084	0.227	1.580	4.635	*Passed*
	N = 2	0.540 0.460	72.285 67.904	0.0824 0.115	1.573	0.651	*Passed*
	N = 3	0.463 0.397 0.140	116.797 85.985 406.051	0.0539 0.093 0.012	1.572	0.813	*Passed*
	N = 4	0.538 0.193 0.137 0.132	100.310 185.751 407.257 120.786	0.063 0.042 0.012 0.072	1.572	0.797	*Passed*
30cm	MLE	1.00	39.060	0.079	1.541	3.881	*Passed*
	N = 2	0.752 0.248	67.765 236.829	0.048 0.010	1.536	0.822	*Passed*
	N = 3	0.388 0.370 0.242	123.224 132.766 259.535	0.024 0.026 0.009	1.536	0.906	*Passed*
	N = 4	0.303 0.250 0.234 0.213	144.100 254.633 104.663 139.345	0.020 0.009 0.032 0.025	1.536	0.903	*Passed*
40cm	MLE	1.00	49.334	0.034	1.497	3.349	*Passed*
	N = 2	0.626 0.374	172.946 269.180	0.010 0.005	1.486	1.118	*Passed*
	N = 3	0.550 0.396 0.054	295.648 240.228 767.221	0.006 0.006 0.002	1.484	1.067	*Passed*
	N = 4	0.532 0.290 0.119 0.059	322.175 160.954 782.971 720.197	0.005 0.009 0.002 0.003	1.483	1.016	*Passed*
60cm	MLE	1.00	49.026	0.014	1.446	2.646	*Passed*
	N = 2	0.626 0.374	196.498 191.389	0.004 0.003	1.435	1.351	*Passed*
	N = 3	0.340 0.399 0.201	449.070 170.063 230.398	0.002 0.035 0.036	1.432	1.226	*Passed*
	N = 4	0.405 0.356 0.189 0.050	163.658 428.192 839.961 541.486	0.004 0.002 0.001 0.0016	1.429	1.213	*Passed*
80cm	MLE	1.00	48.192	0.008	1.392	2.227	*Passed*
	N = 2	0.634 0.366	158.089 135.815	0.003 0.002	1.388	1.725	*Passed*
	N = 3	0.412 0.316 0.272	119.250 462.180 180.013	0.003 0.001 0.0027	1.384	1.590	*Passed*
	N =4	0.432 0.222 0.204 0.142	281.987 441.924 161.406 76.095	0.0016 0.0008 0.003 0.0045	1.381	1.331	*Passed*

environment. Even though this assumption can be reasonable for the cellular communication bands below 60GHz, it is not held for the THz band. As shown in Fig. 4, the mixture models provide more adequate PDFs than MLE for the tx-rx separation between 20cm and 80cm. Fig. 4 shows that three mixtures and four mixtures of Gamma distribution have almost the same performance to fit the actual histograms. It is seen in Fig. 4 that the mean of the received signal power decreases for increasing distance, as expected. Furthermore, Fig. 4 clearly exhibits different clusters in the histograms especially for the distances longer than 30cm, which is consistent with [2, 4].

Moreover, WMRD results and KL-divergence also confirm that mixture models converge to the actual histogram better than MLEs. WMRD results, KL distance, and KS test results are presented in Table I. WMRD results are not accurate enough to show the difference between MLE and mixture models. WMRD can demonstrate that the mixture model is more successful than MLE with only a small variation in its value. However, KL-divergence creates metrics more sensitive to differences between mixture models and MLEs. For instance, KL-divergence is found as 4.635 for MLE at 20cm, whereas it is 0.651 for two mixtures. KL-divergences



(a) (b) (c)

Fig. 5. Gamma mixtures fit much better to the measurement data for (a) 20cm, (b) 60cm, (c) all measurement data compared to Gaussian and Weibull mixtures.

TABLE II
KL Divergence results for Gamma, Gaussian and Weibull mixtures.

Distance	Mixture	KL Divergence
20cm	Gaussian (N = 2)	0.655
	Weibull (N = 2)	1.726
	Weibull (N = 4)	0.978
	Gamma (N = 2)	0.651
60cm	Gaussian (N = 2)	1.344
	Weibull (N = 2)	2.321
	Weibull (N = 4)	2.321
	Gamma (N = 2)	1.351
All	Gaussian (N = 2)	0.582
	Weibull (N = 2)	1.436
	Weibull (N = 4)	1.427
	Gamma (N = 2)	1.023
	Gamma (N = 3)	0.564
	Gamma (N = 4)	0.504

show that the models consisting four mixtures are more similar to the actual PDFs for all measurements except 20cm. Surprisingly, KL-divergence of the model with two mixtures is the smallest for the measurement with the distance of 20cm. Furthermore, the results obtained from goodness-of-fit test with the confidence level $p = 0.05$ imply the suitability of the mixture models to actual PDFs.

B. Comparison with Weibull and Gaussian Distributions

Although it is shown that Gamma mixture distribution is able to describe THz channels with ultra broadband in the Section IV-A, we instigate the accuracy of THz channel models with various mixture of distributions rather than Gamma distribution. We evaluate the mixtures of normal and Weibull distributions. EM simulations have been performed for three measurement data.

Firstly, 20cm measurement data is evaluated. It is observed in Fig. 5(a) that the mixture of two Gamma distributions fits the measurement better. On the other hand, Weibull mixture with two distinct distributions cannot represent the actual data; however, increasing the number of mixtures improves accuracy as seen in Fig. 5(a). Gaussian mixtures are observed to describe the channel almost as well as Gamma mixture. However, Gamma mixture has a slightly smaller KL divergence.

60cm measurements are analyzed with the same scenario as the former. In these measurements, Weibull mixtures are unable to provide adequate fitting for modeling the channel. Surprisingly, increasing the number of mixtures could not improve the accuracy of the channel model as depicted in Fig. 5(b). Particularly, Gaussian mixture slightly outperforms the accuracy of Gamma mixture modeling in terms of KL divergence.

Finally, EM algorithm for all measurement data in Fig. 5 are evaluated. When two distinct distributions are employed, Gaussian mixture shows the best fitting performance. But, it is worth noting the CDF of Gaussian distribution cannot be evaluated as in a closed form while Gamma distribution enables closed form expressions. On the other hand, Weibull distribution cannot leverage fitting to actual data even if four mixtures are employed. It is shown that by increasing the number of Gamma distributions in mixture, fitting to measurements can be done in a more accurate way. KL divergence values between fitted mixtures and measurements are summarized in Table II.

V. Concluding Remarks and Future Works

In this paper, we investigate the channel model for the THz band in between 240GHz-300GHz by using Gamma mixture models. To find the mixture parameters, EM algorithm is utilized. It is visible that the mixture models are better to fit the measurement histogram for all measurements compared to MLEs. The comparison between the mixture models and the actual PDFs is carried out by WMRD, KS and KL-divergence metrics. The metrics adminiculate that the mixture of Gamma distributions can accurately model the THz channels.

Since the average channel capacity, the outage probability, and the symbol error rate are derived for mixture Gamma wireless channels, the analytical analyse can be carried by using mixture parameters given in this study. As known, the EM algorithm requires the number of mixtures as a priori information. However, to determine the number of mixtures, the Dirichlet process mixture model and Bayesian information criterion can be utilized.

Due to limitations of the measurement setup in this study, any mobility could not be considered. However, measurement campaigns which enable to consider mobility in THz band should be carried out. Furthermore, in-vivo channel characteristics are heavily dependent on the density of the materials in the tissue; therefore, the in-vivo channels have greater changes in their behaviors. It can be thought that mixture models are appropriate for also in-vivo channels. As a future work, in-vivo channel can be investigated by utilizing mixture models.

REFERENCES

[1] Cisco, "The Zettabyte era: Trends and analysis," https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.pdf, accessed: 2018-08-10.

[2] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical Communication*, vol. 12, pp. 16–32, 2014.

[3] N. Khalid, N. A. Abbasi, and O. B. Akan, "300 GHz broadband transceiver design for low–THz band wireless communications in indoor internet of things," in *Intl. Conf. on Internet of Things and Green Comp. and Commun. and Cyber, Physical and Social Computing and Smart Data*, Jun. 2017, pp. 770–775.

[4] K. Tekbıyık, A. R. Ekti, G. K. Kurt, and A. Görçin, "Terahertz band communication systems: Challenges, novelties and standardization efforts," *Physical Communication*, vol. 35, p. 100700, 2019.

[5] B. Peng, S. Rey, and T. Kürner, "Channel characteristics study for future indoor millimeter and submillimeter wireless communications," in *European Conf. on Antennas and Propag.*, 2016, pp. 1–5.

[6] B. Peng and T. Kürner, "A stochastic channel model for future wireless THz data centers," in *Intl. Symp. on Wireless Commun. Syst.*, 2015, pp. 741–745.

[7] S. Nie and I. F. Akyildiz, "Three–dimensional dynamic channel modeling and tracking for Terahertz band indoor communications," in *Intl. Symp. on Pers. Indoor and Mobile Radio Commun.*, 2017, pp. 1–5.

[8] D. He, K. Guan, A. Fricke, B. Ai, R. He, Z. Zhong, A. Kasamatsu, I. Hosako, and T. Kürner, "Stochastic channel modeling for kiosk applications in the Terahertz band," *IEEE Trans. Terahertz Sci. Technol.*, vol. 7, no. 5, pp. 502–513, 2017.

[9] S. Priebe and T. Kurner, "Stochastic modeling of THz indoor radio channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4445–4455, 2013.

[10] C. Han, A. O. Bicen, and I. F. Akyildiz, "Multi–ray channel modeling and wideband characterization for wireless communications in the Terahertz band," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2402–2412, 2015.

[11] A. R. Ekti, A. Boyaci, A. Alparslan, I. Unal, S. Yarkan, A. Gorcin, H. Arslan, and M. Uysal, "Statistical modeling of propagation channels for Terahertz band," in *Conf. on Std. for Commun. and Netw.*, 2017, pp. 275–280.

[12] K. Tekbıyık, E. Ulusoy, A. R. Ekti, S. Yarkan, T. Baykaş, A. Görçin, and G. K. Kurt, "Statistical channel modeling for short range line-–of-–sight Terahertz communication," in *Intl. Symp. on Pers. Indoor and Mobile Radio Commun.*, 2019, pp. 1–5.

[13] S. Ju and Y. Xing, "3-D statistical indoor channel model for millimeter-wave and sub-Terahertz bands," in *IEEE Global Communications Conference*, 2020, pp. 1–7.

[14] Y. Chen, Y. Li, C. Han, Z. Yu, and G. Wang, "Channel measurement and ray-tracing-statistical hybrid modeling for low-Terahertz indoor communications," *arXiv preprint arXiv:2101.12436*, 2021.

[15] N. Khalid and O. B. Akan, "Wideband THz communication channel measurements for 5G indoor wireless networks," in *Intl. Conf. on Commun.*, 2016, pp. 1–6.

[16] K. Tsujimura, K. Umebayashi, J. Kokkoniemi, J. Lehtomäki, and Y. Suzuki, "A causal channel model for the Terahertz band," *IEEE Trans. Terahertz Sci. Technol.*, vol. 8, no. 1, pp. 52–62, 2018.

[17] Z. Xu, X. Dong, and J. Bornemann, "Design of a reconfigurable MIMO system for THz communications based on Graphene antennas," *IEEE Trans. Terahertz Sci. Technol.*, vol. 4, no. 5, pp. 609–617, 2014.

[18] N. Khalid and O. B. Akan, "Experimental throughput analysis of low–THz MIMO communication channel in 5G wireless networks," *IEEE Wireless Commun. Lett.*, vol. 5, no. 6, pp. 616–619, 2016.

[19] C.-L. Cheng, S. Sangodoyin, and A. Zajić, "Terahertz MIMO fading analysis and doppler modeling in a data center environment," in *14th European Conference on Antennas and Propagation (EuCAP)*, 2020, pp. 1–5.

[20] ——, "THz cluster-based modeling and propagation characterization in a data center environment," *IEEE Access*, vol. 8, pp. 56 544–56 558, 2020.

[21] I. F. Akyildiz and J. M. Jornet, "Realizing ultra–massive MIMO (1024$\times$ 1024) communication in the (0.06–10) Terahertz band," *Nano Communication Networks*, vol. 8, pp. 46–54, 2016.

[22] L. Zakrajsek, E. Einarsson, N. Thawdar, M. Medley, and J. M. Jornet, "Design of Graphene–based plasmonic nano–antenna arrays in the presence of mutual coupling," in *European Conf. on Antennas and Propag.*, 2017, pp. 1381–1385.

[23] C. Han, J. M. Jornet, and I. Akyildiz, "Ultra–massive MIMO channel modeling for Graphene–enabled Terahertz–band communications," in *Veh. Technol. Conf.*, 2018, pp. 1–5.

[24] S. Priebe, C. Jastrow, M. Jacob, T. Kleine-Ostmann, T. Schrader, and T. Kurner, "Channel and propagation measurements at 300 GHz," *IEEE Trans. Antennas Propag.*, vol. 59, no. 5, pp. 1688–1698, 2011.

[25] M. Hudlička, M. Salhi, T. Kleine-Ostmann, and T. Schrader, "Characterization of a 300–GHz transmission system for digital communications," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 38, no. 8, pp. 1004–1018, 2017.

[26] M. Wiper, D. R. Insua, and F. Ruggeri, "Mixtures of gamma distributions with applications," *Journal of computational and graphical statistics*, vol. 10, no. 3, pp. 440–454, 2001.

[27] V. R. da Silva and A. Yongacoglu, "EM algorithm on the approximation of arbitrary PDFs by gaussian, gamma and lognormal mixture distributions," in *7th IEEE Latin-American Conference on Communications*, 2015, pp. 1–6.

[28] H. Al-Hmood, "A mixture Gamma distribution based performance analysis of switch and stay combining scheme over α—κ—μ shadowed fading channels," in *Conf. on New Trends in Inf. & Communs. Technol. App.*, 2017, pp. 292–297.

[29] O. Alhussein, B. Selim, T. Assaf, S. Muhaidat, J. Liang, and G. K. Karagiannidis, "A generalized mixture of Gaussians for fading channels," in *Veh. Technol. Conf.*, 2015, pp. 1–6.

[30] B. Selim, O. Alhussein, S. Muhaidat, G. K. Karagiannidis, and J. Liang, "Modeling and analysis of wireless channels via the mixture of Gaussian distribution," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8309–8321, 2015.

[31] J. Jung, S.-R. Lee, H. Park, S. Lee, and I. Lee, "Capacity and error probability analysis of diversity reception schemes over generalized-K fading channels using a mixture Gamma distribution," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 9, pp. 4721–4730, 2014.

[32] H. Lei, H. Zhang, I. S. Ansari, C. Gao, Y. Guo, G. Pan, and K. A. Qaraqe, "Performance analysis of physical layer security over generalized-K fading channels using a mixture Gamma distribution," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 408–411, 2015.

[33] S. Atapattu, C. Tellambura, and H. Jiang, "A mixture of Gamma distribution to model the SNR of wireless channels," *IEEE Trans. on Wireless Commun.*, vol. 10, no. 12, pp. 4193–4203, 2011.

[34] V. Cellini and G. Dona, "A novel joint channel and multi-user interference statistics estimator for UWB-IR based on Gaussian mixture model," in *Intl. Conf. on Ultra-Wideband*, 2005, pp. 655–660.

[35] S. Büyükçorak, M. Vural, and G. K. Kurt, "Lognormal mixture shadowing," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4386–4398, 2015.

[36] K. Tekbıyık, A. R. Ekti, G. K. Kurt, and A. Görçin, "THz wireless channel measurements in between 240GHz and 300GHz," 2019. [Online]. Available: http://dx.doi.org/10.21227/2jhd-wp15

[37] K. M. Ramachandran and C. P. Tsokos, *Mathematical statistics with applications*. Elsevier, 2009.

[38] K. P. Murphy, *Machine learning: A probabilistic perspective*. MIT Press, 2012.

[39] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[40] J. A. Bilmes *et al.*, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," *Intl. Computer Science Inst.*, vol. 4, no. 510, p. 126, 1998.

[41] “MILTAL millimeter wave and terahertz technology laboratory,” http://me.mam.tubitak.gov.tr/en/research-areas/millimeter-wave-and-terahertz-technology, accessed: 2020-01-17.

[42] A. Goldsmith, *Wireless communications*. Cambridge University Press, 2005.

Kürşat Tekbıyık [StM’19] (tekbiyik@itu.edu.tr) received his B.Sc. and M.Sc. degrees (with high honors) in electronics and communication engineering from Istanbul Technical University, Istanbul, Turkey, in 2017 and 2019, respectively. He is currently pursuing his Ph.D. degree in telecommunications engineering in Istanbul Technical University. He is also researcher at TUBITAK BILGEM. His research interests are next-generation wireless communication systems, terahertz wireless communications, and machine learning.

Ali Rıza Ekti is from Tarsus, Turkey. He received B.Sc. degree in Electrical and Electronics Engineering from Mersin University, Mersin, Turkey, (September 2002-June 2006), also studied at Universidad Politechnica de Valencia, Valencia, Spain in 2004-2005, received M.Sc. degree in Electrical Engineering from the University of South Florida, Tampa, Florida (August 2008-December 2009) and received Ph.D. degree in Electrical Engineering from Texas A&M University (August 2010-August 2015). He is currently an assistant professor at Balikesir University Electrical and Electronics Engineering Department and also senior researcher at TUBITAK BILGEM. His current research interests include statistical signal processing, convex optimization, machine learning, resource allocation and traffic offloading in wireless communications in 4G and 5G systems and smart grid design and optimization.

Güneş Karabulut Kurt received a Ph.D. degree in electrical engineering from the University of Ottawa, Ottawa, ON, Canada, in 2006. Between 2005 and 2008, she was with TenXc Wireless, and Edgewater Computer Systems, in Ottawa Canada. From 2008 to 2010, she was with Turkcell R&D, Istanbul. Since 2010, she has been with Istanbul Technical University. She is also an Adjunct Research Professor at Carleton University. She is serving as an Associate Technical Editor of *IEEE Communications Magazine*.

Ali Görçin graduated from Istanbul Technical University with B.Sc. in Electronics and Telecommunications Engineering and completed his master’s degree on defense technologies at the same university. After working at Turkish Science Foundation (TUBITAK) on avionics projects for more than six years, he moved to the U.S. to pursue PhD degree in University of South Florida (USF) on wireless communications. He worked for Anritsu Company during his tenure in USF and worked for Reverb Networks and Viavi Solutions after his graduation. He is currently holding an assistant professorship position at Yildiz Technical University in Istanbul and also serving as the president of Informatics and Information Security Research Center (BILGEM) of TUBITAK, responsible from testing and evaluation along with research and development activities on wireless communications technologies.

Serhan Yarkan received the B.S. and M.Sc. degrees in computer science from Istanbul University, Istanbul, Turkey, in 2001 and 2003, respectively, and the Ph.D. degree from the University of South Florida, Tampa, FL, USA, in 2009. He was a Post-Doctoral Research Associate with the Department of Computer and Electrical Engineering, Texas A&M University, College Station, TX, USA, from 2010 to 2012. He is currently an Associate Professor with the Department of Electrical-Electronics Engineering, Istanbul Commerce University, Istanbul. His current research interests include statistical signal processing, cognitive radio, wireless propagation channel measurement and modeling, cross-layer adaptation and optimization, and interference management in next-generation wireless networks and underground mine channels and disaster communications.